

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 January 2003 (03.01.2003)

PCT

(10) International Publication Number
WO 03/000904 A2

(51) International Patent Classification⁷: **C12N 15/82**

(21) International Application Number: PCT/EP02/06968

(22) International Filing Date: 24 June 2002 (24.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/300,112 22 June 2001 (22.06.2001) US
60/325,277 26 September 2001 (26.09.2001) US
60/334,984 30 November 2001 (30.11.2001) US

(71) Applicant (*for all designated States except US*): **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SESSIONS, Allen** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **BRIGGS, Steven** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **COOPER, Bret** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **GOFF, Stephen, A.** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **MOUGHAMER, Todd** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **GLAZEBROOK, Jane** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **KATAGIRI, Fumiaki** [JP/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San

Diego, CA 92121-1125 (US). **KREPS, Joel** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **PROVART, Nicolas** [CA/CA]; 33 Longwood Drive, Toronto, Ontario M3B 1T9 (CA). **RICKE, Darrell** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US). **ZHU, Tong** [US/US]; Torrey Mesa Research Institute, Inc., 3115 Merryfield Row, San Diego, CA 92121-1125 (US).

(74) Agent: **BASTIAN, Werner**; c/o Syngenta Participations AG, Intellectual Property, P.O. Box, CH-4002 Basel (CH).

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

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (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*

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.

(54) Title: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES

(57) Abstract: The present invention relates to nucleic acid molecules obtainable from the rice genome that encode protein products that are involved in the development and timing of flower formation in plants and which can be used to modulate flower development, architecture and flowering time.



WO 03/000904 A2

IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES

The present invention is in the area of plant biotechnology. In particular, the invention relates to nucleic acid molecules obtainable from the rice genome that encode protein products that are involved in the development and timing of flower formation in plants and which can be used to modulate flower architecture and flowering time.

In most angiosperm species flower induction is largely triggered by environmental stimuli such as day length and temperature, and also by internal factors, such as age. Starting point for flower development are groups of specialized stem cells called inflorescence meristems which give rise to flower meristems. From the flower meristems floral organs develop such as sepals and petals.

Flower formation is the result of a series of distinct developmental steps including floral induction, the formation of flower primordia and the production of flower organs. In a first step, floral induction may lead to the transformation of a vegetative shoot meristem into an inflorescens meristem, which then gives rise to a flower meristems. Flower meristems produce floral organ primordia, which develop individually into sepals, petals, stamens or carpels.

Mutations disrupting each of these developmental steps have been isolated in a variety of species, suggesting that a genetic hierarchy directs the flowering process (see for review, Weigel and Meyerowitz, In Molecular Basis of Morphogenesis (ed. M. Bernfield). 51st Annual Symposium of the Society for Developmental Biology, pp. 93-107, New York, 1993).

The ABC model of pattern formation offers a developmental understanding of the flower parts in both monocotyledonous and dicotyledonous plants. It states that genes 'A' and 'B' and 'C' interact with one another to specify different floral identities. Sepals, petals, stamens and carpels are usually arranged in three, four or five concentric whorls. The sepals arise below the petals, and the stamens arise below the carpels. Sepals and petals are sterile appendages that are attached under the fertile floral parts(stamens and carpels). The 'A' genes encode for sepals, the 'A' and 'B' genes together encode petals, the 'B' and 'C' genes together encode stamens, and the 'C' genes encode for carpels. In *Arabidopsis thaliana*, examples of an 'A' gene is *Apetala1* (AP1), 'B' genes are *Apetala3* (AP3) and *Pistillata* (PI) and 'C' gene

is *Agamous* (AG). For a review of the ABC model, we refer to Weigel, D. *et al.*, *Cell* 78:203 (1994) for reference.

The regulatory pathways that control the initial pattern expression of A, B, and C genes, however, are not fully characterized. Only a handful of regulatory genes involved in floral identity have been identified. They are usually homeotic genes that are involved in the control of meristem development. Several of these genes are transcription factors whose conserved DNA-binding domain has been designated the MADS box (Schwarz-Sommer, *et al.*, *supra*). Earlier acting genes that control the identity of flower meristems have also been characterized. Flower meristems are derived from inflorescence meristems in both *Arabidopsis* and *Antirrhinum*. Two factors that control the development of meristematic cells into flowers are known. In *Arabidopsis*, the factors are the products of the *LEAFY* gene (Weigel, *et al.*, *Cell* 69:843, 1992) and the *APETALA1* gene (Mandel, *et al.*, *Nature* 360:273, 1992). The *Leafy* (*LFY*) gene in *Arabidopsis* has been shown to regulate the expression of 'A' gene, *AP1*, and 'C' gene, *AG* (Busch M., *et al.*, *Science* 285:585 (1999) and Wagner D., *et al.*, *Science* 285: 582 (1999)). While *LFY* and *Unusual Floral Organ* (*UFO*) control the expression of B gene, *AP3* (Wilkinson M., *et al.*, *Plant Cell* 1995 7:1485 (1995)).

In *Antirrhinum*, the homologue of the *Arabidopsis* *LEAFY* gene is *FLORICAULA* (Coen, *et al.*, *Cell*, 63:1311, 1990) and that of the *APETALA1* gene is *SQUAMOSA* (Huijser, *et al.*, *EMBO J.*, 11:1239, 1992). The latter pair contains MADS box domains.

Flowering plants exhibit one of two types of inflorescence architecture: indeterminate, in which the inflorescence grows indefinitely, or determinate, in which a terminal flower is produced. In two mutants in distantly related species, terminal flower 1 in *Arabidopsis* and *centroradialis* in *Antirrhinum*, inflorescences that are normally indeterminate are converted to a determinate architecture. The *Antirrhinum* gene *CENTRORADIALIS* (*CEN*) and the *Arabidopsis* gene *TERMINAL FLOWER 1* (*TFL1*) were shown to be homologous, which suggests that a common mechanism underlies indeterminacy in these plants. However, unlike *CEN*, *TFL1* is also expressed during the vegetative phase, where it delays the commitment to inflorescence development and thus affects the timing of the formation of the inflorescence meristem as well as its identity.

Flowering in the long day plant *Arabidopsis* is promoted by a regulatory cascade that is initiated by the flowering time genes, which in turn activate the floral meristem identity

genes, ultimately leading to the patterned expression of the floral organ identity genes (Figure VII). In *Arabidopsis* there are three genetic pathways that control flowering time [192, 193]. The long day pathway represented by *GI* and *CO*, and the autonomous pathway represented by *LD*, *FCA* and *FLC*, are likely integrated through *FT* and *AGL20* to promote activation of
5 meristem identity genes *LFY*, *API* and *CAL*. The vernalization pathway, represented by *FRI*, feeds into the autonomous pathway upstream of *FLC*. The GA pathway, represented by *GAI*, leads to the activation of *LFY*. *TFL* serves to restrict the expression of the meristem identity genes to floral meristems, where they promote the patterned expression of floral organ identity genes *AP2*, *AP3*, *PI* and *AG*, which are also affected by the regulatory genes *ANT*, *UFO*, and
10 *SUP* [194, 195].

The present invention provides and relates to nucleotide sequences encoding at least one polypeptide involved in the development and/or timing of flower formation and/or whole plant architecture, as well as the polypeptides encoded thereby, or antigene sequences thereof, which have numerous applications using techniques that are known to those skilled in the art
15 of molecular biology, biotechnology, biochemistry, genetics, physiology or plant pathology. These techniques include the use of nucleotide molecules as hybridization probes, for chromosome and gene mapping, in PCR technologies, in the production of sense or antisense nucleic acids.

In a further collective embodiment, the present invention provides the ability to
20 modulate development and/or timing of flower formation and/or whole plant architecture in plants, by modulating gene expression (e.g. over-expressing, under-expressing or knocking out) one or more genes involved in or regulating these processes or their gene products, in a host cell, preferably in a plant cell, *in vitro* or *in planta*. Expression vectors comprising at least one nucleotide sequence encoding a polypeptide the activity of which is involved in
25 development and/or timing of flower formation and/or whole plant architecture or the regulation thereof, or its antigene, operably linked to at least one suitable promoter and/or regulatory sequence can be used to study the role of polypeptides encoded by said sequences, for example by transforming a host cell with said expression vector and measuring the effects of overexpression, underexpression or complete knock-out of these sequences. A host cell
30 transformed with at least one expression vector comprising nucleotide sequences encoding a polypeptide the activity of which is involved in the development and/or timing of flower

formation in plants and/or whole plant architecture, or the regulation thereof operably linked to suitable promoters and/or regulatory sequences can be useful to produce a polypeptide having a defined polypeptide profile that results in a plant showing a modified architecture or timing of flower development.

5 The present invention further provides, in a collective embodiment, a transformed plant host cell, or one obtained through breeding, capable of over-expressing, under-expressing, or having a knock out of a flower development or flower time gene or a gene regulating these processes and/or its gene products. In a further embodiment is a plant cell transformed with at least one expression vector comprising nucleotide sequences involved in
10 the development and/or timing of flower formation in plants and/or whole plant architecture, operably linked to suitable promoters and/or regulatory sequences, wherein the plant host cell can be used to regenerate plant tissue or an entire plant, or seed therefrom, in which the effects of expression, including overexpression or underexpression or knock-out, of the introduced sequence or sequences can be measured *in vitro* or in *planta*.

15 In a further embodiment the present invention provides nucleotide sequences including regions of nucleotide sequence encoding polypeptides having homology to at least one functional protein domain (FPD). Embodiments of the invention further provide polypeptides including regions of amino acid sequence having homology to an FPD. In cases where the polypeptide has homology to an FPD in the same or closely related species, the polypeptide
20 may represent a paralogous sequence or paralog, or may represent a variant allele of a gene encoding the FPD. In cases where the polypeptide has homology to an FPD in another species, including other plant species and especially non-plant species, polypeptides may represent orthologous sequences, or orthologs, of the FPD.

 In a further collective embodiment of the invention the nucleic acid molecules disclosed
25 herein or representative parts thereof can be used in hybridization-based assays for detecting and identifying nucleic acid molecules that encode protein products that are involved in the development and/or timing of flower formation in plants and/or whole plant architecture other than rice, but especially in plants belonging to the cereal group.

 Embodiments of the present invention provide a unique oligonucleotide having a
30 sequence identical to or complementary to a region of a polynucleotide sequence encoding at least a portion of a homologue of a protein according to the invention representatives of which

are identified by the SEQ ID NOs provided in the Sequence Listing and/or an FPD thereof, the oligonucleotide being identified by the methods disclosed herein. In one embodiment, the unique oligonucleotide has a length of between 12 and 250 nucleotide bases.

Embodiments of the present invention also provide a nucleotide microarray
5 comprising the unique oligonucleotide having a sequence identical to or complementary to a region of polynucleotide sequence encoding at least a portion of a homologue of a protein according to the invention representatives of which are identified by the SEQ ID NOs provided in the Sequence Listing and/or an FPD thereof. Preferably, the microarray includes a plurality of different, unique oligonucleotides, the sequences corresponding to a plurality of
10 homologues of a protein according to the invention representatives of which are identified by the SEQ ID NOs provided in the Sequence Listing and/or an FPD thereof. Equally preferably, the microarray contains at least about 96 different unique oligonucleotides, wherein each of the 96 different unique oligonucleotides has a sequence that is identical, complementary, or has substantial similarity to a segment of a nucleotide sequence selected from the group
15 consisting of SEQ ID NOs: 1 to SEQ ID NO: 65 and 123 – 145.

Embodiments of the present invention also provide a kit for detecting the presence of a polynucleotide, the kit containing a first nucleotide probe which can hybridize with a region of a nucleotide sequence including the nucleotide sequences of SEQ ID NOs: 1 to SEQ ID NO: 65 and SEQ ID NOs: 123 – 145, a fragment or a variant thereof, and a complementary
20 sequence thereto, the kit further containing at least one additional component such as, for example: a second nucleotide probe, a buffer, an enzyme, a label, a molecular weight standard, a reaction chamber, and a micropipette tip.

Embodiments of the present invention further provide a kit for detecting the presence of a polypeptide, the kit containing a first probe which can hybridize with a region of a
25 polypeptide including the amino acid sequences of SEQ ID NOs: 2 to SEQ ID NOs: 66 and SEQ ID NOs: 124 – 146, a fragment or a variant thereof, and optionally, the kit further containing at least one additional component such as, for example: a probe, a buffer, an enzyme, a label, a molecular weight standard, a reaction chamber, and a micropipette tip. Probes useful in kit embodiments include antibodies, affinity tags, protein A, protein G, or
30 protein-binding substances including chromatographic media.

An additional aspect provides a method for selecting plants, for example cereals, having an altered heading time or whole plant architecture comprising obtaining nucleic acid molecules from the plants to be selected; contacting the nucleic acid molecules with one or more probes that selectively hybridize under stringent or highly stringent conditions to a nucleic acid sequence selected from the group consisting of SEQ ID NOs. 1-66, and 124 - 146; detecting the hybridization of the one or more probes to the nucleic acid sequences wherein the presence of the hybridization indicates the presence of a gene associated with altered heading time or whole plant architecture; and selecting plants on the basis of the presence or absence of such hybridization. In one embodiment, marker-assisted selection is accomplished in rice. In another embodiment, marker assisted selection is accomplished in wheat using one or more probes which selectively hybridize under stringent or highly stringent conditions to sequences selected from the group consisting of SEQ ID NOs: 81 - 105. In yet another embodiment, marker assisted selection is accomplished in maize or corn using one or more probes which selectively hybridize under stringent or highly stringent conditions to sequences selected from the group consisting of SEQ ID NOs: 106 – 122. In still another embodiment, marker assisted selection is accomplished in banana using one or more probes which selectively hybridize under stringent or highly stringent conditions to sequences selected from the group consisting of SEQ ID NOs: 67 – 80. In each case marker-assisted selection can be accomplished using a probe or probes to a single sequence or multiple sequences. If multiple sequences are used they can be used simultaneously or sequentially.

In a further embodiment of the invention a computer readable medium containing one or more of the nucleotide sequences of the invention is provided as well as methods of use for the computer readable medium. This medium allows a nucleotide sequence corresponding to at least one of the sequences selected from the group of SEQ ID NOs: 1-66, and 123 – 146, and 67 – 122 provided in the Sequence Listing (open reading frames or fragments thereof), to be used as a reference sequence to search against a database. This medium also allows for computer-based manipulation of a nucleotide sequence corresponding to at least one of the sequences selected from the group consisting of SEQ ID NOs: 1-66, and 123 - 146, and 67 – 122 provided in the Sequence Listing.

A further aspect provides a computer readable medium having stored thereon computer executable instructions for performing a method comprising receiving data on nucleotide

sequence expression in a test plant of at least one nucleic acid molecule having at least 70%, at least 80%, at least 90% or at least 95%, sequence identity to a nucleotide sequence selected from the group consisting of SEQ ID NOs: SEQ ID NOs. 1-66, and 123 - 146; and 67 – 122 and comparing expression data from said test plant to expression data for the same nucleotide sequence or sequences in a plant during grain filling.

Further aspects, features and advantages of this invention will become apparent from the detailed description of the preferred embodiments that follow.

Brief Description of the Sequence Listing

In the following, a brief description of the sequences in the Sequence Listing is provided:

Odd numbered SEQ ID NOs: 1 – 65 are representing a first sub-group (sub-group I) of polynucleotides comprising nucleotide sequences of rice encoding polypeptides involved in flower timing in plants and/or whole plant architecture or the regulation thereof.

Even numbered SEQ ID NOs: 2 - 66 are polypeptides involved in flower timing in plants and/or whole plant architecture or the regulation thereof encoded by the immediately preceding nucleotide sequence e.g., SEQ ID NO: 2 is the protein encoded by the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:4 is the protein encoded by the nucleotide sequence of SEQ ID NO:3, etc..

SEQ ID NOs: 67 – 80 are banana sequences which show homology to the rice sequences provided in SEQ ID NOs: 1 –65.

SEQ ID NOs: 81 - 105 are wheat sequences which show homology to the rice sequences provided in SEQ ID NOs: 1 –65.

SEQ ID NOs: 106 – 122 are maize sequences which show homology to the rice sequences provided in SEQ ID NOs: 1 –65.

Odd numbered SEQ ID NOs: 123 – 145 are representing a second sub-group (sub-group II) of polynucleotides comprising nucleotide sequences that have homologies between 80% and 99.9% to the nucleotide sequences of sub-group I and possible variants or family members of rice sequences provided in SEQ ID NOs: 1-65. The correlation between the sequences in sub-groups I and II is illustrated in Tables 1 through 3.

Even numbered SEQ ID NOs:124 - 146 are protein sequences encoded by the immediately preceding nucleotide sequence.

Definitions

5 For clarity, certain terms used in the specification are defined and presented as follows:

The term "gene" is used broadly to refer to any segment of nucleic acid with or without its native regulatory sequences such as promoters and polyadenylation sites, signal sequences, etc, associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. For example, gene refers to a nucleic acid fragment that expresses mRNA or functional RNA, or encodes a specific protein, and which includes regulatory sequences. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

The term "native" or "wild type" gene refers to a gene that is present in the genome of an untransformed cell, i.e., a cell not having a known mutation.

A "marker gene" encodes a selectable or screenable trait.

20 The term "chimeric gene" refers to any gene that contains 1) DNA sequences, including regulatory and coding sequences, that are not found together in nature, or 2) sequences encoding parts of proteins not naturally adjoined, or 3) parts of promoters that are not naturally adjoined. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or comprise regulatory sequences and coding sequences derived from the same source, but arranged in a manner different from that found in nature.

A "transgene" refers to a gene that has been introduced into the genome by transformation and is stably maintained. Transgenes may include, for example, genes that are either heterologous or homologous to the genes of a particular plant to be transformed.

30 Additionally, transgenes may comprise native genes inserted into a non-native organism, or chimeric genes. The term "endogenous gene" refers to a native gene in its natural location in

the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism but that is introduced by gene transfer.

An "oligonucleotide" corresponding to a nucleotide sequence of the invention, e.g., for use in probing or amplification reactions, may be about 30 or fewer nucleotides in length (e.g., 9, 12, 15, 18, 20, 21 or 24, or any number between 9 and 30). Generally specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16 to 24 nucleotides in length may be preferred. Those skilled in the art are well versed in the design of primers for use processes such as PCR. If required, probing can be done with entire restriction fragments of the gene disclosed herein which may be 100's or even 1000's of nucleotides in length.

The terms "protein," "peptide" and "polypeptide" are used interchangeably herein.

A "partial-length polypeptide" refers to any part of a given full-length polypeptide that has substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The nucleotide sequences of the invention can be introduced into any plant. The genes to be introduced can be conveniently used in expression cassettes for introduction and expression in any plant of interest. Such expression cassettes will comprise the transcriptional initiation region of the invention linked to a nucleotide sequence of interest. Preferred promoters include constitutive, tissue-specific, developmental-specific, inducible and/or viral promoters. Such an expression cassette is provided with a plurality of restriction sites for insertion of the gene of interest to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes. The cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also, Guerineau et al., 1991; Proudfoot, 1991; Sanfacon et al., 1991; Mogen et al., 1990; Munroe et al., 1990; Ballas et al., 1989; Joshi et al., 1987.

"Coding sequence (CDS)" refers to a DNA or RNA sequence that codes for a specific amino acid sequence and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", i.e., lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a sequence of RNA which is contained in the primary transcript but which is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

The terms "open reading frame" and "ORF" refer to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides ('codon') in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

A "functional RNA" refers to an antisense RNA, ribozyme, or other RNA that is not translated.

The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

"Regulatory sequences" and "suitable regulatory sequences" each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. As is noted above, the term "suitable regulatory sequences" is not limited to promoters.

"5' non-coding sequence" refers to a nucleotide sequence located 5' (upstream) to the coding sequence. It is present in the fully processed mRNA upstream of the initiation codon

and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency (Turner et al., 1995).

"3' non-coding sequence" refers to nucleotide sequences located 3' (downstream) to a coding sequence and include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., 1989.

The term "translation leader sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

"Signal peptide" refers to the amino terminal extension of a polypeptide, which is translated in conjunction with the polypeptide forming a precursor peptide and which is required for its entrance into the secretory pathway. The term "signal sequence" refers to a nucleotide sequence that encodes the signal peptide.

"Promoter" refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. "Promoter" includes a minimal promoter that is a short DNA sequence comprised of a TATA box and other sequences that serve to specify the site of transcription initiation, to which regulatory elements are added for control of expression. "Promoter" also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements that is capable of controlling the expression of a coding sequence or functional RNA. This type of promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other upstream promoter

elements bind sequence-specific DNA-binding proteins that mediate their effects. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also contain DNA sequences that are involved in the binding of
5 protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions.

The "initiation site" is the position surrounding the first nucleotide that is part of the transcribed sequence, which is also defined as position +1. With respect to this site all other sequences of the gene and its controlling regions are numbered. Downstream sequences (i.e.,
10 further protein encoding sequences in the 3' direction) are denominated positive, while upstream sequences (mostly of the controlling regions in the 5' direction) are denominated negative.

Promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation are referred to as "minimal or
15 core promoters." In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription. A "minimal or core promoter" thus consists only of all basal elements needed for transcription initiation, e.g., a TATA box and/or an initiator.

"Constitutive expression" refers to expression using a constitutive or regulated promoter. "Conditional" and "regulated expression" refer to expression controlled by a
20 regulated promoter.

"Constitutive promoter" refers to a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant. Each of the transcription-activating elements do not exhibit an absolute tissue-specificity, but mediate transcriptional activation in most plant parts
25 at a level of $\geq 1\%$ of the level reached in the part of the plant in which transcription is most active.

"Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes both tissue-specific and inducible promoters. It includes natural and synthetic sequences as well as
30 sequences which may be a combination of synthetic and natural sequences. Different promoters may direct the expression of a gene in different tissues or cell types, or at different

stages of development, or in response to different environmental conditions. New promoters of various types useful in plant cells are constantly being discovered, numerous examples may be found in the compilation by Okamuro et al. (1989). Typical regulated promoters useful in plants include but are not limited to safener-inducible promoters, promoters derived from the tetracycline-inducible system, promoters derived from salicylate-inducible systems, promoters derived from alcohol-inducible systems, promoters derived from glucocorticoid-inducible system, promoters derived from pathogen-inducible systems, and promoters derived from ecdysome-inducible systems.

"Tissue-specific promoter" refers to regulated promoters that are not expressed in all plant cells but only in one or more cell types in specific organs (such as leaves or seeds), specific tissues (such as embryo or cotyledon), or specific cell types (such as leaf parenchyma or seed storage cells). These also include promoters that are temporally regulated, such as in early or late embryogenesis, during fruit ripening in developing seeds or fruit, in fully differentiated leaf, or at the onset of senescence.

"Inducible promoter" refers to those regulated promoters that can be turned on in one or more cell types by an external stimulus, such as a chemical, light, hormone, stress, or a pathogen.

"Operably-linked" refers to the association of nucleic acid sequences on single nucleic acid fragment so that the function of one is affected by the other. For example, a regulatory DNA sequence is said to be "operably linked to" or "associated with" a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

"Expression" refers to the transcription and/or translation of an endogenous gene, ORF or portion thereof, or a transgene in plants. For example, in the case of antisense constructs, expression may refer to the transcription of the antisense DNA only. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

"Specific expression" is the expression of gene products which is limited to one or a few plant tissues (spatial limitation) and/or to one or a few plant developmental stages

(temporal limitation). It is acknowledged that hardly a true specificity exists: promoters seem to be preferably switch on in some tissues, while in other tissues there can be no or only little activity. This phenomenon is known as leaky expression. However, with specific expression in this invention is meant preferable expression in one or a few plant tissues.

5 The "expression pattern" of a promoter (with or without enhancer) is the pattern of expression levels which shows where in the plant and in what developmental stage transcription is initiated by said promoter. Expression patterns of a set of promoters are said to be complementary when the expression pattern of one promoter shows little overlap with the expression pattern of the other promoter. The level of expression of a promoter can be
10 determined by measuring the 'steady state' concentration of a standard transcribed reporter mRNA. This measurement is indirect since the concentration of the reporter mRNA is dependent not only on its synthesis rate, but also on the rate with which the mRNA is degraded. Therefore, the steady state level is the product of synthesis rates and degradation rates.

15 The rate of degradation can however be considered to proceed at a fixed rate when the transcribed sequences are identical, and thus this value can serve as a measure of synthesis rates. When promoters are compared in this way techniques available to those skilled in the art are hybridization S1-RNase analysis, northern blots and competitive RT-PCR. This list of techniques in no way represents all available techniques, but rather describes commonly used
20 procedures used to analyze transcription activity and expression levels of mRNA.

 The analysis of transcription start points in practically all promoters has revealed that there is usually no single base at which transcription starts, but rather a more or less clustered set of initiation sites, each of which accounts for some start points of the mRNA. Since this distribution varies from promoter to promoter the sequences of the reporter mRNA in each of
25 the populations would differ from each other. Since each mRNA species is more or less prone to degradation, no single degradation rate can be expected for different reporter mRNAs. It has been shown for various eukaryotic promoter sequences that the sequence surrounding the initiation site ('initiator') plays an important role in determining the level of RNA expression directed by that specific promoter. This includes also part of the transcribed sequences. The
30 direct fusion of promoter to reporter sequences would therefore lead to suboptimal levels of transcription.

A commonly used procedure to analyze expression patterns and levels is through determination of the 'steady state' level of protein accumulation in a cell. Commonly used candidates for the reporter gene, known to those skilled in the art are β -glucuronidase (GUS), chloramphenicol acetyl transferase (CAT) and proteins with fluorescent properties, such as green fluorescent protein (GFP) from *Aequora victoria*. In principle, however, many more proteins are suitable for this purpose, provided the protein does not interfere with essential plant functions. For quantification and determination of localization a number of tools are suited. Detection systems can readily be created or are available which are based on, e.g., immunochemical, enzymatic, fluorescent detection and quantification. Protein levels can be determined in plant tissue extracts or in intact tissue using *in situ* analysis of protein expression.

Generally, individual transformed lines with one chimeric promoter reporter construct will vary in their levels of expression of the reporter gene. Also frequently observed is the phenomenon that such transformants do not express any detectable product (RNA or protein). The variability in expression is commonly ascribed to 'position effects', although the molecular mechanisms underlying this inactivity are usually not clear.

"Overexpression" refers to the level of expression in transgenic cells or organisms that exceeds levels of expression in normal or untransformed (nontransgenic) cells or organisms.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from an endogenous gene or a transgene.

"Gene silencing" refers to homology-dependent suppression of viral genes, transgenes, or endogenous nuclear genes. Gene silencing may be transcriptional, when the suppression is due to decreased transcription of the affected genes, or post-transcriptional, when the suppression is due to increased turnover (degradation) of RNA species homologous to the affected genes (English et al., 1996). Gene silencing includes virus-induced gene silencing (Ruiz et al. 1998).

The terms "heterologous DNA sequence," "exogenous DNA segment" or "heterologous nucleic acid," as used herein, each refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling.

The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides. A "homologous" DNA sequence is a DNA sequence that is naturally associated with a host cell into which it is introduced.

"Homologous to" in the context of nucleotide sequence identity refers to the similarity between the nucleotide sequence of two nucleic acid molecules or between the amino acid sequences of two protein molecules. As used herein, "homology" and "homologous" refer to an evaluation of the similarity between two sequences based on measurements of sequence identity adjusted for variables including gaps, insertions, frame shifts, conservative substitutions, and sequencing errors, as described below. Two nucleotide sequences or polypeptides are to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the sequence can form a Watson-Crick base pair with a reference polynucleotide sequence. Complementary sequences can include nucleotides, such as inosine, that neither disrupt Watson-Crick base pairing nor contribute to the pairing. A "reverse complement" of a sequence corresponds to the complementary sequence, but in the opposite orientation of bases from 5' to 3', or to the complement of the primary sequence, if the primary sequence is in a reverse orientation of bases from 5' to 3'.

Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, *Proc Natl Acad Sci (USA)* 85:2444 (1988); Altschul *et al.*, *J Mol Biol* 215:403 (1990)). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (Karlin and Altschul, *Proc Natl Acad Sci USA* 87:2264 (1990); Altschul *et al.* (1990) *supra*, Altschul *et al.*, *Nucleic Acids Res* 25:3389 (1997)). In particular, five specific BLAST programs are used to perform the following task:

- (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;
- (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;
- 5 (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;
- (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and
- (5) TBLASTX compares the six-frame translations of a nucleotide query sequence
- 10 against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as “high-scoring segment pairs,” between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (aligned) by means

15 of a scoring matrix selected from the many scoring matrices known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al.*, *Science* 256:1443 (1992); Henikoff and Henikoff, *Proteins* 17:49 (1993)). Likewise, the PAM or PAM250 matrices may also be used (Schwartz and Dayhoff, In *Atlas of Protein Sequence and Structure*, Dayhoff, ed., Natl. Biomed. Res. Found., pp. 353-358 (1978)). The BLAST programs

20 evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (Karlin and Altschul (1990) *supra*).

25 “Percentage of sequence identity” can be determined from alignments performed using algorithms known in the art. Alignment of nucleotide or polypeptide sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman (*Add APL Math* 2:482 (1981)), by the homology alignment algorithm of Needleman and Wunsch (*J Mol Biol* 48:443 (1970)), by the search for similarity method of Pearson and Lipman (*Proc Natl Acad Sci USA* 85:2444 (1988)), by computerized implementations of these algorithms

30 (GAP, BESTFIT, BLAST, PASTA, and TFASTA in the Wisconsin Genetics Software

Package, Genetics Computer Group), or by inspection. When two sequences have been identified for comparison, GAP and BESTFIT are preferably employed to determine their optimal alignment. Typically, the default values of 5.00 for gap weight and 0.30 for gap weight length are used. In a preferred embodiment, percent identity is determined using the
5 GAP program for global alignment using default parameters, using the version of GAP found in the GCG package (Wisconsin Package Version 10.1, Genetics Computer Group, 575 Science Dr., Madison, Wisconsin).

“Percentage of sequence identity” is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the sequence in the comparison
10 window may include additions or deletions, including for example gaps or overhangs, as compared to the reference sequence (which does not include additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleotide base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched
15 positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

In a broad sense, the term “substantially similar”, when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence, wherein the corresponding sequence encodes a polypeptide having substantially the
20 same structure as the polypeptide encoded by the reference nucleotide sequence. Desirably, the substantially similar nucleotide sequence encodes the polypeptide encoded by the reference nucleotide sequence. Preferably, “substantially similar” refers to nucleotide sequences having at least 50% sequence identity, preferably at least 60%, 70%, 80% or 85%, more preferably at least 90% or 95%, and even more preferably, at least 96%, 97% or 99%
25 sequence identity compared to a reference sequence containing nucleotide sequences of Table I, that encode a protein having at least 50% identity, more preferably at least 85% identity, yet still more preferably at least 90% identity to a region of sequence of a BIOPATH protein and/or an FPD, wherein the protein sequence comparisons are conducted using GAP analysis as described below. Also, “substantially similar” preferably also refers to nucleotide
30 sequences having at least 50% identity, more preferably at least 80% identity, still more preferably 95% identity, yet still more preferably at least 99% identity, to a region of

nucleotide sequence encoding a BIOPATH protein and/or an FPD, wherein the nucleotide sequence comparisons are conducted using GAP analysis as described below. The term “substantially similar” is specifically intended to include nucleotide sequences wherein the sequence has been modified to optimize expression in particular cells.

5 A polynucleotide including a nucleotide sequence “substantially similar” to the reference nucleotide sequence preferably hybridizes to a polynucleotide including the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1%
10 SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

15 The term “substantially similar”, when used herein with respect to a protein or polypeptide, means a protein or polypeptide corresponding to a reference protein, wherein the protein has substantially the same structure and function as the reference protein, where only changes in amino acids sequence that do not materially affect the polypeptide function occur. When used for a protein or an amino acid sequence the percentage of identity between the
20 substantially similar and the reference protein or amino acid sequence desirably is preferably at least 30%, more preferably at least 40%, 50%, 60%, 70%, 80%, 85%, or 90%, still more preferably at least 95% , still more preferably at least 99% with every individual number falling within this range of at least 30% to at least 99% also being part of the invention, using default GAP analysis parameters with the University of Wisconsin GCG (version 10),
25 SEQWEB application of GAP, based on the algorithm of Needleman and Wunsch (1970), *supra*. As used herein the term “polypeptide of the present invention,” or any similar term refers to an amino acid sequence encoded by a DNA molecule including a nucleotide sequence substantially similar to an Nucleotide sequence according to the invention. Homologs of BIOPATH protein and/or FPDs include amino acid sequences that are at least
30 30% identical to BIOPATH protein and/or FPD sequences found in searchable databases, as measured using the parameters described above.

"Target gene" refers to a gene on the replicon that expresses the desired target coding sequence, functional RNA, or protein. The target gene is not essential for replicon replication. Additionally, target genes may comprise native non-viral genes inserted into a non-native organism, or chimeric genes, and will be under the control of suitable regulatory sequences.

5 Thus, the regulatory sequences in the target gene may come from any source, including the virus. Target genes may include coding sequences that are either heterologous or homologous to the genes of a particular plant to be transformed. However, target genes do not include native viral genes. Typical target genes include, but are not limited to genes encoding a structural protein, a seed storage protein, a protein that conveys herbicide resistance, and a
10 protein that conveys insect resistance. Proteins encoded by target genes are known as "foreign proteins". The expression of a target gene in a plant will typically produce an altered plant trait.

The term "altered plant trait" means any phenotypic or genotypic change in a transgenic plant relative to the wild-type or non-transgenic plant host.

15 "Chromosomally-integrated" refers to the integration of a foreign gene or DNA construct into the host DNA by covalent bonds. Where genes are not "chromosomally integrated" they may be "transiently expressed." Transient expression of a gene refers to the expression of a gene that is not integrated into the host chromosome but functions independently, either as part of an autonomously replicating plasmid or expression cassette,
20 for example, or as part of another biological system such as a virus.

The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to as "transgenic" cells, and organisms comprising transgenic cells are referred to as "transgenic organisms". Examples of methods
25 of transformation of plants and plant cells include *Agrobacterium*-mediated transformation (De Blaere et al., 1987) and particle bombardment technology (Klein et al. 1987; U.S. Patent No. 4,945,050). Whole plants may be regenerated from transgenic cells by methods well known to the skilled artisan (see, for example, Fromm et al., 1990).

"Transformed," "transgenic," and "recombinant" refer to a host organism such as a
30 bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome generally known in the art

and are disclosed in Sambrook et al., 1989. See also Innis et al., 1995 and Gelfand, 1995; and Innis and Gelfand, 1999. Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially mismatched primers, and the like. For example, "transformed," "transformant," and "transgenic" plants or calli have been through the transformation process and contain a foreign gene integrated into their chromosome. The term "untransformed" refers to normal plants that have not been through the transformation process.

"Transiently transformed" refers to cells in which transgenes and foreign DNA have been introduced (for example, by such methods as *Agrobacterium*-mediated transformation or biolistic bombardment), but not selected for stable maintenance.

"Stably transformed" refers to cells that have been selected and regenerated on a selection media following transformation.

"Transient expression" refers to expression in cells in which a virus or a transgene is introduced by viral infection or by such methods as *Agrobacterium*-mediated transformation, electroporation, or biolistic bombardment, but not selected for its stable maintenance.

"Genetically stable" and "heritable" refer to chromosomally-integrated genetic elements that are stably maintained in the plant and stably inherited by progeny through successive generations.

"Primary transformant" and "T0 generation" refer to transgenic plants that are of the same genetic generation as the tissue which is initially transformed (i.e., not having gone through meiosis and fertilization since transformation).

"Secondary transformants" and the "T1, T2, T3, etc. generations" refer to transgenic plants derived from primary transformants through one or more meiotic and fertilization cycles. They may be derived by self-fertilization of primary or secondary transformants or crosses of primary or secondary transformants with other transformed or untransformed plants.

"Wild-type" refers to a virus or organism found in nature without any known mutation.

"Genome" refers to the complete genetic material of an organism.

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, composed of monomers

(nucleotides) containing a sugar, phosphate and a base which is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise

5 indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzner et al., 1991; Ohtsuka et al., 1985; Rossolini et al. 1994). A "nucleic acid fragment" is a fraction of a given

10 nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered

15 nucleotide bases capable of incorporation into DNA or RNA polymers. The terms "nucleic acid" or "nucleic acid sequence" may also be used interchangeably with gene, cDNA, DNA and RNA encoded by a gene.

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. In the context of the present invention, an "isolated" or "purified" DNA

20 molecule or an "isolated" or "purified" polypeptide is a DNA molecule or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or polypeptide may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof,

25 is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the

30 nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of

nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein or polypeptide having less than about 30%, 20%, 10%, 5%, (by dry weight) of contaminating protein. When the protein of the invention, or
5 biologically active portion thereof, is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein of interest chemicals.

The nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant (variant) forms. Such variants will continue to possess the
10 desired activity, i.e., either promoter activity or the activity of the product encoded by the open reading frame of the non-variant nucleotide sequence.

Thus, by "variants" is intended substantially similar sequences. For nucleotide sequences comprising an open reading frame, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical amino acid sequence of the native
15 protein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis and for open reading frames, encode the native protein, as well as those that
20 encode a polypeptide having amino acid substitutions relative to the native protein. Generally, nucleotide sequence variants of the invention will have at least 40, 50, 60, to 70%, e.g., preferably 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, generally at least 80%, e.g., 81%-84%, at least 85%, e.g., 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, to 98% and 99% nucleotide sequence identity to the native (wild type or endogenous)
25 nucleotide sequence.

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or essentially identical amino acid sequences, or where the nucleic acid sequence does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number
30 of functionally identical nucleic acids encode any given polypeptide. For instance the codons CGT, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every

position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations" which are one species of "conservatively modified variations." Every nucleic acid sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

The nucleic acid molecules of the invention can be "optimized" for enhanced expression in plants of interest. See, for example, EPA 035472; WO 91/16432; Perlak et al., 1991; and Murray et al., 1989. In this manner, the open reading frames in genes or gene fragments can be synthesized utilizing plant-preferred codons. See, for example, Campbell and Gowri, 1990 for a discussion of host-preferred codon usage. Thus, the nucleotide sequences can be optimized for expression in any plant. It is recognized that all or any part of the gene sequence may be optimized or synthetic. That is, synthetic or partially optimized sequences may also be used. Variant nucleotide sequences and proteins also encompass sequences and protein derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different coding sequences can be manipulated to create a new polypeptide possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer, 1994; Stemmer, 1994; Crameri et al., 1997; Moore et al., 1997; Zhang et al., 1997; Crameri et al., 1998; and U.S. Patent Nos. 5,605,793 and 5,837,458.

By "variant" polypeptide is intended a polypeptide derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more

sites in the native protein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

Thus, the polypeptides may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the polypeptides can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel, 1985; Kunkel et al., 1987; U. S. Patent No. 4,873,192; Walker and Gaastra, 1983 and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff et al. (1978). Conservative substitutions, such as exchanging one amino acid with another having similar properties, are preferred.

Individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations," where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (I); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). See also, Creighton, 1984. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

"Expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide

sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. The expression of the nucleotide sequence in the expression
5 cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, the promoter can also be specific to a particular tissue or organ or stage of development.

"Vector" is defined to include, inter alia, any plasmid, cosmid, phage or
10 *Agrobacterium* binary vector in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g. autonomous replicating plasmid with an origin of replication).

Specifically included are shuttle vectors by which is meant a DNA vehicle capable,
15 naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g. higher plant, mammalian, yeast or fungal cells).

Preferably the nucleic acid in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell such as a
20 microbial, e.g. bacterial, or plant cell. The vector may be a bi-functional expression vector which functions in multiple hosts. In the case of genomic DNA, this may contain its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell.

"Cloning vectors" typically contain one or a small number of restriction endonuclease
25 recognition sites at which foreign DNA sequences can be inserted in a determinable fashion without loss of essential biological function of the vector, as well as a marker gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Marker genes typically include genes that provide tetracycline resistance, hygromycin resistance or ampicillin resistance.

30 A "transgenic plant" is a plant having one or more plant cells that contain an expression vector.

"Plant tissue" includes differentiated and undifferentiated tissues or plants, including but not limited to roots, stems, shoots, leaves, pollen, seeds, tumor tissue and various forms of cells and culture such as single cells, protoplast, embryos, and callus tissue. The plant tissue may be in plants or in organ, tissue or cell culture.

5 The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

 (a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified
10 sequence; for example, as a segment of a full length cDNA or gene sequence, or the complete cDNA or gene sequence.

 (b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the
15 reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the
20 number of matches.

 Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller, 1988; the local homology algorithm of Smith et al.
25 1981; the homology alignment algorithm of Needleman and Wunsch 1970; the search-for-similarity-method of Pearson and Lipman 1988; the algorithm of Karlin and Altschul, 1990, modified as in Karlin and Altschul, 1993.

 Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but
30 are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST,

FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. 1988; Higgins et al. 1989; Corpet et al. 1988; Huang et al. 1992; and Pearson et al. 1994. The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul et al., 1990, are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. 1997. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al., *supra*. When utilizing BLAST, Gapped BLAST, PSI-
5 BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the
10 BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989). See <http://www.ncbi.nlm.nih.gov>. Alignment may also be performed manually by inspection.

For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default parameters
15 or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid
20 or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical
25 properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to
30 those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for

example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

5 (d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is
10 calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a
15 polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will
20 recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, more preferably at least 80%, 90%, and most preferably at least 95%.

25 Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions (see below). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired
30 degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode

are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, or even more preferably, 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (1970). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

“Stringent hybridization conditions” and “stringent hybridization wash conditions” in the context of nucleic acid hybridization experiments such as Southern and Northern hybridization are sequence dependent, and are different under different environmental parameters. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, 1984; $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$; where M is the molarity of monovalent cations, $\%GC$ is the percentage of guanosine and cytosine nucleotides in the DNA, $\% \text{ form}$ is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m , hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $>90\%$ identity are sought, the T_m can be decreased 10°C . Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point T_m for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point T_m ; moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point T_m ; low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point T_m . Using the equation, hybridization and wash compositions, and desired T , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, 1993. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point T_m for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2X SSC wash at 65°C for 15 minutes

(see, Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1X SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6X SSC at 40°C for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, more preferably about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long robes (e.g., >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2X (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

The following are examples of sets of hybridization/wash conditions that may be used to clone orthologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1%

SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

“DNA shuffling” is a method to introduce mutations or rearrangements, preferably randomly, in a DNA molecule or to generate exchanges of DNA sequences between two or more DNA molecules, preferably randomly. The DNA molecule resulting from DNA shuffling is a shuffled DNA molecule that is a non-naturally occurring DNA molecule derived from at least one template DNA molecule. The shuffled DNA preferably encodes a variant polypeptide modified with respect to the polypeptide encoded by the template DNA, and may have an altered biological activity with respect to the polypeptide encoded by the template DNA.

“Recombinant DNA molecule” is a combination of DNA sequences that are joined together using recombinant DNA technology and procedures used to join together DNA sequences as described, for example, in Sambrook et al., 1989.

The word “plant” refers to any plant, particularly to seed plant, and “plant cell” is a structural and physiological unit of the plant, which comprises a cell wall but may also refer to a protoplast. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, a plant tissue, or a plant organ.

“Significant increase” is an increase that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater.

“Significantly less” means that the decrease is larger than the margin of error inherent in the measurement technique, preferably a decrease by about 2-fold or greater.

25

The transition between vegetative growth and flowering (reproduction) represents a major developmental shift in the plant life cycle. Identification and modulation of flowering time genes can enable the timing of flowering and whole plant architecture to be altered. Altering the timing of flowering has important agricultural implications. The period to produce plant seeds and fruits can be extended when crop plants flower earlier in the growing season. This can enhance the development of plant fruits and seeds by increasing the growing

30

period, or increasing crop yield with multiple crops per season. Rice, for example, requires more than six months to grow and mature in some agricultural areas. Early flowering can provide two rice crops per year. Also, early flowering is a helpful adaptation to some cold climates (Johansen, U. *et al.*, *Science* 5490: 344 (2000)). When crop plants flower later in the
5 growing season, plant crops such as spinach and lettuce tend to bolt. The delay allows them to dedicate less energy for reproduction and more energy towards vegetative growth. Larger and greener leaves and stems are commercially valuable.

The flowering locus C (FLC) gene and relatives of FLC, from *Arabidopsis thaliana* and *Brassica rapa*, have been shown to play a significant role in the delay of flowering. In
10 experiments with *Arabidopsis*, genetically modified plants lacking FLC activity or activity of an FLC relative flowered significantly earlier than wild-type plants. In addition, plants over-expressing the FLC gene or relatives took substantially longer to flower or did not flower at all, and showed large (i.e. 10-fold) increases in biomass.

The genetics of flowering time in the short day plant rice has only recently begun to be
15 understood. Rice and other cereal flowers are distinct from those of *Arabidopsis* in the perianth and inflorescence structure, but are otherwise likely to develop under similar genetic control mechanisms.

Flowering in *Arabidopsis* is initiated by flowering time genes that activate floral meristem identity genes leading to the patterned expression of floral organ identity genes.

20 The rice genome contains clear single copy homologs of the *Arabidopsis* flowering time genes *GI*, *CO*, *LD*, and *FCA* (e.g. SEQ ID NOs: 33, 17, 49, 27).

The LUMINIDEPENDENS (LD) gene is demonstrated (Plant Cell 1994 Jan;6(1):75-83)) to be involved in the timing of flowering in *Arabidopsis*. Mutations in this gene render *Arabidopsis* late flowering and appear to affect light perception. The late-flowering phenotype
25 of the *ld* mutation is partially suppressed by vernalization. Genomic and cDNA clones of the LD gene were characterized. The predicted amino acid sequence of the LD protein contains 953 residues and includes two putative bipartite nuclear localization signals and a glutamine-rich region.

The CO gene from *Arabidopsis* is isolated, and two zinc fingers that show a similar
30 spacing of cysteines, but little direct homology, to members of the GATA1 family were identified in the amino acid sequence. *co* mutations were shown to affect amino acids that are

conserved in both fingers. Some transgenic plants containing extra copies of CO flowered earlier than wild type, suggesting that CO activity is limiting on flowering time. Double mutants were constructed containing co and mutations affecting gibberellic acid responses, meristem identity, or phytochrome function, and their phenotypes suggested a model for the role of CO in promoting flowering (*Cell* 80, 847-857 (1995)).

The rice genome contains a family of nucleotide sequences that encode homologs (SEQ ID NOs: 18; 34; and 50) of the Arabidopsis CO flowering time protein.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity and abundance of which can alter flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NO: 18; 34; and 50, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity and abundance of which can alter flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO: 18; 34; and 50, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity and abundance of which can alter flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 18; 34; and 50, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NO: 17; 33; and 49, or a part thereof which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO: 17; 33; and 49 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

The Arabidopsis FCA genes has been cloned and shown (*Cell* 89, 737-745 (1997)) to encode a protein containing two RNA-binding domains and a WW protein interaction domain. This suggests that FCA functions in the posttranscriptional regulation of transcripts involved in the flowering process. The FCA transcript is alternatively spliced with only one form encoding the entire FCA protein. Plants carrying the FCA gene fused to the strong constitutive 35S promoter flowered earlier, and the ratio and abundance of the different FCA transcripts were altered. Thus, FCA appears to be a component of a posttranscriptional cascade involved in the control of flowering time.

The rice genome encodes at least two homologs of the Arabidopsis FCA protein which are given in SEQ ID NOs: 14, 28 and 52, respectively. One of these homologs is encoded by a nucleotide sequence that maps to rice QTL Hd-3, which controls heading date in rice.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which functions in the posttranscriptional regulation of transcripts involved in the flowering process and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given

in SEQ ID NO: 14, 28 and 52, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

5 The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which functions in the posttranscriptional regulation of transcripts involved in the flowering process and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO: 14, 28 and 52, with any individual number within
10 this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a
15 nucleotide sequence encoding a polypeptide the activity of which functions in the posttranscriptional regulation of transcripts involved in the flowering process and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 14, 28 and 52 or a partial-length polypeptide having substantially the same activity as the full-length
20 polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- 25 a) as given in SEQ ID NO: 13, 27 and 51, or a part thereof which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide...;
- b) having substantial similarity to (a);
- 30 c) capable of hybridizing to (a) or the complement thereof;

- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO: 13, 27 and 51, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- 5 f) which is the reverse complement of (a), (b) or (c).

The process of vernalization (the acceleration of flowering by a cold period of 3-8 weeks at 4°-8°C) is an important characteristic in many plants. The FRIGIDA (FRI) gene ensures that flowering is delayed until after winter so that the plant flowers in the favorable
10 conditions of spring.

The Arabidopsis FRIGIDA gene was isolated (Mol Gen Genet 1994 Jan;242(1):81-9; Science 290, 344-347 (2000)) and can now be used to alter the vernalization requirements of crop plants to benefit farmers.

Different types of Arabidopsis occur in nature. The majority require a long period of
15 cold temperatures to accelerate flowering (winter type). Others flower early without the need for a cold temperature treatment (summer type). It was demonstrated that different forms (alleles) of the same gene -- FRIGIDA -- determined whether the Arabidopsis plants were winter or summer types.

The FRI genes were isolated from different lines of Arabidopsis and their DNA
20 sequences compared. It was found that the genes isolated from the summer Arabidopsis lines all had parts of the gene sequence missing -- the genes were in fact damaged.

It was thus concluded that the activity of the normal FRI gene delays flowering, but cold treatment antagonizes the effect of the gene and so accelerates flowering. If the FRI gene is damaged and not functioning, then it is unable to delay flowering and plants flower early,
25 without the need for a cold treatment.

Two forms of disruption of the FRI gene were found in the early-flowering types, indicating that on at least two different occasions, summer types had arisen by mutation from winter types.

The independent evolution of early-flowering types suggests that there has been strong
30 selection in some environments for ecotypes that do not require vernalization. This analysis is

one of a few examples where the molecular basis of adaptively significant variation has been determined.

A requirement for vernalization has been bred into, and out of, many crops (eg. wheat, oilseed rape and sugarbeet) to produce winter or spring varieties. 95% of all crops grown in the UK are winter types that are sown in the autumn and will not flower until the following spring. These are high yielding, due to growth in the autumn and mild winter periods giving them a head start relative to spring-sown crops.

At least two *FRI* homologs exist in the rice genome in a tandem duplication. By providing rice homologs of the *FRI* gene, manipulation of *FRI* in a range of cereal crops will now allow to develop additional varieties of the winter/spring type.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements and flowering time of plants and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 6 and 8, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements and flowering time of plants and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 6 and 8, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements and flowering time of plants and, optionally, a regulatory region,

which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 6 and 8, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 5 and 7, or a part thereof, which still encodes, a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NOs: 5 and 7 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

Another gene that can potentially used to alter the vernalization requirements of crop plants is the rice homolog of the Arabidopsis VRN1 gene, which is involved in the vernalization response of Arabidopsis plants.

As mentioned previously, flowering in many plant species is accelerated by a long period of cold temperature, known as a vernalization period. Mutagenesis of the late-flowering, vernalization-responsive, Arabidopsis mutant, fca, has yielded five independent mutations (termed vrn mutations) conferring an altered vernalization response. Allelism tests showed that these mutations fall into at least three complementation groups defining three loci named VRN 1, 2 and 3. The vrn1 and vrn2 mutations did not affect the acclimation response as judged by expression of cold-induced transcripts and freezing tolerance assays. vrn1-1 affected the short-day vernalization response of Landsberg erecta and reduced the

vernalization response of other late-flowering *Arabidopsis* mutants. The acceleration of flowering by GA3 was not affected by *vrn1-1*. The VRN 1 locus was mapped to chromosome 3.

5 The rice genome contains at least two nucleotide sequence which show homology to the *Arabidopsis VRN1* gene (SEQ ID NOs: 3, 55).

10 In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements of plants and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 4 and 56, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

15 The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements of plants and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 4 and 56, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

25 The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements of plants and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 4 and 56, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

30

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 3 and 55, or a part thereof, which still encodes , a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NOs: 3 and 55 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

15

The *FT/TFL* gene family in *Arabidopsis* encodes proteins with homology to RAF kinase inhibitor protein (Science 286, 1962 (1999))originally described in humans and contains 6 members. Orthologs from other plants have been shown to be involved in major aspects of whole plant architecture (Pnueli, L., et al., Development, 1998. 125(11): p. 1979-89 ;Bradley, D., et al., Nature, 1996. 379(6568): p. 791-7).

The overall aerial architecture of flowering plants depends on a group of meristematic cells in the shoot apex. It was demonstrated (*Development* 125, 1609-1615 (1998) that the *Arabidopsis* TERMINAL FLOWER 1 gene has a unified effect on the rate of progression of the shoot apex through different developmental phases. In transgenic *Arabidopsis* plants which ectopically express TERMINAL FLOWER 1, both the vegetative and reproductive phases are greatly extended. As a consequence, these plants exhibit dramatic changes in their overall morphology, producing an enlarged vegetative rosette of leaves, followed by a highly branched inflorescence which eventually forms normal flowers. Activity of the floral meristem identity genes *LEAFY* and *APETALA 1* is not directly inhibited by TERMINAL FLOWER 1, but their upregulation is markedly delayed compared to wild-type controls. These phenotypic and molecular effects complement those observed in the *tfl1* mutant, where

25

30

all phases are shortened. The results suggest that TERMINAL FLOWER 1 participates in a common mechanism underlying major shoot apical phase transitions, rather than there being unrelated mechanisms which regulate each specific transition during the life cycle.

The rice genome encodes for at least 17 members of this gene family, one member is most similar to *TFL* (SEQ ID NO 19), and nine are more similar to *FT* (including, for example, SEQ ID NOs: 11, 15, 25, 31, 37 and 61). At least two of the rice genes (SEQ ID NOs: 37 and 61) were shown to map to rice QTL Hd-3, which controls rice heading date. SEQ ID NOs: 37 and 61, which map to another rice QTL that controls heading date, Hd-6, also show homology to the Arabidopsis *TFL* gene.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulating flowering time, and whole plant architecture by modulating plant branching patterns, and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NO: 20, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulating flowering time, and whole plant architecture by modulating plant branching patterns, and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO: 20, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulating flowering time, and whole plant architecture by modulating plant branching patterns, and, optionally, a regulatory region, which induces transcription of the coding region, which

polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 20, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

5 In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NO: 19, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- 10 b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID
- 15 NO: 19 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

Using activation tagging mutagenesis in whole plants, another key gene in regulation of flowering, FT, has been identified and cloned in Arabidopsis. Ectopic expression of FT leads to early flowering regardless of daylength, and causes formation of a terminal flower. Loss-of-function EMS-induced mutant alleles ft-1, 2 and 3 have been described previously as late flowering (Korneef et al, 1991). FT protein is 50% identical to TFL. Missense mutations in TFL cause early flowering and formation of a terminal flower, whereas overexpression of TFL dramatically extends vegetative stage. So it is remarkable that gain-of-function and loss-of-function phenotypes of FT and TFL mutants are completely opposite. From this and previous studies it is concluded that FT is acting in parallel with LEAFY, probably by regulating competence of the apical meristem to respond to LEAFY activity, and TFL protein acts as antagonist of FT in promoting flowering. Temporal regulation of FT expression together with the phenotype of mutants suggest that FT is one of the master genes necessary and sufficient for commitment to flowering in response to day length..

20

25

30

The rice genome encodes for at least 17 members of this FT/TFL gene family, nine of which are most similar to the Arabidopsis *FT* gene (see, for example; SEQ ID NOs: 12, 16, 26, 32, 38 and 62).

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 12, 16, 26, 32, 38 and 62, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 12, 16, 26, 32, 38 and 62, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 12, 16, 26, 32, 38 and 62, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 11, 15, 25, 31, 37 and 61, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NOs: 11, 15, 25, 31, 37 and 61 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

15

In a more recent study (Science 288, 1613-1616 (2000)) it was shown that the *FLOWERING LOCUS T (FT)* gene and also a second gene, called *SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1 or AGL20)* are early target genes of CO and are required for CO to promote flowering. *SOC1* or *AGL20* is an Arabidopsis MADS box transcription factor which is at the intersection of 2 distinct flowering pathways (the "long day" and the "constitutive" pathways) and is likely to be a central regulator of flowering in Arabidopsis. The *SOC1* or *AGL20* and the *FT* genes are also regulated by a second flowering-time pathway that acts independently of CO. Thus, early target genes of CO define common components of distinct flowering-time pathways.

The rice genome encodes for several members of this SOC1/AGL20 gene family, 4 of which (SEQ ID NOs: 9, 47, 57, and 59) are most similar to the Arabidopsis SOC1 (AGL20) gene. At least one of the nucleotide sequences encoding a homolog of the Arabidopsis AGL20 protein (SEQ ID NO: 60) was shown to map to rice QTL Hd-6, controlling heading date in rice by photoperiod response. Also SEQ ID NO: 1, which was shown to map to rice QTL HD-3, has high homology to the Arabidopsis SOC1 (AGL20) gene.

30

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time, and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2, 10, 48, 58, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 2, 10, 48, 58, and 60, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide..

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 2, 10, 48, 58, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 1, 9, 47, 57, and 59, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least

80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;

- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- 5 d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NOs: 1, 9, 47, 57, and 59, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

10

MADS-box genes encode putative transcription factors that have been shown to play important regulatory roles in plants, yeast and animals. In plants, MADS-box genes regulate key events in flower development, such as the specification of flower-meristem and flower-organ identity, as well as the differentiation of each flower organ. In Arabidopsis, more than two dozen MADS-box genes have been cloned. The *APETALA1* (*API*) and *CAULIFLOWER* (*CAL*) MADS-box genes have overlapping functions that specify flower-meristem identity. *cal* single mutant plants are indistinguishable from wild-type plants, however, mutations in the *CAL* gene dramatically enhance the phenotype of *ap1* single mutants (Bowman, J.L., et al, *Development* 119, 721-743 (1993)). The functional redundancy observed between the *CAL* and *API* genes has been attributed to their closely related sequences and expression patterns, suggesting that they regulate many of the same targets.

20

Homologs of *CAL*, a MADS domain gene, could be identified among the large rice MADS domain gene family including those given in SEQ ID NOs: 1, 47 and 57. At least one of the rice homologs (SEQ ID NO: 1) was shown to map to rice QTL Hd-3, which controls rice heading date. Also SEQ ID NO: 59, which maps to another rice controlling rice heading date, QTL Hd-6, shows homology to the *CAL* gene.

25

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in specifying flower-meristem identity and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2, 48, 58, and 60, or a

30

partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a
5 nucleotide sequence encoding a polypeptide the activity of which is involved in specifying flower-meristem identity and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 2, 48, 58, and 60, with any individual number within this range of between 70% and 99% also
10 being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in specifying
15 flower-meristem identity and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 2, 48, 58, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-
20 length polypeptide.4858

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 1, 47, 57, and 59, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity
25 as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;

- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NOs: 1, 47, 57, and 59 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

The phenotype caused by mutations that affect the timing of flowering in *Arabidopsis thaliana* has been most extensively analyzed in the Landsberg erecta (Ler) genetic background. In Ler, the late-flowering phenotype of FRIGIDA and mutations in LUMINIDEPENDENS is suppressed by the Ler allele of FLC. The interactions of nine mutations conferring late flowering with the FLC allele of the Columbia ecotype (FLC-Col), which does not suppress late flowering, were examined. The effect on flowering time of combining six of the mutations with FLC-Col was shown to be additive; plants containing FLC-Col with fd, gi, fwa, fha, ft, and fe flowered slightly later than plants containing these mutations with the Ler allele of FLC. In contrast, a synergistic effect was observed between FLC-Col and three mutations; fca, fpa, and fve plants became extremely late flowering when combined with FLC-Col. Maximum delay in flowering for the majority of the mutant strains required FLC-Col in a homozygous state, although for fpa and fe a single copy of FLC-Col allowed maximum lateness. In addition, the fd and fe mutations became more dominant in the presence of FLC-Col (Mol Gen Genet 1996 Apr 24;251(1):69-74).

Homologs of *FLC*, a MADS domain gene, could be identified among the large rice MADS domain gene family including those given in SEQ ID NOs: 9; 29, 47; and 57.

Also SEQ ID NOs: 1 and 59, which were shown to map to rice QTL Hd-3 and Hd-6, respectively, which controls heading date in rice, have high homology to the Arabidopsis FLC gene.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is capable of modulating flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2, 10; 30, 48; 58 and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least

50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is capable of modulating flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 2, 10; 30, 48; 58 and 60, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is capable of modulating flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 2, 10; 30, 48; 58 and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 1, 9; 29, 47; 57 and 59, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 1, 9; 29, 47; 57 and 59, or the complement thereof;
- e) complementary to (a), (b) or (c); and

f) which is the reverse complement of (a), (b) or (c).

Genes from other cereals that have been shown to affect flowering time, including *ZmId1* in maize (Colasanti, J., Z. Yuan, and V. Sundaresan, *Cell*, 1998. **93**(4): p. 593-603) and *Rht-B1/Zmd8* in wheat and maize [J. Peng *et al.*, *Nature* **400**, 256-61 (Jul 15, 1999); J. M. Thornsberry *et al.*, *Nat Genet* **28**, 286-9 (Jul, 2001).] are also represented by clear homologs in the rice genome.

Flowering in plants is a consequence of the transition of the shoot apex from vegetative to reproductive growth in response to environmental and internal signals. The *indeterminate1* gene (*id1*) was shown to control the transition to flowering in maize (Colasanti, J., Z. Yuan, and V. Sundaresan, *Cell*, 1998. **93**(4): p. 593-603). By cloning the *id1* gene it was demonstrated that it encodes a protein with zinc finger motifs, suggesting that the *id1* gene product functions as a transcriptional regulator of the floral transition. *id1* mRNA expression studies and analyses of transposon-induced chimeric plants indicate that *id1* acts non-cell-autonomously to regulate the production of a transmissible signal in the leaf that elicits the transformation of the shoot apex to reproductive development.

The rice genome encodes for at least 1 members of this gene family (SEQ ID NO: 21).

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NO:22, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO: 22, with any individual number within this range of

between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

5 The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 22, or a
10 partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- 15 a) as given in SEQ ID NO: 21, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.,
- b) having substantial similarity to (a);
- 20 c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NO: 21 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- 25 f) which is the reverse complement of (a), (b) or (c).

The dominant late flowering Arabidopsis mutant *lhy* (*late elongated hypocotyl* (*lhy*)) was shown to disrupt circadian clock regulation of gene expression and leaf movements and caused flowering to occur independently of photoperiod. *LHY* was shown to encode a MYB
30 DNA-binding protein. In wild-type plants, the *LHY* mRNA showed a circadian pattern of expression with a peak around dawn but in the mutant was expressed constantly at high levels.

Increased *LHY* expression from a transgene caused the endogenous gene to be expressed at a constant level, suggesting that *LHY* was part of a feedback circuit that regulated its own expression. Thus, constant expression of *LHY* disrupts several distinct circadian rhythms in *Arabidopsis*, and *LHY* may be closely associated with the central oscillator of the circadian clock (Schaffer et al, Cell 93, 1219-1229 (1998)).

The rice genome encodes for at least 2 members of this gene family (SEQ ID NOs: 43 and 45).

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 44 and 46, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 44 and 46, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 44 and 46, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 43 and 45, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 43 and 45 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

FPA is a gene that regulates flowering time in Arabidopsis via a pathway that is independent of daylength (the autonomous pathway). Mutations in *FPA* result in extremely delayed flowering. *FPA* was identified by means of positional cloning. The predicted *FPA* protein contains three RNA recognition motifs in the N-terminal region. *FPA* is expressed most strongly in developing tissues, similar to the expression of *FCA* and *LUMINIDEPENDENS*, two components of the autonomous pathway discussed herein previously. Overexpression of *FPA* in Arabidopsis causes early flowering in noninductive short days and creates plants that exhibit a more day-neutral flowering behavior. (Schomburg et al, The Plant Cell, Vol. 13, 1427-1436, June 2001)

The rice genome contains at least two homologs of the Arabidopsis *FPA* gene, which map to rice QTLs (Hd-3 and Hd-6, respectively) that are regulate heading time in rice.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 14 and 52, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least

50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 14 and 52, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 14 and 52, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 13 and 51, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 13 and 51 or the complement thereof;
- e) complementary to (a), (b) or (c); and

f) which is the reverse complement of (a), (b) or (c).

Further genes placed in the autonomous pathway involved in the regulation of flowering time in plants is *FVE*, which encodes a WD40 repeat protein with similarity to proteins involved in transcription repression complexes and *FY*, which interacts with this WW domain in FCA.

The rice genome contains at least two nucleotide sequence which shows homology to either *FVE* or *FY*, or to both. *FVE* and *FY* map to rice QTLs Hd-3 and Hd-6, respectively, which regulates heading time in rice.

In a further specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 24 and 36, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 24 and 36, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 24 and 36, or a partial-length polypeptide having substantially

the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- 5 a) as given in SEQ ID NOs: 23 and 35, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- 10 b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 23 and 35 or the complement thereof;
- 15 e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

FWA was identified by loss-of-function mutations in normally flowering revertants of the *fwa* mutant, and it encodes a homeodomain-containing transcription factor. The DNA
20 sequence of wild-type and *fwa* mutant alleles was identical in the genomic region of *FWA*.

The transition to flowering in *Arabidopsis thaliana* is delayed in *fwa* mutant plants. Furthermore, the *FWA* gene is ectopically expressed in *fwa* mutants and silenced in mature wild-type plants. This silencing is associated with extensive methylation of two direct repeats in the 5' region of the gene. (Soppe et al, Molecular Cell, Vol. 6, 791–802, October, 2000

25 The rice genome contains at least 5 nucleotide sequences that show homology to the the *Arabidopsis FWA* gene.

In a further specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces
30 transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 40, 42, 54, 64 and 66, or a

partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a
5 nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 40, 42, 54, 64 and 66, with any individual number within this range of between 70% and 99% also being part of
10 the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of
15 flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 40, 42, 54, 64 and 66, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-
20 length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 39, 41, 53, 63 and 65, or a part thereof, which still encodes a partial-length polypeptide having substantially the same
25 activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;

- d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 39, 41, 53, 63 and 65, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- 5 f) which is the reverse complement of (a), (b) or (c).

Many important crop traits including those relating to flower development and timing are quantitative traits and result from the combined interactions of several genes. These genes reside at different loci in the genome, often on different chromosomes, and generally exhibit
10 multiple alleles at each locus. Developing markers, tools, and methods to identify and isolate the QTL involved in a trait, enables marker-assisted breeding to enhance desirable traits or suppress undesirable traits.

The ability to align regions of heterologous cereal chromosomes to the rice genome provides the opportunity to connect mapped cereal trait genes (quantitative trait loci or QTL) to the rice genome, and assign candidate genes to the traits that these genes influence. There
15 are approximately 2,000 mapped cereal QTL available in the public domain, and many more in private databases. The ability to use the rice genome framework to assign candidate genes to maize, wheat, barley, or other cereal QTL allows association of genes with their molecular functions. Mapped cereal QTL can be placed on the rice genome map. Positioning of
20 molecular markers within the genome will further facilitate mapping and identification of trait-controlling genes.

A comparison of mapped rice QTL controlling flowering time with the positions of rice flowering time gene homologs make it possible to identify and/or confirm functions for many of these rice genes.

25 One such set of QTL from rice is known to control heading of rice. Heading date is a critical trait for adaptation to different cultivation areas and cropping seasons and therefore a major objective in rice breeding programs. Heading in rice is basically determined by two factors, duration of the basic vegetative growth (BVG) and photoperiod-sensitivity (PS). Several genes are known to be involved in controlling these two factors .

30 Yano et al (Theoretical and Applied Genetics (1997) 95: 1025-1032) reported five QTL, designated Hd1-Hd5, which control rice heading date in an F2 population from a cross

between a *japonica* variety, Nipponbare, and an *indica* variety, Kasalath. Scan analysis revealed that two QTL, Hd-1 and Hd-2, had large effects, while three others, Hd-3, Hd-4 and Hd-5, were found to cause minor effects.

Two additional QTL were found by Yamamoto et al (Genetics (February 2000) 154: 885-891). A backcrossed population (BC₄F₂) derived from a cross between a *japonica* rice variety, Nipponbare, as the recurrent parent and an *indica* rice variety, Kasalath, as the donor parent showed a long-range variation in days to heading. Quantitative trait loci (QTL) analysis revealed that two QTL, one on chromosome 3, designated *Hd6*, and another on chromosome 2, designated *Hd7*, were involved in this variation; and *Hd6* was precisely mapped as a single Mendelian factor by using progeny testing (BC₄F₃). The nearly isogenic line with QTL (QTL-NIL) that carries the chromosomal segment from Kasalath for the *Hd6* region in Nipponbare's genetic background was developed by marker-assisted selection. In a day-length treatment test, the QTL-NIL for *Hd6* prominently increased days to heading under a 13.5-hr day length compared with the recurrent parent, Nipponbare, suggesting that *Hd6* controls photoperiod sensitivity. QTL analysis of the F₂ population derived from a cross between the QTL-NILs revealed existence of an epistatic interaction between *Hd2*, which is one of the photoperiod sensitivity genes detected in a previous analysis, and *Hd6*. The day-length treatment tests of these QTL-NILs, including the line introgressing both *Hd2* and *Hd6*, also indicated an epistatic interaction for photoperiod sensitivity between them.

The present invention provides nucleotide sequences as given in the Sequence Listing with the SEQ ID NOs being identified in column C of table 2 corresponding to genes which are located between markers C844 and S1205S on rice chromosome 7 that flank the Hd-2 QTL controlling flowering time in rice (see TAG 95:1025-1032 (1997)) and these are candidate genes controlling this QTL.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in the Sequence Listing with the SEQ ID NOs of the corresponding nucleotide sequences being identified in column C of table 2; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more

preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide given in the Sequence Listing with the SEQ ID NOs of the corresponding nucleotide sequences being identified in column C of table 2, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide as given in the Sequence Listing with the SEQ ID NOs of the corresponding nucleotide sequences being identified in column C of table 2; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- a) as given in the Sequence Listing with the SEQ ID NOs being identified in column C of table 2 or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide,
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;

- d) capable of hybridizing to a nucleic acid as given in the Sequence Listing with the SEQ ID NOs being identified in column C of table 2 comprising 50 to 200 or more consecutive nucleotides or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

The invention further provides nucleotide sequences as given in the Sequence Listing with the SEQ ID NOs being identified in column D of table 2 corresponding to genes which are located between markers R1952 and S1520 on rice chromosome 6 that flank the Hd-3 QTL controlling flowering time in rice (see TAG 95:1025-1032 (1997)) and these are candidate genes controlling this QTL. A subset of these genes being represented by SEQ ID NOs: 41, 13, 23, 37 and 61 are homologous to known Arabidopsis genes involved in control of flowering time as discussed hereinbefore.

In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in the Sequence Listing with the SEQ ID NOs of the corresponding nucleic acid molecules being identified in column D of table 2, but especially a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 42, 14, 24, 38 and 62; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide as given in the Sequence Listing with the SEQ ID NOs of the corresponding nucleic acid molecules being identified in column D of table 2, but especially to a polypeptide as given in SEQ ID NOs: 42, 14, 24, 38 and 62, with any individual number within this range of between 70% and 99% also being part of the

invention; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a
 5 nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide as given in the Sequence Listing with the SEQ ID NOs of the corresponding nucleic acid molecules being identified in column D of table 2, but especially a polypeptide of SEQ ID
 10 NOs: 42, 14, 24, 38 and 62; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- 15 a) as given in the Sequence Listing with the SEQ ID NOs being identified in column D of table 2; but especially in SEQ ID NOs: 41, 13, 23, 37 and 61 or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to
 20 95% the activity of the full-length polypeptide,
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule as given in the Sequence Listing with the SEQ ID NOs being identified in column D of table 2, but
 25 especially in SEQ ID NOs: 41, 13, 23, 37 and 61 and , comprising 50 to 200 or more consecutive nucleotides., or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

The invention also provides nucleotide sequences as given in the Sequence Listing
 30 with the SEQ ID NOs being identified in column E of table 2 corresponding to genes located between markers R2462 and R2966 on rice chromosome 3 that flank the Hd-6 QTL

controlling flowering time in rice (see Genetics 154:885-891 (2000)) and these are candidate genes controlling this QTL. A subset of these genes being represented by SEQ ID NOs: 35, 51, and 59 are homologous to known Arabidopsis genes involved in control of flowering time as discussed hereinbefore.

5 In a specific embodiment the present invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in the Sequence Listing with the SEQ ID NOs of the corresponding
10 nucleic acid molecule being identified in column E of table 2, but especially to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 36, 52, and 60; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

15 The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide as given in the Sequence Listing with the
20 SEQ ID NOs of the corresponding nucleic acid molecule being identified in column E of table 2, but especially to a polypeptide as given in SEQ ID NOs: 36, 52, and 60, with any individual number within this range of between 70% and 99% also being part of the invention; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the
25 activity of the full-length polypeptide.

The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide as
30 given in the Sequence Listing with the SEQ ID NOs of the corresponding nucleic acid molecule being identified in column E of table 2, but especially a polypeptide of SEQ ID

NOs: 36, 52, and 60; or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

In particular, the invention relates to an isolated nucleic acid molecule comprising a
5 nucleotide sequence

- a) as given in the Sequence Listing with the SEQ ID NOs being identified in column E of table 2, but especially in SEQ ID NOs: 35, 51, and 59 or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least
10 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide,
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule as given in the Sequence
15 Listing with the SEQ ID NOs being identified in column E of table 2, but especially in SEQ ID NOs: 35, 51, and 59 comprising 50 to 200 or more consecutive nucleotides, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

20

A further embodiment of the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a
25 polypeptide as given in SEQ ID NOs: 4, 6, 8, 10, 12, 16, 18, 20, 22, 26, 28, 30, 32, 34, 38, 40, 42, 44, 46, 48, 50, 54, 56, 58, 62, 64, and 66, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

30 The invention further relates to an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in

rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide as given in SEQ ID NOs: 4, 6, 8, 10, 12, 16, 18, 20, 22, 26, 28, 30, 32, 34, 38, 40, 42, 44, 46, 48, 50, 54, 56, 58, 62, 64, and 66, or a
5 partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide, with any individual number within this range of between 70% and 99% also being part of the invention.

The invention further relates to an isolated nucleic acid molecule comprising a
10 nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide as given SEQ ID NOs: 4, 6, 8, 10, 12, 16, 18, 20, 22, 26, 28, 30, 32, 34, 38, 40, 42, 44, 46, 48, 50, 54, 56, 58, 62, 64, and 66, or a partial-length polypeptide having substantially the same
15 activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide

In particular, the invention relates to an isolated nucleic acid molecule comprising a nucleotide sequence

- 20 g) as given in SEQ ID NOs 3, 5, 7, 9, 11, 15, 17, 376478, 21, 25, 27, 29, 31, 33, 37, 745356, 39, 43, 45, 47, 49, 53, 55, 57, 61, 63, and 65 or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to
25 95% the activity of the full-length polypeptide,
- a) having substantial similarity to (a);
 - b) capable of hybridizing to (a) or the complement thereof;
 - c) capable of hybridizing to a nucleic acid molecule as given in SEQ ID NOs
30 3, 5, 7, 9, 11, 15, 17, 376478, 21, 25, 27, 29, 31, 33, 37, 39, 41, 43, 45, 47, 49, 53, 55, 57, 61, 63, and 65 comprising 50 to 200 or more consecutive nucleotides or the complement thereof;

- d) complementary to (a), (b) or (c); and
- e) which is the reverse complement of (a), (b) or (c).

5 A further subset of nucleic acid molecules according to the invention comprises at least one polynucleotide which contains one or more Simple Sequence Repeats (SSRs) in the coding region.

10 Microsatellites are tandem repeats of short units of DNA that occur with high frequency throughout the genomes of many organisms. SSRs are more or less evenly distributed over the plant genome with a frequency of 1 SSR every 9 Kb on the average. SSRs are present in coding as well as non-coding regions. Their length is dependent on the number of repetitions of the simple di-, tri-, or tetranucleotide. Thus, the length of an SSR may vary from species to species or from gene to gene. Microsatellite loci have a high degree of variability that is due to a high rate of mutations that alter microsatellite length. The primary mutational mechanism leading to changes in microsatellite length is polymerase template slippage. During replication of a repetitive region, DNA strands may dissociate, and reassociate incorrectly. Renewed replication in this misaligned state leads to insertion or deletion of repeat units, thus altering allele length. There are several major reasons for interest in microsatellite loci.

- The abundance and high level of allelic variation at microsatellite loci in the genomes of many organisms has made them popular genetic markers.
- 20 • Genetic distance measures based on microsatellites can be used to answer questions concerning population structure and divergence.

This size change can be captured by PCR with primers flanking the SSRs. The SSR becomes a marker to distinguish any two individual species with different copy number of that particular SSR. SSRs are abundant, highly variable, and easy to assay. Using a number of SSR markers to genotype a segregating population, as described herein, one can generate a genetic map, map a trait of interest, and clone the underlying gene by map-based cloning.

25 In a specific embodiment, the invention provides an isolated nucleic acid molecule including a repeat region, the repeat region having at least about 3 consecutive dinucleotide, trinucleotide, or tetranucleotide repeat units, the nucleic acid further including either a first flanking region of at least about 4 nucleotides at a first end of the repeat region, or a second flanking region of at least about 4 nucleotides at a second end of the repeat region, or both a

first and a second flanking region. The SSR can be either some or all of the repeat units, with one or both of the flanking regions. The flanking region can be as little as one base adjacent to the repeat units, and can be typically several bases in length. The size of the flanking region is preferably in a range of between about 4 to 250 nucleotides, with each individual number of
5 nucleotides within this range being also part of the invention, or can even be longer than that. Preferred is a flanking region where the nucleotide sequence is between about 5 to about 9, preferably between about 10 to about 30, and more preferably between about 31 to 50 nucleotides in length.

The nucleic acid molecule according to the invention can be adapted for a use such as,
10 for example, use as a chromosomal marker to identify the location of a corresponding or complementary nucleic acid on a native or artificial chromosome; use as a marker for RFLP analysis; use as a marker for quantitative trait linked breeding; use as a marker for marker-assisted breeding; use as a bait sequence in a two-hybrid system to identify sequence encoding polypeptides interacting with the polypeptide encoded by the bait sequence; use as a
15 diagnostic tool for genotyping or identifying an individual or population of individuals; use for genetic analysis to identify boundaries of genes or exons; and the like.

The invention further provides a method of modifying the frequency of a gene in a plant population, including the steps of: identifying an SSR within a coding region of a gene; screening a plurality of plants using the SSR as a marker to determine the presence or absence
20 of the gene in an individual plant; selecting at least one individual plant for breeding based on the presence or absence of the gene; and breeding at least one plant thus selected to produce a population of plants having a modified frequency of the gene. Further provided is a method of plant breeding to select for or against a trait of interest, including the steps of: identifying the trait of interest; identifying at least one SSR that can be used as a marker for the trait;
25 screening at least one plant for the presence of the at least one SSR; selecting at least one plant based on presence or absence of the at least one SSR; breeding at least one plant thus selected to produce a population of plants having a modified frequency of the at least one SSR; and screening at least one plant of the population for the presence or absence of the trait. The trait of interest can be, for example, disease resistance, drought resistance, weather resistance, early
30 ripening, pest resistance, salt resistance, cold tolerance, yield, flavor, nutritional value, texture, ease of harvesting, ease of processing, and the like.

In other aspects, the invention provides a method of mapping a genome of a plant species, including screening nucleic acid sequences from the plant species to identify an SSR therein, the SSR having a first and a second flanking region; and identifying a first position in the genome, the first position corresponding to the SSR, the first position being relative to a second position in the genome. In preferred embodiments, the second position can correspond to a location on a known natural or artificial chromosome, such as a location of an SSR, or a location of a gene. The second position can also correspond to a map location of a trait.

The invention also provides a genetic map of a plant, the map including map positions, wherein the map positions can include positions of a plurality of SSRs, such as at least about 100 SSRs; at least about 500 SSRs; at least about 1000 SSRs; or more. The plurality of SSRs can include at least about 10 SSRs, each SSR having a repeat region and at least a first or a second flanking region, or both a first and a second flanking region.

The nucleotide sequences according to the invention can also be used to generate markers, including single-sequence repeats (SSRs) and microsatellite markers for QTL to assist marker-assisted breeding. The nucleotide sequences of the invention can further be used to identify QTL and isolate alleles as described by Li *et al.* in a study of QTL involved in resistance to a pathogen of rice. (Li *et al.*, *Mol Gen Genet* 261:58 (1999)). In addition to isolating QTL alleles in rice, other cereals, and other monocot and dicot crop species, the nucleotide sequences of the invention can also be used to isolate alleles from the corresponding QTL(s) of wild relatives. Transgenic plants having various combinations of QTL alleles can then be created and the effects of the combinations measured. Once an ideal allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed conventional breeding programs. (Flowers *et al.*, *J Exp Bot* 51:99 (2000); Tanksley and McCouch, *Science* 277:1063 (1997)).

Nucleotides of the present invention can also be used for nucleic acid analysis in a variety of ways. One preferred use involves microbead technology and capture probe hybridization. This technology is described in copending U.S. Patent Application No. 09/565,214, entitled "Novel Assay for Nucleic Acid Analysis," which was filed on May 4, 2000, and which is hereby incorporated by reference in its entirety.

Obtaining sufficient levels of transgene expression in the appropriate plant tissues is an important aspect in the production of genetically engineered crops. Expression of

heterologous DNA sequences in a plant host is dependent upon the presence of an operably linked promoter that is functional within the plant host. Choice of the promoter sequence will determine when and where within the organism the heterologous DNA sequence is expressed.

According to one embodiment, the present invention is directed to a nucleic acid molecule comprising a nucleotide sequence isolated or obtained from any plant which encodes a polypeptide having at least 70% amino acid sequence identity to a polypeptide encoded by a gene comprising any one of SEQ ID NOs provided in the Sequence Listing.

Based on the *Oryza* nucleic acid sequences of the present invention as given in the SEQ ID NOs provided in the Sequence Listing, orthologs may be identified or isolated from the genome of any desired organism, preferably from another plant, according to well known techniques based on their sequence similarity to the *Oryza* nucleic acid sequences, e.g., hybridization, PCR or computer generated sequence comparisons. For example, all or a portion of a particular *Oryza* nucleic acid sequence is used as a probe that selectively hybridizes to other gene sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen source organism. Further, suitable genomic and cDNA libraries may be prepared from any cell or tissue of an organism. Such techniques include hybridization screening of plated DNA libraries (either plaques or colonies; see, e.g., Sambrook et al., 1989) and amplification by PCR using oligonucleotide primers preferably corresponding to sequence domains conserved among related polypeptide or subsequences of the nucleotide sequences provided herein (see, e.g., Innis et al., 1990). These methods are particularly well suited to the isolation of gene sequences from organisms closely related to the organism from which the probe sequence is derived. The application of these methods using the *Oryza* sequences as probes is well suited for the isolation of gene sequences from any source organism, preferably other plant species. In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any plant of interest. Methods for designing PCR primers and PCR cloning are generally known in the art.

In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA

libraries) from a chosen organism. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labeled with a detectable group such as ^{32}P , or any other detectable marker. Thus, for example, probes for hybridization can be made by labeling synthetic oligonucleotides based on the sequence of the invention. Methods for preparation of probes for hybridization and for construction of cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook et al. (1989). In general, sequences that hybridize to the sequences disclosed herein will have at least 40% to 50%, about 60% to 70% and even about 80% 85%, 90%, 95% to 98% or more identity with the disclosed sequences. That is, the sequence similarity of sequences may range, sharing at least about 40% to 50%, about 60% to 70%, and even about 80%, 85%, 90%, 95% to 98% sequence similarity, with each individual number within the ranges given above also being part of the invention.

The nucleic acid molecules of the invention can also be identified by, for example, a search of known databases for genes encoding polypeptides having a specified amino acid sequence identity or DNA having a specified nucleotide sequence identity. Methods of alignment of sequences for comparison are well known in the art and are described hereinabove.

The present invention further provides a composition, an expression cassette or a recombinant vector containing the nucleic acid molecule of the invention, and host cells comprising the expression cassette or vector, e.g., comprising a plasmid. In particular, the present invention provides an expression cassette or a recombinant vector comprising a promoter as disclosed herein linked to a nucleic acid segment which, when present in a plant, plant cell or plant tissue, results in transcription of the linked nucleic acid segment. The invention also provides an expression cassette or a recombinant vector comprising a plant nucleotide sequence comprising an open reading frame of the invention such as, for example, those given in the SEQ ID NOs provided in the Sequence Listing which, when present in a plant, plant cell or plant tissue, results in expression of the protein product encoded by the open reading frame. Further, the invention provides isolated polypeptides encoded by any one of the open reading frames given in the SEQ ID NOs provided in the Sequence Listing, a fragment thereof which encodes a polypeptide which has substantially the same activity as the

corresponding polypeptide encoded by an ORF listed in the SEQ ID NOs provided in the Sequences Listing, or the orthologs thereof.

Virtually any DNA composition may be used for delivery to recipient plant cells, e.g., monocotyledonous cells, to ultimately produce fertile transgenic plants in accordance with the present invention. For example, DNA segments or fragments in the form of vectors and plasmids, or linear DNA segments or fragments, in some instances containing only the DNA element to be expressed in the plant, and the like, may be employed. The construction of vectors which may be employed in conjunction with the present invention will be known to those of skill of the art in light of the present disclosure (see, e.g., Sambrook et al., 1989; Gelvin et al., 1990).

It is one of the objects of the present invention to provide recombinant DNA molecules comprising a nucleotide sequence which directs transcription according to the invention operably linked to a nucleic acid segment or sequence of interest.

In one specific embodiment, the invention provides an expression cassette or vector containing an isolated nucleic acid molecule having a nucleotide sequence that directs transsscription of a linked nucleic acid segment of interest in a cell, which nucleotide sequence is from a gene which encodes a polypeptide having at least 70% identity to an *Oryza* polypeptide encoded by a gene having one of the promoters listed in the SEQ ID NOs provided in the Sequence Listing, and which nucleotide sequence is optionally operably linked to other suitable regulatory sequences, e.g., a transcription terminator sequence, operator, repressor binding site, transcription factor binding site and/or an enhancer.

Furthermore, the nucleic acid segment of interest can, for example, code for a ribosomal RNA, an antisense RNA or any other type of RNA that is not translated into protein. In another preferred embodiment of the invention, the nucleic acid segment of interest is translated into a protein product. The nucleotide sequence which directs transcription and/or the nucleic acid segment may be of homologous or heterologous origin with respect to the plant to be transformed. A recombinant DNA molecule useful for introduction into plant cells includes that which has been derived or isolated from any source, that may be subsequently characterized as to structure, size and/or function, chemically altered, and later introduced into plants. An example of a nucleotide sequence or segment of interest "derived" from a source, would be a nucleotide sequence or segment that is identified as a useful fragment within a

given organism, and which is then chemically synthesized in essentially pure form. An example of such a nucleotide sequence or segment of interest "isolated" from a source, would be nucleotide sequence or segment that is excised or removed from said source by chemical means, e.g., by the use of restriction endonucleases, so that it can be further manipulated, e.g., amplified, for use in the invention, by the methodology of genetic engineering. Such a nucleotide sequence or segment is commonly referred to as "recombinant."

Therefore a useful nucleotide sequence, segment or fragment of interest includes completely synthetic DNA, semi-synthetic DNA, DNA isolated from biological sources, and DNA derived from introduced RNA. Generally, the introduced DNA is not originally resident in the plant genotype which is the recipient of the DNA, but it is within the scope of the invention to isolate a gene from a given plant genotype, and to subsequently introduce multiple copies of the gene into the same genotype, e.g., to enhance production of a given gene product as given in the SEQ ID NOs provided in the Sequence Listing.

The introduced recombinant DNA molecule includes but is not limited to, DNA from plant genes, and non-plant genes such as those from bacteria, yeasts, animals or viruses. The introduced DNA can include modified genes, portions of genes, or chimeric genes, including genes from the same or different genotype. The term "chimeric gene" or "chimeric DNA" is defined as a gene or DNA sequence or segment comprising at least two DNA sequences or segments from species which do not combine DNA under natural conditions, or which DNA sequences or segments are positioned or linked in a manner which does not normally occur in the native genome of untransformed plant.

The introduced recombinant DNA molecule used for transformation herein may be circular or linear, double-stranded or single-stranded. Generally, the DNA is in the form of chimeric DNA, such as plasmid DNA, that can also contain coding regions flanked by regulatory sequences which promote the expression of the recombinant DNA present in the resultant plant.

Generally, the introduced recombinant DNA molecule will be relatively small, i.e., less than about 30 kb to minimize any susceptibility to physical, chemical, or enzymatic degradation which is known to increase as the size of the nucleotide molecule increases. As noted above, the number of proteins, RNA transcripts or mixtures thereof which is introduced

into the plant genome is preferably preselected and defined, e.g., from one to about 5-10 such products of the introduced DNA may be formed.

This expression cassette or vector may be contained in a host cell. The expression cassette or vector may augment the genome of a transformed plant or may be maintained
5 extrachromosomally. The expression cassette may be operatively linked to a structural gene, the open reading frame thereof, or a portion thereof. The expression cassette may further comprise a Ti plasmid and be contained in an *Agrobacterium tumefaciens* cell; it may be carried on a microparticle, wherein the microparticle is suitable for ballistic transformation of a plant cell; or it may be contained in a plant cell or protoplast. Further, the expression
10 cassette or vector can be contained in a transformed plant or cells thereof, and the plant may be a dicot or a monocot. In particular, the plant may be a cereal plant.

Two principal methods for the control of expression are known, viz.: overexpression and underexpression. Overexpression can be achieved by insertion of one or more than one extra copy of the selected gene. It is, however, not unknown for plants or their progeny,
15 originally transformed with one or more than one extra copy of a nucleotide sequence, to exhibit the effects of underexpression as well as overexpression. For underexpression there are two principle methods which are commonly referred to in the art as “antisense downregulation” and “sense downregulation” (sense downregulation is also referred to as “cosuppression”). Generically these processes are referred to as “gene silencing”. Both of
20 these methods lead to an inhibition of expression of the target gene.

The invention hence also provides sense and anti-sense nucleic acid molecules corresponding to the open reading frames identified in the SEQ ID NOs provided in the Sequence Listing as well as their orthologs.

The genes and open reading frames according to the present invention which are
25 substantially similar to a nucleotide sequence encoding a polypeptide as identified in the SEQ ID NOs provided in the Sequence Listing including any corresponding anti-sense constructs can be operably linked to any promoter that is functional within the plant host including the promoter sequences according to the invention or mutants thereof.

Further promoters which are useful for plant transgene expression include those that
30 are inducible, viral, synthetic, constitutive (Odell et al., 1985), temporally regulated, spatially regulated, tissue-specific, and spatio-temporally regulated.

Where expression in specific tissues or organs is desired, tissue-specific promoters may be used. In contrast, where gene expression in response to a stimulus is desired, inducible promoters are the regulatory elements of choice. Where continuous expression is desired throughout the cells of a plant, constitutive promoters are utilized. Additional regulatory sequences upstream and/or downstream from the core promoter sequence may be included in expression constructs of transformation vectors to bring about varying levels of expression of heterologous nucleotide sequences in a transgenic plant.

A variety of 5N and 3N transcriptional regulatory sequences are available for use in the present invention. Transcriptional terminators are responsible for the termination of transcription and correct mRNA polyadenylation. The 3N nontranslated regulatory DNA sequence preferably includes from about 50 to about 1,000, more preferably about 100 to about 1,000, nucleotide base pairs and contains plant transcriptional and translational termination sequences. Appropriate transcriptional terminators and those which are known to function in plants include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator, the pea *rbcS* E9 terminator, the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*, and the 3N end of the protease inhibitor I or II genes from potato or tomato, although other 3N elements known to those of skill in the art can also be employed. Alternatively, one also could use a gamma coixin, oleosin 3 or other terminator from the genus *Coix*.

Preferred 3N elements include those from the nopaline synthase gene of *Agrobacterium tumefaciens* (Bevan et al., 1983), the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*, and the 3' end of the protease inhibitor I or II genes from potato or tomato.

As the DNA sequence between the transcription initiation site and the start of the coding sequence, i.e., the untranslated leader sequence, can influence gene expression, one may also wish to employ a particular leader sequence. Preferred leader sequences are contemplated to include those which include sequences predicted to direct optimum expression of the attached gene, i.e., to include a preferred consensus leader sequence which may increase or maintain mRNA stability and prevent inappropriate initiation of translation.

The choice of such sequences will be known to those of skill in the art in light of the present

disclosure. Sequences that are derived from genes that are highly expressed in plants will be most preferred.

Other sequences that have been found to enhance gene expression in transgenic plants include intron sequences (e.g., from *Adh1*, *bronze1*, *actin1*, *actin 2* (WO 00/760067), or the sucrose synthase intron) and viral leader sequences (e.g., from TMV, MCMV and AMV). For example, a number of non-translated leader sequences derived from viruses are known to enhance expression. Specifically, leader sequences from Tobacco Mosaic Virus (TMV), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (e.g., Gallie et al., 1987; Skuzeski et al., 1990). Other leaders known in the art include but are not limited to: Picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5 noncoding region) (Elroy-Stein et al., 1989); Potyvirus leaders, for example, TEV leader (Tobacco Etch Virus); MDMV leader (Maize Dwarf Mosaic Virus); Human immunoglobulin heavy-chain binding protein (BiP) leader, (Macejak et al., 1991); Untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling et al., 1987; Tobacco mosaic virus leader (TMV), (Gallie et al., 1989; and Maize Chlorotic Mottle Virus leader (MCMV) (Lommel et al., 1991. See also, Della-Cioppa et al., 1987.

Regulatory elements such as *Adh* intron 1 (Callis et al., 1987), sucrose synthase intron (Vasil et al., 1989) or TMV omega element (Gallie, et al., 1989), may further be included where desired.

Examples of enhancers include elements from the CaMV 35S promoter, octopine synthase genes (Ellis et al., 1987), the rice actin I gene, the maize alcohol dehydrogenase gene (Callis et al., 1987), the maize shrunken I gene (Vasil et al., 1989), TMV Omega element (Gallie et al., 1989) and promoters from non-plant eukaryotes (e.g. yeast; Ma et al., 1988).

The present invention further provides a method of augmenting a plant genome by contacting plant cells with a nucleic acid molecule of the invention, e.g., one having a nucleotide sequence that directs constitutive, tissue-specific, or tissue-preferential and/or induced transcription of a linked nucleic acid segment encoding a polypeptide that is substantially similar to a polypeptide encoded by the an *Oryza* gene having a sequence according to any one of the SEQ ID NOs provided in the Sequence Listing so as to yield

transformed plant cells; and regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the nucleic acid molecule in the cells of the plant. The nucleic acid molecule may be present in the nucleus, chloroplast, mitochondria and/or plastid of the cells of the plant.

5 Plant species may be transformed with the DNA construct of the present invention by the DNA-mediated transformation of plant cell protoplasts and subsequent regeneration of the plant from the transformed protoplasts in accordance with procedures well known in the art.

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a vector of the present invention. The term
10 “organogenesis,” as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers; the term “embryogenesis,” as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being
15 transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and utilane meristem).

Plants of the present invention may take a variety of forms. The plants may be
20 chimeras of transformed cells and non-transformed cells; the plants may be clonal transformants (e.g., all cells transformed to contain the expression cassette); the plants may comprise grafts of transformed and untransformed tissues (e.g., a transformed root stock grafted to an untransformed scion in citrus species). The transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding
25 techniques. For example, first generation (or T1) transformed plants may be selfed to give homozygous second generation (or T2) transformed plants, and the T2 plants further propagated through classical breeding techniques. A dominant selectable marker (such as npt II) can be associated with the expression cassette to assist in breeding.

Thus, the present invention provides a transformed (transgenic) plant cell, *in planta* or
30 *ex planta*, including a transformed plastid or other organelle, e.g., nucleus, mitochondria or chloroplast. The present invention may be used for transformation of any plant species,

including, but not limited to, cells from corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*),
 5 foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*),
 10 pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilane*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed
 15 (*Lemna*), barley, vegetables, ornamentals, and conifers.

Duckweed (*Lemna*, see WO 00/07210) includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as follows: genus *Lemna* (*L. aequinoctialis*, *L. disperma*, *L. ecuadoriensis*, *L. gibba*, *L. japonica*, *L. minor*, *L. miniscula*, *L. obscura*, *L. perpusilla*, *L. tenera*, *L. trisulca*, *L. turionifera*, *L. valdiviana*); genus *Spirodela* (*S. intermedia*, *S. polyrrhiza*, *S. punctata*); genus *Woffia* (*Wa. Angusta*, *Wa. Arrhiza*, *Wa. Australina*, *Wa. Borealis*, *Wa. Brasiliensis*, *Wa. Columbiana*, *Wa. Elongata*, *Wa. Globosa*, *Wa. Microscopica*, *Wa. Neglecta*) and genus *Wofliella* (*Wl. utila*, *Wl. utilanen*, *Wl. gladiata*, *Wl. utila*, *Wl. lingulata*, *Wl. repunda*, *Wl. rotunda*, and *Wl. neotropica*). Any other genera or species of *Lemnaceae*, if they exist, are also aspects of the present invention. *Lemna gibba*,
 25 *Lemna minor*, and *Lemna miniscula* are preferred, with *Lemna minor* and *Lemna miniscula* being most preferred. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of *Lemnaceae* – A Monograph Study. Geobotanischen Institut ETH, Stiftung Rubel, Zurich (1986)).

30 Vegetables within the scope of the invention include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans

(*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*).

Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga utililane*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo. Preferred forage and turf grass for use in the methods of the invention include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redbtop.

Other plants within the scope of the invention include *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, cauliflower, Brassica, e.g., broccoli, cabbage, utilan sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, utililane, and zucchini.

Ornamental plants within the scope of the invention include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum,

Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia. Other plants within the scope of the invention are shown in Table 2 (above).

Preferably, transgenic plants of the present invention are crop plants and in particular
5 cereals (for example, corn, alfalfa, sunflower, rice, *Brassica*, canola, soybean, barley, soybean, sugarbeet, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, etc.), and even more preferably corn, rice and soybean.

The present invention also provides a transgenic plant prepared by this method, a seed
from such a plant and progeny plants from such a plant including hybrids and inbreds.
10 Preferred transgenic plants are transgenic maize, soybean, barley, alfalfa, sunflower, canola, soybean, cotton, peanut, sorghum, tobacco, sugarbeet, rice, wheat, rye, turfgrass, millet, sugarcane, tomato, or potato.

A transformed (transgenic) plant of the invention includes plants, the genome of which
is augmented by a nucleic acid molecule of the invention, or in which the corresponding gene
15 has been disrupted, e.g., to result in a loss, a decrease or an alteration, in the function of the product encoded by the gene, which plant may also have increased yields and/or produce a better-quality product than the corresponding wild-type plant. The nucleic acid molecules of the invention are thus useful for targeted gene disruption, as well as markers and probes.

The invention also provides a method of plant breeding, e.g., to prepare a crossed
20 fertile transgenic plant. The method comprises crossing a fertile transgenic plant comprising a particular nucleic acid molecule of the invention with itself or with a second plant, e.g., one lacking the particular nucleic acid molecule, to prepare the seed of a crossed fertile transgenic plant comprising the particular nucleic acid molecule. The seed is then planted to obtain a crossed fertile transgenic plant. The plant may be a monocot or a dicot. In a particular
25 embodiment, the plant is a cereal plant.

The crossed fertile transgenic plant may have the particular nucleic acid molecule inherited through a female parent or through a male parent. The second plant may be an inbred plant. The crossed fertile transgenic may be a hybrid. Also included within the present invention are seeds of any of these crossed fertile transgenic plants.

30 Transformation of plants can be undertaken with a single DNA molecule or multiple DNA molecules (i.e., co-transformation), and both these techniques are suitable for use with

the expression cassettes of the present invention. Numerous transformation vectors are available for plant transformation, and the expression cassettes of this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the preferred transformation technique and the target species for transformation.

5 A variety of techniques are available and known to those skilled in the art for introduction of constructs into a plant cell host. These techniques generally include transformation with DNA employing *A. tumefaciens* or *A. rhizogenes* as the transforming agent, liposomes, PEG precipitation, electroporation, DNA injection, direct DNA uptake, microprojectile bombardment, particle acceleration, and the like (See, for example, EP
10 295959 and EP 138341) (see below). However, cells other than plant cells may be transformed with the expression cassettes of the invention. The general descriptions of plant expression vectors and reporter genes, and *Agrobacterium* and *Agrobacterium*-mediated gene transfer, can be found in Gruber et al. (1993).

 Expression vectors containing genomic or synthetic fragments can be introduced into
15 protoplasts or into intact tissues or isolated cells. Preferably expression vectors are introduced into intact tissue. General methods of culturing plant tissues are provided for example by Maki et al., (1993); and by Phillips et al. (1988). Preferably, expression vectors are introduced into maize or other plant tissues using a direct gene transfer method such as microprojectile-mediated delivery, DNA injection, electroporation and the like. More
20 preferably expression vectors are introduced into plant tissues using the microprojectile media delivery with the biolistic device. See, for example, Tomes et al. (1995). The vectors of the invention can not only be used for expression of structural genes but may also be used in exon-trap cloning, or promoter trap procedures to detect differential gene expression in varieties of tissues, (Lindsey et al., 1993; Auch & Reth et al.).

25 It is particularly preferred to use the binary type vectors of Ti and Ri plasmids of *Agrobacterium spp.* Ti-derived vectors transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants, such as soybean, cotton, rape, tobacco, and rice (Pacciotti et al., 1985; Byrne et al., 1987; Sukhapinda et al., 1987; Lorz et al., 1985; Potrykus, 1985; Park et al., 1985; Hiei et al., 1994). The use of T-DNA to transform plant cells has
30 received extensive study and is amply described (EP 120516; Hoekema, 1985; Knauf, et al.,

1983; and An et al., 1985). For introduction into plants, the chimeric genes of the invention can be inserted into binary vectors as described in the examples.

Other transformation methods are available to those skilled in the art, such as direct uptake of foreign DNA constructs (see EP 295959), techniques of electroporation (Fromm et al., 1986) or high velocity ballistic bombardment with metal particles coated with the nucleic acid constructs (Kline et al., 1987, and U.S. Patent No. 4,945,050). Once transformed, the cells can be regenerated by those skilled in the art. Of particular relevance are the recently described methods to transform foreign genes into commercially important crops, such as rapeseed (De Block et al., 1989), sunflower (Everett et al., 1987), soybean (McCabe et al., 1988; Hinchee et al., 1988; Chee et al., 1989; Christou et al., 1989; EP 301749), rice (Hiei et al., 1994), and corn (Gordon Kamm et al., 1990; Fromm et al., 1990).

Those skilled in the art will appreciate that the choice of method might depend on the type of plant, i.e., monocotyledonous or dicotyledonous, targeted for transformation. Suitable methods of transforming plant cells include, but are not limited to, microinjection (Crossway et al., 1986), electroporation (Riggs et al., 1986), *Agrobacterium*-mediated transformation (Hinchee et al., 1988), direct gene transfer (Paszkowski et al., 1984), and ballistic particle acceleration using devices available from Agracetus, Inc., Madison, Wis. And BioRad, Hercules, Calif. (see, for example, Sanford et al., U.S. Pat. No. 4,945,050; and McCabe et al., 1988). Also see, Weissinger et al., 1988; Sanford et al., 1987 (onion); Christou et al., 1988 (soybean); McCabe et al., 1988 (soybean); Datta et al., 1990 (rice); Klein et al., 1988 (maize); Klein et al., 1988 (maize); Klein et al., 1988 (maize); Fromm et al., 1990 (maize); and Gordon-Kamm et al., 1990 (maize); Svab et al., 1990 (tobacco chloroplast); Koziel et al., 1993 (maize); Shimamoto et al., 1989 (rice); Christou et al., 1991 (rice); European Patent Application EP 0 332 581 (orchardgrass and other Pooideae); Vasil et al., 1993 (wheat); Weeks et al., 1993 (wheat). In one embodiment, the protoplast transformation method for maize is employed (European Patent Application EP 0 292 435, U. S. Pat. No. 5,350,689).

In another embodiment, a nucleotide sequence of the present invention is directly transformed into the plastid genome. Plastid transformation technology is extensively described in U.S. Patent Nos. 5,451,513, 5,545,817, and 5,545,818, in PCT application no. WO 95/16783, and in McBride et al., 1994. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable

marker together with the gene of interest into a suitable target tissue, e.g., using biolistics or protoplast transformation (e.g., calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate orthologous recombination with the plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab et al., 1990; Staub et al., 1992). This resulted in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub et al., 1993). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial *aadA* gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3N-adenyltransferase (Svab et al., 1993). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention. Typically, approximately 15-20 cell division cycles following transformation are required to reach a homoplastidic state. Plastid expression, in which genes are inserted by orthologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclear-expressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein. In a preferred embodiment, a nucleotide sequence of the present invention is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants homoplasmic for plastid genomes containing a nucleotide sequence of the present invention are obtained, and are preferentially capable of high expression of the nucleotide sequence.

Agrobacterium tumefaciens cells containing a vector comprising an expression cassette of the present invention, wherein the vector comprises a Ti plasmid, are useful in methods of making transformed plants. Plant cells are infected with an *Agrobacterium tumefaciens* as described above to produce a transformed plant cell, and then a plant is regenerated from the transformed plant cell. Numerous *Agrobacterium* vector systems useful in carrying out the present invention are known.

For example, vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, 1984). In one preferred embodiment, the expression cassettes of the present invention may be inserted into either of the binary vectors pCIB200 and pCIB2001 for use with *Agrobacterium*. These vector cassettes for *Agrobacterium*-mediated transformation wear constructed in the following manner. PTJS75kan was created by NarI digestion of pTJS75 (Schmidhauser & Helinski, 1985) allowing excision of the tetracycline-resistance gene, followed by insertion of an AccI fragment from pUC4K carrying an NPTII (Messing & Vierra, 1982; Bevan et al., 1983; McBride et al., 1990). XhoI linkers were ligated to the EcoRV fragment of pCIB7 which contains the left and right T-DNA borders, a plant selectable nos/nptII chimeric gene and the pUC polylinker (Rothstein et al., 1987), and the XhoI- digested fragment was cloned into Sall-digested pTJS75kan to create pCIB200 (see also EP 0 332 104, example 19). PCIB200 contains the following unique polylinker restriction sites: EcoRI, SstI, KpnI, BglII, XbaI, and Sall. The plasmid pCIB2001 is a derivative of pCIB200 which was created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001 are EcoRI, SstI, KpnI, BglII, XbaI, Sall, MluI, BclI, AvrII, ApaI, HpaI, and StuI. PCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-DNA borders for *Agrobacterium*-mediated transformation, the RK2-derived trfA function for mobilization between *E. coli* and other hosts, and the OriT and OriV functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

An additional vector useful for *Agrobacterium*-mediated transformation is the binary vector pCIB 10, which contains a gene encoding kanamycin resistance for selection in plants, T-DNA right and left border sequences and incorporates sequences from the wide host- range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein et al., 1987. Various derivatives of pCIB10 have been constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz et al., 1983. These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

Methods using either a form of direct gene transfer or *Agrobacterium*-mediated transfer usually, but not necessarily, are undertaken with a selectable marker which may provide resistance to an antibiotic (e.g., kanamycin, hygromycin or methotrexate) or a herbicide (e.g., phosphinothricin). The choice of selectable marker for plant transformation is not, however, critical to the invention.

For certain plant species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the nptII gene which confers resistance to kanamycin and related antibiotics (Messing & Vierra, 1982; Bevan et al., 1983), the bar gene which confers resistance to the herbicide phosphinothricin (White et al., 1990, Spencer et al., 1990), the hph gene which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann), and the dhfr gene, which confers resistance to methotrexate (Bourouis et al., 1983).

One such vector useful for direct gene transfer techniques in combination with selection by the herbicide Basta (or phosphinothricin) is pCIB3064. This vector is based on the plasmid pCIB246, which comprises the CaMV 35S promoter in operational fusion to the *E. coli* GUS gene and the CaMV 35S transcriptional terminator and is described in the PCT published application WO 93/07278, herein incorporated by reference. One gene useful for conferring resistance to phosphinothricin is the bar gene from *Streptomyces viridochromogenes* (Thompson et al., 1987). This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

An additional transformation vector is pSOG35 which utilizes the *E. coli* gene dihydrofolate reductase (DHFR) as a selectable marker conferring resistance to methotrexate. PCR was used to amplify the 35S promoter (about 800 bp), intron 6 from the maize Adh1 gene (about 550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250 bp fragment encoding the *E. coli* dihydrofolate reductase type II gene was also amplified by PCR and these two PCR fragments were assembled with a SacI-PstI fragment from pBI221 (Clontech) which comprised the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generated pSOG19 which contains the 35S promoter in fusion with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus check (MCMV) generated the vector pSOG35.

pSOG19 and pSOG35 carry the pUC-derived gene for ampicillin resistance and have HindIII, SphI, PstI and EcoRI sites available for the cloning of foreign sequences.

Transgenic plant cells are then placed in an appropriate selective medium for selection of transgenic cells which are then grown to callus. Shoots are grown from callus and plantlets
5 generated from the shoot by growing in rooting medium. The various constructs normally will be joined to a marker for selection in plant cells. Conveniently, the marker may be resistance to a biocide (particularly an antibiotic, such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, herbicide, or the like). The particular marker used will allow for selection of transformed cells as compared to cells lacking the DNA which has been
10 introduced. Components of DNA constructs including transcription cassettes of this invention may be prepared from sequences which are native (endogenous) or foreign (exogenous) to the host. By "foreign" it is meant that the sequence is not found in the wild-type host into which the construct is introduced. Heterologous constructs will contain at least one region which is not native to the gene from which the transcription-initiation-region is derived.

15 To confirm the presence of the transgenes in transgenic cells and plants, a variety of assays may be performed. Such assays include, for example, "molecular biological" assays well known to those of skill in the art, such as Southern and Northern blotting, *in situ* hybridization and nucleic acid-based amplification methods such as PCR or RT-PCR; "biochemical" assays, such as detecting the presence of a protein product, e.g., by
20 immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as seed assays; and also, by analyzing the phenotype of the whole regenerated plant, e.g., for disease or pest resistance.

DNA may be isolated from cell lines or any plant parts to determine the presence of the preselected nucleic acid segment through the use of techniques well known to those skilled in
25 the art. Note that intact sequences will not always be present, presumably due to rearrangement or deletion of sequences in the cell.

The presence of nucleic acid elements introduced through the methods of this invention may be determined by polymerase chain reaction (PCR). Using this technique discreet fragments of nucleic acid are amplified and detected by gel electrophoresis. This type of
30 analysis permits one to determine whether a preselected nucleic acid segment is present in a stable transformant, but does not prove integration of the introduced preselected nucleic acid

segment into the host cell genome. In addition, it is not possible using PCR techniques to determine whether transformants have exogenous genes introduced into different sites in the genome, i.e., whether transformants are of independent origin. It is contemplated that using PCR techniques it would be possible to clone fragments of the host genomic DNA adjacent to an introduced preselected DNA segment.

Positive proof of DNA integration into the host genome and the independent identities of transformants may be determined using the technique of Southern hybridization. Using this technique specific DNA sequences that were introduced into the host genome and flanking host DNA sequences can be identified. Hence the Southern hybridization pattern of a given transformant serves as an identifying characteristic of that transformant. In addition it is possible through Southern hybridization to demonstrate the presence of introduced preselected DNA segments in high molecular weight DNA, i.e., confirm that the introduced preselected DNA segment has been integrated into the host cell genome. The technique of Southern hybridization provides information that is obtained using PCR, e.g., the presence of a preselected DNA segment, but also demonstrates integration into the genome and characterizes each individual transformant.

It is contemplated that using the techniques of dot or slot blot hybridization which are modifications of Southern hybridization techniques one could obtain the same information that is derived from PCR, e.g., the presence of a preselected DNA segment.

Both PCR and Southern hybridization techniques can be used to demonstrate transmission of a preselected DNA segment to progeny. In most instances the characteristic Southern hybridization pattern for a given transformant will segregate in progeny as one or more Mendelian genes (Spencer et al., 1992); Laursen et al., 1994) indicating stable inheritance of the gene. The nonchimeric nature of the callus and the parental transformants (R_0) was suggested by germline transmission and the identical Southern blot hybridization patterns and intensities of the transforming DNA in callus, R_0 plants and R_1 progeny that segregated for the transformed gene.

Whereas DNA analysis techniques may be conducted using DNA isolated from any part of a plant, RNA may only be expressed in particular cells or tissue types and hence it will be necessary to prepare RNA for analysis from these tissues. PCR techniques may also be used for detection and quantitation of RNA produced from introduced preselected DNA segments.

In this application of PCR it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique will demonstrate the presence of an RNA species and give information about the integrity of that RNA. The presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and will only demonstrate the presence or absence of an RNA species.

While Southern blotting and PCR may be used to detect the preselected DNA segment in question, they do not provide information as to whether the preselected DNA segment is being expressed. Expression may be evaluated by specifically identifying the protein products of the introduced preselected DNA segments or evaluating the phenotypic changes brought about by their expression.

Assays for the production and identification of specific proteins may make use of physical-chemical, structural, functional, or other properties of the proteins. Unique physical-chemical or structural properties allow the proteins to be separated and identified by electrophoretic procedures, such as native or denaturing gel electrophoresis or isoelectric focusing, or by chromatographic techniques such as ion exchange or gel exclusion chromatography. The unique structures of individual proteins offer opportunities for use of specific antibodies to detect their presence in formats such as an ELISA assay. Combinations of approaches may be employed with even greater specificity such as Western blotting in which antibodies are used to locate individual gene products that have been separated by electrophoretic techniques. Additional techniques may be employed to absolutely confirm the identity of the product of interest such as evaluation by amino acid sequencing following purification. Although these are among the most commonly employed, other procedures may be additionally used.

Assay procedures may also be used to identify the expression of proteins by their functionality, especially the ability of enzymes to catalyze specific chemical reactions involving specific substrates and products. These reactions may be followed by providing and

quantifying the loss of substrates or the generation of products of the reactions by physical or chemical procedures. Examples are as varied as the enzyme to be analyzed.

Very frequently the expression of a gene product is determined by evaluating the phenotypic results of its expression. These assays also may take many forms including but not
5 limited to analyzing changes in the chemical composition, morphology, or physiological properties of the plant. Morphological changes may include greater stature or thicker stalks. Most often changes in response of plants or plant parts to imposed treatments are evaluated under carefully controlled conditions termed bioassays.

The present invention also provides a method to identify a nucleotide sequence that
10 encodes a protein product that is involved in the development and/or timing of flower formation in a respective plant by contacting a probe of plant nucleic acid, e.g., cRNA from rice, isolated from various tissues of a plant, with a plurality of isolated nucleic acid samples on one or more, i.e., a plurality of, solid substrates so as to form a complex between at least a portion of the probe and a nucleic acid sample(s) having sequences that are structurally related
15 to the sequences in the probe. Each sample comprises one or a plurality of oligonucleotides corresponding to at least a portion of a plant gene. Then complex formation is compared between samples contacted with a particular tissue, e.g., a seed-specific, probe and samples contacted with a different tissue, e.g., a non-seed specific probe, so as to determine which RNAs are expressed in the particular tissue of the plant. The probe and/or samples may be
20 nucleic acid from a dicot or from a monocot.

The compositions of the invention include plant nucleic acid molecules, and the amino acid sequences for the polypeptides or partial-length polypeptides encoded by the nucleic acid molecule which comprises an open reading frame. These sequences can be employed to alter
25 expression of a particular gene corresponding to the open reading frame by decreasing or eliminating expression of that plant gene or by overexpressing a particular gene product. Methods of this embodiment of the invention include stably transforming a plant with the nucleic acid molecule which includes an open reading frame operably linked to a promoter capable of driving expression of that open reading frame (sense or antisense) in a plant cell. By "portion" or "fragment", as it relates to a nucleic acid molecule which comprises an open
30 reading frame or a fragment thereof encoding a partial-length polypeptide having the activity of the full length polypeptide, is meant a sequence having at least 80 nucleotides, more

preferably at least 150 nucleotides, and still more preferably at least 400 nucleotides. If not employed for expressing, a "portion" or "fragment" means at least 9, preferably 12, more preferably 15, even more preferably at least 20, consecutive nucleotides, e.g., probes and primers (oligonucleotides), corresponding to the nucleotide sequence of the nucleic acid molecules of the invention such as those identified in the SEQ ID NOs provided in the Sequence Listing. Thus, to express a particular gene product, the method comprises introducing to a plant, plant cell, or plant tissue an expression cassette comprising a promoter linked to an open reading frame so as to yield a transformed differentiated plant, transformed cell or transformed tissue. Transformed cells or tissue can be regenerated to provide a transformed differentiated plant. The transformed differentiated plant or cells thereof preferably expresses the open reading frame in an amount that alters the amount of the gene product in the plant or cells thereof, which product is encoded by the open reading frame. The present invention also provides a transformed plant prepared by the method, progeny and seed thereof.

The invention further includes a nucleotide sequence which is complementary to one (hereinafter "test" sequence) which hybridizes under stringent conditions with a nucleic acid molecule of the invention as well as RNA which is transcribed from the nucleic acid molecule. When the hybridization is performed under stringent conditions, either the test or nucleic acid molecule of the invention is preferably supported, e.g., on a membrane or DNA chip. Thus, either a denatured test or nucleic acid molecule of the invention is preferably first bound to a support and hybridization is effected for a specified period of time at a temperature of, e.g., between 55 and 70°C, in double strength citrate buffered saline (SC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SC concentration. Depending upon the degree of stringency required such reduced concentration buffers are typically single strength SC containing 0.1% SDS, half strength SC containing 0.1% SDS and one-tenth strength SC containing 0.1% SDS.

Polynucleotides derived from nucleotide sequences of the present invention having any of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 1-66, and 123 - 146, and 67 - 122 provided in the Sequence Listing are useful to detect the presence in a test sample of at least one copy of a nucleotide sequence containing the same or substantially the same sequence, or a fragment, complement, or variant thereof. The sequence

of the probes and/or primers of the instant invention need not be identical to those provided in the Sequence Listing or the complements thereof. Some variation in probe or primer sequence and/or length can allow additional family members to be detected, as well as orthologous genes and more taxonomically distant related sequences. Similarly probes and/or primers of the invention can include additional nucleotides that serve as a label for detecting duplexes, for isolation of duplexed polynucleotides, or for cloning purposes.

Preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides containing a contiguous span of between at least 12 to at least 1000 nucleotides of any of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 1-66, and 123 - 146, and 67 - 122 provided in the Sequence Listing or the complements thereof, with each individual number of nucleotides within this range also being part of the invention. Preferred are isolated, purified, or recombinant polynucleotides containing a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 750, or 1000 nucleotides of any of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 1-66, and 123 - 146, and 67 - 122 provided in the Sequence Listing or the complements thereof. The appropriate length for primers and probes will vary depending on the application. For use as PCR primers, probes are 12-40 nucleotides, preferably 18-30 nucleotides long. For use in mapping, probes are 50 to 500 nucleotides, preferably 100-250 nucleotides long. For use in Southern hybridizations, probes as long as several kilobases can be used. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art.

The primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang *et al.* (*Meth Enzymol* 68: 90 (1979)), the diethylphosphoramidite method, the triester method of Matteucci *et al.* (*J Am Chem Soc* 103: 3185 (1981)), or according to Urdea *et al.* (*Proc Natl Acad* 80: 7461 (1981)), the solid support method described in EP 0 707 592, or using commercially available automated oligonucleotide synthesizers.

Detection probes are generally nucleotide sequences or uncharged nucleotide analogs such as, for example peptide nucleotides which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patent Nos.

5,185,444, 5,034,506 and 5,142,047. The probe may have to be rendered “non-extendable” such that additional dNTPs cannot be added to the probe. Analogs are usually non-extendable, and nucleotide probes can be rendered non-extendable by modifying the 3’ end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3’ end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3’ hydroxyl group simply can be cleaved, replaced or modified so as to render the probe non-extendable.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (^{32}P , ^{35}S , ^3H , ^{125}I), fluorescent dyes (5-bromodesoxyuridine, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3’ and 5’ ends. Examples of non-radioactive labeling of nucleotide fragments are described in the French patent No. FR-7810975 and by Urdea *et al.* (*Nuc Acids Res* 16:4937 (1988)). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as described in EP 0 225 807.

A label can also be used to capture the primer so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member that forms a binding pair with the solid’s phase reagent’s specific binding member, for example biotin and streptavidin. Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent’s binding member is a nucleotide sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or “tail” that is not complementary to the target. In the case where a polynucleotide primer itself serves as the

capture label, at least a portion of the primer will be free to hybridize with a nucleotide on a solid phase. DNA labeling techniques are well known in the art.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleotides on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material that is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor that has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

The polynucleotides of the invention that are expressed or repressed in response to environmental stimuli such as, for example, stress or treatment with chemicals or pathogens or at different developmental stages can be identified by employing an array of nucleic acid samples, *e.g.*, each sample having a plurality of oligonucleotides, and each plurality
5 corresponding to a different plant gene, on a solid substrate, *e.g.*, a DNA chip, and probes corresponding to nucleic acid expressed in, for example, one or more plant tissues and/or at one or more developmental stages, *e.g.*, probes corresponding to nucleic acid expressed in seed of a plant relative to control nucleic acid from sources other than seed. Thus, genes that are upregulated or downregulated in the majority of tissues at a majority of developmental
10 stages, or upregulated or downregulated in one tissue such as in seed, can be systematically identified. The probes may also correspond to nucleic acid expressed in response to a defined treatment such as, for example, a treatment with a variety of plant hormones or the exposure to specific environmental conditions involving, for example, an abiotic stress or exposure to light.

15 Specifically, labeled rice cRNA probes were hybridized to the rice DNA array, expression levels were determined by laser scanning and then rice genes were identified that had a particular expression pattern. The rice oligonucleotide probe array consists of probes from over 18,000 unique rice genes, which covers approximately 40-50% of the genome. This genome array permits a broader, more complete and less biased analysis of gene
20 expression.

Consequently, the invention also deals with a method for detecting the presence of a polynucleotide including a nucleotide sequence which is substantially similar to a nucleotide sequence given in SEQ ID NOs: 1 to SEQ ID NO: 66, or a fragment or a variant thereof, or a complementary sequence thereto, in a sample, the method including the following steps of:

25 (a) bringing into contact a nucleotide probe or a plurality of nucleotide probes which can hybridize with a polynucleotide having a nucleotide sequence which is substantially similar to a nucleotide sequence given in SEQ ID NOs: 1 to SEQ ID NO: 66, a fragment or a variant thereof, or a complementary sequence thereto and the sample to be assayed.

30 (b) detecting the hybrid complex formed between the probe and a nucleotide in the sample.

The invention further concerns a kit for detecting the presence of a polynucleotide including a nucleotide sequence which is substantially similar to a nucleotide sequence given in SEQ ID NOs: 1 to SEQ ID NO: 66, a fragment or a variant thereof, or a complementary sequence thereto, in a sample, the kit including a nucleotide probe or a plurality of nucleotide probes which can hybridize with a nucleotide sequence included in a polynucleotide, which nucleotide sequence is substantially similar to a nucleotide sequence given in of SEQ ID NOs: 1 to SEQ ID NO: 66, a fragment or a variant thereof, or a complementary sequence thereto and, optionally, the reagents necessary for performing the hybridization reaction.

In a first preferred embodiment of this detection method and kit, the nucleotide probe or the plurality of nucleotide probes are labeled with a detectable molecule. In a second preferred embodiment of the method and kit, the nucleotide probe or the plurality of nucleotide probes has been immobilized on a substrate.

The isolated polynucleotides of the invention can be used to create various types of genetic and physical maps of the genome of rice or other plants. Such maps are used to devise positional cloning strategies for isolating novel genes from the mapped crop species. The sequences of the present invention are also useful for chromosome mapping, chromosome identification, tagging of genes involved in regulating flowering time, and whole plant architecture.

The isolated polynucleotides of the invention can further be used as probes for identifying polymorphisms associated with phenotypes of interest such as, for example, early or delayed flowering time. Briefly, total DNA is isolated from an individual or isogenic line, cleaved with one or more restriction enzymes, separated according to mass, transferred to a solid support, and hybridized with a probe molecule according to the invention. The pattern of fragments hybridizing to a probe molecule is compared for DNA from different individuals or lines, where differences in fragment size signals a polymorphism associated with a particular nucleotide sequence according to the present invention. After identification of polymorphic sequences, linkage studies can be conducted. After identification of many polymorphisms using a nucleotide sequence according to the invention, linkage studies can be conducted by using the individuals showing polymorphisms as parents in crossing programs. Recombinants, F₂ progeny recombinants or recombinant inbreds, can then be analyzed using the same restriction enzyme/hybridization procedure. The order of DNA polymorphisms

along the chromosomes can be inferred based on the frequency with which they are inherited together versus inherited independently. The closer together two polymorphisms occur in a chromosome, the higher the probability that they are inherited together. Integration of the relative positions of polymorphisms and associated marker sequences produces a genetic map of the species, where the distances between markers reflect the recombination frequencies in that chromosome segment. Preferably, the polymorphisms and marker sequences are sufficiently numerous to produce a genetic map of sufficiently high resolution to locate one or more loci of interest.

The use of recombinant inbred lines for such genetic mapping is described for rice (Oh *et al.*, *Mol Cells* 8:175 (1998); Nandi *et al.*, *Mol Gen Genet* 255:1 (1997); Wang *et al.*, *Genetics* 136:1421 (1994)), sorghum (Subudhi *et al.*, *Genome* 43:240 (2000)), maize (Burr *et al.*, *Genetics* 118:519 (1998); Gardiner *et al.*, *Genetics* 134:917 (1993)), and *Arabidopsis* (*Methods in Molecular Biology*, Martinez-Zapater and Salinas, eds., 82:137-146, (1998)). However, this procedure is not limited to plants and can be used for other organisms such as yeast or other fungi, or for oomycetes or other protistans.

The nucleotide sequences of the present invention can also be used for simple sequence repeat identification, also known as single sequence repeat, (SSR) mapping. SSR mapping in rice has been described by Miyao *et al.* (*DNA Res* 3:233 (1996)) and Yang *et al.* (*Mol Gen Genet* 245:187 (1994)), and in maize by Ahn *et al.* (*Mol Gen Genet* 241:483 (1993)). SSR mapping can be achieved using various methods. In one instance, polymorphisms are identified when sequence specific probes flanking an SSR contained within an sequence of the invention are made and used in polymerase chain reaction (PCR) assays with template DNA from two or more individuals or, in plants, near isogenic lines. A change in the number of tandem repeats between the SSR-flanking sequence produces differently sized fragments (U.S. Patent No. 5,766,847). Alternatively, polymorphisms can be identified by using the PCR fragment produced from the SSR-flanking sequence specific primer reaction as a probe against Southern blots representing different individuals (Refseth *et al.*, *Electrophoresis* 18:1519 (1997)). Rice SSRs were used to map a molecular marker closely linked to a nuclear restorer gene for fertility in rice as described by Akagi *et al.* (*Genome* 39:205 (1996)).

The nucleotide sequences of the present invention can be used to identify and develop a variety of microsatellite markers, including the SSRs described above, as genetic markers for comparative analysis and mapping of genomes. The nucleotide sequences of the present invention can be used in a variation of the SSR technique known as inter-SSR (ISSR), which
5 uses microsatellite oligonucleotides as primers to amplify genomic segments different from the repeat region itself (Zietkiewicz *et al.*, *Genomics* 20:176 (1994)). ISSR employs oligonucleotides based on a simple sequence repeat anchored or not at their 5'- or 3'-end by two to four arbitrarily chosen nucleotides, which triggers site-specific annealing and initiates PCR amplification of genomic segments which are flanked by inversely orientated and closely
10 spaced repeat sequences. In one embodiment of the present invention, microsatellite markers derived from the nucleotide sequences disclosed in the Sequence Listing, or substantially similar sequences or allelic variants thereof, may be used to detect the appearance or disappearance of markers indicating genomic instability as described by Leroy *et al.* (*Electron. J Biotechnol*, 3(2), at <http://www.ejb.org> (2000)), where alteration of a fingerprinting pattern
15 indicated loss of a marker corresponding to a part of a gene involved in the regulation of cell proliferation. Microsatellite markers derived from nucleotide sequences as provided in the Sequence Listing will be useful for detecting genomic alterations such as the change observed by Leroy *et al.* (*Electron. J Biotechnol*, 3(2), *supra* (2000)) which appeared to be the consequence of microsatellite instability at the primer binding site or modification of the
20 region between the microsatellites, and illustrated somaclonal variation leading to genomic instability. Consequently, the nucleotide sequences of the present invention are useful for detecting genomic alterations involved in somaclonal variation, which is an important source of new phenotypes.

In addition, because the genomes of closely related species are largely syntenic (that is,
25 they display the same ordering of genes within the genome), these maps can be used to isolate novel alleles from wild relatives of crop species by positional cloning strategies. This shared synteny is very powerful for using genetic maps from one species to map genes in another. For example, a gene mapped in rice provides information for the gene location in maize and wheat.

30 The various types of maps discussed above can be used with the nucleotide sequences of the invention to identify Quantitative Trait Loci (QTLs) for a variety of uses, including marker-

assisted breeding. Many important crop traits are quantitative traits and result from the combined interactions of several genes. These genes reside at different loci in the genome, often on different chromosomes, and generally exhibit multiple alleles at each locus.

Developing markers, tools, and methods to identify and isolate the QTLs involved in heading time and/or whole plant architecture, enables marker-assisted breeding to modify the timing of flower development or suppress undesirable traits that interfere with a desired heading time.

The nucleotide sequences as provided in the Sequence Listing can be used to generate markers, including single-sequence repeats (SSRs) and microsatellite markers for QTLs involved in heading time and/or whole plant architecture, to assist marker-assisted breeding.

The nucleotide sequences of the invention can be used to identify QTLs involved in heading time and/or whole plant architecture, and isolate alleles as described by Li *et al.* in a study of QTLs involved in resistance to a pathogen of rice. (Li *et al.*, *Mol Gen Genet* 261:58 (1999)). In addition to isolating QTL alleles in rice, other cereals, and other monocot and dicot crop species, the nucleotide sequences of the invention can also be used to isolate alleles from the corresponding QTL(s) of wild relatives. Transgenic plants having various combinations of QTL alleles can then be created and the effects of the combinations measured. Once an ideal allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed conventional breeding programs. (Flowers *et al.*, *J Exp Bot* 51:99 (2000); Tanksley and McCouch, *Science* 277:1063 (1997)).

In another embodiment the nucleotide sequences of the invention can be used to help create physical maps of the genome of maize, *Arabidopsis* and related species. Where the nucleotide sequences of the invention have been ordered on a genetic map, as described above, then the nucleotide sequences of the invention can be used as probes to discover which clones in large libraries of plant DNA fragments in YACs, PACs, etc. contain the same nucleotide sequences of the invention or similar sequences, thereby facilitating the assignment of the large DNA fragments to chromosomal positions. Subsequently, the large BACs, YACs, etc. can be ordered unambiguously by more detailed studies of their sequence composition and by using their end or other sequence to find the identical sequences in other cloned DNA fragments (Mozo *et al.*, *Nat Genet* 22:271 (1999)). Overlapping DNA sequences in this way allows assembly of large sequence contigs that, when sufficiently extended, provide a

complete physical map of a chromosome. The nucleotide sequences of the invention themselves may provide the means of joining cloned sequences into a contig, and are useful for constructing physical maps.

In another embodiment, the nucleotide sequences of the present invention may be useful
5 in mapping and characterizing the genomes of other cereals. Rice has been proposed as a model for cereal genome analysis (Havukkala, *Curr Opin Genet Devel* 6:711 (1996)), based largely on its smaller genome size and higher gene density, combined with the considerable conserved gene order among cereal genomes (Ahn *et al.*, *Mol Gen Genet* 241:483 (1993)). The cereals demonstrate both general conservation of gene order (synteny) and considerable
10 sequence homology among various cereal gene families. This suggests that studies on the functions of genes or proteins from rice involved in regulating flowering time, and whole plant architecture could lead to elucidation of the functions of orthologous genes or proteins in other cereals, including maize, wheat, secale, sorghum, barley, millet, teff, milo, triticale, flax, gramma grass, *Tripsacum* sp., and teosinte. The nucleotide sequences according to the
15 invention can also be used to physically characterize homologous chromosomes in other cereals, as described by Sarma *et al.* (*Genome* 43:191 (2000)), and their use can be extended to non-cereal monocots such as sugarcane, grasses, and lilies.

Given the synteny between rice and other cereal genomes, the nucleotide sequences of the present invention can be used to obtain molecular markers for mapping and, potentially,
20 for positional cloning. Kilian *et al.* described the use of probes from the rice genomic region of interest to isolate a saturating number of polymorphic markers in barley, which were shown to map to syntenic regions in rice and barley, suggesting that the nucleotide sequences of the invention derived from the rice genome would be useful in positional cloning of syntenic genes of interest from other cereal species that are involved in regulating flowering time, and
25 whole plant architecture. (Kilian, *et al.*, *Nucl Acids Res* 23:2729 (1995); Kilian, *et al.*, *Plant Mol Biol* 35:187 (1997)). Synteny between rice and barley has recently been reported in the area of the carrying malting quality QTLs (Han, *et al.*, *Genome* 41:373 (1998)), and use of synteny between cereals for positional cloning efforts is likely to add considerable value to rice genome analysis. Likewise, mapping of the ligules region of sorghum was facilitated
30 using molecular markers from a syntenic region of the rice genome. (Zwick, *et al.*, *Genetics* 148:1983 (1998)).

Rice marker technology utilizing the nucleotide sequences of the present invention can also be used to identify QTL alleles for modified heading time from a wild relative of cultivated rice, for example as described by Xiao, *et al.* (*Genetics* 150:899 (1998)). Wild relatives of domesticated plants represent untapped pools of genetic resources for abiotic and biotic stress resistance, apomixis and other breeding strategies, plant architecture, determinants of yield, secondary metabolites, and other valuable traits. In rice, Xiao *et al.* (supra) used molecular markers to introduce an average of approximately 5% of the genome of a wild relative, and the resulting plants were scored for phenotypes such as plant height, panicle length and 1000-grain weight. Trait-improving alleles were found for all phenotypes except plant height, where any change is considered negative. Of the 35 trait-improving alleles, Xiao *et al.* found that 19 had no effect on other phenotypes whereas 16 had deleterious effects on other traits. The nucleotide sequences of the invention such as those provided in the Sequence Listing can be employed as molecular markers to identify QTL alleles regulating heading time in plants from a wild relative, by which these valuable traits can be introgressed from wild relatives using methods including, but not limited to, that described by Xiao *et al.* ((1998) supra). Accordingly, the nucleotide sequences of the invention can be employed in a variety of molecular marker technologies for yield improvement.

Following the procedures described above to identify polymorphisms, and using a plurality of the nucleotide sequences of the invention, any individual (or line) can be genotyped. Genotyping a large number of DNA polymorphisms such as single nucleotide polymorphisms (SNPs), in breeding lines makes it possible to find associations between certain polymorphisms or groups of polymorphisms, and certain phenotypes. In addition to sequence polymorphisms, length polymorphisms such as triplet repeats are studied to find associations between polymorphism and phenotype. Genotypes can be used for the identification of particular cultivars, varieties, lines, ecotypes, and genetically modified plants or can serve as tools for subsequent genetic studies of complex traits involving multiple phenotypes.

The patent publication WO95/35505 and U.S. Patent Nos. 5,445,943 and 5,410,270 describe scanning multiple alleles of a plurality of loci using hybridization to arrays of oligonucleotides. The nucleotide sequences of the invention are suitable for use in genotyping techniques useful for each of the types of mapping discussed above.

In a preferred embodiment, the nucleotide sequences of the invention are useful for identifying and isolating a least one unique stretch of protein-encoding nucleotide sequence. The nucleotide sequences of the invention are compared with other coding sequences having sequence similarity with the sequences provided in the Sequence Listing, using a program such as BLAST. Comparison of the nucleotide sequences of the invention with other similar coding sequences permits the identification of one or more unique stretches of coding sequences encoding proteins that are involved in regulating flowering time, and whole plant architecture and that are not identical to the corresponding coding sequence being screened. Preferably, a unique stretch of coding sequence of about 25 base pairs (bp) long is identified, more preferably 25 bp, or even more preferably 22 bp, or 20 bp, or yet even more preferably 18 bp or 16 bp or 14 bp. In one embodiment, a plurality of nucleotide sequences is screened to identify unique coding sequences according to the invention. In one embodiment, one or more unique coding sequences according to the invention can be applied to a chip as part of an array, or used in a non-chip array system. In a further embodiment, a plurality of unique coding sequences according to the invention is used in a screening array. In another embodiment, one or more unique coding sequences according to the invention can be used as immobilized or as probes in solution. In yet another embodiment, one or more unique coding sequences according to the invention can be used as primers for PCR. In a further embodiment, one or more unique coding sequences according to the invention can be used as organism-specific primers for PCR in a solution containing DNA from a plurality of sources.

In another embodiment unique stretches of nucleotide sequences according to the invention are identified that are preferably about 30 bp, more preferably 50 bp or 75 bp, yet more preferably 100 bp, 150 bp, 200 bp, 250, 500 bp, 750 bp, or 1000 bp. The length of an unique coding sequence may be chosen by one of skill in the art depending on its intended use and on the characteristics of the nucleotide sequence being used. In one embodiment, unique coding sequences according to the invention may be used as probes to screen libraries to find homologs, orthologs, or paralogs. In another embodiment, unique coding sequences according to the invention may be used as probes to screen genomic DNA or cDNA to find homologs, orthologs, or paralogs. In yet another embodiment, unique coding sequences according to the invention may be used to study gene evolution and genome evolution.

Probes and primers of the invention can be used to identify and/or isolate polynucleotides related to the nucleotide sequences according to the invention provided in the Sequence Listing, or allelic variants of the nucleotide sequences according to the invention. Generally, related polynucleotides have similar sequences or encode polypeptides with similar biological activity, but are found at other loci within an organism, or are found in other organisms. Identification and isolation of related sequences can provide important tools for functional genomics, to study the evolution of genomes, and to predict gene and protein function, interaction, and regulation. Related sequences including paralogs and orthologs are particularly important in identifying Clusters of Orthologous Groups (COGs) of proteins, which aids protein function prediction and the functional and phylogenetic annotation of newly sequenced genomes.

Hybridization of the nucleotide sequences according to the invention to nucleotides obtained from other organisms can be used to identify and isolate paralogous sequences, or paralogs, which are additional members of gene families. The terms “paralogous sequence” and “paralog” as used herein encompass both full-length genes and regions and fragments thereof. Paralogs may be located in the same or a different region of the genome in which the sequence used as a probe is located. Paralogs generally have a high sequence identity with the probe sequence or the gene from which the probe was prepared; however, paralogs may have overall sequence identity with a probe sequence as low as 20 to 30% and still be recognizable as members of the same gene family with similar functions, as reported by Takata *et al.* for RAD51B paralogs (*Mol Cell Biol* 20:6476 (2000)). When overall sequence identity is not high, the sequence similarity among paralogs of a gene family is often concentrated into one or a few portions of the sequence, notably in a portion encoding a protein or RNA that has an enzymatic or structural function. The degree of identity in the amino acid sequence of the domain that defines the gene family can be as low as 20%, and is preferably at least 50%, more preferably 75%, even more preferably 80 to 95%, most preferably 85 to 99%. Paralogs may differ in their expression profiles, indicating that they may act at different time, a different place, or at a different developmental stage, even when their function appears to be similar. Differences in function among paralogs may suggest that paralogs encode polypeptides that are “remodeled” during plant evolution, for example to create new forms of

oxidized carotenoids in tomato as described by Bouvier *et al.* (*Eur J Biochem* 267:6346 (2000)).

In one embodiment, paralogs may be isolated by hybridizing an nucleotide sequence according to the invention probe to a Southern blot containing the appropriate genomic DNA or cDNA of the organism. To search for paralogs within a species, low stringency
5 hybridization is usually performed, that this will depend on size, distribution and degree of sequence divergence of domains that define the gene family. Given the resulting hybridization data, one of ordinary skill in the art could distinguish and isolate the correct DNA fragments by size, restriction sites and stated hybridization conditions from a gel or
10 from a library. Alternately, paralogs may be isolated by large-scale sequencing followed by BLAST analysis of sequences to identify putative paralogs, as described by Ospina-Giraldo *et al.* (*Fungal Genet Biol* 29:81 (2000)). In another embodiment, paralogs may be isolated using reverse-transcriptase polymerase chain reaction (RT-PCR) using primers to conserved regions of sequence. Paralogs may be cloned using standard techniques described below to screen
15 libraries using at least one Nucleotide sequence according to the invention as a probe.

An orthologous sequence, or orthologous gene, or ortholog, has a high degree of sequence similarity to a known sequence or gene of interest, with the similarity often occurring along the entire length of the coding portion of the gene. The terms “orthologous sequence” and “ortholog” as used herein encompass both full-length genes and regions and
20 fragments thereof. An ortholog often encodes a gene product that performs a similar function in the organism. Functions for orthologous genes are expected to be the same as or very similar to that of the gene from which the probe was prepared. The degree of identity is a function of evolutionary separation and, in closely related species, the degree of sequence identity can be 98 to 100%. Orthologous sequences sometimes have significantly lower levels
25 of sequence identity, for example as described by Weise *et al.* (*Plant Cell* 12:1345 (2000)) where orthologs of sucrose transporters from *Arabidopsis*, tomato, and potato had 47% similarity to the previously characterized sucrose transporter. The amino acid sequence of a protein encoded by an orthologous gene can be less than 50% identical, but tends to be at least 50%, or at least 70% or at least 80% identical, more preferably at least 90%, most preferably
30 at least 95% identical to the amino acid sequence of the reference protein.

To find orthologs, probes are hybridized to nucleotides from a species of interest under low stringency conditions and blots are then washed under conditions of increasing stringency. It is preferable that the wash stringency be such that sequences that are 85 to 100% identical will hybridize. More preferably, sequences 90 to 100% identical will hybridize and most preferably only sequences greater than 95% identical will hybridize. The low stringency condition is preferably one where sequences containing as much as 40-45% mismatches will be able to hybridize. This condition is established by $T_m - 40^\circ\text{C}$ to $T_m - 48^\circ\text{C}$ (see below). One of ordinary skill in the art will recognize that, due to degeneracy in the genetic code, amino acid sequences that are identical can be encoded by DNA sequences as little as 67% identical. Thus, it is preferable to make an overlapping series of shorter probes, on the order of 24 to 45 nucleotides, and individually hybridize them to the same arrayed library to avoid the problem of degeneracy introducing large numbers of mismatches.

In one embodiment, orthologous sequences, or orthologs, may be isolated by hybridizing a nucleotide sequence according to the invention probe to a Southern blot containing the appropriate genomic DNA or cDNA of a different organism. Alternately, orthologs may be isolated by large-scale sequencing followed by BLAST analysis of sequences to identify putative orthologous sequences and full-length orthologs. In another embodiment, orthologs may be isolated using reverse-transcriptase polymerase chain reaction (RT-PCR) using primers to conserved regions of sequence in a Nucleotide sequence according to the invention. Orthologs and/or orthologous sequences may be cloned using standard techniques described below to screen libraries using at least one nucleotide sequence according to the invention as a probe.

As evolutionary divergence increases, genome sequences also tend to diverge. Thus, one of skill will recognize that searches for orthologous genes between more divergent species will require the use of lower stringency conditions compared to searches between closely related species. Also, degeneracy is more of a problem for searches in the genome of a species more distant evolutionarily from the species that is the source of the AC probe sequences.

Identification of the relationship of nucleotide or amino acid sequences among plant species can be done by comparison of the subject nucleotide or amino acid sequence with the

nucleotide sequences according to the invention of the present application presented in the Sequence Listing.

Probes for Southern blotting to distinguish individual restriction fragments can range in size from 15 to 20 nucleotides to several thousand nucleotides. More preferably, the probe is 100 to 1000 nucleotides long for identifying members of a gene family when it is found that repetitive sequences would complicate the hybridization. For identifying an entire corresponding gene in another species, the probe is more preferably the length of the gene, typically 2000 to 10,000 nucleotides, but probes 50-1,000 nucleotides long might be used. Some genes, however, might require probes up to 15,000 nucleotides long or overlapping probes constituting the full-length sequence to span their lengths.

In one embodiment, the probe derived from nucleotide sequences according to the invention of the present invention is homogeneous, having a single sequence. In another embodiment, a probe designed to represent or identify members of a gene family having diverse sequences can be generated using PCR to amplify genomic DNA or RNA templates using primers derived from nucleotide sequences according to the invention that include sequences that define the gene family.

For identifying corresponding genes in another species, the probe for Southern blotting most preferably would be the genomic copy of the probe gene. This allows all elements of the gene to be identified in the other species. The next most preferable probe is a cDNA spanning the entire coding sequence (for example, as indicated by CDS coordinates in the Sequence Listing) which allows the entire mRNA-coding portion of the gene to be identified; in this case it is possible that some introns in the gene might be missed. Probes for Southern blotting can easily be generated from Nucleotide sequences according to the invention by making primers having the sequence at the ends of the Nucleotide sequence according to the invention and using rice (*Oryza sativa*) genomic DNA as a template. In instances where the Nucleotide sequence according to the invention includes sequence conserved among species, primers including the conserved sequence can be used for PCR with genomic DNA from a species of interest to obtain a probe. Similarly, if the Nucleotide sequence according to the invention includes a domain of interest, that portion of the Nucleotide sequence according to the invention can be used to make primers and, with appropriate template DNA, used to make a probe to identify genes containing the domain. Alternatively, the PCR products can be

resolved, for example by gel electrophoresis, and cloned and/or sequenced. In this manner, the variants of the domain among members of a gene family, both within and across species, can be examined.

The nucleotide sequences according to the invention can be used to isolate the
5 corresponding DNA from the same organism or other organisms. Either cDNA or genomic DNA can be isolated. Libraries of genomic DNA, or lambda, cosmid, BAC or YAC, or other large insert genomic library from the plant of interest can be constructed using standard molecular biology techniques as described in detail by Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989)) and by Ausubel *et al.*
10 (*Current Protocols in Molecular Biology*, Greene Publishing (1992)) with updates).

To screen a phage library, recombinant lambda clones are plated out on appropriate bacterial medium using an appropriate *E. coli* host strain. The resulting plaques are lifted from the plates using nylon or nitrocellulose filters. The plaque lifts are processed through denaturation, neutralization, and washing treatments following the standard protocols outlined
15 by Ausubel *et al.* ((1992) *supra*). The plaque lifts are hybridized to either radioactively labeled or non-radioactively labeled AC DNA at room temperature for about 16 hours, usually in the presence of 50% formamide and 5X SSC (sodium chloride and sodium citrate) buffer and blocking reagents. The plaque lifts are then washed at 42°C with 1% sodium dodecyl sulfate (SDS) and at a particular concentration of SSC. The SSC concentration used is
20 dependent upon the stringency at which hybridization occurred in the initial Southern blot analysis performed. For example, if a fragment hybridized under medium stringency such as $T_m - 20^\circ\text{C}$, then this condition is maintained or preferably adjusted to a less stringent condition such as $T_m - 30^\circ\text{C}$, to wash the plaque lifts. Positive clones showing hybridization to the probe are detected by exposure to X-ray films or chromogen formation or any other suitable
25 detection method, and subsequently isolated for purification using the same general protocol outlined above. Once the clone is purified, restriction analysis can be conducted to narrow the region corresponding to the gene of interest. Restriction analysis and succeeding subcloning steps can be done using procedures described by, for example, Sambrook *et al.* ((1989) *supra*).

30 To screen a YAC library, the procedures outlined for the lambda library are essentially similar except the YAC clones are harbored in bacterial colonies. The YAC clones are plated

out at reasonable density on nitrocellulose or nylon filters supported by appropriate bacterial medium in petri plates. Following the growth of the bacterial clones, the filters are processed through the denaturation, neutralization, and washing steps following the procedures of Ausubel *et al.* ((1992) *supra*). The same hybridization procedures for lambda library
5 screening are followed.

To isolate cDNA, similar procedures using appropriately modified vectors are employed. For instance, the library can be constructed in a lambda vector appropriate for cloning cDNA such as λ gt11. Alternatively, the cDNA library can be made in a plasmid vector. cDNA for cloning can be prepared by any of the methods known in the art, but is
10 preferably prepared as described above. Preferably, a cDNA library will include a high proportion of full-length clones.

Identification and isolation of alleles and paralogs within a species, and orthologs from other species, is particularly desirable because of their potential use as a tool for crop improvement. Many important crop traits, such as the solids content of tomatoes, result from
15 the combined interactions of the products of several genes residing at different loci in the genome. Generally, different alleles at each of these loci can make quantitative differences to the trait. By identifying and isolating numerous alleles for each locus from a single species or from different species, transgenic plants having various combinations of alleles can be created and the effects of the combinations measured. Once a more favorable ideal allele combination
20 has been identified, crop improvement can be accomplished either through biotechnological means or by traditional (conventional) breeding programs. (Tanksley *et al.*, *Science* 277:1063 (1997)). In a similar manner, substitution of at least one paralogous or orthologous sequence in at least one locus will introduce diversity at each substituted locus, and the substituted
25 sequences will contribute to a trait that is influenced by combined interactions of the products of several genes residing at different loci in the genome. When favorable combinations of substituted sequences and endogenous sequences at loci whose products interact are identified, crop improvement can be accomplished through further biotechnological manipulation or by traditional breeding programs, including sexual crossing and also apomixis.

30 The results from hybridization of the nucleotide sequences according to the invention of the invention to Southern blots containing DNA from another species can be used to

generate restriction fragment maps for the corresponding genomic regions. These maps provide additional information about the relative positions of restriction sites within fragments, further distinguishing mapped DNA from the remainder of the genome. Physical maps can be made by digesting genomic DNA with different combinations of restriction enzymes.

Sequence analysis and mapping of related sequences (paralogs and orthologs) can be used in phylogenetic analyses of the evolution of the sequences in question, including the determination of gene duplication and rearrangements. In addition, expression studies of related sequences can be used to further understand the evolutionary history and function of the paralogs and orthologs, and to suggest future uses for the sequences.

Nucleotide sequences according to the invention encompassing regulatory regions can be used to identify coordinately expressed genes by using the regulatory region of the portion of the Nucleotide sequence according to the invention as a probe. In the instances where the Nucleotide sequences according to the invention are identified as being expressed from genes that confer a particular phenotype, then the Nucleotide sequences according to the invention can also be used as probes to assay plants of different species for those phenotypes. When using Nucleotide sequences according to the invention to identify orthologous genes in other species, the practitioner will preferably adjust the amount of target DNA of each species so that, as nearly as is practical, the same number of genome equivalents is present for each species examined. This prevents faint signals from species having large genomes, and thus small numbers of genome equivalents per mass of DNA, from erroneously being interpreted as absence of the corresponding gene in the genome.

Nucleotide sequences according to the invention can be applied to substrates for use in array applications such as, but not limited to, assays of global gene expression, including to measure expression profiles of organisms under specific conditions of development, growth conditions, or for diagnostic or forensic methods. (Schena *et al.*, *Science* 270:467 (1995); DeRisi *et al.*, *Science* 278:680 (1997); Chee *et al.*, *Science* 274:610 (1996)). Array applications are crucial to all aspects of functional genomics, as arrays permit the easy, rapid, and reproducible detection of the presence and amount of large numbers of gene transcripts to generate large data sets for complex analyses of the function of genes. In one embodiment, a substrate including a plurality of Nucleotide sequences according to the invention, or

oligonucleotide primers or probes derived from Nucleotide sequences according to the invention of the invention may be used for detecting targeted sequences in a polynucleotide including any of SEQ ID NOs: 1 to SEQ ID NO: 66 for amplifying targeted sequences in a polynucleotide including any of SEQ ID NOs: 1 to SEQ ID NO: 66 or for detecting mutations
5 in the coding or in the non-coding sequences of polynucleotide including any of SEQ ID NO:1 to SEQ ID NO:66.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the
10 solid support that does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically include a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise
15 location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854 and PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced
20 using mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis. The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density
25 array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. Further presentation strategies aimed at providing arrays of nucleotides immobilized on solid supports were developed to order and display the
30 oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and

sequence information as disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in a polynucleotide including any of SEQ ID NOs: 1 to SEQ ID NO: 66 and preferably in the regulatory region of these sequences. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By “known mutations” it is meant, mutations of a polynucleotide including any of SEQ ID NOs: 1 to SEQ ID NO: 66 that have been identified using techniques known in the art.

Another technique that is used to detect mutations in a polynucleotide including any of SEQ ID NOs: 1 to SEQ ID NO: 66 is the use of a high-density DNA array, where single base mutations are encompassed by this technique. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of the genomic DNA or cDNA of interest. Thus, an array containing oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the “wild-type” nucleotide sequence, measure its amount, and detect differences between the target sequence and the reference wild-type nucleotide sequence. One such design termed a “4L tiled array”, is implemented using a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleotide target of length L is scanned for mutations with a tiled array containing 4L probes; the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a “footprint” for the probes flanking a mutation position. This technique was described by Chee *et al.* (*Science* 274:610 (1996)).

Another embodiment provides an array containing any of a variety of types of oligonucleotides, either alone or in a mixture of oligonucleotides types on an array, including for example, non-unique oligonucleotides, oligonucleotides that are identical to or complementary to a portion of a nucleotide sequence in a public database, oligonucleotides

that do not correspond to a BIOPATH protein and/or an FPD, degenerate oligonucleotides, or random oligonucleotides alone or in a mixture.

Consequently, the invention concerns an array of nucleotide molecules including one or more nucleotide sequences according to the invention useful as probes and primers.

5 Preferably, the invention concerns an array of nucleotides including at least two polynucleotides described above, or fragments or variants thereof, as probes and primers.

A computer readable medium containing one or more of the nucleotide sequences of the invention as well as methods of use for the computer readable medium are provided. This medium allows a nucleotide sequence corresponding to at least one of SEQ ID NOs provided
10 in the Sequence Listing (open reading frames or fragments thereof), to be used as a reference sequence to search against a database. This medium also allows for computer-based manipulation of a nucleotide sequence corresponding to at least one of SEQ ID NOs provided in the Sequence Listing.

15 EXAMPLES

The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified. Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described in detail in Sambrook *et al.*
20 (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989)) and by Ausubel *et al.* (*Current Protocols in Molecular Biology*, Greene Publishing (1992)).

Example 1: Isolation and sequencing of DNA fragments

25 1.1 Isolation and sequencing of genomic DNA fragments

Genomic DNA is isolated from nuclei of *Oryza sativa* L. ssp *japonica* cv Nipponbare and then sheared to produce fragments of approximately 500 bp. Using a method derived from the method of Mao *et al.* (*Genome Res* 10:982 (2000)), seeds are germinated on cheese cloth immersed in water and grown for 4-6 weeks under greenhouse conditions. After plants
30 reached a height of approximately 5-8 inches, the upper parts of the green leaves are harvested

and wrapped in aluminum foil at 4°C overnight. Leaf material is then stored at -80°C or directly used for extraction of nuclei. Intact nuclei are isolated by homogenization (in a blender for fresh material or by grinding with mortar and pestle for frozen material) in a buffer containing 10 mM Trizma base, 80 mM KCl, 10 mM EDTA, 1 mM spermidine, 1 mM spermine, 0.5 M sucrose, 0.5% Triton-X-100, 0.15% β -mercaptoethanol pH 9.5. The homogenate is filtered and nuclei recovered by gentle centrifugation using a fixed-angle rotor at 1,800 g at 4 C for 20 minutes. The pellet recovered after centrifugation is gently resuspended with the assistance of a small paint brush soaked in ice cold wash buffer and wash buffer added. Particulate matter remaining in the suspension is removed by filtering the resuspended nuclei into a 50 ml centrifuge tube through two layers of miracloth by gravity and centrifuging the filtrate at 57 g (500 rpm), 4 C for 2 minutes to remove intact cells and tissue residues. The supernatant fluid is transferred into a fresh centrifuge tube and nuclei are pelleted by centrifugation at 1,800 g, 4 C for 15 minutes in a swinging bucket centrifuge.

DNA is isolated from the nuclear preparation by phenol/chloroform extraction, as in Sambrook *et al* (supra). Isolated total genomic DNA is physically sheared (Hydroshear) to generate for generating random DNA fragments, and fragments of approximately 500 bp are recovered. DNA is eluted and the ends filled in using T₄ DNA polymerase, Klenow fragments, and dNTPs. Double-stranded DNA is linkered and cloned into a Novartis proprietary medium-copy vector derived from pSC101.

Vector inserts are amplified by PCR and sequenced using the MegaBACE sequencing system (Molecular Dynamics, Amersham). The amplification reaction is diluted before use and is not purified using an exonuclease/alkaline phosphatase procedure. Sequencing reactions are performed using DYEnamic ET Terminator Kit. The reactions contained approximately 50 ng of amplicon, DYEnamic ET Terminator premix, and 5 pmol of -40 M13 forward primer. The sequencing reaction is amplified for 30 cycles, and reaction products are concentrated and purified using ethanol precipitation. The sample is electrokinetically injected into the capillary at 3 kV for 45 sec and separated via electrophoresis at 9 kV for 120 min.

1.2 Isolation and sequencing of cDNA fragments

Construction of rice cDNA library. Total RNA is purified from rice plant tissue using standard total RNA purification methods. PolyA⁺ RNA is isolated from the total RNA using the Qiagen Oligotex mRNA purification system (Qiagen, Valencia, CA), and cDNA is
5 generated using cDNA synthesis reagents from Life Technologies (Rockville, MD). First strand cDNA synthesis is catalyzed by reverse transcriptase using oligo dT primers with a NotI restriction site. Second strand synthesis is catalyzed by DNA polymerase. An oligonucleotide linker with a SalI restriction endonuclease site is attached to the 5' end of the cDNAs using DNA ligase. The cDNAs are digested with NotI and SalI restriction
10 endonucleases and inserted into an *E. coli*-replicating plasmid harboring a selectable marker. *E. coli* is transfected with the recombinant plasmids and grown on selectable media. *E. coli* colonies are individually picked off the selectable media and placed into storage plates.

Sequencing the rice cDNA library. The DNA sequence of the cDNA cloned into the plasmid purified from an *E. coli* colony is determined using standard dideoxy sequencing
15 methods. Oligonucleotide primers respectively corresponding to plasmid DNA regions upstream of the 5' end of the cDNA insert (Forward reaction) and downstream of the 3' end of the cDNA insert (Reverse reaction) are used in the dideoxy sequencing reactions. If the DNA sequence determined as a result of the Forward and Reverse reactions from the cDNA overlapped, the two sequences could be merged into a contig using computerized analysis
20 software (Consed, University of Washington, Seattle), to assemble a full-length sequence of the cDNA. In cases case where DNA sequence from the Forward and Reverse reactions from a single clone did not overlap sufficiently to be assembled into a contig, such that there is a region of unsequenced DNA to bridge the DNA from the Forward and Reverse reaction in order to form a contig, the DNA sequence of the separating region is determined using one of
25 two dideoxy sequencing methods. In a "primer walking" approach, a primer specifically corresponding to the 3' end of the DNA sequence determined from the Forward reaction is used in a second dedeoxy sequencing reaction. The primer walking procedure is repeated until the DNA sequence that separated the original Forward and Reverse is resolved and a contig could be assembled. Alternatively, the clone harboring the cDNA is subjected to
30 transposon in vitro insertion dideoxysequencing (Epicentre, Madison, WI). In this procedure,

the insertion process is random and the result is multiple DNA sequence coverage over the targeted cDNA, where the sequences thus obtained are assembled into a contig.

5 **Example 2:** Data Analysis

Several approaches are used to identify potential candidate genes for the construction of a unigene dataset. Six different gene prediction program and training model combinations are used to identify candidate transcripts. The programs used are Genscan trained on arabidopsis and maize, Gmhmm trained on rice and Arabidopsis, and Fgenesh and Glimmer trained on
10 rice. All of the predicted CDSs are blasted against the genechip sequence again to extract the top hit predicted CDS.

Candidate transcript translations are examined for known Prosite motifs and full-length Pfam domains as evidence of accurate predictions. Additional candidate transcripts are identified from plant and fungal mRNA (BLASTN and TBLASTX) and protein (TBLASTN)
15 searches against the rice genome. Frameshifts caused by InDels and gaps between assembled sequence contigs are considered during the analysis, and software is adjusted to compensate for these potential problems. The percent coverage of each candidate transcript is summed across all overlapping similarities as the percent "evidence" for each candidate transcript. The results of gene prediction and homology searches are combined into an integrated dataset. For
20 overlapping candidate transcripts, the transcript with the highest evidence is selected. By blasting *Arabidopsis* genes that are know to be involved in flower development and timing of flower formation or the regulation of these processes against the above rice dataset the following rice homologs could be identified:

25 **Table 1:** Table 1 provides provides a subset of rice genes that show homology to known *Arabidopsis* flowering time genes and further identifies those *Arabidopsis* genes. Further provided are SEQ ID NOs of banana, wheat and maize representing nucleotide sequences that are homologous to the rice sequences show in column 1.

Rice Sub-group I (SEQ ID NO)	Rice Sub-group II (SEQ ID NO)	Banana (SEQ ID NO)	Wheat (SEQ ID NO)	Maize (SEQ ID NO)	Known Gene Name	AGI#
9	131	67	81	106	<i>AGL20 (a)</i>	At2g45660

57	133	68	82	107	<i>AGL20</i>	At2g45660
47		68	82	107	<i>AGL20</i>	At2g45660
47		68	82	107	<i>CAL (b)</i>	At1g26310
57	133	68	82	107	<i>CAL</i>	At1g26310
33		68	82	107	<i>CO</i>	At1g51600
17		69	83	-	<i>CO</i>	At1g51600
49	135	70	84	-	<i>CO</i>	At1g51600
27	129	71	85	-	<i>FCA</i>	At4g16280
9	131	67	81	106	<i>FLC</i>	At1g22770
57	133	68	82	107	<i>FLC</i>	At1g22770
47		68	82	107	<i>FLC</i>	At1g22770
29		67	86	108	<i>FLC</i>	At1g22770
5		72	87	109	<i>FRI</i>	At4g00650
7	139	73	87	109	<i>FRI</i>	At4g00650
61		74	88	-	<i>FT</i>	At1g65480
31	137	74	88	-	<i>FT</i>	At1g65480
25		75	-	-	<i>FT</i>	At1g65480
15		74	88	-	<i>FT</i>	At1g65480
11		76	89	-	<i>FT</i>	At1g65480
37		74	88	-	<i>FT</i>	At1g65480
39	127	-	90	110	<i>FWA</i>	At4g25530
41		77	91	111	<i>FWA</i>	At4g25530
53		-	92	112	<i>FWA</i>	At4g25530
65	125	-	93	113	<i>FWA</i>	At4g25530
63		-	94	114	<i>FWA</i>	At4g25530
43	141	-	95	115	<i>LHY</i>	At1g01060
45		-	95	-	<i>LHY</i>	At1g01060
19	123	76	96	-	<i>TFL</i>	At5g03840
55		-	97	-	<i>VRN1</i>	At3g18990
3		-	98	116	<i>VRN1</i>	At3g18990
21		78	99	117	<i>ZmID1</i>	maize

Example 3: Isolation and sequencing of DNA fragments

5 The ability to align regions of heterologous cereal chromosomes to the rice genome provides the opportunity to connect mapped cereal trait genes (quantitative trait loci or QTL) to the rice genome, and assign candidate genes to the traits that these genes influence. There are approximately 2,000 mapped cereal QTL available in the public domain, and many more in private databases. The ability to use the rice genome framework to assign candidate genes to maize, wheat, barley, or other cereal QTL allows association of genes with their molecular

functions. Mapped cereal QTL can be placed on the rice genome map. Positioning of molecular markers within the genome will further facilitate mapping and identification of trait-controlling genes.

A comparison of mapped rice QTL controlling flowering time with the positions of rice flowering time gene homologs make it possible to identify functions for many of these rice genes.

One such set of QTL from rice is known to control heading of rice. Heading date is a critical trait for adaptation to different cultivation areas and cropping seasons and therefore a major objective in rice breeding programs. Heading in rice is basically determined by two factors, duration of the basic vegetative growth (BVG) and photoperiod-sensitivity (PS). Several genes are known to be involved in controlling these two factors.

Seven QTL, designated Hd1-Hd7, are known to control rice heading date (Yano et al (Theoretical and Applied Genetics (1997) 95: 1025-1032); Yamamoto et al (Genetics (February 2000) 154: 885-891)).

The following Tables 2 and 3 provides a first set of rice genes that can be found in QTL Hd-3 (Table 2) and Hd-6 (Table 3) and also show homology to known *Arabidopsis* flowering time genes.

Table 2 provides a subset of rice genes from the Hd-3 QTL that show homology to known *Arabidopsis* flowering time genes and further identifies those *Arabidopsis* genes. Further provided are SEQ ID NOs of banana, wheat and maize representing nucleotide sequences that are homologous to the rice sequences show in column 1.

Rice Sub-group I (SEQ ID NO)	Rice Sub-group II (SEQ ID NO)	Banana (SEQ ID NO)	Wheat (SEQ ID NO)	Maize (SEQ ID NO)	Known Gene Name	AGI#
61		74	100	-	FT	At1g65480
37	145	74	100	-	FT	At1g65480
1		68	101	118	CAL	At1g26310
1		68	101	118	AP1	
61		74	100	-	TFL	At5g03840
37		74	100	-	TFL	At5g03840
13	145	122	102	119	FCA	At4g16280
13	145	122	102	119	FCA	At4g16280
1		68	101	118	AGL20	At2g45660

1		68	101	118	FLC(F)	At1g22770
13	145	79	102	119	FPA	AT2G43410
13		79	102	119	FPA	AT2G43410
23		-	103	120	FY	At5g13480

Table 3 provides a subset of rice genes from the Hd-6 QTL that show homology to known *Arabidopsis* flowering time genes and further identifies those *Arabidopsis* genes. Further provided are SEQ ID NOs of banana, wheat and maize representing nucleotide sequences that are homologous to the rice sequences show in column 1.

Rice Sub-group I (SEQ ID NO)	Rice Sub-group II (SEQ ID NO)	Banana (SEQ ID NO)	Wheat (SEQ ID NO)	Maize (SEQ ID NO)	Known Gene Name	AGI#
59					CAL	At1g26310
59					AP1	AtAP1
51		80	104	121	FCA	At4g16280
59					AGL20	At2g45660
59					FLC(F)_	At1g22770
51		90	104	121	FPA	AT2G43410
35	143	-	105	-	FY	At5g13480
35	143	-	105	-	FVE	At2g19520

Example 4 : *Rice Orthologs of Arabidopsis Flowering Genes Identified by Reverse Genetics*

Understanding the function of every gene is the major challenge in the age of completely sequenced eukaryotic genomes. Sequence homology can be helpful in identifying possible functions of many genes. However, reverse genetics, the process of identifying the function of a gene by obtaining and studying the phenotype of an individual containing a mutation in that gene, is another approach to identify the function of a gene.

Reverse genetics in *Arabidopsis* has been aided by the establishment of large publicly available collections of insertion mutants (Krysan et al., (1999) Plant Cell **11**, 2283-2290; Tisser et al., (1999) Plant Cell **11**, 1841-1852; Speulman et al., (1999) . Plant Cell **11**, 1853-1866; Parinov et al., (1999) . Plant Cell **11**, 2263-2270; Parinov and Sundaresan, 2000; Biotechnology **11**, 157-161). Mutations in genes of interest are identified by screening the population by PCR amplification using primers derived from sequences near the insert border

and the gene of interest to screen through large pools of individuals. Pools producing PCR products are confirmed by Southern hybridization and further deconvoluted into subpools until the individual is identified (Sussman et al., (2000) *Plant Physiology* **124**, 1465-1467).

Recently, some groups have begun the process of sequencing insertion site flanking regions from individual plants in large insertion mutant populations, in effect prescreening a subset of lines for genomic insertion sites (Parinov et al., (1999) . *Plant Cell* **11**, 2263-2270; Tisser et al., (1999) . *Plant Cell* **11**, 1841-1852). The advantage to this approach is that the laborious and time-consuming process of PCR-based screening and deconvolution of pools is avoided.

A large database of insertion site flanking sequences from approximately 100,000 T-DNA mutagenized *Arabidopsis* plants of the Columbia ecotype (GARLIC lines) is prepared. T-DNA left border sequences from individual plants are amplified using a modified thermal asymmetric interlaced-polymerase chain reaction (TAIL-PCR) protocol (Liu et al., (1995) . *Plant J.* **8**, 457-463). Left border TAIL-PCR products are sequenced and assembled into a database that associates sequence tags with each of the approximately 100,000 plants in the mutant collection. Screening the collection for insertions in genes of interest involves a simple gene name or sequence BLAST query of the insertion site flanking sequence database, and search results point to individual lines. Insertions are confirmed using PCR.

Analysis of the GARLIC insert lines suggests that there are 76,856 insertions that localize to a subset of the genome representing coding regions and promoters of 22,880 genes. Of these, 49,231 insertions lie in the promoters of over 18,572 genes, and an additional 27,625 insertions are located within the coding regions of 13,612 genes. Approximately 25,000 T-DNA left border mTAIL-PCR products (25% of the total 102,765) do not have significant matches to the subset of the genome representing promoters and coding regions, and are therefore presumed to lie in noncoding and/or repetitive regions of the genome.

The *Arabidopsis* T-DNA GARLIC insertion collection is used to investigate the roles of certain genes in the timing of flowering and whole plant architecture. Target genes are chosen using a variety of criteria, including public reports of mutant phenotypes, RNA profiling experiments, and sequence similarity to genes implicated in the timing of flowering and whole plant architecture. Plant lines with insertions in genes of interest are then identified. Each T-DNA insertion line is represented by a seed lot collected from a plant that

is hemizygous for a particular T-DNA insertion. Plants homozygous for insertions of interest are identified using a PCR assay. The seed produced by these plants is homozygous for the T-DNA insertion mutation of interest.

Homozygous mutant plants are tested for altered flowering time or whole plant architecture. The genes interrupted in these mutants contribute to the observed phenotype. The genes interrupted in these mutants interfere with the normal timing and development of the plant flower.

Rice orthologs of the *Arabidopsis* genes affecting flowering time and flower architecture in plants are identified by similarity searching of a rice database using the Double-Affine Smith-Waterman algorithm (BLASP with e values better than 10^{-10}).

Example 5 : *Cloning and Sequencing of Nucleic Acid Molecules from Rice*

Genomic DNA: Plant genomic DNA samples can be isolated from frozen tissues, according to one of the three procedures, e.g., standard procedures described by Ausubel et al. (1995), a quick leaf prep described by Klimyuk et al. (1993), or using FTA paper (Life Technologies). For the latter procedure, a piece of leaf is excised from the plant, placed on top of the FTA paper and covered with a small piece of parafilm that serves as a barrier material to prevent contamination of the crushing device. In order to drive the sap and cells from the plant tissue into the FTA paper matrix for effective cell lysis and nucleic acid entrapment, a crushing device is used to mash the tissue into the FTA paper. The FTA paper is air dried for an hour. For analysis of DNA, the samples can be archived on the paper until analysis. Two mm punches are removed from the specimen area on the FTA paper using a 2 mm Harris Micro PunchTM and placed into PCR tubes. Two hundred (200) microliters of FTA purification reagent is added to the tube containing the punch and vortexed at low speed for 2 seconds. The tube is then incubated at room temperature for 5 minutes. The solution is removed with a pipette so as to repeat the wash one more time. Two hundred (200) microliters of TE (10 mM Tris, 0.1 mM EDTA, pH 8.0) is added and the wash is repeated two more times. The PCR mix is added directly to the punch for subsequent PCR reactions.

Cloning of Candidate cDNA: A candidate cDNA is amplified from total RNA isolated from rice tissue after reverse transcription using primers designed against the computationally predicted cDNA. Primers designed based on the genomic sequence can be

used to PCR amplify the full-length cDNA (start to stop codon) from first strand cDNA prepared from rice cultivar Nipponbare tissue.

The Qiagen RNeasy kit (Qiagen, Hilden, Germany) is used for extraction of total RNA. The Superscript II kit (Invitrogen, Carlsbad, USA) is used for the reverse transcription reaction. PCR amplification of the candidate cDNA is carried out using the reverse primer sequence located at the translation start of the candidate gene in 5' - 3' direction. This is performed with high-fidelity Taq polymerase (Invitrogen, Carlsbad, USA).

The PCR fragment is then cloned into pCR2.1-TOPO (Invitrogen) or the pGEM-T easy vector (Promega Corporation, Madison, Wis., USA) per the manufacturer's instructions, and several individual clones are subjected to sequencing analysis.

DNA sequencing: DNA preps for 2-4 independent clones are minipreped following the manufacturer's instructions (Qiagen). DNA is subjected to sequencing analysis using the BigDye™ Terminator Kit according to manufacturer's instructions (ABI).

Sequencing makes use of primers designed to both strands of the predicted gene of interest.

DNA sequencing is performed using standard dye-terminator sequencing procedures and automated sequencers (models 373 and 377; Applied Biosystems, Foster City, CA). All sequencing data are analyzed and assembled using the Phred/Phrap/Consed software package (University of Washington) to an error ratio equal to or less than 10^{-4} at the consensus sequence level.

The consensus sequence from the sequencing analysis is then to be validated as being intact and the correct gene in several ways. The coding region is checked for being full length (predicted start and stop codons present) and uninterrupted (no internal stop codons). Alignment with the gene prediction and BLAST analysis is used to ascertain that this is in fact the right gene.

The clones are sequenced to verify their correct amplification.

Example 6: Functional analysis in plants

A plant complementation assay can be used for the functional characterization of the flowering time genes according to the invention.

Rice and Arabidopsis putative orthologue pairs are identified using BLAST comparisons, TFASTXY comparisons, and Double-Affine Smith-Waterman similarity

searches. Constructs containing a rice cDNA or genomic clone inserted between the promoter and terminator of the Arabidopsis orthologue are generated using overlap PCR (Gene 77, 61-68 (1989)) and GATEWAY cloning (Life Technologies Invitrogen). For ease of cloning, rice cDNA clones are preferred to rice genomic clones. A three stage PCR strategy is used to make these constructs.

(1) In the first stage, primers are used to PCR amplify: (i) 2Kb upstream of the translation start site of the Arabidopsis orthologue, (ii) the coding region or cDNA of the rice orthologue, and (iii) the 500 bp immediately downstream of the Arabidopsis orthologue's translation stop site. Primers are designed to incorporate onto their 5' ends at least 16 bases of the 3' end of the adjacent fragment, except in the case of the most distal primers which flank the gene construct (the forward primer of the promoter and the reverse primer of the terminator). The forward primer of the promoters contains on their 5' ends partial AttB1 sites, and the reverse primer of the terminators contains on their 5' ends partial AttB2 sites, for Gateway cloning.

(2) In the second stage, overlap PCR is used to join either the promoter and the coding region, or the coding region and the terminator.

(3) In the third stage either the promoter-coding region product can be joined to the terminator or the coding region-terminator product can be joined to the promoter, using overlap PCR and amplification with full Att site-containing primers, to link all three fragments, and put full Att sites at the construct termini.

The fused three-fragment piece flanked by Gateway cloning sites are introduced into the LTI donor vector pDONR201 using the BP clonase reaction, for confirmation by sequencing. Confirmed sequenced constructs are introduced into a binary vector containing Gateway cloning sites, using the LR clonase reaction such as, for example, pAS200.

The pAS200 vector was created by inserting the Gateway cloning cassette RfA into the Acc65I site of pNOV3510.

pNOV3510 was created by ligation of inverted pNOV2114 VSI binary into pNOV3507, a vector containing a PTX5' Arab Prottox promoter driving the PPO gene with the Nos terminator.

pNOV2114 was created by insertion of virGN54D (Pazour *et al.* 1992, J. Bacteriol. 174:4169-4174) from pAD1289 (Hansen *et al.* 1994, PNAS 91:7603-7607) into pHINK085.

pHiNK085 was created by deleting the 35S:PMI cassette and M13 ori in pVictorHiNK.

pPVictorHiNK was created by modifying the T-DNA of pVictor (described in WO 97/04112) to delete M13 derived sequences and to improve its cloning versatility by introducing the BIGLINK polylinker.

5 The sequence of the pVictor HiNK vector is disclosed in SEQ ID NO: 5 in WO 00/6837, which is incorporated herein by reference. The pVictorHiNK vector contains the following constituents that are of functional importance:

- The origin of replication (ORI) functional in *Agrobacterium* is derived from the *Pseudomonas aeruginosa* plasmid pVS1 (Itoh *et al.* 1984. Plasmid 11: 206-220; Itoh and Haas, 1985. Gene 36: 27-36). The pVS1 ORI is only functional in *Agrobacterium* and can be mobilised by the helper plasmid pRK2013 from *E.coli* into *A. tumefaciens* by means of a triparental mating procedure (Ditta *et al.*, 1980. Proc. Natl. Acad. Sci USA 77: 7347-7351).

- The ColE1 origin of replication functional in *E. coli* is derived from pUC19 (Yannisch-Perron *et al.*, 1985. Gene 33: 103-119).

- The bacterial resistance to spectinomycin and streptomycin encoded by a 0.93 kb fragment from transposon Tn7 (Fling *et al.*, 1985. Nucl. Acids Res. 13: 7095) functions as selectable marker for maintenance of the vector in *E. coli* and *Agrobacterium*. The gene is fused to the *tac* promoter for efficient bacterial expression (Amman *et al.*, 1983. Gene 25: 167-178).

- The right and left T-DNA border fragments of 1.9 kb and 0.9 kb that comprise the 24 bp border repeats, have been derived from the Ti-plasmid of the nopaline type *Agrobacterium tumefaciens* strains pTiT37 (Yadav *et al.*, 1982. Proc. Natl. Acad. Sci. USA. 79: 6322-6326).

25 The plasmid is introduced into *Agrobacterium tumefaciens* GV3101pMP90 by electroporation. The positive bacterial transformants are selected on LB medium containing 50 µg/µl kanamycin and 25 µg/µl gentamycin. Plants are transformed by standard methodology (e.g., by dipping flowers into a solution containing the *Agrobacterium*) except that 0.02% Silwet –77 (Lehle Seeds, Round Rock, TX) is added to the bacterial suspension

30

and the vacuum step omitted. Five hundred (500) mg of seeds are planted per 2 ft² flat of soil and , and progeny seeds are selected for transformants using PPO selection.

Primary transformants are analyzed for complementation. Primary transformants are genotyped for the Arabidopsis mutation and presence of the transgene. When possible, >50
5 mutants harboring the transgene should be phenotyped to observe variation due to transgene copy number and expression.

Example 7: Vector construction for overexpression and gene “knockout” experiments

Overexpression

10 Vectors used for expression of full-length “candidate genes” of interest in plants (overexpression) are designed to overexpress the protein of interest and are of two general types, biolistic and binary, depending on the plant transformation method to be used.

For biolistic transformation (biolistic vectors), the requirements are as follows:

- 15 1. a backbone with a bacterial selectable marker (typically, an antibiotic resistance gene) and origin of replication functional in *Escherichia coli* (*E. coli*; eg. ColE1), and
2. a plant-specific portion consisting of:
- a. a gene expression cassette consisting of a promoter (eg. ZmUBIint MOD), the gene of interest (typically, a full-length cDNA) and a transcriptional terminator
20 (eg. *Agrobacterium tumefaciens nos* terminator);
- b. a plant selectable marker cassette, consisting of a promoter (eg. rice Act1D-BV MOD), selectable marker gene (eg. phosphomannose isomerase, PMI) and transcriptional terminator (eg. CaMV terminator).

Vectors designed for transformation by *Agrobacterium tumefaciens* (*A. tumefaciens*; binary
25 vectors) consist of:

1. a backbone with a bacterial selectable marker functional in both *E. coli* and *A. tumefaciens* (eg. spectinomycin resistance mediated by the *aadA* gene) and two origins of replication, functional in each of aforementioned bacterial hosts, plus the *A. tumefaciens virG* gene;

2. a plant-specific portion as described for biolistic vectors above, except in this instance this portion is flanked by *A. tumefaciens* right and left border sequences which mediate transfer of the DNA flanked by these two sequences to the plant.

Knock out vectors

5 Vectors designed for reducing or abolishing expression of a single gene or of a family or related genes (knockout vectors) are also of two general types corresponding to the methodology used to downregulate gene expression: antisense or double-stranded RNA interference (dsRNAi).

(a) Anti-sense

10 For antisense vectors, a full-length or partial gene fragment (typically, a portion of the cDNA) can be used in the same vectors described for full-length expression, as part of the gene expression cassette. For antisense-mediated down-regulation of gene expression, the coding region of the gene or gene fragment will be in the opposite orientation relative to the promoter; thus, mRNA will be made from the non-coding (antisense) strand *in planta*.

(b) dsRNAi

15 For dsRNAi vectors, a partial gene fragment (typically, 300 to 500 basepairs long) is used in the gene expression cassette, and is expressed in both the sense and antisense orientations, separated by a spacer region (typically, a plant intron, eg. the OsSH1 intron 1, or a selectable marker, eg. conferring kanamycin resistance). Vectors of this type are designed to
20 form a double-stranded mRNA stem, resulting from the basepairing of the two complementary gene fragments *in planta*.

25 Biolistic or binary vectors designed for overexpression or knockout can vary in a number of different ways, including eg. the selectable markers used in plant and bacteria, the transcriptional terminators used in the gene expression and plant selectable marker cassettes, and the methodologies used for cloning in gene or gene fragments of interest (typically, conventional restriction enzyme-mediated or GatewayTM recombinase-based cloning). An important variant is the nature of the gene expression cassette promoter driving expression of the gene or gene fragment of interest in most tissues of the plants (constitutive, eg. ZmUBIint

MOD), in specific plant tissues (eg. maize ADP-gpp for endosperm-specific expression), or in an inducible fashion (eg. GAL4bsBz1 for estradiol-inducible expression in lines constitutively expressing the cognate transcriptional activator for this promoter).

5 **Example 8:** Insertion of a “candidate gene” involved in flower timing into a Plant Expression Vector

 A validated rice cDNA clone in pCR2.1-TOPO or the pGEM-T easy vector is subcloned using conventional restriction enzyme-based cloning into a vector, downstream of the maize ubiquitin promoter and intron, and upstream of the *Agrobacterium tumefaciens* nos 3' end
10 transcriptional terminator. The resultant gene expression cassette (promoter, “candidate gene” and terminator) is further subcloned, using conventional restriction enzyme-based cloning, into the pNOV2117 binary vector (Negrotto et al (2000) Plant Cell Reports 19, 798-803; plasmid pNOV117 disclosed in this article corresponds to pNOV2117 described herein), generating pNOVCAND.

15 The pNOVCAND binary vector is designed for transformation and over-expression of the “candidate gene” in monocots. It consists of a binary backbone containing the sequences necessary for selection and growth in *Escherichia coli* DH-5alpha (Invitrogen) and *Agrobacterium tumefaciens* LBA4404 (pAL4404; pSB1), including the bacterial spectinomycin antibiotic resistance *aadA* gene from *E. coli* transposon Tn7, origins of
20 replication for *E. coli* (ColE1) and *A. tumefaciens* (VS1), and the *A. tumefaciens* *virG* gene. In addition to the binary backbone, which is identical to that of pNOV2114 described herein previously (see Example 6 above) pNOV2117 contains the T-DNA portion flanked by the right and left border sequences, and including the PositechTM (Syngenta) plant selectable marker (WO 94/20627) and the “candidate gene” gene expression cassette. The PositechTM
25 plant selectable marker confers resistance to mannose and in this instance consists of the maize ubiquitin promoter driving expression of the PMI (phosphomannose isomerase) gene, followed by the cauliflower mosaic virus transcriptional terminator.

 Plasmid pNOV2117 is introduced into *Agrobacterium tumefaciens* LBA4404 (pAL4404; pSB1) by electroporation. Plasmid pAL4404 is a disarmed helper plasmid (Ooms
30 et al (1982) Plasmid 7, 15-29). Plasmid pSB1 is a plasmid with a wide host range that

contains a region of homology to pNOV2117 and a 15.2 kb KpnI fragment from the virulence region of pTiBo542 (Ishida et al (1996) Nat Biotechnol 14, 745-750). Introduction of plasmid pNOV2117 into Agrobacterium strain LBA4404 results in a co-integration of pNOV2117 and pSB1.

- 5 Alternatively, plasmid pCIB7613, which contains the hygromycin phosphotransferase (*hpt*) gene (Gritz and Davies, Gene 25, 179-188, 1983) as a selectable marker, may be employed for transformation.

Plasmid pCIB7613 (see WO 98/06860, incorporated herein by reference in its entirety) is selected for rice transformation. In pCIB7613, the transcription of the nucleic acid
10 sequence coding hygromycin-phosphotransferase (HYG gene) is driven by the corn ubiquitin promoter (ZmUbi) and enhanced by corn ubiquitin intron 1. The 3' polyadenylation signal is provided by NOS 3' nontranslated region.

Other useful plasmids include pNADII002 (GAL4-ER-VP16) which contains the yeast GAL4 DNA Binding domain (Keegan et al., Science, 231:699 (1986)), the mammalian
15 estrogen receptor ligand binding domain (Greene et al., Science, 231 :1150 (1986)) and the transcriptional activation domain of the HSV VP16 protein (Triezenberg et al., 1988). Both *hpt* and GAL4-ER-VP16 are constitutively expressed using the maize Ubiquitin promoter, and pSGCDL1 (GAL4BS Bz1 Luciferase), which carries the firefly luciferase reporter gene under control of a minimal maize Bronze1 (Bz1) promoter with 10 upstream synthetic GAL4
20 binding sites. All constructs use termination signals from the nopaline synthase gene.

Example 9 Plant Transformation

9.1 Rice Transformation

- 25 pNOVCAND is transformed into a rice cultivar (Kaybonnet) using Agrobacterium-mediated transformation, and mannose-resistant calli are selected and regenerated.

Agrobacterium is grown on YPC solid plates for 2-3 days prior to experiment initiation. Agrobacterial colonies are suspended in liquid MS media to an OD of 0.2 at $\lambda 600\text{nm}$. Acetosyringone is added to the agrobacterial suspension to a concentration of 200 μM and agro
30 is induced for 30min.

Three-week-old calli which are induced from the scutellum of mature seeds in the N6 medium (Chu, C.C. et al., Sci. Sin., 18, 659-668(1975)) are incubated in the agrobacterium solution in a 100 x 25 petri plate for 30 minutes with occasional shaking. The solution is then removed with a pipet and the callus transferred to a MSAs medium which is overlaid with sterile filter paper.

Co-Cultivation is continued for 2 days in the dark at 22°C.

Calli are then placed on MS-Timetin plates for 1 week. After that they are transferred to PAA + mannose selection media for 3 weeks.

Growing calli (putative events) are picked and transferred to PAA+ mannose media and cultivated for 2 weeks in light.

Colonies are transferred to MS20SorbKinTim regeneration media in plates for 2 weeks in light. Small plantlets are transferred to MS20SorbKinTim regeneration media in GA7 containers. When they reach the lid, they are transferred to soil in the greenhouse.

Expression of the "candidate gene" in transgenic T₀ plants is analyzed. Additional rice cultivars, such as but not limited to, Nipponbare, Taipei 309 and Fuzisaka 2 are also transformed and assayed for expression of the "candidate gene" product and enhanced protein expression.

9.2 Maize transformation

Alternatively, pNOVCAND is transformed into immature maize embryos. Transformation of immature maize embryos is performed essentially as described in Negrotto et al., (2000) Plant Cell Reports 19: 798-803. For this example, all media constituents are as described in Negrotto et al., *supra*. However, various media constituents described in the literature may be substituted.

1. Transformation plasmids and selectable marker

The genes used for transformation are cloned into a vector suitable for maize transformation as described in Example 17. Vectors used contain the phosphomannose isomerase (PMI) gene (Negrotto et al. (2000) Plant Cell Reports 19: 798-803).

2. Preparation of *Agrobacterium tumefaciens*

Agrobacterium strain LBA4404 (pSB1) containing the plant transformation plasmid is grown on YEP (yeast extract (5 g/L), peptone (10g/L), NaCl (5g/L), 15g/l agar, pH 6.8) solid medium for 2 to 4 days at 28°C. Approximately 0.8×10^9 *Agrobacteria* are suspended in LS-inf media supplemented with 100 µM acetosyringone (As) (Negrotto *et al.*, (2000) Plant Cell Rep 19: 798-803). Bacteria are pre-induced in this medium for 30-60 minutes.

3. Inoculation

Immature embryos from A188 or other suitable maize genotypes are excised from 8 – 12 day old ears into liquid LS-inf + 100 µM As. Embryos are rinsed once with fresh infection medium. *Agrobacterium* solution is then added and embryos are vortexed for 30 seconds and allowed to settle with the bacteria for 5 minutes. The embryos are then transferred scutellum side up to LSAs medium and cultured in the dark for two to three days. Subsequently, between 20 and 25 embryos per petri plate are transferred to LSDc medium supplemented with cefotaxime (250 mg/l) and silver nitrate (1.6 mg/l) and cultured in the dark for 28°C for 10 days.

4. Selection of transformed cells and regeneration of transformed plants

Immature embryos producing embryogenic callus are transferred to LSD1M0.5S medium. The cultures are selected on this medium for 6 weeks with a subculture step at 3 weeks. Surviving calli are transferred either to LSD1M0.5S medium to be bulked-up or to Reg1 medium. Following culturing in the light (16 hour light/ 8 hour dark regiment), green tissues are then transferred to Reg2 medium without growth regulators and incubated for 1-2 weeks. Plantlets are transferred to Magenta GA-7 boxes (Magenta Corp, Chicago Ill.) containing Reg3 medium and grown in the light. Plants that are PCR positive for the candidate gene cassette are transferred to soil and grown in the greenhouse.

Example 10: Method of modifying the gene frequency

The invention further provides a method of modifying the frequency of a gene in a plant population, including the steps of: identifying an SSR within a coding region of a gene; screening a plurality of plants using the SSR as a marker to determine the presence or absence

of the gene in an individual plant; selecting at least one individual plant for breeding based on the presence or absence of the gene; and breeding at least one plant thus selected to produce a population of plants having a modified frequency of the gene. The identification of the SSR within the coding region of a gene can be accomplished based on sequence similarity between the nucleic acid molecules of the invention and the region within the gene of interest flanking the SSR.

Example 11: Chromosomal Markers to Identify the Location of a Nucleic Acid Sequence

The sequences of the present invention can also be used for SSR mapping. SSR mapping in rice has been described by Miyao *et al.* (*DNA Res* 3:233 (1996)) and Yang *et al.* (*Mol Gen Genet* 245:187 (1994)), and in maize by Ahn *et al.* (*Mol Gen Genet* 241:483 (1993)). SSR mapping can be achieved using various methods. In one instance, polymorphisms are identified when sequence specific probes flanking an SSR contained within a sequence are made and used in polymerase chain reaction (PCR) assays with template DNA from two or more individuals or, in plants, near isogenic lines. A change in the number of tandem repeats between the SSR-flanking sequence produces differently sized fragments (U.S. Patent No. 5,766,847). Alternatively, polymorphisms can be identified by using the PCR fragment produced from the SSR-flanking sequence specific primer reaction as a probe against Southern blots representing different individuals (Refseth *et al.*, *Electrophoresis* 18:1519 (1997)). Rice SSRs can be used to map a molecular marker closely linked to functional gene, as described by Akagi *et al.* (*Genome* 39:205 (1996)).

The sequences of the present invention can be used to identify and develop a variety of microsatellite markers, including the SSRs described above, as genetic markers for comparative analysis and mapping of genomes.

Many of the polynucleotides listed in Tables 1 to 4 contain at least 3 consecutive di-, tri- or tetranucleotide repeat units in their coding region that can potentially be developed into SSR markers. Trinucleotide motifs that can be commonly found in the coding regions of said polynucleotides and easily identified by screening the polynucleotides sequences for said motifs are, for example: CGG; GCC, CGC, GGC, etc (see Table 4 for further details). Once such a repeat unit has been found, primers can be designed which are complementary to the region flanking the repeat unit and used in any of the methods described below.

Sequences of the present invention can also be used in a variation of the SSR technique known as inter-SSR (ISSR), which uses microsatellite oligonucleotides as primers to amplify genomic segments different from the repeat region itself (Zietkiewicz *et al.*, *Genomics* 20:176 (1994)). ISSR employs oligonucleotides based on a simple sequence repeat anchored or not at their 5'- or 3'-end by two to four arbitrarily chosen nucleotides, which triggers site-specific annealing and initiates PCR amplification of genomic segments which are flanked by inversely orientated and closely spaced repeat sequences. In one embodiment of the present invention, microsatellite markers as disclosed herein, or substantially similar sequences or allelic variants thereof, may be used to detect the appearance or disappearance of markers indicating genomic instability as described by Leroy *et al.* (*Electron. J Biotechnol*, 3(2), at <http://www.ejb.org> (2000)), where alteration of a fingerprinting pattern indicated loss of a marker corresponding to a part of a gene involved in the regulation of cell proliferation. Microsatellite markers are useful for detecting genomic alterations such as the change observed by Leroy *et al.* (*Electron. J Biotechnol*, 3(2), *supra* (2000)) which appeared to be the consequence of microsatellite instability at the primer binding site or modification of the region between the microsatellites, and illustrated somaclonal variation leading to genomic instability. Consequently, sequences of the present invention are useful for detecting genomic alterations involved in somaclonal variation, which is an important source of new phenotypes.

In addition, because the genomes of closely related species are largely syntenic (that is, they display the same ordering of genes within the genome), these maps can be used to isolate novel alleles from wild relatives of crop species by positional cloning strategies. This shared synteny is very powerful for using genetic maps from one species to map genes in another. For example, a gene mapped in rice provides information for the gene location in maize and wheat.

Example 12: Quantitative Trait Linked Breeding

Various types of maps can be used with the sequences of the invention to identify Quantitative Trait Loci (QTLs) for a variety of uses, including marker-assisted breeding. Many important crop traits are quantitative traits and result from the combined interactions of several genes. These genes reside at different loci in the genome, often on different chromosomes, and generally exhibit multiple alleles at each locus. Developing markers, tools,

and methods to identify and isolate the QTLs involved in a trait, enables marker-assisted breeding to enhance desirable traits or suppress undesirable traits. The sequences disclosed herein can be used as markers for QTLs to assist marker-assisted breeding. The sequences of the invention can be used to identify QTLs and isolate alleles as described by Li *et al.* in a study of QTLs involved in resistance to a pathogen of rice. (Li *et al.*, *Mol Gen Genet* 261:58 (1999)). In addition to isolating QTL alleles in rice, other cereals, and other monocot and dicot crop species, the sequences of the invention can also be used to isolate alleles from the corresponding QTL(s) of wild relatives. Transgenic plants having various combinations of QTL alleles can then be created and the effects of the combinations measured. Once an ideal allele combination has been identified, crop improvement can be accomplished either through biotechnological means or by directed conventional breeding programs. (Flowers *et al.*, *J Exp Bot* 51:99 (2000); Tanksley and McCouch, *Science* 277:1063 (1997)).

Example 13: Marker-Assisted Breeding

Markers or genes associated with specific desirable or undesirable traits are known and used in marker assisted breeding programs. It is particularly beneficial to be able to screen large numbers of markers and large numbers of candidate parental plants or progeny plants. The methods of the invention allow high volume, multiplex screening for numerous markers from numerous individuals simultaneously.

Markers or genes associated with specific desirable or undesirable traits are known and used in marker assisted breeding programs. It is particularly beneficial to be able to screen large numbers of markers and large numbers of candidate parental plants or progeny plants. The methods of the invention allow high volume, multiplex screening for numerous markers from numerous individuals simultaneously.

A multiplex assay is designed providing SSRs specific to each of the markers of interest. The SSRs are linked to different classes of beads. All of the relevant markers may be expressed genes, so RNA or cDNA techniques are appropriate. RNA is extracted from root tissue of 1000 different individual plants and hybridized in parallel reactions with the different classes of beads. Each class of beads is analyzed for each sample using a microfluidics analyzer. For the classes of beads corresponding to qualitative traits, qualitative measures of presence or absence of the target gene are recorded. For the classes of beads

corresponding to quantitative traits, quantitative measures of gene activity are recorded. Individuals showing activity of all of the qualitative genes and highest expression levels of the quantitative traits are selected for further breeding steps. In procedures wherein no individuals have desirable results for all the measured genes, individuals having the most
5 desirable, and fewest undesirable, results are selected for further breeding steps. In either case, progeny are screened to further select for homozygotes with high quantitative levels of expression of the quantitative traits.

Supporting TABLES**Table 4:** This table identifies the start and end points and the nucleotide sequence of tri-nucleotide repeat units in the coding sequence of selected flowering time genes

SeqID	Start	End	Sequence
247	612	626	CAG
805	1542	1556	GCA
1015	42	56	CGC
1223	57	80	CGG
	1484	1498	CGT
1317	588	602	TGA
1615	596	610	CGG

Table 5: Swiss-Prot Data

	Seq ID: 1		Seq ID: 15
	Accession: P29383		Accession: Q9S7R5
	Swissprot_id: AGL3_ARATH		Swissprot_id: TSF_ARATH
5	Gi_number: 3915599	50	Gi_number: 17433194
	Description: AGAMOUS-LIKE MADS BOX PROTEIN AGL3		Description: TWIN SISTER OF FT protein (TFL1 like protein)
10	Seq ID: 5	55	Seq ID: 17
	Accession: Q9FDW0		Accession: Q96502
	Swissprot_id: FRI_ARATH		Swissprot_id: COL2_ARATH
	Gi_number: 17432946		Gi_number: 17433030
	Description: FRIGIDA protein		Description: Zinc finger protein
		60	CONSTANS-LIKE 2
15	Seq ID: 7		
	Accession: Q9FDW0		Seq ID: 19
	Swissprot_id: FRI_ARATH		Accession: O82088
	Gi_number: 17432946	65	Swissprot_id: SP_LYCES
20	Description: FRIGIDA protein		Gi_number: 17367328
			Description: SELF-PRUNING protein
	Seq ID: 9		
	Accession: Q38836	70	
25	Swissprot_id: AG11_ARATH		Seq ID: 21
	Gi_number: 12229648		Accession: Q03309
	Description: Agamous-like MADS box protein AGL11		Swissprot_id: ZF46_MOUSE
			Gi_number: 547909
		75	Description: Zinc finger protein 46 (Zfp-46) (Zinc finger protein MLZ-4)
30	Seq ID: 11		
	Accession: Q9XH42		
	Swissprot_id: CET4_TOBAC		
	Gi_number: 17367229	80	Seq ID: 23
35	Description: CEN-like protein 4		Accession: Q00808
			Swissprot_id: HET1_PODAN
			Gi_number: 3023956
			Description: Vegetatible incompatibility protein HET-E-1
		85	
	Seq ID: 13		
	Accession: O59800		Seq ID: 25
40	Swissprot_id: CWF5_SCHPO		Accession: Q9SXZ2
	Gi_number: 18202094		Swissprot_id: FT_ARATH
	Description: Cell cycle control protein cwf5		Gi_number: 17432933
45		90	

	Description: FLOWERING LOCUS T protein		Description: FLOWERING LOCUS T protein
5	Seq ID: 27 Accession: O04425 Swissprot_id: FCA_ARATH Gi_number: 3023721 Description: Flowering time control protein FCA	50	Seq ID: 39 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein
10		55	ATHB-10) (HD-ZIP protein ATHB-10)
15	Seq ID: 29 Accession: Q9SZJ6 Swissprot_id: AG1Y_ARATH Gi_number: 12229674 Description: Agamous-like MADS box protein At4g37940	60	Seq ID: 41 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein
20	Seq ID: 31 Accession: Q9SXZ2 Swissprot_id: FT_ARATH Gi_number: 17432933 Description: FLOWERING LOCUS T protein	65	ATHB-10) (HD-ZIP protein ATHB-10)
25		70	
30	Seq ID: 33 Accession: P13728 Swissprot_id: SGS3_DROYA Gi_number: 134469 Description: Salivary glue protein SGS-3 precursor	75	Seq ID: 43 Accession: P97399 Swissprot_id: DSPP_MOUSE Gi_number: 17865460 Description: Dentin sialophosphoprotein precursor (Dentin matrix protein-3) (DMP-3) [Contains: Dentin phosphoprotein (Dentin phosphophoryn) (DPP); Dentin sialoprotein (DSP)]
35		80	
40	Seq ID: 35 Accession: O15736 Swissprot_id: TIPD_DICDI Gi_number: 3122952 Description: TIPD protein	85	Seq ID: 47 Accession: P29386 Swissprot_id: AGL6_ARATH Gi_number: 1351899 Description: Agamous-like MADS box protein AGL6
45	Seq ID: 37 Accession: Q9SXZ2 Swissprot_id: FT_ARATH Gi_number: 17432933	90	

5	Seq ID: 49 Accession: Q9LUA9 Swissprot_id: COLE_ARATH Gi_number: 17433067 Description: Zinc finger protein constans-like.14	40	Accession: P35632 Swissprot_id: AP3_ARATH Gi_number: 543815 Description: FLORAL HOMEOTIC PROTEIN APETALA3
10	Seq ID: 51 Accession: P98175 Swissprot_id: RBMA_HUMAN Gi_number: 12644371 Description: RNA-binding protein 10 (RNA binding motif protein 10) (DXS8237E)	45	Seq ID: 61 Accession: Q9SXZ2 Swissprot_id: FT_ARATH Gi_number: 17432933 Description: FLOWERING LOCUS T protein
15	Seq ID: 53 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10) (HD-ZIP protein ATHB-10)	50	Seq ID: 63 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10) (HD-ZIP protein ATHB-10)
20	Seq ID: 53 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10) (HD-ZIP protein ATHB-10)	55	Seq ID: 63 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10) (HD-ZIP protein ATHB-10)
25	Seq ID: 57 Accession: P29386 Swissprot_id: AGL6_ARATH Gi_number: 1351899 Description: Agamous-like MADS box protein AGL6	60	Seq ID: 65 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10) (HD-ZIP protein ATHB-10)
30	Seq ID: 57 Accession: P29386 Swissprot_id: AGL6_ARATH Gi_number: 1351899 Description: Agamous-like MADS box protein AGL6	65	Seq ID: 65 Accession: P46607 Swissprot_id: HGL2_ARATH Gi_number: 2506525 Description: Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10) (HD-ZIP protein ATHB-10)
35	Seq ID: 59	70	ATHB-10) (HD-ZIP protein ATHB-10)

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

What is claimed is:

1. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide involved in control of flowering time in rice, which nucleic acid molecule is
5 substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2 – 66 and 124 - 146.
2. The isolated nucleic acid molecule of claim 1 comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice, which
10 polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide given in SEQ ID NOs: 2 – 66 and 124 - 146.
3. The isolated nucleic acid molecule of claim 1 comprising a nucleotide sequence encoding a polypeptide which is involved in control of flowering time in rice, which induces
15 transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide as given in SEQ ID NOs: 2 – 66 and 124 - 146.
4. The isolated nucleic acid molecule of claim 1 comprising a nucleotide sequence
 - a) as given in SEQ ID NOs: 1 – 65 and and 123 - 145, or a fragment thereof
20 encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.
 - b) having substantial similarity to (a);
 - 25 c) capable of hybridizing to (a) or the complement thereof;
 - d) capable of hybridizing to a nucleic acid as given in the Sequence Listing with the SEQ ID NOs being identified in column C of table 2 or the complement thereof;
 - e) complementary to (a), (b) or (c); and
 - 30 f) which is the reverse complement of (a), (b) or (c).

5. An isolated nucleic acid molecule comprising a nucleotide sequence which maps to rice QTL Hd-3, which controls heading date in rice and encodes a polypeptide involved in control of flowering time in rice, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2, 14; 24; 38, and 62, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

6. The isolated nucleic acid molecule of claim 5 comprising a nucleotide sequence which maps to rice QTL Hd-3, which controls heading date in rice and encodes a polypeptide involved in control of flowering time in rice, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide as given in SEQ ID NOs: 2, 14; 24; 38, and 62, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

7. The isolated nucleic acid molecule of claim 5 comprising a nucleotide sequence which maps to rice QTL Hd-3, which controls heading date in rice and encodes a polypeptide involved in control of flowering time in rice, which polypeptide is immunologically reactive with antibodies raised against a polypeptide as given in SEQ ID NOs: 2, 14; 24; 38, and 62, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

8. The isolated nucleic acid molecule of claim 5 comprising a nucleotide sequence

a) as given in SEQ ID NOs: 1, 13; 23; 37, and 61, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence as given in SEQ ID NOs: 1, 13; 23; 37, and 61, , , or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

9. An isolated nucleic acid molecule comprising a nucleotide sequence which maps to rice QTL Hd-6, which controls heading date in rice and encodes a polypeptide involved in control of flowering time in rice, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 36, 52, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

10. The isolated nucleic acid molecule of claim 9 comprising a nucleotide sequence which maps to rice QTL Hd-6, which controls heading date in rice and encodes a polypeptide which is involved in control of flowering time in rice, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide as given in SEQ ID NOs: 36, 52, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide..

11. The nucleic acid molecule of claim 9 comprising a nucleotide sequence which maps to rice QTL Hd-6, which controls heading date in rice and encodes a polypeptide which is involved in control of flowering time in rice, which polypeptide is immunologically reactive with antibodies raised against a polypeptide as given in SEQ ID NOs: 36, 52, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

12. The isolated nucleic acid molecule of claim 9 comprising a nucleotide sequence
- a) as given in SEQ ID NOs: 35, 51, and 59, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.
 - b) having substantial similarity to (a);
 - c) capable of hybridizing to (a) or the complement thereof;
 - d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence as given in SEQ ID NOs: 35, 51, and 59, or the complement thereof;
 - e) complementary to (a), (b) or (c); and
 - f) which is the reverse complement of (a), (b) or (c).
13. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity and abundance of which can alter flowering time, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NO: 18; 34; and 50, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.
14. The nucleic acid molecule of claim 13 comprising a nucleotide sequence encoding a polypeptide the activity and abundance of which can alter flowering time, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO: 18; 34; and 50, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.
15. The nucleic acid molecule of claim 13 comprising a nucleotide sequence encoding a polypeptide the activity and abundance of which can alter flowering time, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 18;

34;and 50, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

- 5 16. The nucleic acid molecule of claim 13 comprising a nucleotide sequence
- a) as given in SEQ ID NO: 17; 33;and 49, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length
 - 10 polypeptide;
 - b) having substantial similarity to (a);
 - c) capable of hybridizing to (a) or the complement thereof;
 - d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of a nucleotide sequence given in SEQ ID NO: 17;
 - 15 33;and 49, or the complement thereof;
 - e) complementary to (a), (b) or (c); and
 - f) which is the reverse complement of (a), (b) or (c).

17. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a
- 20 polypeptide the activity of which functions in the posttranscriptional regulation of transcripts involved in the flowering process, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NO: 14, 28 and 52, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of
- 25 the full-length polypeptide..

18. The nucleic acid molecule of claim 17 comprising a nucleotide sequence encoding a polypeptide the activity of which functions in the posttranscriptional regulation of transcripts involved in the flowering process, which polypeptide has at least between 70%, and 99%
- 30 amino acid sequence identity to the polypeptide of SEQ ID NO: 14, 28 and 52, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at

least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide..

19. The nucleic acid molecule of claim 17 comprising a nucleotide sequence encoding
5 a polypeptide the activity of which functions in the posttranscriptional regulation of transcripts involved in the flowering process, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 14, 28 and 52, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of
10 the full-length polypeptide.

20. The isolated nucleic acid molecule of claim 17 comprising a nucleotide sequence

- a) as given in SEQ ID NO: 13, 27 and 51, or a fragment thereof encoding a
15 partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- 20 d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NO: 13, 27 and 51, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

25

21. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements and flowering time of plants, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOS: 6 and 8, or a partial-length polypeptide having
30 substantially the same activity as the full-length polypeptide, e.g., at least 50%, more

preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

22. The nucleic acid molecule of claim 21 comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements and flowering time of plants, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 6 and 8, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

23. The nucleic acid molecule of claim 21 comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements and flowering time of plants, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 6 and 8, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide

24. The isolated nucleic acid molecule of claim 21 comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 5 and 7, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 5 and 7, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

25. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulating flowering time, and whole plant architecture by modulating plant branching patterns, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NO: 20, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

26. The nucleic acid molecule of claim 25 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulating flowering time, and whole plant architecture by modulating plant branching patterns, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO: 20, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

27. The nucleic acid molecule of claim 25 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulating flowering time, and whole plant architecture by modulating plant branching patterns, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 20, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

28. The isolated nucleic acid molecule of claim 25 comprising a nucleotide sequence

- a) as given in SEQ ID NO: 19, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;

- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NO: 19, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- 5 f) which is the reverse complement of (a), (b) or (c).

29. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, which nucleic acid molecule is substantially similar
10 to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: : 12, 16, 26, 32, 38 and 62, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

15 30. The nucleic acid molecule of claim 29 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: : 12, 16, 26, 32, 38 and 62, or a partial-length polypeptide having substantially the same activity as the full-length
20 polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

31. The nucleic acid molecule claim 29 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early
25 flowering and terminal flower formation, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: : 12, 16, 26, 32, 38 and 62, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

30

32. The isolated nucleic acid molecule of claim 29 comprising a nucleotide sequence

- 5 a) as given in SEQ ID NOs : 11, 15, 25, 31, 37 and 61, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- 10 d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 11, 15, 25, 31, 37 and 61, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

15 33. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2, 10, 48, 58, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

20

34. The nucleic acid molecule of claim 33 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 2, 10, 48, 58, and 60, or a partial-length polypeptide having

25 substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

35. The nucleic acid molecule of claim 33 comprising a nucleotide sequence encoding

30 a polypeptide the activity of which is involved in regulation of flowering time, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ

ID NOs: 2, 10, 48, 58, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

- 5 36. The isolated nucleic acid molecule of claim 33 comprising a nucleotide sequence
- a) as given in SEQ ID NOs: 1, 9, 47, 57, and 59, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length
 - 10 polypeptide;
 - b) having substantial similarity to (a);
 - c) capable of hybridizing to (a) or the complement thereof;
 - d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 1, 9, 47,
 - 15 57, and 59; or the complement thereof;
 - e) complementary to (a), (b) or (c); and
 - f) which is the reverse complement of (a), (b) or (c).

- 20 37. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in specifying flower-meristem identity, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2, 48, 58, and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

- 25 38. The nucleic acid molecule of claim 37 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in specifying flower-meristem identity, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 2, 48, 58, and 60, or a partial-length polypeptide having
- 30 substantially the same activity as the full-length polypeptide, e.g., at least 50%, more

preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

39. The nucleic acid molecule of claim 37 comprising a nucleotide sequence
5 encoding a polypeptide the activity of which is involved in specifying flower-meristem
identity, which polypeptide is immunologically reactive with antibodies raised against a
polypeptide of SEQ ID NOs: 2, 48, 58, and 60, or a partial-length polypeptide having
substantially the same activity as the full-length polypeptide, e.g., at least 50%, more
preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-
10 length polypeptide.

40. The isolated nucleic acid molecule of claim 37 comprising a nucleotide sequence
a) as given in SEQ ID NOs: 1, 47, 57, and 59, or a fragment thereof encoding a
partial-length polypeptide having substantially the same activity as the full-
15 length polypeptide, e.g., at least 50%, more preferably at least 80%, even
more preferably at least 90% to 95% the activity of the full-length
polypeptide;
b) having substantial similarity to (a);
c) capable of hybridizing to (a) or the complement thereof;
20 d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or
more consecutive nucleotides of nucleotides given in SEQ ID NOs: 1, 47, 57,
and 59, or the complement thereof;
e) complementary to (a), (b) or (c); and
f) which is the reverse complement of (a), (b) or (c).

25
41. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a
polypeptide which is capable of modulating flowering time, which nucleic acid molecule is
substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 2, 10;
30, 48; 58 and 60, or a partial-length polypeptide having substantially the same activity as the
30 full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably
at least 90% to 95% the activity of the full-length polypeptide.

42. The nucleic acid molecule of claim 41 comprising a nucleotide sequence encoding a polypeptide which is capable of modulating flowering time, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 2, 10; 30, 48; 58 and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

43. The nucleic acid molecule of claim 41 comprising a nucleotide sequence encoding a polypeptide which is capable of modulating flowering time, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 2, 10; 30, 48; 58 and 60, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

44. The isolated nucleic acid molecule of claim 41 comprising a nucleotide sequence
- a) as given in SEQ ID NOs: 1, 9; 29, 47; 57 and 59, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
 - b) having substantial similarity to (a);
 - c) capable of hybridizing to (a) or the complement thereof;
 - d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 1, 9; 29, 47; 57 and 59, or the complement thereof;
 - e) complementary to (a), (b) or (c); and
 - f) which is the reverse complement of (a), (b) or (c).

45. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early

flowering and terminal flower formation, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NO:22, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

46. The nucleic acid molecule of claim 45 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NO: 22, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

47. The nucleic acid molecule of claim 45 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in flowering time and capable of inducing early flowering and terminal flower formation, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NO: 22, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

48. The isolated nucleic acid molecule of claim 45 comprising a nucleotide sequence

- a) as given in SEQ ID NO: 21, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;

- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NO: 21, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- 5 f) which is the reverse complement of (a), (b) or (c).

49. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ
10 ID NOs: 44 and 46, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

50. The nucleic acid molecule of claim 49 comprising a nucleotide sequence encoding
15 a polypeptide the activity of which is involved in regulation of flowering time, which polypeptide has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 44 and 46, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

20

51. The nucleic acid molecule of claim 49 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 44 and 46, or a partial-length polypeptide having substantially the same activity as
25 the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

52. The isolated nucleic acid molecule of claim 49 comprising a nucleotide sequence
30 a) as given in SEQ ID NOs: 43 and 45, or a fragment thereof encoding a partial-length polypeptide having substantially the same activity as the full-length

polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide,

- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- 5 d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 43 and 45, or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

10

53. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements of plants and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given
15 in SEQ ID NOs: 4 and 56, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

20

54. The nucleic acid molecule of claim 53 comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements of plants and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence identity to the polypeptide of SEQ ID NOs: 4 and 56, with any individual number within this range of between 70% and 99% also being part of the invention, or a
25 partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

30

55. The nucleic acid molecule of claim 53 comprising a nucleotide sequence encoding a polypeptide the activity of which can modulate the vernalization requirements of plants and, optionally, a regulatory region, which induces transcription of the coding region, which

polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 4 and 56, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

5

56. The nucleic acid molecule of claim 53 comprising a nucleotide sequence

- a) as given in SEQ ID NOs: 3 and 55, or a part thereof, which still encodes, a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even
10 more preferably at least 90% to 95% the activity of the full-length polypeptide.;
- b) having substantial similarity to (a);
- c) capable of hybridizing to (a) or the complement thereof;
- d) capable of hybridizing to a nucleic acid molecule comprising 50 to 200 or
15 more consecutive nucleotides of a nucleotide sequence given in SEQ ID NOs: 3 and 55 or the complement thereof;
- e) complementary to (a), (b) or (c); and
- f) which is the reverse complement of (a), (b) or (c).

- 20 57. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which nucleic acid molecule is substantially similar to a nucleic acid encoding a polypeptide as given in SEQ ID NOs: 14 and 52, or a partial-length polypeptide having substantially the same activity as the
25 full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

- 30 58. The nucleic acid molecule of claim 57 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is substantially similar, and preferably has at least between 70%, and 99% amino acid sequence

identity to the polypeptide of SEQ ID NOs: 14 and 52, with any individual number within this range of between 70% and 99% also being part of the invention, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

59. The nucleic acid molecule of claim 57 comprising a nucleotide sequence encoding a polypeptide the activity of which is involved in regulation of flowering time and, optionally, a regulatory region, which induces transcription of the coding region, which polypeptide is immunologically reactive with antibodies raised against a polypeptide of SEQ ID NOs: 14 and 52, or a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide.

60. The nucleic acid molecule of claim 57 comprising a nucleotide sequence
- a) as given in SEQ ID NOs: 13 and 51, or a part thereof, which still encodes a partial-length polypeptide having substantially the same activity as the full-length polypeptide, e.g., at least 50%, more preferably at least 80%, even more preferably at least 90% to 95% the activity of the full-length polypeptide;
 - b) having substantial similarity to (a);
 - c) capable of hybridizing to (a) or the complement thereof;
 - d) capable of hybridizing to a nucleic acid comprising 50 to 200 or more consecutive nucleotides of nucleotides given in SEQ ID NOs: 13 and 51 or the complement thereof;
 - e) complementary to (a), (b) or (c); and
 - f) which is the reverse complement of (a), (b) or (c).

61. A polypeptide which has an amino acid sequence encoded by any one of the polynucleotides according to claims 1 to 60.

62. A polypeptide according to claim 61, which has an amino acid sequence encoded by a polynucleotide which has a nucleotide sequence that is substantially similar to any one of the sequences selected from the group consisting of SEQ ID NOs: 1 to 65 and 124 - 146.

5 63. A polypeptide according to claim 62 wherein said polypeptide has at least 90% amino acid sequence identity to the polypeptide of SEQ ID NOs: 2 - 66 or 124 - 146 .

64. A recombinant vector comprising a polynucleotide of any of claims 1 to 60.

10 65. An expression cassette comprising as operably linked components, a promoter, a polynucleotide of any of claims 1-60 and a termination sequence.

66. A host cell comprising all or parts of a vector and/or an expression cassette of claims 64 or 65.

15

67. The host cell of claim 66 wherein said host cell is a bacterial cell, a yeast cell, an animal cell or a plant cell.

68. The host cell of claim 67, wherein said plant cell is from a cereal plant.

20

69. A plant comprising a host cell of any of claims 66 - 68.

70. A plant according to claim 69, wherein said plant is selected from the group consisting of maize, soybean, barley, alfalfa, sunflower, tomato, banana, canola, cotton,
25 peanut, sorghum, tobacco, sugarbeet, wheat, and rice.

71. A seed which is derived from a plant according to any one of claims 69 and 70.

72. A method of modulating flowering time and/or whole plant architecture of a plant,
30 comprising functionally integrating an isolated nucleic acid molecule according to any one of claims 1 to 60 comprising a nucleic acid sequence encoding a polypeptide, which is involved

in or associated with regulation of flowering time and/or whole plant architecture and the expression of which is up-regulated or down-regulated during flowering time, into a cell, group of cells, tissue or organ of a plant.

5 73. A method of modulating flowering time and/or whole plant architecture of a plant, comprising:

a) functionally integrating

i. an isolated nucleic acid molecule according to anyone of claims 1 to 56, or a portion thereof in an anti-sense orientation; or

10 ii. an dsRNAi construct comprising an isolated nucleic acid molecule according to anyone of claims 1 to 56, or a portion thereof in both a sense and an anti-sense orientation, which portions, optionally, may be separated by a spacer region;

b) under the transcriptional control of regulatory sequences required for

15 expression in plants, into a cell, group of cells, tissue or organ of a plant; and

c) expressing the constructs as provided in a) above in a cell, group of cells, a tissue or organ of a plant to produce a RNA transcript.

20 74. A method of identifying or isolating polynucleotide sequences that are orthologous to a nucleic acid molecule according to anyone of claims 1 to 60 comprising a nucleic acid fragment encoding a polypeptide that involved in or associated with the regulation of flowering time and/or whole plant architecture, from the genome of another plant, wherein all or a portion of a particular nucleic acid sequence according to anyone of claims 1 to 56 is used as a probe that selectively hybridizes to gene sequences present in a
25 population of cloned genomic DNA fragments or cDNA fragments from a chosen source organism.

30 75. A method for detecting the presence of a polynucleotide according to any one of claims 1 to 60, or a fragment or a variant thereof, or a complementary sequence thereto in a sample, the method including the following steps of:

- a) bringing into contact a nucleotide probe or a plurality of nucleotide probes which can hybridize with a polynucleotide according to any one of claims 1 to 56, or a fragment or a variant thereof, or a complementary sequence thereto and the sample to be assayed.
- 5 b) detecting the hybrid complex formed between the probe and a nucleotide in the sample.

76. A kit for detecting the presence of a polynucleotide according to any one of claims 1 to 60, or a fragment or a variant thereof, or a complementary sequence thereto in a sample, the kit including a nucleotide probe or a plurality of nucleotide probes which can hybridize with a nucleotide sequence comprised within a polynucleotide according to any one of claims 1 to 56, or a fragment or a variant thereof, or a complementary sequence thereto and, optionally, the reagents necessary for performing the hybridization reaction.

- 15 77. A method of modifying the frequency of a flowering time gene in a plant population, comprising the steps of:
- d) screening a plurality of plants using an oligonucleotide as a marker to determine the presence or absence of a flowering time gene in an individual plant, the oligonucleotide consisting of not more than 300 bases of a nucleotide sequence according to any one of claims 1 to 56,
- 20 b) selecting at least one individual plant for breeding based on the presence or absence of the flowering time gene; and
- c) breeding at least one plant thus selected to produce a population of plants having a modified frequency of the flowering time gene.

25

78. A method according to claim 77, wherein the oligonucleotide comprises a simple sequence repeat (SSR) sequence comprising at least two consecutive repeat units of an SSR, the start and end points of which are provided in Table 4, and a flanking sequence of at least about 14 nucleic acids immediately adjacent to said at least two consecutive repeat units.

30

79. A method of plant breeding to select for or against a trait of interest which is associated with the regulation of flowering time and/or whole plant architecture, comprising the steps of:

- a) identifying the trait of interest; identifying at least one oligonucleotide that can be used as a marker for the trait, the oligonucleotide consisting of not more than 300 bases of a nucleotide sequence according to any one of claims 1-60,
- b) screening at least one plant for the presence of the at least one oligonucleotide;
- c) selecting at least one plant based on presence or absence of the at least one oligonucleotide;
- d) breeding at least one plant thus selected to produce a population of plants having a modified frequency of the at least one oligonucleotide; and
- e) screening at least one plant of the population for the presence or absence of the flowering time trait.

80. A method according to claim 79, wherein the oligonucleotide comprises a simple sequence repeat (SSR) sequence comprising at least two consecutive repeat units of an SSR, the start and end points of which are provided in Table 4, and a flanking sequence of at least about 14 nucleic acids immediately adjacent to said at least two consecutive repeat units.

81. A method of determining a varietal identity of a plant, comprising:

- a) obtaining a nucleic acid sample from a plant;
- b) identifying at least one oligonucleotide to obtain an oligonucleotide profile for the plant, wherein the oligonucleotide consists of not more than 300 bases of a nucleotide sequence according to any one of claims 1 to 56, the oligonucleotide comprising a simple sequence repeat (SSR) sequence comprising at least two consecutive repeat units of an SSR, the start and end points of which are provided in Table 4, and a flanking sequence of at least about 14 nucleic acids immediately adjacent to said at least two consecutive repeat units in the sample; and

- c) comparing the SSR profile to at least one known SSR profile corresponding to at least one known variety to determine the varietal identity of the plant.

5 82. An oligonucleotide primer consisting of between 8 and 150 bases which comprises at least 14 bases selected from the group of flanking sequences obtainable from a nucleotide sequence provided in SEQ ID NOs: 1 to 65, which at least 14 bases are immediately adjacent to at least two consecutive repeat units of an SSR.

10 83. A computer-readable medium having stored thereon a data structure comprising:
a) sequence information of a polynucleotide according to any one of claims 1 - 60; ; and
b) a module receiving the nucleic acid molecule which compares the nucleic acid sequence of the molecule to at least one other nucleic acid sequence.

15

SEQUENCE LISTING

<110> Syngenta Participations AG

<120> IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES

<130> S-60147A

<140>

<141>

<150> US 60/300,112

<151> 2001-06-22

<150> US 60/325,277

<151> 2001-09-26

<150> US 60/334,984

<151> 2001-11-30

<160> 146

<170> PatentIn Ver. 2.2

<210> 1

<211> 747

<212> DNA

<213> Oryza sativa

<400> 1

atggggcgag ggaaagtaga gctgaaagcg gatcgagaac aagataagcc ggcaggtgac	60
gttcgcgaag aggaggaacg ggctgctgaa gaaggcgtag gagctgtccg tgctctgcga	120
cgccgaggtc gccctcatca tcttctccac ccgcggccgc ctcttcgagt tctccacctc	180
ctcctggtat gtacaagaca ctggagcgat accgcagttg caactacaac cttaactcat	240
gtgaagcatc tgctgcactg gaaactgaac taagcaatta ccaagagtag ttaaagttaa	300
agacaagagt tgagttccta caaacaactc agagcagaaa tcttcttggc gaggacttgg	360
ttccacttag cttgaaggag ctcgagcaac ttgagaacca gatcgagata tccctcatga	420
atatcaggtc atcaaagaga atcaacagtt gcttgatcaa gtatttgagc tcaaacgtaa	480
ggaacaacaa cttcaagatg ctaataaaga cttaaaaagg aagttaaaaa gggaattgca	540
gttctttcag cttaactaaa aatctttttt attgcaacag atacaagaaa ctagtggaga	600
aaatatgctt catatatctt gccaaagatgt agggcccagt ggccatgcta gtgaagctaa	660

ccaagagttt ctccatcatg caatttgtga cccttccctg catatagggg tatcaagctt 720

acatggatca cctcaaccaa tgaatga 747

<210> 2

<211> 248

<212> PRT

<213> Oryza sativa

<220>

<221> X_region

<222> (146)..(146)

<223> Xaa = any naturally occuring amino acid

<220>

<221> X_region

<222> (186)..(186)

<223> Xaa = any naturally occuring amino acid

<400> 2

Met Gly Arg Gly Lys Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Ser
1 5 10 15

Arg Gln Val Thr Phe Ala Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
20 25 30

Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
35 40 45

Ser Thr Arg Gly Arg Leu Phe Glu Phe Ser Thr Ser Ser Gly Met Tyr
50 55 60

Lys Thr Leu Glu Arg Tyr Arg Ser Cys Asn Tyr Asn Leu Asn Ser Cys
65 70 75 80

Glu Ala Ser Ala Ala Leu Glu Thr Glu Leu Ser Asn Tyr Gln Glu Tyr
85 90 95

Leu Lys Leu Lys Thr Arg Val Glu Phe Leu Gln Thr Thr Gln Ser Arg
100 105 110

Asn Leu Leu Gly Glu Asp Leu Val Pro Leu Ser Leu Lys Glu Leu Glu
115 120 125

Gln Leu Glu Asn Gln Ile Glu Ile Ser Leu Met Asn Ile Arg Ser Ser
130 135 140

Lys Xaa Asn Gln Gln Leu Leu Asp Gln Val Phe Glu Leu Lys Arg Lys
145 150 155 160

Glu Gln Gln Leu Gln Asp Ala Asn Lys Asp Leu Lys Arg Lys Leu Lys

	165		170		175
Arg Glu Leu Gln Phe Phe Gln Leu Asn Xaa Lys Ser Phe Leu Leu Gln					
	180		185		190
Gln Ile Gln Glu Thr Ser Gly Glu Asn Met Leu His Ile Ser Cys Gln					
	195		200		205
Asp Val Gly Pro Ser Gly His Ala Ser Glu Ala Asn Gln Glu Phe Leu					
	210		215		220
His His Ala Ile Cys Asp Pro Ser Leu His Ile Gly Val Ser Ser Leu					
	225		230		235
					240
His Gly Ser Pro Gln Pro Met Asn					
	245				

<210> 3

<211> 693

<212> DNA

<213> Oryza sativa

<400> 3

atggcggagg cgcatcgta cgaggagcag cggaggcggc aggttgaggc gaacaagcgg	60
aagctggagg agctccgctt ccaccacctc tccgccgccc tccgcgaatc cgcctccaag	120
ccgtcgccgg tgaagcagcg gaagaggaag gcgcgtgccc tgccggggcgc cggcgaggat	180
gccccgctcc ggcggtcggc ccgcgtcgcc aacctccccg agaagcccaa gtaccgagac	240
gagttccaag atttcgagaa gaggataagg aggtcttatg gcggcaagag gagggatttg	300
tccaatcgag ttacgcgac cgacgagcaa agggactacg ccatcaacgc tgctcaagag	360
ctggaagagg agctcggctc cgactaccca atcttcgtca agcccatgct tcagtcccat	420
gtcactggag gggtctggct gagcctgccg acgcatttca gccggaagta tctgccgaag	480
cgcgacgaga ctatccgttt ggtggatgag gaggatgatg agtttgacac cctctatctt	540
gccaacaaga gggggctcag tggaggggtg aggggggtttt ccattgctca taagctagtt	600
gatggagatt gtttggtggt ccaacttatc cagcgcacga agttcaaggt ctacataatc	660
agagcaagtt cctactatga aactgacgac tga	693

<210> 4

<211> 230

<212> PRT

<213> Oryza sativa

<400> 4

Met Ala Glu Ala Ala Ser Tyr Glu Glu Gln Arg Arg Arg Gln Val Glu
 1 5 10 15
 Ala Asn Lys Arg Lys Leu Glu Glu Leu Arg Leu His His Leu Ser Ala
 20 25 30
 Ala Val Arg Glu Ser Ala Ser Lys Pro Ser Pro Val Lys Gln Arg Lys
 35 40 45
 Arg Lys Ala Arg Ala Leu Pro Gly Ala Gly Glu Asp Ala Pro Leu Arg
 50 55 60
 Arg Ser Gly Arg Val Ala Asn Leu Pro Glu Lys Pro Lys Tyr Arg Asp
 65 70 75 80
 Glu Phe Gln Asp Phe Glu Lys Arg Ile Arg Arg Ser Tyr Gly Gly Lys
 85 90 95
 Arg Arg Asp Leu Ser Asn Arg Val Tyr Ala Thr Asp Glu Gln Arg Asp
 100 105 110
 Tyr Ala Ile Asn Ala Ala Gln Glu Leu Glu Glu Glu Leu Gly Ser Asp
 115 120 125
 Tyr Pro Ile Phe Val Lys Pro Met Leu Gln Ser His Val Thr Gly Gly
 130 135 140
 Phe Trp Leu Ser Leu Pro Thr His Phe Ser Arg Lys Tyr Leu Pro Lys
 145 150 155 160
 Arg Asp Glu Thr Ile Arg Leu Val Asp Glu Glu Asp Asp Glu Phe Asp
 165 170 175
 Thr Leu Tyr Leu Ala Asn Lys Arg Gly Leu Ser Gly Gly Trp Arg Gly
 180 185 190
 Phe Ser Ile Ala His Lys Leu Val Asp Gly Asp Cys Leu Val Phe Gln
 195 200 205
 Leu Ile Gln Arg Thr Lys Phe Lys Val Tyr Ile Ile Arg Ala Ser Ser
 210 215 220
 Tyr Tyr Glu Thr Asp Asp
 225 230

<210> 5

<211> 1632

<212> DNA

<213> Oryza sativa

<400> 5

atgtccgaca tggagtccat gacagccctt atggagtcta caggctccaa gctgcagctg	60
cttcagcgtg catttgccga gcttgagagc cagagtgtg tttcattgaa cttaaagtgg	120
aagcaacttg aggaccattt ccatggcctt gagcagtcgc ttaagaaaaa gtttgatgag	180
ctgaaagagc aggaaaagga gttcaaggag acagtcacga aatcggagca gatgcttgag	240
caacaagagg ctgctgtcgt ggcaaaggaa ctgacttctc ttgaaagact gcagcagaaa	300
agggatgtg cgttggctat gatcttcggc aagtcaaagc tgaacttgct tatgcctgta	360
atcaacccaa tcagcaaata tgtgagtaac aatgctgtgt taaatggcaa catcgctagc	420
ctgtggccca aacctgcaac tgcgcatggg gcatacctgc aagatggaaa cactgctgtg	480
aagcctcgtt ctcaacttgt tatcctctgt gaagaaatga atgttaacgg gcttcataag	540
ttcatatcag acaatcgcaa agacttaaca tccatccgtg aggaaattcc cgttgcaatt	600
aggggagcaa ctgatccata tggcttggtg ttagcttctc tagaagactt ctattttgga	660
gacaatctta tattggatgg taaaaaggat ggaaaccttc tgggtgtacg gaggacgtgc	720
ctcatgttga tggaatcact tgctcagctg caaactgatg ctactactgg tttcatatca	780
aaaggccaaa tgcttactgc tagtattaag gagagagcga aaaagatagc gcttgagtgg	840
aaatccaagt tggatagcct tgactttgat gctagcaatg ggaactgcct ggaagcacat	900
gcgtttcttc aactcctggc tacctttgct atattttctg aatttgcgga agacgagctg	960
tgtaagcttc tcccatcagt cagtcgtcgt cgccagacac ctgagctttg tagaatactt	1020
gggttgctac agaacatgcc aggtgtcatt ggagttttga ttgaaaatgg gagaactatt	1080
gatgcaatta atttggctta tgcatttgag ctgactgatc agtttgaacc agtcgaatta	1140
ctgaaagctt atctgaagga ggtcaagagt atgtcacatg tcaagaccgg aaagatgtct	1200
ccgggagtac agaatgaaat taatgaacgc gagttgtctg ctttgaaagc tgttatcaaa	1260
tgtattgagg agcacaaact tgatgagaaa taccctatcg accacttca gagaagagtc	1320
atccagctag agaaagccaa ggcagacaag aggagggctg ttgaagctgg aaagccacag	1380
tctaagaggc ctcgtgccaa cggatcagta tatgcgccac atattactag cttttctgac	1440
aaaagcttct accaggccgc tgcgccacag aggatttct acccttacga gaggcagtat	1500
gtatatggtg ctgaggccca tcatcaccca acaatgataa gctcagctcc ctatggcatg	1560
tcacctgccc acacaacata ctatggtaat ggctatcagg ttcagtacca gttgtatcaa	1620

tccactagct ag

1632

<210> 6
 <211> 543
 <212> PRT
 <213> Oryza sativa

<400> 6

```

Met Ser Asp Met Glu Ser Met Thr Ala Leu Met Glu Ser Thr Gly Ser
1           5           10           15

Lys Leu Gln Leu Leu Gln Arg Ala Phe Ala Glu Leu Glu Ser Gln Ser
          20           25           30

Ala Val Ser Leu Asn Leu Lys Trp Lys Gln Leu Glu Asp His Phe His
          35           40           45

Gly Leu Glu Gln Ser Leu Lys Lys Lys Phe Asp Glu Leu Lys Glu Gln
          50           55           60

Glu Lys Glu Phe Lys Glu Thr Val Thr Lys Ser Glu Gln Met Leu Glu
65           70           75           80

Gln Gln Glu Ala Ala Val Val Ala Lys Glu Leu Thr Ser Leu Glu Arg
          85           90           95

Leu Gln Gln Lys Arg Asp Ala Ala Leu Ala Met Ile Phe Gly Lys Ser
          100          105          110

Lys Leu Asn Leu Ser Met Pro Val Ile Asn Pro Ile Ser Lys Ser Val
          115          120          125

Ser Asn Asn Ala Val Leu Asn Gly Asn Ile Ala Ser Leu Trp Pro Lys
          130          135          140

Pro Ala Thr Ala His Gly Ala Tyr Leu Gln Asp Gly Asn Thr Ala Val
145          150          155          160

Lys Pro Arg Ser Gln Leu Val Ile Leu Cys Glu Glu Met Asn Val Asn
          165          170          175

Gly Leu His Lys Phe Ile Ser Asp Asn Arg Lys Asp Leu Thr Ser Ile
          180          185          190

Arg Glu Glu Ile Pro Val Ala Leu Arg Gly Ala Thr Asp Pro Tyr Gly
          195          200          205

Leu Val Leu Ala Ser Leu Glu Asp Phe Tyr Phe Gly Asp Asn Leu Ile
          210          215          220

Leu Asp Gly Lys Lys Asp Gly Asn Leu Leu Gly Val Arg Arg Thr Cys
225          230          235          240

```

Leu Met Leu Met Glu Ser Leu Ala Gln Leu Gln Thr Asp Ala Thr Thr
 245 250 255
 Gly Phe Ile Ser Lys Gly Gln Met Leu Thr Ala Ser Ile Lys Glu Arg
 260 265 270
 Ala Lys Lys Ile Ala Leu Glu Trp Lys Ser Lys Leu Asp Ser Leu Asp
 275 280 285
 Phe Asp Ala Ser Asn Gly Asn Cys Leu Glu Ala His Ala Phe Leu Gln
 290 295 300
 Leu Leu Ala Thr Phe Ala Ile Phe Ser Glu Phe Ala Glu Asp Glu Leu
 305 310 315 320
 Cys Lys Leu Leu Pro Ser Val Ser Arg Arg Arg Gln Thr Pro Glu Leu
 325 330 335
 Cys Arg Ile Leu Gly Leu Ser Gln Asn Met Pro Gly Val Ile Gly Val
 340 345 350
 Leu Ile Glu Asn Gly Arg Thr Ile Asp Ala Ile Asn Leu Ala Tyr Ala
 355 360 365
 Phe Glu Leu Thr Asp Gln Phe Glu Pro Val Glu Leu Leu Lys Ala Tyr
 370 375 380
 Leu Lys Glu Val Lys Ser Met Ser His Val Lys Thr Gly Lys Met Ser
 385 390 395 400
 Pro Gly Val Gln Asn Glu Ile Asn Glu Arg Glu Leu Ser Ala Leu Lys
 405 410 415
 Ala Val Ile Lys Cys Ile Glu Glu His Lys Leu Asp Glu Lys Tyr Pro
 420 425 430
 Ile Asp Pro Leu Gln Arg Arg Val Ile Gln Leu Glu Lys Ala Lys Ala
 435 440 445
 Asp Lys Arg Arg Ala Val Glu Ala Gly Lys Pro Gln Ser Lys Arg Pro
 450 455 460
 Arg Ala Asn Gly Ser Val Tyr Ala Pro His Ile Thr Ser Phe Ser Asp
 465 470 475 480
 Lys Ser Phe Tyr Gln Ala Ala Ala Pro Gln Arg His Ser Tyr Pro Tyr
 485 490 495
 Glu Arg Gln Tyr Val Tyr Gly Ala Glu Ala His His His Pro Thr Met
 500 505 510
 Ile Ser Ser Ala Pro Tyr Gly Met Ser Pro Ala His Thr Thr Tyr Tyr
 515 520 525

Gly Asn Gly Tyr Gln Val Gln Tyr Gln Leu Tyr Gln Ser Thr Ser
 530 535 540

<210> 7
 <211> 1377
 <212> DNA
 <213> Oryza sativa

<400> 7
 atgtccgaca tggagtccat ggctacactt atggagtcca caggctccaa gctgcagcag 60
 cttcagcgtg catttgcaga gctcgagagc cagagcgtg tttccttgaa cttaaagtgg 120
 aagcaactgg aggaccattt ccatgggctt gagcagtcgc tcaagaaaaa gtttgatgat 180
 ctgaaaagggc aggaagagga gtctgaggag acggttgcaa agtcagagca gatgctggag 240
 caacaagagg ctgttgtcgt ggcaaaggaa ctgacttctc ttgaaaaact gcagcagaaa 300
 agggatgccg cgttggctgt gatctttggc aagtccaagc tgaacttgct tacgcctctt 360
 atcaacccaa tcagcaaadc cgtgaataac aatgctgtgt ttaatggaaa catcgggtgg 420
 agtctttctg ttaagtggcc caaacctgca actgcgcgat gtgcatacct gcaagatgaa 480
 aacactgctg tgaagcctcg ttctcaactt gttgtcctct gtgaagagat gaatgttaac 540
 gggcttcata agttcatatc agacaatcgc aaagacttaa catccatccg tgaggaaaac 600
 cccgttgcac ttaggggagc aactgatcca tatggcttgg tgtagcttc tctagaagac 660
 ttctattttg gagacaatct tatattggat ggcaaaaagg atggaaacct tctgggtgta 720
 cggaggacgt gcctcatgtt gatggaatca cttgctcagc tgcaaactga tgctactact 780
 ggtttcatac cggaagggtc agtgcttact gcaagtatca aggagcgtgc aaaaaagatt 840
 gctcttgagt ggaaatccaa gttggatagc cttgactttg atgctagcaa tgggaactgc 900
 ctggaagcac atgcgtttct tcaactcctg gctacctttg gtatatttgc tgaatttgca 960
 caagatgagc tgtgtaagct acttcctca gtcagtcgtc gtcgtcagac acctgagctt 1020
 tgtagaatac ttggattgtc ccagaacatg ccagggtgtc ttggagttct gggtgaaaat 1080
 gggagaacta ttgatgcaat taatttggct tatgcatttg agctgactaa tcagtttgaa 1140
 ccggtggaat tactgaaagc ctatctacag gaggtcaaga gtgtgccaca tttcaagact 1200
 ggaaagatat ctctgcaagt acagaatgaa atgaatgaac gagagctgtc tgccttgaaa 1260
 gctgctatta agtgatcga ggagcacaaa cttgatgaga aatatccgat tgacctgctt 1320

cagaaaagag ttatccagct agagaaagct aaggcagaac aagaggaggg ctgttga 1377

```
<210> 8
<211> 458
<212> PRT
<213> Oryza sativa
```

<400> 8

Met 1	Ser	Asp	Met	Glu 5	Ser	Met	Ala	Thr	Leu 10	Met	Glu	Ser	Thr	Gly	Ser
Lys	Leu	Gln	Gln 20	Leu	Gln	Arg	Ala	Phe 25	Ala	Glu	Leu	Glu	Ser	Gln	Ser
Ala	Val	Ser 35	Leu	Asn	Leu	Lys	Trp 40	Lys	Gln	Leu	Glu	Asp 45	His	Phe	His
Gly	Leu 50	Glu	Gln	Ser	Leu	Lys 55	Lys	Lys	Phe	Asp	Asp 60	Leu	Lys	Arg	Gln
Glu 65	Glu	Glu	Phe	Glu	Glu 70	Thr	Val	Ala	Lys	Ser 75	Glu	Gln	Met	Leu	Glu 80
Gln	Gln	Glu	Ala	Val 85	Val	Val	Ala	Lys	Glu 90	Leu	Thr	Ser	Leu	Glu 95	Lys
Leu	Gln	Gln	Lys 100	Arg	Asp	Ala	Ala	Leu 105	Ala	Val	Ile	Phe	Gly 110	Lys	Ser
Lys	Leu	Asn 115	Leu	Ser	Thr	Pro	Leu 120	Ile	Asn	Pro	Ile	Ser 125	Lys	Ser	Val
Asn 130	Asn	Asn	Ala	Val	Phe	Asn 135	Gly	Asn	Ile	Gly	Gly 140	Ser	Leu	Ser	Val
Lys 145	Trp	Pro	Lys	Pro	Ala 150	Thr	Ala	His	Gly	Ala 155	Tyr	Leu	Gln	Asp	Glu 160
Asn	Thr	Ala	Val	Lys 165	Pro	Arg	Ser	Gln	Leu 170	Val	Val	Leu	Cys	Glu 175	Glu
Met	Asn	Val	Asn 180	Gly	Leu	His	Lys	Phe 185	Ile	Ser	Asp	Asn	Arg 190	Lys	Asp
Leu	Thr	Ser 195	Ile	Arg	Glu	Glu	Ile 200	Pro	Val	Ala	Leu	Arg 205	Gly	Ala	Thr
Asp 210	Pro	Tyr	Gly	Leu	Val	Leu 215	Ala	Ser	Leu	Glu	Asp 220	Phe	Tyr	Phe	Gly
Asp	Asn	Leu	Ile	Leu	Asp	Gly	Lys	Lys	Asp	Gly	Asn	Leu	Leu	Gly	Val

```
<210> 9
<211> 824
<212> DNA
<213> Oryza sativa

<400> 9
atggggaggg gcaggattga gatcaagagg atcgagaaca cgacaagccg ccaggtgacc 60
ttctgcaagc gccgcaacgg acttctcaag aaggcgtatg agctctccgt cctctgcgat 120
```



```

gccgaggtgg ctctcatcgt cttctccagc cgtggccgcc tctacgagta ctccaacaac 180
aacaatgtga aggctacaat tgacaggtac aagaaggcgc atgcttgtgg ctcaacttct 240
ggtgcacctc tcatagaggt caatgctcag agcaatacta ccagcaggag tctgccaaac 300
tgcgccacca gattcagatg ctgcaaaaca ccaacaaaag gcacctgggtt ggcgataatg 360
tgagcaacct gtcactgaag gagctgaagc aacttgaaag ccgcctggag aaaggcattt 420
caaagatcag agccaggaag agaatgaact gctggcttca gagatcaatt acatggccaa 480
aaggggagat tgagcttcag aacgacaaca tggacctcag aaccaagcag attgctgagg 540
aggagcagca gctgcagcag gtgacggtgg cccggtcggc cgccatggag ctgcaggctg 600
cggcggcggc gcagcagcag cagcagaatc cgttcgcggt ggcggcggcg cagctggaca 660
tgaagtgctt cttcccgttg aacctgttcg aggcggcggc gcaggtgcag gccgtggcgg 720
cgcagcgcca gcagatcatc cccaccgagc tcaacctcgg ctaccaccac caccttgcca 780
ttcccggcgc cgccgccgcc gacgcgccgc ctctcactt ctga 824

```

```

<210> 10
<211> 274
<212> PRT
<213> Oryza sativa

```

```

<220>
<221> X_region
<222> (91)..(91)
<223> Xaa = any naturally occurring amino acid

```

```

<220>
<221> X_region
<222> (148)..(148)
<223> Xaa = any naturally occurring amino acid

```

```

<400> 10

```

```

Met Gly Arg Gly Arg Ile Glu Ile Lys Arg Ile Glu Asn Thr Thr Ser
1           5           10           15
Arg Gln Val Thr Phe Cys Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
          20           25           30
Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Val Phe
          35           40           45
Ser Ser Arg Gly Arg Leu Tyr Glu Tyr Ser Asn Asn Asn Asn Val Lys

```

50	55	60
Ala Thr Ile Asp Arg Tyr Lys Lys Ala His Ala Cys Gly Ser Thr Ser		
65	70	75 80
Gly Ala Pro Leu Ile Glu Val Asn Ala Gln Xaa Gln Tyr Tyr Gln Gln		
	85	90 95
Glu Ser Ala Lys Leu Arg His Gln Ile Gln Met Leu Gln Asn Thr Asn		
	100	105 110
Lys Arg His Leu Val Gly Asp Asn Val Ser Asn Leu Ser Leu Lys Glu		
	115	120 125
Leu Lys Gln Leu Glu Ser Arg Leu Glu Lys Gly Ile Ser Lys Ile Arg		
	130	135 140
Ala Arg Lys Xaa Asn Glu Leu Leu Ala Ser Glu Ile Asn Tyr Met Ala		
	145	150 155 160
Lys Arg Glu Ile Glu Leu Gln Asn Asp Asn Met Asp Leu Arg Thr Lys		
	165	170 175
Gln Ile Ala Glu Glu Glu Gln Gln Leu Gln Gln Val Thr Val Ala Arg		
	180	185 190
Ser Ala Ala Met Glu Leu Gln Ala Ala Ala Ala Ala Gln Gln Gln Gln		
	195	200 205
Gln Asn Pro Phe Ala Val Ala Ala Ala Gln Leu Asp Met Lys Cys Phe		
	210	215 220
Phe Pro Leu Asn Leu Phe Glu Ala Ala Ala Gln Val Gln Ala Val Ala		
	225	230 235 240
Ala Gln Arg Gln Gln Ile Ile Pro Thr Glu Leu Asn Leu Gly Tyr His		
	245	250 255
His His Leu Ala Ile Pro Gly Ala Ala Ala Ala Asp Ala Pro Pro Pro		
	260	265 270

His Phe

<210> 11
 <211> 520
 <212> DNA
 <213> Oryza sativa

 <220>
 <221> misc_feature
 <222> (493)..(493)
 <223> y = cytosine or thymine

```

<400> 11
atgtcgtcgg cgaacagcct ggtgctgggg cgggtgatcg gcgacgtggt ggacctgttc      60
tcgccggagg tgacgtccg ggtgatgtac aacggcgtgc gggtcgtcaa cggcgaggac      120
ctccggccgt cggcggtgtc ggcgaggccc agcgtcgagg tgatggtgga tccagatgct      180
ccaaacccaa gcaatccgac gttgagagag tacttacact gggtgcacag gttggtgaca      240
gatattcctg gaacaactga tgcgaactat ggggcgcgag gtggtgtgct acgagagccc      300
cgggccagcg gcggggatcc accgggtggc ggtggtgctg ttccggcaga tggcgcgcgg      360
cggcgtggac cagccgccgc tgctccgcc caacttctcc acccgcggtc tcgccgacga      420
ccacgccctc ggggcccccg tcgcgcgcgc cttcttcacc tgcaagcccg agggcggcac      480
cggcggccgc cgyttccggc caccgtcacg gcatagctag                               520

```

```

<210> 12
<211> 173
<212> PRT
<213> Oryza sativa

```

```

<220>
<221> X_region
<222> (165)..(165)
<223> Xaa = any naturally occurring amino acid

```

```

<400> 12

Met Ser Ser Ala Asn Ser Leu Val Leu Gly Arg Val Ile Gly Asp Val
1          5          10          15

Val Asp Leu Phe Ser Pro Glu Val Thr Leu Arg Val Met Tyr Asn Gly
20          25          30

Val Arg Val Val Asn Gly Glu Asp Leu Arg Pro Ser Ala Val Ser Ala
35          40          45

Arg Pro Ser Val Glu Val Met Val Asp Pro Asp Ala Pro Asn Pro Ser
50          55          60

Asn Pro Thr Leu Arg Glu Tyr Leu His Trp Leu His Arg Leu Val Thr
65          70          75          80

Asp Ile Pro Gly Thr Thr Asp Ala Asn Tyr Gly Ala Arg Gly Gly Val
85          90          95

Leu Arg Glu Pro Pro Ala Ser Gly Gly Asp Pro Pro Gly Gly Gly Gly
100         105         110

```

Ala Val Pro Ala Asp Gly Ala Arg Arg Arg Gly Pro Ala Ala Ala Ala
 115 120 125

Pro Ala Thr Thr Ser Pro Pro Ala Ala Ser Pro Thr Thr Thr Pro Ser
 130 135 140

Gly Pro Pro Ser Pro Pro Pro Ser Ser Pro Ala Ser Pro Arg Ala Ala
 145 150 155 160

Pro Ala Ala Ala Xaa Ser Gly His Arg His Gly Ile Ala
 165 170

<210> 13
 <211> 1486
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (1350)..(1350)
 <223> s = guanine or cytosine

<400> 13
 atggctcacc ggctgctgcg agacgcgcag gcggacgggt gggagcggtc cgacttcccc 60
 atcatctgcg agtcctgcct cggcgacaac ccctacgtcc gcatgattta ctgtctttct 120
 tgtttgtctt ggcagttgag agcagaatat gataaagagt gcaaaatatg tgcacgtccg 180
 ttactgtct tccgttggcg acctggccgg gatgcaagat acaagaagac agagatctgc 240
 cagacatggt gcaaactgaa aaatgtctgc caggtctgcc tgctagatct tgaatatggt 300
 ctaccgggcc aggttaggga tacagcactc gctattaatt caaatgatgc aatcccaagg 360
 agtgatgtga accgtgagta ctttgcagaa gagcatgata gcaagaggcc agggctggaa 420
 tagattatga ttcttcccat gggaaggcac gtcccaatga caccattctg aagcttcaga 480
 ggacggcacc atattacaag aggaatcgag ctcatgtctg cagtttctat gtgcgtggtg 540
 aatgtacaag aggtgctgaa tgcccttata gacatgagat gcctgaaact ggagagttat 600
 ctcagcagaa cattaaagat cgttactatg gtacagagtt aatgatccag ttgctttgaa 660
 acttttgggt aaggccggtg agatgccatc cctgacacct ccagacgatg agagtataag 720
 gacactttac attggtggcc ttaataatcg aatcaccgag caggatttga gggatcaatt 780
 ttacgcacat ggtgaaattg agtccataag gatggttctg caacgtgcct gtgcatttgt 840
 gacgtacact acaagagaag gtgctgagaa agctgcagag gagcttgcca acaagttagt 900

```

catcaagggg atacgcctga agctcatgtg gggtaagcct caagcaccaa agccagaaga 960
cgatgaagct ggaaggcaag ggcatgttgc acatggagga atgctcccca gggctgttat 1020
atctcagcag cagagtgggtg accagcctca accacctgga atggaggggac aacagcaagc 1080
accatcgggg tcatactact tcaacattcc agcaccgcca ggagcagagc ggacactgta 1140
cccttcaatg gatccccaga ggatgggtgc cttggtcaag tcacaggagg gtgatggcaa 1200
acccggtcca cagcaggcag cccaagctca ggcatcaagc agctcaggac agagttaccc 1260
tatgccacca caatattacc atggtcagta cccaccctac taccaccgt atggcggtta 1320
catgccccca cctggcatgc catacccacs gccaccgcag tatccgcctt atcagccgat 1380
gttggcaaca ccagcacagt cgcaagctag ctcatcaca caacctgcac cagcaacgct 1440
gcaccaggct caagtcctc ctccgcagca gacgaccag aactga 1486

```

```

<210> 14
<211> 495
<212> PRT
<213> Oryza sativa

```

```

<220>
<221> X_region
<222> (73)..(73)
<223> Xaa = any naturally occurring amino acid

```

```

<220>
<221> X_region
<222> (451)..(451)
<223> Xaa = any naturally occurring amino acid

```

```

<400> 14

```

```

Met Ala His Arg Leu Leu Arg Asp Ala Gln Ala Asp Gly Trp Glu Arg
1           5           10           15

```

```

Ser Asp Phe Pro Ile Ile Cys Glu Ser Cys Leu Gly Asp Asn Pro Tyr
          20           25           30

```

```

Val Arg Met Ile Tyr Cys Leu Ser Cys Leu Ser Trp Gln Leu Arg Ala
          35           40           45

```

```

Glu Tyr Asp Lys Glu Cys Lys Ile Cys Ala Arg Pro Phe Thr Val Phe
          50           55           60

```

```

Arg Trp Arg Pro Gly Arg Asp Ala Xaa Arg Tyr Lys Lys Thr Glu Ile
          65           70           75           80

```

Cys Gln Thr Cys Cys Lys Leu Lys Asn Val Cys Gln Val Cys Leu Leu
 85 90 95
 Asp Leu Glu Tyr Gly Leu Pro Val Gln Val Arg Asp Thr Ala Leu Ala
 100 105 110
 Ile Asn Ser Asn Asp Ala Ile Pro Arg Ser Asp Val Asn Arg Glu Tyr
 115 120 125
 Phe Ala Glu Glu His Asp Arg Lys Ala Arg Ala Gly Ile Asp Tyr Asp
 130 135 140
 Ser Ser His Gly Lys Ala Arg Pro Asn Asp Thr Ile Leu Lys Leu Gln
 145 150 155 160
 Arg Thr Ala Pro Tyr Tyr Lys Arg Asn Arg Ala His Val Cys Ser Phe
 165 170 175
 Tyr Val Arg Gly Glu Cys Thr Arg Gly Ala Glu Cys Pro Tyr Arg His
 180 185 190
 Glu Met Pro Glu Thr Gly Glu Leu Ser Gln Gln Asn Ile Lys Asp Arg
 195 200 205
 Tyr Tyr Gly Tyr Arg Val Asn Asp Pro Val Ala Leu Lys Leu Leu Gly
 210 215 220
 Lys Ala Gly Glu Met Pro Ser Leu Thr Pro Pro Asp Asp Glu Ser Ile
 225 230 235 240
 Arg Thr Leu Tyr Ile Gly Gly Leu Asn Asn Arg Ile Thr Glu Gln Asp
 245 250 255
 Leu Arg Asp Gln Phe Tyr Ala His Gly Glu Ile Glu Ser Ile Arg Met
 260 265 270
 Val Leu Gln Arg Ala Cys Ala Phe Val Thr Tyr Thr Thr Arg Glu Gly
 275 280 285
 Ala Glu Lys Ala Ala Glu Glu Leu Ala Asn Lys Leu Val Ile Lys Gly
 290 295 300
 Ile Arg Leu Lys Leu Met Trp Gly Lys Pro Gln Ala Pro Lys Pro Glu
 305 310 315 320
 Asp Asp Glu Ala Gly Arg Gln Gly His Val Ala His Gly Gly Met Leu
 325 330 335
 Pro Arg Ala Val Ile Ser Gln Gln Gln Ser Gly Asp Gln Pro Gln Pro
 340 345 350
 Pro Gly Met Glu Gly Gln Gln Gln Ala Pro Ser Gly Ser Tyr Tyr Phe
 355 360 365

Asn Ile Pro Ala Pro Pro Gly Ala Glu Arg Thr Leu Tyr Pro Ser Met
370 375 380

Asp Pro Gln Arg Met Gly Ala Leu Val Lys Ser Gln Glu Gly Asp Gly
385 390 395 400

Lys Pro Gly Pro Gln Gln Ala Ala Gln Ala Gln Ala Ser Ser Ser Ser
405 410 415

Gly Gln Ser Tyr Pro Met Pro Pro Gln Tyr Tyr His Gly Gln Tyr Pro
420 425 430

Pro Tyr Tyr Pro Pro Tyr Gly Gly Tyr Met Pro Pro Pro Gly Met Pro
435 440 445

Tyr Pro Xaa Pro Pro Gln Tyr Pro Pro Tyr Gln Pro Met Leu Ala Thr
450 455 460

Pro Ala Gln Ser Gln Ala Ser Ser Ser Gln Gln Pro Ala Pro Ala Thr
465 470 475 480

Leu His Gln Ala Gln Val Pro Pro Pro Gln Gln Thr Thr Gln Asn
485 490 495

<210> 15

<211> 522

<212> DNA

<213> Oryza sativa

<400> 15

atggccaacg attcattggc tacagggcgt gtgatcggag atgtcctgga tcccttcac 60

agcaccgtcg atctcaccgt catgtatggg gatgatggca tgccgggtcat aagcggcgtg 120

gagcttcgcg caccggcggg cgcgagaaaa ccgggtggtcg aagtcggggg agacgatctt 180

cgcgtcgcat atactctggg aggtgatggg tgatcctgat gcacctaacc ctagcaatcc 240

aactctgagg gaatacctcc actggggagg gaggtggtgt gctacgagag cccgaaccgg 300

acgacgggga tccacaggat ggtgctggtg ctgttccggc agctggggag ggagacggtg 360

tacgcgccgg cgggtgcgca caacttcacc acccgcgctt tcgcccgcgg ctacaacctc 420

ggcgcgcccc tcgcccgggt ctacttcaac tgccagcgcc aggcgggctc cggcggccgg 480

aggttcaccg gaccttacac ctcccggcgc cgccaagcct aa 522

<210> 16

<211> 80

<212> PRT

<213> Oryza sativa

<220>
 <221> X_region
 <222> (45)..(45)
 <223> Xaa = any naturally occurring amino acid

<400> 16

Gly	Leu	Ala	Ala	Ala	Gly	Gly	Val	Arg	Ser	Gly	Glu	Pro	Pro	Ala	Ala
1				5					10					15	
Gly	Ala	Gly	Leu	Ala	Leu	Ala	Val	Glu	Val	Asp	Gly	Gly	Asp	Gly	Arg
			20					25					30		
Ala	Glu	Val	Val	Ala	Ala	Gly	Glu	Gly	Ala	Gly	Gly	Xaa	Ser	Cys	Gly
		35					40					45			
Ala	Pro	Pro	Ala	Arg	Thr	Pro	Ser	Pro	Ser	Pro	Ala	Ala	Gly	Thr	Ala
	50					55					60				
Pro	Ala	Pro	Ser	Cys	Gly	Ser	Pro	Ser	Ser	Gly	Ser	Gly	Ser	Arg	Ser
65					70					75				80	

<210> 17
 <211> 699
 <212> DNA
 <213> Oryza sativa

<400> 17

atgacgttcg	tttcggcggg	tggcggagcc	ggtcagcggc	ggccgcggag	gagatcgtcg	60
ctggctcagt	ggtctttggc	ttcgacgatt	ggagtaaggg	gtgagacagc	atcacgggtg	120
ccgcggcggg	gagcggggcg	acgcttgccg	tggcagggga	tggaggaggt	cacggtgggtg	180
aggggagggg	gatcgatctg	ggcgggagga	gaggatttgg	ggaaggagag	cacccgtccc	240
cgagcttggtg	gcggcgatc	cactggcctg	gaggaggagg	cgctacatgt	ggcggcggtg	300
gagatgacaa	ggggcgtttc	gatgtcatcg	ctggagaaca	tgagcacggt	cccagactgc	360
ggtgtgccag	acatcaccac	ctcctacttg	aggtcatcca	agagcaccat	cgacctcttc	420
acggcagcag	ctggctcccc	tgtggcggcg	cacagcataa	tgctgccacc	gcagttcatg	480
ggcgccattg	acagggaggg	caggggtgcac	cggtacaggg	agaagaggaa	gacgaggcgg	540
ttcgagaaga	ccatcaggta	cgcgtcgcgg	aaggcgtatg	ccgagacccg	gccgaggatc	600
aagggacgct	tcgccaagcg	cagcgacacc	gaccttgagg	tcgaccagta	cttctccacc	660
accgccgact	ccagctgcgg	cgtcgtgccc	acgttctag			699

<210> 18
 <211> 232
 <212> PRT
 <213> Oryza sativa

<400> 18

```

Met Thr Phe Val Ser Ala Gly Gly Gly Ala Gly Gln Arg Arg Pro Arg
1          5          10          15

Arg Arg Ser Ser Leu Ala Gln Trp Ser Leu Ala Ser Thr Ile Gly Val
          20          25          30

Arg Gly Glu Thr Ala Ser Arg Val Pro Arg Arg Gly Ala Gly Arg Arg
          35          40          45

Leu Arg Trp Gln Gly Met Glu Glu Val Thr Val Val Arg Gly Gly Gly
          50          55          60

Ser Ile Trp Ala Gly Gly Glu Asp Leu Gly Lys Glu Ser Thr Arg Pro
65          70          75          80

Arg Ala Cys Gly Gly Val Ser Thr Gly Leu Glu Glu Glu Ala Leu His
          85          90          95

Val Ala Ala Val Glu Met Thr Arg Gly Val Ser Met Ser Ser Leu Glu
          100          105          110

Asn Met Ser Thr Val Pro Asp Cys Gly Val Pro Asp Ile Thr Thr Ser
          115          120          125

Tyr Leu Arg Ser Ser Lys Ser Thr Ile Asp Leu Phe Thr Ala Ala Ala
          130          135          140

Gly Ser Pro Val Ala Ala His Ser Ile Met Ser Pro Pro Gln Phe Met
145          150          155          160

Gly Ala Ile Asp Arg Glu Ala Arg Val His Arg Tyr Arg Glu Lys Arg
          165          170          175

Lys Thr Arg Arg Phe Glu Lys Thr Ile Arg Tyr Ala Ser Arg Lys Ala
          180          185          190

Tyr Ala Glu Thr Arg Pro Arg Ile Lys Gly Arg Phe Ala Lys Arg Ser
          195          200          205

Asp Thr Asp Leu Glu Val Asp Gln Tyr Phe Ser Thr Thr Ala Asp Ser
          210          215          220

Ser Cys Gly Val Val Pro Thr Phe
225          230

```

<210> 19
 <211> 525
 <212> DNA
 <213> Oryza sativa

<400> 19
 atgtcgaggg tgctggagcc tctcattgtg gggaaggtga tcggcgaggt gctggacaac 60
 ttcaacccca cgggtgaagat gacggccacc tacggcgcca acaagcaggt gttcaacggc 120
 cagcagttct tcccctccgc cgtcgccggc aagccgcgcg tcgaggtcca gggcggcgac 180
 ctcaggtcct tcttcacatt ggtgatgact gaccctgatg tgccagggcc tagtgatcca 240
 tacctgaggg agcatcttca ctgggattgt tactgatatt cctggggact actgatgcct 300
 cttttgggga gggaggtggt gagctacgag agcccgcggc caaacatccg gctccacagg 360
 ttcacccggg tgctgttccg gcagaagcgc cggcaggcgg tgagcccgcc gccgtcgagg 420
 gaccgcttca gcaccgcga gttcgccgag gacaacgacc tcggcctccc cgtcgccgcc 480
 gtctacttca acgcgagcgc cgagaccgcc gctcgccgcc gctaa 525

<210> 20
 <211> 174
 <212> PRT
 <213> Oryza sativa

<220>
 <221> X_region
 <222> (41)..(42)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (55)..(55)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (71)..(71)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (80)..(80)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (107)..(107)
 <223> Xaa = any naturally occurring amino acid

<400> 20

Arg Pro Pro Thr Ala Pro Thr Ser Arg Cys Ser Thr Ala Thr Ser Ser
 1 5 10 15

Ser Pro Pro Pro Ser Pro Ala Ser Arg Ala Ser Arg Ser Arg Ala Ala
 20 25 30

Thr Ser Gly Pro Ser Ser His Trp Xaa Xaa Leu Thr Leu Met Cys Gln
 35 40 45

Gly Leu Val Ile His Thr Xaa Gly Ser Ile Phe Thr Gly Ile Val Thr
 50 55 60

Asp Ile Pro Gly Asp Tyr Xaa Cys Leu Phe Trp Gly Gly Arg Trp Xaa
 65 70 75 80

Ala Thr Arg Ala Arg Gly Gln Thr Ser Gly Ser Thr Gly Ser Ser Gly
 85 90 95

Cys Cys Ser Gly Arg Ser Ala Gly Arg Arg Xaa Ala Arg Arg Arg Arg
 100 105 110

Gly Thr Ala Ser Ala Pro Ala Ser Ser Pro Arg Thr Thr Thr Ser Ala
 115 120 125

Ser Pro Ser Pro Pro Ser Thr Ser Thr Arg Ser Ala Thr Ala Ser Ala
 130 135 140

Pro Ala Ser Ser Pro Arg Thr Thr Thr Ser Ala Ser Pro Ser Pro Pro
 145 150 155 160

Ser Thr Ser Thr Arg Ser Ala Arg Pro Pro Leu Ala Ala Ala
 165 170

<210> 21
 <211> 1479
 <212> DNA
 <213> Oryza sativa

<400> 21
 atgtcacaca catcagagga agaatccttg aacaacctcc agcagcaacc gaagctagag 60
 gcttgcgctg ctggatcaag caaaggtgac actgtcatgc ctgtggtgaa gaagagacga 120
 ggtcacccag gaaacccaga tccagatgtt gaggtcgtag ctttgtcacc taagacactc 180

```

ctggcaacaa acaggtatat atgcgaggtc tgccacaagg gcttccagag agaccagaac 240
ctccaacttc ataggagagg ccacaacct ccatggaagc tgaagcagag aagcagcaca 300
gaagctaaga agaaggtgta tgtctgccct gagattacct gccctcatca tgatgcaacc 360
cgagcttttg gcgatctcac aggcattaag aaacactact ccaggaagca tggggagaag 420
aagtggaagt gcgacagggtg ttcaaagaag tatgctgttc aatcagattg gaaggctcac 480
acaaagatat gtggcaccaa ggagtaccga tgtgactgtg ggacaatctt ctcaagaaag 540
gatagtttca tcacacaccg agcctttctgt gatgcattag ctgaagacac ttctaggggt 600
aaccacagcc tggcaacaat ggttgggaagt ctgcacggtc aacagcaaga tatgttttca 660
catggagtg ctagcttcag ctcttcacca actgatatga tcgcaaactt agctagtaac 720
gaccataact cagacagcca tctgagatca ctatccccct atgctctcgt tacgaggaac 780
actacattgt tctcaaacca gatatcacc aaggaatcag gattcccact tgatggaagt 840
gtgacaagtt atccttatat gtccatgaac tccccatata tgtctgcaac agccttgcta 900
cagaaagctg cagagatggg agcaaagacc agccaggatc ctatttcgcc attacttctt 960
aaaagtttcc caaataactt gactagttct agagatcata tgaacatatc ttctagaagc 1020
cagggagatt cccttgggaa ttcaacagtt aatagtgatt gcatgaagac aactgaagat 1080
gagaactctt acatgattgg ccgtggtaat attctgatca acgcaccgtg gtccagtggc 1140
atcatgaggc caggtactgt gcctttaatt ggactaatga accatccatt ttccatgaga 1200
gaagaaaagg ataaccctag cattttccct gaaaaccaga cacagcaca cagacaggac 1260
aacatttcag gtgtagtaga tgcagatgca ggactaacc aagactttct gggtttagga 1320
gggagtggga ctttgatata gagttccgag acctacaacg cggatgtgac agcattaagc 1380
tattctgatg agcaggagaa acctcaggaa catatctact cttaccatca gtcattctct 1440
gacccactg cattggaaaa acccatttgg gattcttaa 1479

```

<210> 22

<211> 492

<212> PRT

<213> Oryza sativa

<400> 22

```

Met Ser His Thr Ser Glu Glu Glu Ser Leu Asn Asn Leu Gln Gln Gln
1           5           10           15

```

Pro Lys Leu Glu Ala Cys Ala Ala Gly Ser Ser Lys Gly Asp Thr Val
 20 25 30
 Met Pro Val Val Lys Lys Arg Arg Gly His Pro Gly Asn Pro Asp Pro
 35 40 45
 Asp Val Glu Val Val Ala Leu Ser Pro Lys Thr Leu Leu Ala Thr Asn
 50 55 60
 Arg Tyr Ile Cys Glu Val Cys His Lys Gly Phe Gln Arg Asp Gln Asn
 65 70 75 80
 Leu Gln Leu His Arg Arg Gly His Asn Leu Pro Trp Lys Leu Lys Gln
 85 90 95
 Arg Ser Ser Thr Glu Ala Lys Lys Lys Val Tyr Val Cys Pro Glu Ile
 100 105 110
 Thr Cys Pro His His Asp Ala Thr Arg Ala Leu Gly Asp Leu Thr Gly
 115 120 125
 Ile Lys Lys His Tyr Ser Arg Lys His Gly Glu Lys Lys Trp Lys Cys
 130 135 140
 Asp Arg Cys Ser Lys Lys Tyr Ala Val Gln Ser Asp Trp Lys Ala His
 145 150 155 160
 Thr Lys Ile Cys Gly Thr Lys Glu Tyr Arg Cys Asp Cys Gly Thr Ile
 165 170 175
 Phe Ser Arg Lys Asp Ser Phe Ile Thr His Arg Ala Phe Cys Asp Ala
 180 185 190
 Leu Ala Glu Asp Thr Ser Arg Val Asn His Ser Leu Ala Thr Met Val
 195 200 205
 Gly Ser Leu His Gly Gln Gln Gln Asp Met Phe Ser His Gly Val Pro
 210 215 220
 Ser Phe Ser Ser Ser Pro Thr Asp Met Ile Ala Asn Leu Ala Ser Asn
 225 230 235 240
 Asp His Asn Ser Asp Ser His Leu Arg Ser Leu Ser Pro Tyr Ala Leu
 245 250 255
 Val Thr Arg Asn Thr Thr Leu Phe Ser Asn Gln Ile Ser Pro Lys Glu
 260 265 270
 Ser Gly Phe Pro Leu Asp Gly Ser Val Thr Ser Tyr Pro Tyr Met Ser
 275 280 285
 Met Asn Ser Pro Tyr Met Ser Ala Thr Ala Leu Leu Gln Lys Ala Ala
 290 295 300

Glu Met Gly Ala Lys Thr Ser Gln Asp Pro Ile Ser Pro Leu Leu Leu
 305 310 315 320
 Lys Ser Phe Pro Asn Asn Leu Thr Ser Ser Arg Asp His Met Asn Ile
 325 330 335
 Ser Ser Arg Ser Gln Gly Asp Ser Leu Gly Asn Ser Thr Val Asn Ser
 340 345 350
 Asp Cys Met Lys Thr Thr Glu Asp Glu Asn Ser Tyr Met Ile Gly Arg
 355 360 365
 Gly Asn Ile Leu Ile Asn Ala Pro Trp Ser Ser Gly Ile Met Arg Pro
 370 375 380
 Gly Thr Val Pro Leu Ile Gly Leu Met Asn His Pro Phe Ser Met Arg
 385 390 395 400
 Glu Glu Lys Asp Asn Pro Ser Ile Phe Pro Glu Asn Gln Thr Gln His
 405 410 415
 Asn Arg Gln Asp Asn Ile Ser Gly Val Val Asp Ala Asp Ala Gly Leu
 420 425 430
 Thr Gln Asp Phe Leu Gly Leu Gly Gly Ser Gly Thr Leu Asp Met Ser
 435 440 445
 Ser Glu Thr Tyr Asn Ala Asp Val Thr Ala Leu Ser Tyr Ser Asp Glu
 450 455 460
 Gln Glu Lys Pro Gln Glu His Ile Tyr Ser Tyr His Gln Ser Ser Leu
 465 470 475 480
 Asp Pro Thr Ala Leu Glu Lys Pro Ile Trp Asp Ser
 485 490

<210> 23
 <211> 1449
 <212> DNA
 <213> Oryza sativa

<400> 23
 atggcgccgg cctgtcgcca cggagcggcg gtggcggggt acatggcgga ggatgaccct 60
 gacggggcgg cttcggaaga cggcgatatg gacgtggagg tgggcggcga ggagtcgcag 120
 gcgcgcgacg gcgacaggag ggacggcggc gacggcgacg acgagtacgc gctgggtcgaa 180
 cttccctcaa tcaggagagc tcatggataa gagatatgca agcaagtgga taagatcagc 240
 aagtattatg aatttcacta caacactaga ttagtgaagc catcaatcct ccactttcag 300

```

agcttaggaa tttgttatgg gcaacgtcga agcatgatgt ctacttcatg tcaaattcca 360
cagtaggccca ctggatcatca ttgtctcaca atctatcaga gcagaaacac cctggcagtt 420
tactggaagg gtttagcggg gttcaagtta gcacactttc agtgaatgag ggtttattgg 480
ttgctggtgg ttttcaggga gaactaattt gcaagagggt gtgggagacc gtgatgttaa 540
gttctgcaca aggacaactt taagcgacaa tgctatcaca aatgcgatag atatacacag 600
atctgcaagt agtggaaagct tgcgtgttac cgtgtccaac aatgattgtg gttgttcgtg 660
aattttgaca tggaaacatt tcagctcttg aaccacttca gttataactg gccagtaaat 720
cacacatctg tgagccctga taggaaactt ctggcagtag ttggagatga tcgggatgct 780
cttcttggtg attcaagaaa tggaaagcag tggcgaagcc acgggggggtg ccggggaggg 840
cctgacacct acgttaccct gataggctga ggctgggatc gaaggaagac aacggcactg 900
ggcgccgggtc gaagcggaca ggggaggacg gatcagagac ggcggaagg caacgcagcg 960
acggtgtagg cggcggcgct gcgatgggga ttggggtagt gaggagagag gggcagtagt 1020
tggccatctg gactattcgt ttgcttcggc atggcaccct gatggccgca ccttcgcaac 1080
cggtaaccaa gacaagacct gccgcgtctg ggacgtacga aacctgtcaa cctccctctc 1140
agtacttagg ggcaacattg gcgctatccg ttgcatccgc tactcctcgg atggccagtt 1200
catgctcttc tcggagcccg ctgacttcgt gcacgtctac agcgccgctg cggactacaa 1260
gaagaggcag gagatagatt tcttcggcga gatctccggc atctcgctta gccagacga 1320
cgagtcaactg ttcgtcggag tgtgcgaccg tgtgtacgcc agcttgctgc actacagaat 1380
gggccactcc tttggatacc tcgactcgtt catgtagaga gggaataaaa gtggtgtttc 1440
tgtatgtga 1449

```

<210> 24

<211> 46

<212> PRT

<213> Oryza sativa

<400> 24

```

Glu Gly Ser Ser Thr Ser Ala Tyr Ser Ser Ser Pro Ser Pro Pro Ser
1           5           10           15

```

```

Leu Leu Ser Pro Ser Arg Ala Cys Asp Ser Ser Pro Pro Thr Ser Thr
          20           25           30

```

Ser Ile Ser Pro Ser Ser Glu Ala Ala Pro Ser Gly Ser Ser
 35 40 45

<210> 25
 <211> 491
 <212> DNA
 <213> Oryza sativa

<400> 25
 atgtcacgag gtagggatcc tttggcattg agccaggtaa ttggcgatgt gttggatcct 60
 ttcataaagt cagctgcaat gaggattaat tatgggtgaga aggagattac aaatggaact 120
 ggagtacgat catctgctgt tttcactgca ccacatgttg agattgaagg tcgtgaccaa 180
 acgaagctct acacactttt ttataagggtt atgggtggatc ctgatgcgcc aagtccaagc 240
 aaaccagaat acaggggaata tttgcattgg gcaatgaaat agttccgtac gaagctccac 300
 ggccaccggc tggaattcat cggcttggtt ttgtgctatt caaacaggaa gcacgacaaa 360
 cagtttatgc tccaggatgg cggcaaaatt tcaacgtcag agatttctct gcattttaca 420
 atcttgacc acctgttgct gcattatact tcaactgcca gaaggagagt ggtgttggtg 480
 gcagaaggta g 491

<210> 26
 <211> 163
 <212> PRT
 <213> Oryza sativa

<220>
 <221> X_region
 <222> (121)..(121)
 <223> Xaa = any naturally occurring amino acid

<400> 26

Met Ser Arg Gly Arg Asp Pro Leu Ala Leu Ser Gln Val Ile Gly Asp
 1 5 10 15
 Val Leu Asp Pro Phe Ile Lys Ser Ala Ala Met Arg Ile Asn Tyr Gly
 20 25 30
 Glu Lys Glu Ile Thr Asn Gly Thr Gly Val Arg Ser Ser Ala Val Phe
 35 40 45
 Thr Ala Pro His Val Glu Ile Glu Gly Arg Asp Gln Thr Lys Leu Tyr
 50 55 60
 Thr Leu Phe Tyr Lys Val Met Val Asp Pro Asp Ala Pro Ser Pro Ser

<400> 27
atgcaccgcg gcggcgaccg ttccacsgac cctctctcgc gccccgcgcc cggttctcgc 60
ggcggcgggg acggccgatt cggccgcggc cttctcgcgt ggtcgtccgc cggtkgcggc 120

ggcggtaacg gcagcccacc ccacggcttc tcccgtggtk ggggttgggg cggcggcgac	180
ggtggcggag gcgagatca tgataacaga aacaactatg ttaaactttt tattgggtca	240
gttccaagga cagcgactga ggatgatgta cgtcctttat ttgaggagca tggagatggt	300
gttgaagttg ctttgatcaa ggataggaag actggtgaac agcaaggttg ttgttttgtt	360
aaatatgcta cttctgaaga ggctgagcga gccatcagag ctcttcacaa ccagtacact	420
ttacctgggg caatggggcc tattcaagtc agatatgctg atggtgaaag ggaacgccat	480
ggggctattg agcacaagct gtttgcgca tcaactcaata agcaagctac tgcgaaggag	540
attgaagaga tttttgcgcc gtatggtcat gtggaagatg tctacattat gaaagatggc	600
atgaggcaga gccgaggttg tggctttgtc aaattctcat cacgagaacc ggcgctggca	660
gcatgagtg ctctgagtgg gaattatgta atgaggggt gcgagcaacc attaataatt	720
cggttcgctg atcctaagag acctagacct ggagaatcaa ggggtgggtc tgcatttggg	780
ggtcctgggt tcagtccacg atctgatgca gctcttgta tcaggccgac cgccaacctt	840
gatgaaccaa ggggtcggca tatgcctcct gactcttggc atccatcaag cccaaggcca	900
gcacctcatc agtttaataa ctttggatct gacaatccta tgggtccgaa agggagtact	960
gttacatcaa caacagatac ggccactttt cgacctcaga tgttcagcgg aaatggctct	1020
ttgtcaagcc aaacagctgt gccaagtctg tcacatatgg gcatgaacc tcctcccatg	1080
gcacaaggcc atcatctggg aggccagcag atcccacctt tgcaaaagct accaggggtg	1140
cctcagaatt tccctgttca attgcagaat aatcagctgg gacaacctt gcagggctct	1200
gctcaacaga ttggtcagct tcaagttcca caatctatgg gtcttggtt ttttggccag	1260
aataggttat ctggtcagct tcctgtgtca cagccgttga tgcaacaaa tgcctccgtc	1320
agcgctgtgc aggtgccttc agccgtgtcc aattccatgc aagctattcc cgggcaacag	1380
catcttccct caaatgttgc accacaaatg cttcagcagc cagtgcagca gatgccatcg	1440
caagcaccac aattgctact ccaacagcaa gctgctttgc aatccagtta tcaatcttca	1500
cagcaggcaa tttatcagct tcaacaacag ctgcaactaa tgcagcagca gcagcaatcg	1560
aatctaaacc atcaacaacc cacacaggtc gctgagcaac atgggtcagcc agtgcaatct	1620
agtaaccctg gtgccccaaa tgccattatt ccatcaaaca ttaatacaat tccccagcag	1680
gctacttcac ctgcagttcc tttaacctgc aattggactg aacatacgtc tcctgaaggt	1740

```

tttaaatact actacaatag cataactcga gagagtaaga tcgcaacatt tgtgggtctc 1800
actcttagag ttattgtttt acagtgggat aagcctgaag agtatgtgct gtatgagcaa 1860
cagcaacagc agcagcaaca gcagaaacta cttttgcttc aacagcacca acaaaaactt 1920
gccatgcagc aacttcagtc accacctmag gctcagacac atccagcaat gcaacctgtc 1980
caacaaattc ctcaggcaca acaaggacag caacaaatgc agatgaaaca gcaggaattg 2040
aattatactc agctacagac tcctgggtgct attgatccca gtaggataca acaggtgaat 2100
gataaattgg caactgccag agattttatg cccaaaattg ccaggcatga tgtacggaaa 2160
gcaacactgg cgattaaact atctaccatg aaagagacca acaagggccc cgggttaaac 2220
agtttgcttg atcaataa 2238

```

```

<210> 28
<211> 745
<212> PRT
<213> Oryza sativa

```

```

<400> 28

```

```

Met His Arg Gly Gly Asp Arg Ser Thr Asp Pro Ser Ser Gly Pro Ala
1          5          10          15

Pro Gly Ser Arg Gly Gly Gly Asp Gly Arg Phe Gly Arg Gly Pro Ser
          20          25          30

Arg Trp Ser Ser Gly Gly Arg Gly Gly Gly Asn Gly Ser Pro Pro His
          35          40          45

Gly Phe Ser Arg Gly Arg Gly Trp Gly Gly Gly Asp Gly Gly Gly Gly
50          55          60

Gly Asp His Asp Asn Arg Asn Asn Tyr Val Lys Leu Phe Ile Gly Ser
65          70          75          80

Val Pro Arg Thr Ala Thr Glu Asp Asp Val Arg Pro Leu Phe Glu Glu
          85          90          95

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

Glu Gln Gln Gly Cys Cys Phe Val Lys Tyr Ala Thr Ser Glu Glu Ala
115          120          125

Glu Arg Ala Ile Arg Ala Leu His Asn Gln Tyr Thr Leu Pro Gly Ala
130          135          140

Met Gly Pro Ile Gln Val Arg Tyr Ala Asp Gly Glu Arg Glu Arg His

```

145		150		155		160									
Gly	Ala	Ile	Glu	His	Lys	Leu	Phe	Val	Ala	Ser	Leu	Asn	Lys	Gln	Ala
				165					170					175	
Thr	Ala	Lys	Glu	Ile	Glu	Glu	Ile	Phe	Ala	Pro	Tyr	Gly	His	Val	Glu
			180					185					190		
Asp	Val	Tyr	Ile	Met	Lys	Asp	Gly	Met	Arg	Gln	Ser	Arg	Gly	Cys	Gly
		195					200					205			
Phe	Val	Lys	Phe	Ser	Ser	Arg	Glu	Pro	Ala	Leu	Ala	Ala	Met	Ser	Ala
	210					215					220				
Leu	Ser	Gly	Asn	Tyr	Val	Met	Arg	Gly	Cys	Glu	Gln	Pro	Leu	Ile	Ile
225					230					235				240	
Arg	Phe	Ala	Asp	Pro	Lys	Arg	Pro	Arg	Pro	Gly	Glu	Ser	Arg	Gly	Gly
				245					250					255	
Pro	Ala	Phe	Gly	Gly	Pro	Gly	Phe	Ser	Pro	Arg	Ser	Asp	Ala	Ala	Leu
			260					265					270		
Val	Ile	Arg	Pro	Thr	Ala	Asn	Leu	Asp	Glu	Pro	Arg	Gly	Arg	His	Met
		275					280					285			
Pro	Pro	Asp	Ser	Trp	His	Pro	Ser	Ser	Pro	Arg	Ser	Ala	Pro	His	Gln
	290					295					300				
Phe	Asn	Asn	Phe	Gly	Ser	Asp	Asn	Pro	Met	Ala	Pro	Lys	Gly	Ser	Thr
305					310					315				320	
Val	Thr	Ser	Thr	Thr	Asp	Thr	Ala	Thr	Phe	Arg	Pro	Gln	Met	Phe	Ser
				325					330					335	
Gly	Asn	Gly	Ser	Leu	Ser	Ser	Gln	Thr	Ala	Val	Pro	Ser	Ser	Ser	His
			340					345					350		
Met	Gly	Met	Asn	Pro	Pro	Pro	Met	Ala	Gln	Gly	His	His	Leu	Gly	Gly
		355					360					365			
Gln	Gln	Ile	Pro	Pro	Leu	Gln	Lys	Leu	Pro	Gly	Leu	Pro	Gln	Asn	Phe
	370					375					380				
Pro	Val	Gln	Leu	Gln	Asn	Asn	Gln	Leu	Gly	Gln	Pro	Leu	Gln	Gly	Pro
385					390					395				400	
Ala	Gln	Gln	Ile	Gly	Gln	Leu	Gln	Val	Pro	Gln	Ser	Met	Gly	Pro	Gly
				405					410					415	
Ser	Phe	Gly	Gln	Asn	Arg	Leu	Ser	Gly	Gln	Leu	Pro	Val	Ser	Gln	Pro
			420					425					430		
Leu	Met	Gln	Gln	Asn	Ala	Ser	Val	Ser	Ala	Val	Gln	Val	Pro	Ser	Ala

435	440	445																	
Val	Ser	Asn	Ser	Met	Gln	Ala	Ile	Pro	Gly	Gln	Gln	His	Leu	Pro	Ser				
450						455					460								
Asn	Val	Ala	Pro	Gln	Met	Leu	Gln	Gln	Pro	Val	Gln	Gln	Met	Pro	Ser				
465					470					475					480				
Gln	Ala	Pro	Gln	Leu	Leu	Leu	Gln	Gln	Gln	Ala	Ala	Leu	Gln	Ser	Ser				
				485					490					495					
Tyr	Gln	Ser	Ser	Gln	Gln	Ala	Ile	Tyr	Gln	Leu	Gln	Gln	Gln	Leu	Gln				
			500					505						510					
Leu	Met	Gln	Gln	Gln	Gln	Gln	Ser	Asn	Leu	Asn	His	Gln	Gln	Pro	Thr				
		515					520					525							
Gln	Val	Ala	Glu	Gln	His	Gly	Gln	Pro	Val	Gln	Ser	Ser	Asn	Pro	Gly				
530						535					540								
Ala	Pro	Asn	Ala	Ile	Ile	Pro	Ser	Asn	Ile	Asn	Thr	Ile	Pro	Gln	Gln				
545					550					555					560				
Ala	Thr	Ser	Pro	Ala	Val	Pro	Leu	Thr	Cys	Asn	Trp	Thr	Glu	His	Thr				
				565					570					575					
Ser	Pro	Glu	Gly	Phe	Lys	Tyr	Tyr	Tyr	Asn	Ser	Ile	Thr	Arg	Glu	Ser				
			580					585					590						
Lys	Ile	Ala	Thr	Phe	Val	Gly	Leu	Thr	Leu	Arg	Val	Ile	Val	Leu	Gln				
		595					600					605							
Trp	Asp	Lys	Pro	Glu	Glu	Tyr	Val	Leu	Tyr	Glu	Gln	Gln	Gln	Gln	Gln				
610						615					620								
Gln	Gln	Gln	Gln	Lys	Leu	Leu	Leu	Leu	Gln	Gln	His	Gln	Gln	Lys	Leu				
625					630					635					640				
Ala	Met	Gln	Gln	Leu	Gln	Ser	Pro	Pro	Lys	Ala	Gln	Thr	His	Pro	Ala				
				645					650					655					
Met	Gln	Pro	Val	Gln	Gln	Ile	Pro	Gln	Ala	Gln	Gln	Gly	Gln	Gln	Gln				
			660					665					670						
Met	Gln	Met	Lys	Gln	Gln	Glu	Leu	Asn	Tyr	Thr	Gln	Leu	Gln	Thr	Pro				
		675					680					685							
Gly	Ala	Ile	Asp	Pro	Ser	Arg	Ile	Gln	Gln	Val	Asn	Asp	Lys	Leu	Ala				
690						695					700								
Thr	Ala	Arg	Asp	Phe	Met	Pro	Lys	Ile	Ala	Arg	His	Asp	Val	Arg	Lys				
705					710					715					720				
Ala	Thr	Leu	Ala	Ile	Lys	Leu	Ser	Thr	Met	Lys	Glu	Thr	Asn	Lys	Gly				

725 730 735

Pro Gly Leu Asn Ser Leu Leu Asp Gln
740 745

<210> 29
<211> 494
<212> DNA
<213> Oryza sativa

<400> 29
atgggggagag ggaagattgc catcaagagg atcgacaaca cgatgaaccg gcaggtgacc 60
ttctcgaagc ggcgcggcgg gctgatgaag aaggcccgagg agctggccat cctctgcgac 120
gccgacgtcg gcctcattgt cttctcctgc accggccgcc tctacgactt ctccagctca 180
agagcatgaa atcaataata gagcgggtacc aggaggcagg agaggagcat tgctcggttgc 240
tgaacccaat gtcagaggct aagaggcaac ttttgggaga ggaaatctcc aacttcacag 300
ttagagatct gcagcttctc cagaaccaag ttgagatgag cctacattcc ataagaaata 360
aaaagcaaaa ggagaacagt gaacttcgca agaagttcaa cattgctcat caacgcaaca 420
tagaattaca caagaaggta aagtttctat tcgatctact cgaaacatac aaaaatgtcg 480
tagaatatta ttga 494

<210> 30
<211> 164
<212> PRT
<213> Oryza sativa

<220>
<221> X_region
<222> (61)..(61)
<223> Xaa = any naturally occurring amino acid

<400> 30

Met Gly Arg Gly Lys Ile Ala Ile Lys Arg Ile Asp Asn Thr Met Asn
1 5 10 15
Arg Gln Val Thr Phe Ser Lys Arg Arg Gly Gly Leu Met Lys Lys Ala
20 25 30
Arg Glu Leu Ala Ile Leu Cys Asp Ala Asp Val Gly Leu Ile Val Phe
35 40 45
Ser Cys Thr Gly Arg Leu Tyr Asp Phe Ser Ser Ser Xaa Ser Met Lys
50 55 60

Ser Ile Ile Glu Arg Tyr Gln Glu Ala Gly Glu Glu His Cys Arg Leu
 65 70 75 80
 Leu Asn Pro Met Ser Glu Ala Lys Arg Gln Leu Leu Gly Glu Glu Ile
 85 90 95
 Ser Asn Phe Thr Val Arg Asp Leu Gln Leu Leu Gln Asn Gln Val Glu
 100 105 110
 Met Ser Leu His Ser Ile Arg Asn Lys Lys Gln Lys Glu Asn Ser Glu
 115 120 125
 Leu Arg Lys Lys Phe Asn Ile Ala His Gln Arg Asn Ile Glu Leu His
 130 135 140
 Lys Lys Val Lys Phe Leu Phe Asp Leu Leu Glu Thr Tyr Lys Asn Val
 145 150 155 160
 Val Glu Tyr Tyr

<210> 31
 <211> 494
 <212> DNA
 <213> Oryza sativa

<400> 31
 atgtctaggg acccattggg tgtcgggtcat gtcgtcggcg atatcgtgga cccgttcgtc 60
 accaccgctt cgcttagggg cttctacaac agcaaggaga tgacaaatgg gtctgagctc 120
 aagccatctc aggtgttgaa ccaaccaagg atttatatcg aaggtcgcga catgaggacg 180
 ctctacacgc ttgtaatggg ggaccctgat gcaccaagcc ccagcaaccc tactaaaaga 240
 gagtaccttc attggtgtca tcaggtaatg agattgtccc ctatgagagc ccacgcccaa 300
 ctgcaggcat ccatcgcttc gtgttcatcc tattcaggca gtcagtcagg cagaccacct 360
 atgcaccagg gtggcgccaa aacttcaata caagggactt tgctgagctc tacaacctcg 420
 gttcgccggg cgccgcgctc ttcttcaact gccagaggga gaacggctgt ggaggaagaa 480
 ggtgtgttag atga 494

<210> 32
 <211> 163
 <212> PRT
 <213> Oryza sativa

<400> 32

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

<210> 33
 <211> 615
 <212> DNA
 <213> Oryza sativa

<400> 33
 atggacaacc cgacgccgtc gtcgagctcc ggcggttgcg gcgccgccgc cgccgccaat 60
 gccagctccg acgagcaccg caagacgtcg ccggcgacgg ccgctgattt cttcgacgag 120
 gaggcacacct tcgacgatat cctgctcccg gtggacttgt tcgacttcac gcacggactc 180
 gacgccggcg acggctgtta cgtagacgat caggctgcga ttgggcagaa gccgctgtcg 240
 ccggcgccgg agccggcggc agagcagcag ccgtcgccgg cgccggagca cggcgacgat 300
 gagcggctgg tgatgtacta ccaaggacag gactacatct tcgaccccggt gcagccgcag 360
 aaggatatctc ttggtcttgc tgcacaatgt gcttgtttct tggttaagacc catcactggt 420

cccgaggact tgcacaggtt tgctgcgctg acgcggtata gggagaaaaa gaggaacatc 480
 aagttcatca agaaggcgga ttacagtgcc cgaaaggaag tcgctttgcg gatgaaacgc 540
 agcaaagggga aatttgcacc taggggttcag acctctgaaa attccttggc acacagaaaag 600
 gggataactt tgtaa 615

<210> 34
 <211> 204
 <212> PRT
 <213> Oryza sativa

<400> 34

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

195

200

<210> 35
 <211> 1530
 <212> DNA
 <213> *Oryza sativa*

<400> 35
 atggtggagg cagaggccgg caaggaggcg atccggcgag ccctgcggtc tctgcggcgg 60
 aggcatctcg tcgaggaggg cggccaccgc ccggccatcg aggcgctcgc cgggcccttc 120
 gccgcccagg cgggtggagtg gaaagagaag gctgagaagc atgaattgga gctgcagcaa 180
 tgctacaagg ctcatgcacg tttgtccgag cagctcgta cagagataga agaggggaaa 240
 gcatcaaaag cgctgcttaa agagaaagaa aacttgatta ctaccatgca gactgaactc 300
 gagcaaacca gggaagaaaa tacgcagctc aagcagtcgc ttgaagaaaa aacaagcgct 360
 ttagatctta taattcagga gcaccaggca gttaaagctg aacttgaaca ggcattaaca 420
 aaacagaaag tggcagaaga tgaaaacagg aatctgattg atcgatggat gcttgagaag 480
 atgaaggatg cagagaggct taatgaggca aatgcaatgt atgaggagat ggttctaaag 540
 ctgaagtcag ctggtgttgg tggatataca cacaatgcac tacaagaagc tgatggcatc 600
 attcgtcggt ctgaagctgg gtacatggac atcatggaaa caccaatccc atctacatgc 660
 agaatcacta tccgtgctca tgatggtggc tgtggatcta taattttcca gcataacaca 720
 gacaagctaa tcagtgttgg ccaggatcaa actgtcaaga tttgggtctgc acataccggt 780
 gctttaacct ccactctaca aggetgtttg gggctctgtc atgatcttgc tgtgactaat 840
 gacaacaagt ttgtgattgc tgctgtagc tctaacaat tgtttgtatg ggaagttaat 900
 gggggggcgc cacgtcacac tctgactggc cacacaaaa atgtttcatc cgttgatgca 960
 agctgggtga agagctgtgt tcttgctagt tcatccaatg atcataccat caagatttgg 1020
 gatcttcagt caggcttctg caaaagcacc ataatgtcag ggagcaacgc caactcacta 1080
 gctttcattg atggtgtcac cctttgctcg ggtcataggg atggacacct tcggttatgg 1140
 gacatccgaa gtgcaaaatg cactagtcag acattcgctc atcttgacgt aagctctgtc 1200
 agtgtgtcac ggaatagaaa tttcattctt acaagtggaa aggataacgt gcacaacctt 1260
 tttgatccta gaacaatgga agtttgtgga aaattcaagg ccatgggtaa cagagtgggt 1320
 agcagctggg gcaggccttg cataagccct gatgaaaact ccatagctgc tgggtgctaat 1380

gatgggtctg tttacatatg gtcaagattg aaaaaggatg gggtagccgac gatactacag 1440
 ggccattctt cgtcggtagt ttcaagttca tgggtgtgggt taggacctct tgctaccgct 1500
 gacaagcatc atatctatat ctggacttga 1530

<210> 36
 <211> 509
 <212> PRT
 <213> Oryza sativa

<400> 36

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

Met Asp Ile Met Glu Thr Pro Ile Pro Ser Thr Cys Arg Ile Thr Ile
 210 215 220
 Arg Ala His Asp Gly Gly Cys Gly Ser Ile Ile Phe Gln His Asn Thr
 225 230 235 240
 Asp Lys Leu Ile Ser Gly Gly Gln Asp Gln Thr Val Lys Ile Trp Ser
 245 250 255
 Ala His Thr Gly Ala Leu Thr Ser Thr Leu Gln Gly Cys Leu Gly Ser
 260 265 270
 Val Asn Asp Leu Ala Val Thr Asn Asp Asn Lys Phe Val Ile Ala Ala
 275 280 285
 Cys Ser Ser Asn Lys Leu Phe Val Trp Glu Val Asn Gly Gly Arg Pro
 290 295 300
 Arg His Thr Leu Thr Gly His Thr Lys Asn Val Ser Ser Val Asp Ala
 305 310 315 320
 Ser Trp Val Lys Ser Cys Val Leu Ala Ser Ser Ser Asn Asp His Thr
 325 330 335
 Ile Lys Ile Trp Asp Leu Gln Ser Gly Phe Cys Lys Ser Thr Ile Met
 340 345 350
 Ser Gly Ser Asn Ala Asn Ser Leu Ala Phe Ile Asp Gly Val Thr Leu
 355 360 365
 Cys Ser Gly His Arg Asp Gly His Leu Arg Leu Trp Asp Ile Arg Ser
 370 375 380
 Ala Lys Cys Thr Ser Gln Thr Phe Ala His Leu Asp Val Ser Ser Val
 385 390 395 400
 Ser Val Ser Arg Asn Arg Asn Phe Ile Leu Thr Ser Gly Lys Asp Asn
 405 410 415
 Val His Asn Leu Phe Asp Pro Arg Thr Met Glu Val Cys Gly Lys Phe
 420 425 430
 Lys Ala Met Gly Asn Arg Val Val Ser Ser Trp Gly Arg Pro Cys Ile
 435 440 445
 Ser Pro Asp Glu Asn Ser Ile Ala Ala Gly Ala Asn Asp Gly Ser Val
 450 455 460
 Tyr Ile Trp Ser Arg Leu Lys Lys Asp Gly Val Pro Thr Ile Leu Gln
 465 470 475 480
 Gly His Ser Ser Ser Val Val Ser Ser Ser Trp Cys Gly Leu Gly Pro
 485 490 495

Leu Ala Thr Ala Asp Lys His His Ile Tyr Ile Trp Thr
 500 505

<210> 37
 <211> 559
 <212> DNA
 <213> Oryza sativa

<400> 37
 atggccggaa gtggcaggga cagggaccct cttgtggttg gtagggttgt gggatgatgtg 60
 ctggacgcgt tcgtccggag caccaacctc aagggtcacct atgggtccaa gaccgtgtcc 120
 aatggctgcg agctcaagcc gtccatggtc acccaccagc ctagggtcga ggtcggcggc 180
 aatgacatga ggacattcta cacccttgta tggatgatgtg agaccagat gcaccaagcc 240
 caagtgacct taaccttagg gagtatctac attgggtcgg cgaagttcct ggtgctgaag 300
 ttctggcgcc acccggctc cagtgggtacc aggaatatcg gtgacctacg agagcccaag 360
 gccaacctat gggatccacc ggctgggtgtt cgtgctgttc cagcagctgg ggcgtcagac 420
 ggtgtacgca ccgggggtggc gccagaactt cagcaccagg aacttcgccg agctctacaa 480
 cctcggctcg ccgggtcgcca ccgtctactt caactgccag cgcgaggccg gctccggcgg 540
 caggaggggtc tacccttag 559

<210> 38
 <211> 53
 <212> PRT
 <213> Oryza sativa

<400> 38

Ala Arg Thr Ser Ala Pro Gly Thr Ser Pro Ser Ser Thr Thr Ser Ala
 1 5 10 15
 Arg Arg Ser Pro Pro Ser Thr Ser Thr Ala Ser Ala Arg Pro Ala Pro
 20 25 30
 Ala Ala Gly Gly Ser Thr Val Gln Trp Leu Arg Ala Gln Ala Val His
 35 40 45
 Gly His Pro Pro Ala
 50

<210> 39
 <211> 2386
 <212> DNA

<213> Oryza sativa

<220>

<221> misc_feature

<222> (1793)..(1793)

<223> r = adenine or guanine

<220>

<221> misc_feature

<222> (1820)..(1820)

<223> k = guanine or thymine

<400> 39

atgcagttcc ctttcgccag tggcttcgcg tcctcgccgg cgctctccct cgcgctggga	60
caacgccggc ggtggcatcg gggggcggat gctcgccggc ggcgctggcg ctggttctag	120
cgcgggcggc gcgatgacga gggacactga ggcggagaac gacagccggt cgggaagcga	180
ccatctcgac gccatctcgg ccgccggcga ggacgacgtc gaggacgccg agcccagcaa	240
ctcccgaag aggaagaagc gataccaccg ccacacgccg cagcagatcc aagagctcga	300
agcggctgtt caaggagtgc ccccatccgg acgagaagca gcgcgccgag ctacgcagga	360
ggctctccct cgatgcgcgg caggtcaagt tctggttcca gaatcgccgc acgcagatga	420
agaagacgca actggagcgg cagcagaacg cgcttctcaa gcaggagaac gacaagctgc	480
gcgccgagaa catgacgac cgggaggcga tgcgcagccc aatgtgcggc agctgcggga	540
gccccgccat gctcggggag gtgtccctag aggagcagca cctgcgcacg gagaatgcgc	600
gcctcaagga cgagctcaac cgcgtgtgcg ccctcgccac caagttccta ggcaagccca	660
tctccctcct gtcgccgccg cccttgctgc agccgcacct ctgctgcct atgcctaact	720
cgctcgtgga gctagcgatc ggaggcatcg gtggcttagg gtctctcggg aactaccgc	780
gctgtatgaa tgagttcgcc ggaggcgtgt cgagcccgat gggcacgggtg atcacgcctg	840
cgcgcgcaac cggggctgct ataccatcgc tgggtgggtaa catcgacagg tccgtgttct	900
tagagcttgc aatcagcgcg atggacgaac ttgtcaagat ggcacagatg gacgatccat	960
tgtgggttcc ggactgcct ggttctccca gcaaggagggt gctcaacttt gaggagtatc	1020
ttcactcctt cctaccgtgc attggcatga agcctgccgg ctacgtctct gaggcctcga	1080
gggaatccgg tctagtcac atcgacaaca gccttgccct tgtagagacc ctcatggatg	1140
aggagacggg ggtctgacat gttctcgtgc atgattgcta aggcaacagt gcttgaggag	1200

```

gtgtctaccg gcattgcagg aagcagaaat ggcgcgttgc tgctggatga aggctgagct 1260
acagggtgctc tcacctctag taccaattag ggagggttaca tttctccggt tttgtaagca 1320
gctggctgag ggtgcatggg cagtagttga tgtgtccatt gatggtttag tgagagatca 1380
taattctgga acggcaccca cagggtgaaa tgtgaagtg aggagggtag cttccggctg 1440
tgtgatgcaa gacactccca atgggtattg caagaggtca catgggttga gcatacggaa 1500
tacgatgagg catcagtgc ccagctctac cgtccactcc ttcggtctgg tctgccttt 1560
ggcgccaggc gttggcttgc gacgctgcag cgccaatgcg aatgccttgc catcctcatg 1620
tcctctgcta cagtgcagc aaatgactcg actggtacag atatcgcaag agggcaagcg 1680
aagcatgctg aagctggcac gccggatgac ggagaacttc tgtgctgggg tgagcgcac 1740
gtctgcacgt gaatggagca agctggatgg tgcgaccggt agcatcgggg agracgtgcg 1800
cgtgatggca cggaagagck tgagcgagcc tggagagcca ccgggcgtgg tgctgagcgc 1860
ggccacctcg gtgtgggtgc ccgtggctcc ggagaagctc ttcaacttcc tgcgcgatga 1920
gcagctgcgt gcggagtggg acatcctcag caatggaggc cccatgcagg agatgacca 1980
gatcgccaag gggcagcggg atgggaactc ggtttctact ctcagggcca gtgctgtgag 2040
tgccaaccag agcagcatgc tgatacttca ggagacgtgc accgacgctt ctggctccat 2100
tgttgtgtac gctccggtag acatcccggc aatgcagctt gtcatgaatg gtggcgactc 2160
cacctatgtt gctctgcttc catccggatt cgccatcctg cccgatgggc ctcgcatcgg 2220
cgccaccggg tacgagacgg gcggctctct gctcaccgtg gcgttccaga ttcttgtcaa 2280
caaccagccc accgccaagc tcaccgtgga atcggtcgag accgtgaaca atctcatctc 2340
ctgcaccatc aagaagatca agacggcgct gcagtgcgac gcctga 2386

```

<210> 40

<211> 793

<212> PRT

<213> Oryza sativa

<220>

<221> X_region

<222> (491)..(491)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (597)..(597)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (606)..(606)

<223> Xaa = any naturally occurring amino acid

<400> 40

Met	Gln	Phe	Pro	Phe	Ala	Ser	Gly	Phe	Ala	Ser	Ser	Pro	Ala	Leu	Ser
1				5					10					15	
Leu	Ala	Leu	Asp	Asn	Ala	Gly	Gly	Gly	Ile	Gly	Gly	Arg	Met	Leu	Gly
			20					25					30		
Gly	Gly	Ala	Gly	Ala	Gly	Ser	Ser	Ala	Gly	Gly	Ala	Met	Thr	Arg	Asp
		35					40					45			
Thr	Glu	Ala	Glu	Asn	Asp	Ser	Arg	Ser	Gly	Ser	Asp	His	Leu	Asp	Ala
	50					55					60				
Ile	Ser	Ala	Ala	Gly	Glu	Asp	Asp	Val	Glu	Asp	Ala	Glu	Pro	Ser	Asn
65					70					75					80
Ser	Arg	Lys	Arg	Lys	Lys	Arg	Tyr	His	Arg	His	Thr	Pro	Gln	Gln	Ile
				85					90					95	
Gln	Glu	Leu	Glu	Ala	Leu	Phe	Lys	Glu	Cys	Pro	His	Pro	Asp	Glu	Lys
			100					105					110		
Gln	Arg	Ala	Glu	Leu	Ser	Arg	Arg	Leu	Ser	Leu	Asp	Ala	Arg	Gln	Val
		115					120					125			
Lys	Phe	Trp	Phe	Gln	Asn	Arg	Arg	Thr	Gln	Met	Lys	Lys	Thr	Gln	Leu
	130					135					140				
Glu	Arg	His	Glu	Asn	Ala	Leu	Leu	Lys	Gln	Glu	Asn	Asp	Lys	Leu	Arg
145					150					155					160
Ala	Glu	Asn	Met	Thr	Ile	Arg	Glu	Ala	Met	Arg	Ser	Pro	Met	Cys	Gly
				165					170					175	
Ser	Cys	Gly	Ser	Pro	Ala	Met	Leu	Gly	Glu	Val	Ser	Leu	Glu	Glu	Gln
			180					185					190		
His	Leu	Arg	Ile	Glu	Asn	Ala	Arg	Leu	Lys	Asp	Glu	Leu	Asn	Arg	Val
		195					200					205			
Cys	Ala	Leu	Ala	Thr	Lys	Phe	Leu	Gly	Lys	Pro	Ile	Ser	Leu	Leu	Ser
	210					215					220				
Pro	Pro	Pro	Leu	Leu	Gln	Pro	His	Leu	Ser	Leu	Pro	Met	Pro	Asn	Ser

225		230		235		240									
Ser	Leu	Glu	Leu	Ala	Ile	Gly	Gly	Ile	Gly	Gly	Leu	Gly	Ser	Leu	Gly
				245					250					255	
Thr	Leu	Pro	Gly	Cys	Met	Asn	Glu	Phe	Ala	Gly	Gly	Val	Ser	Ser	Pro
			260					265					270		
Met	Gly	Thr	Val	Ile	Thr	Pro	Ala	Arg	Ala	Thr	Gly	Ala	Ala	Ile	Pro
		275					280					285			
Ser	Leu	Val	Gly	Asn	Ile	Asp	Arg	Ser	Val	Phe	Leu	Glu	Leu	Ala	Ile
	290					295					300				
Ser	Ala	Met	Asp	Glu	Leu	Val	Lys	Met	Ala	Gln	Met	Asp	Asp	Pro	Leu
305					310					315					320
Trp	Val	Pro	Ala	Leu	Pro	Gly	Ser	Pro	Ser	Lys	Glu	Val	Leu	Asn	Phe
			325						330					335	
Glu	Glu	Tyr	Leu	His	Ser	Phe	Leu	Pro	Cys	Ile	Gly	Met	Lys	Pro	Ala
		340						345					350		
Gly	Tyr	Val	Ser	Glu	Ala	Ser	Arg	Glu	Ser	Gly	Leu	Val	Ile	Ile	Asp
	355						360					365			
Asn	Ser	Leu	Ala	Leu	Val	Glu	Thr	Leu	Met	Asp	Arg	Arg	Arg	Trp	Ser
	370					375					380				
Asp	Met	Phe	Ser	Cys	Met	Ile	Ala	Lys	Ala	Thr	Val	Leu	Glu	Glu	Val
385					390					395					400
Ser	Thr	Gly	Ile	Ala	Gly	Ser	Arg	Asn	Gly	Ala	Leu	Leu	Trp	Met	Lys
			405						410					415	
Ala	Glu	Leu	Gln	Val	Leu	Ser	Pro	Leu	Val	Pro	Ile	Arg	Glu	Val	Thr
			420					425					430		
Phe	Leu	Arg	Phe	Cys	Lys	Gln	Leu	Ala	Glu	Gly	Ala	Trp	Ala	Val	Val
	435					440						445			
Asp	Val	Ser	Ile	Asp	Gly	Leu	Val	Arg	Asp	His	Asn	Ser	Gly	Thr	Ala
	450					455					460				
Pro	Thr	Gly	Gly	Asn	Val	Lys	Cys	Arg	Arg	Val	Pro	Ser	Gly	Cys	Val
465					470					475					480
Met	Gln	Asp	Thr	Pro	Asn	Gly	Tyr	Cys	Lys	Xaa	Val	Thr	Trp	Val	Glu
			485						490					495	
His	Thr	Glu	Tyr	Asp	Glu	Ala	Ser	Val	His	Gln	Leu	Tyr	Arg	Pro	Leu
		500						505					510		
Leu	Arg	Ser	Gly	Leu	Ala	Phe	Gly	Ala	Arg	Arg	Trp	Leu	Ala	Thr	Leu

515	520	525													
Gln Arg Gln Cys Glu Cys Leu Ala Ile Leu Met Ser Ser Ala Thr Val															
530	535	540													
Thr Ala Asn Asp Ser Thr Gly Thr Asp Ile Ala Arg Gly Lys Arg Ser															
545	550	555													560
Met Leu Lys Leu Ala Arg Arg Met Thr Glu Asn Phe Cys Ala Gly Val															
	565	570													575
Ser Ala Ser Ser Ala Arg Glu Trp Ser Lys Leu Asp Gly Ala Thr Gly															
	580	585													590
Ser Ile Gly Glu Xaa Val Arg Val Met Ala Arg Lys Ser Xaa Ser Glu															
	595	600													605
Pro Gly Glu Pro Pro Gly Val Val Leu Ser Ala Ala Thr Ser Val Trp															
	610	615													620
Val Pro Val Ala Pro Glu Lys Leu Phe Asn Phe Leu Arg Asp Glu Gln															
	625	630													640
Leu Arg Ala Glu Trp Asp Ile Leu Ser Asn Gly Gly Pro Met Gln Glu															
	645	650													655
Met Thr Gln Ile Ala Lys Gly Gln Arg Asp Gly Asn Ser Val Ser Leu															
	660	665													670
Leu Arg Ala Ser Ala Val Ser Ala Asn Gln Ser Ser Met Leu Ile Leu															
	675	680													685
Gln Glu Thr Cys Thr Asp Ala Ser Gly Ser Ile Val Val Tyr Ala Pro															
	690	695													700
Val Asp Ile Pro Ala Met Gln Leu Val Met Asn Gly Gly Asp Ser Thr															
	705	710													720
Tyr Val Ala Leu Leu Pro Ser Gly Phe Ala Ile Leu Pro Asp Gly Pro															
	725	730													735
Arg Ile Gly Ala Thr Gly Tyr Glu Thr Gly Gly Ser Leu Leu Thr Val															
	740	745													750
Ala Phe Gln Ile Leu Val Asn Asn Gln Pro Thr Ala Lys Leu Thr Val															
	755	760													765
Glu Ser Val Glu Thr Val Asn Asn Leu Ile Ser Cys Thr Ile Lys Lys															
	770	775													780
Ile Lys Thr Ala Leu Gln Cys Asp Ala															
	785	790													

<210> 41
 <211> 2288
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (539)..(539)
 <223> s = guanine or cytosine

<220>
 <221> misc_feature
 <222> (760)..(760)
 <223> k = guanine or thymine

<220>
 <221> misc_feature
 <222> (773)..(773)
 <223> r = adenine or guanine

<220>
 <221> misc_feature
 <222> (896)..(896)
 <223> m = adenine or cytosine

<220>
 <221> misc_feature
 <222> (2082)..(2082)
 <223> d = adenine, guanine, or thymine

<400> 41
 atgatcccg cgaggcatat gccgtcgatg atcggccgga acggcgcagc gtacggatcc 60
 tcatcggcac tgtcactcag ccagctcctt gaccagatcc cggcgacgac ggcggagagc 120
 ggcgacaaca tgatccgcag ccgtgcgctc gaccctctcg gcggcgacga gttcgagagc 180
 aagtccggca gcgagaacgt cgacggcgct tccgtcgacg accaggaccc caaccagcgg 240
 ccgcgcaaga agcgctacca ccgccacacc cagcatcaga tccaggagat ggaagcgttt 300
 cttcaaggaa tgtccacacc cagatgataa gcagcgcaag gagctgagcc gggagctcgg 360
 cctggagcct ctccaagtga agttctggtt ccagaacaag cgcacccaaa tgaagagaac 420
 cagcatgaga ggcacgagaa ctgcgagctg cggtcggata acgagaagct ccgggcccag 480
 aacatgcggt acaaggaggc cctgagcagc gcctcctgcc caactgcggc ggccccgcsg 540
 ccctcggcga gatgtccttc gacgagcacc acctccgcat cgagaacgcc cgctcccgcg 600

aagagatcga taggatatcg gcgatcgcg cgaagtacgt ggggaagcca atggtgccgt 660
tcccggtgct gtogaacccg atggcgcgcg cggcgtcgcg ggcgcgctg gacctccccg 720
tggcgccgta cggcgctccc ggcgacatgt tcggcgggcgk tggccccggg garctgctgc 780
gcggggtgca gtcggaggtc gacaagccga tgatcgtgga gtcgcccgtg gcggcgatgg 840
aggagctgcy catggcgag ctggacgagc cgctctggag cgtcgcgccg ccgctmgacg 900
ccacggcggc ggcgatggag acgctcagcg aggaggagta cgcgcgcatg ttcccccgag 960
ggctcgcccc caagcagtac gggctcaggt cggaggcgct ccgcgacagc gccgtcgtca 1020
tcatgacgca cgccaacctc gtcgagatcc tcatggacgc gcgcggcagt gttctcgaac 1080
atcgtgtcga gggccatcac gctggaggtg ctgtccaccg gcgtcgcggg gaactacaac 1140
ggtgcattgc aagtgatgtc agtggagttt caggtgccat cgccactggg gccaacgaga 1200
gagagctact tcgtgagata ctgcaagcag aacgcagacg ggacatgggc cgtcgtcgac 1260
gtctcgtggt acagcctccg ccctagcccc gtctctaaat gtcgcccggg tccgtcaggc 1320
tgctcatcc aagaaatgcc caatggctac tccaagaggt gacgtgggtg gagcatgtgg 1380
aggtggacga caggtcggtg cacaacatct acaagctgct ggtgaactcc gggctggcgt 1440
tcggcgcgcg gcggtgggtc ggcacgctgg accgccagt cgagcgccct gccagcgtca 1500
tggccagcaa catccccacc agcgacatcg gcggtgatta cgagctcgga ggggaggaag 1560
agcatgctga agctggcgga gaggatggtg gtgagcttct gcggcgcggt gacggcgctcg 1620
gtggcgcatc agtggacgac gctgtccggc agcgcgcccg aggacgtccg cgtcatgacg 1680
aggaagagcy tcgacgacc cgggagggcc cccggcatcg tcctcaacgc cgccacctcc 1740
ttctggctcc ccgtccctcc caagcgcgtc ttcgacttcc tccgcgacga atcctctcgc 1800
agtgagagtg ggacatcctc tccaatggcg gtatcgttca agaaatggcc cacatcgcca 1860
atggccggga tcagggaac tgctgtctac ttcttcgctg caacagctcg aactcgaacc 1920
agagcaacat gctgatcctg caggagagct gcacggacgc gtcggggctg tacgtcatct 1980
acgcgccggt ggacgtggtt gcgatgaacg tcgtgctaaa cggcgggcgac ccgactacgt 2040
cgctgcatgc tgccacctcg ggcttcgcca tcctgccga tdggccggcg cacgacggtt 2100
gcgacggcga cggcgcgctc ggcgtcggt cggcggggtc gctgctgacg gtggcggttc 2160
agatactggt ggactccgtc ccgacggcga agctctcgct ggggtcggtg gcgacggtga 2220

acagcctcat cgcctgcacg gtggagcgc tcaaggcggc cgtctccggc gagagcaatc 2280

cgcagtag 2288

<210> 42

<211> 769

<212> PRT

<213> Oryza sativa

<220>

<221> X_region

<222> (253)..(253)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (377)..(377)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (425)..(429)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (459)..(459)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (608)..(608)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (701)..(702)

<223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (718)..(718)

<223> Xaa = any naturally occurring amino acid

<400> 42

Met Ile Pro Ala Arg His Met Pro Ser Met Ile Gly Arg Asn Gly Ala
 1 5 10 15
 Ala Tyr Gly Ser Ser Ser Ala Leu Ser Leu Ser Gln Leu Leu Asp Gln
 20 25 30
 Ile Pro Ala Thr Thr Ala Glu Ser Gly Asp Asn Met Ile Arg Ser Arg
 35 40 45
 Ala Ser Asp Pro Leu Gly Gly Asp Glu Phe Glu Ser Lys Ser Gly Ser
 50 55 60
 Glu Asn Val Asp Gly Val Ser Val Asp Asp Gln Asp Pro Asn Gln Arg
 65 70 75 80
 Pro Arg Lys Lys Arg Tyr His Arg His Thr Gln His Gln Ile Gln Glu
 85 90 95
 Met Glu Ala Phe Phe Lys Glu Cys Pro His Pro Asp Asp Lys Gln Arg
 100 105 110
 Lys Glu Leu Ser Arg Glu Leu Gly Leu Glu Pro Leu Gln Val Lys Phe
 115 120 125
 Trp Phe Gln Asn Lys Arg Thr Gln Met Lys Arg Asn Gln His Glu Arg
 130 135 140
 His Glu Asn Ser Gln Leu Arg Ser Asp Asn Glu Lys Leu Arg Ala Glu
 145 150 155 160
 Asn Met Arg Tyr Lys Glu Ala Leu Ser Ser Ala Ser Cys Pro Cys Gly
 165 170 175
 Gly Pro Ala Ala Leu Gly Glu Met Ser Phe Asp Glu His His Leu Arg
 180 185 190
 Ile Glu Asn Ala Arg Leu Arg Glu Glu Ile Asp Arg Ile Ser Ala Ile
 195 200 205
 Ala Ala Lys Tyr Val Gly Lys Pro Met Val Pro Phe Pro Val Leu Ser
 210 215 220
 Asn Pro Met Ala Ala Ala Ala Ser Arg Ala Pro Leu Asp Leu Pro Val
 225 230 235 240
 Ala Pro Tyr Gly Val Pro Gly Asp Met Phe Gly Gly Xaa Gly Pro Gly
 245 250 255
 Glu Leu Leu Arg Gly Val Gln Ser Glu Val Asp Lys Pro Met Ile Val
 260 265 270
 Glu Leu Ala Val Ala Ala Met Glu Glu Leu Arg Met Ala Gln Leu Asp
 275 280 285

Glu Pro Leu Trp Ser Val Ala Pro Pro Leu Asp Ala Thr Ala Ala Ala
 290 295 300
 Met Glu Thr Leu Ser Glu Glu Glu Tyr Ala Arg Met Phe Pro Arg Gly
 305 310 315 320
 Leu Gly Pro Lys Gln Tyr Gly Leu Arg Ser Glu Ala Ser Arg Asp Ser
 325 330 335
 Ala Val Val Ile Met Thr His Ala Asn Leu Val Glu Ile Leu Met Asp
 340 345 350
 Ala Arg Ala Ala Val Phe Ser Asn Ile Val Ser Arg Ala Ile Thr Leu
 355 360 365
 Glu Val Leu Ser Thr Gly Val Ala Xaa Gly Asn Tyr Asn Gly Ala Leu
 370 375 380
 Gln Val Met Ser Val Glu Phe Gln Val Pro Ser Pro Leu Val Pro Thr
 385 390 395 400
 Arg Glu Ser Tyr Phe Val Arg Tyr Cys Lys Gln Asn Ala Asp Gly Thr
 405 410 415
 Trp Ala Val Val Asp Val Ser Leu Xaa Xaa Xaa Xaa Xaa Asp Ser Leu
 420 425 430
 Arg Pro Ser Pro Val Leu Lys Cys Arg Arg Arg Pro Ser Gly Cys Leu
 435 440 445
 Ile Gln Glu Met Pro Asn Gly Tyr Ser Lys Xaa Val Thr Trp Val Glu
 450 455 460
 His Val Glu Val Asp Asp Arg Ser Val His Asn Ile Tyr Lys Leu Leu
 465 470 475 480
 Val Asn Ser Gly Leu Ala Phe Gly Ala Arg Arg Trp Val Gly Thr Leu
 485 490 495
 Asp Arg Gln Cys Glu Arg Leu Ala Ser Val Met Ala Ser Asn Ile Pro
 500 505 510
 Thr Ser Asp Ser Ala Val Ile Thr Ser Ser Glu Gly Arg Lys Ser Met
 515 520 525
 Leu Lys Leu Ala Glu Arg Met Val Val Ser Phe Cys Gly Gly Val Thr
 530 535 540
 Ala Ser Val Ala His Gln Trp Thr Thr Leu Ser Gly Ser Gly Ala Glu
 545 550 555 560
 Asp Val Arg Val Met Thr Arg Lys Ser Val Asp Asp Pro Gly Arg Pro
 565 570 575

Pro Gly Ile Val Leu Asn Ala Ala Thr Ser Phe Trp Leu Pro Val Pro
 580 585 590

Pro Lys Arg Val Phe Asp Phe Leu Arg Asp Glu Ser Ser Arg Ser Xaa
 595 600 605

Glu Trp Asp Ile Leu Ser Asn Gly Gly Ile Val Gln Glu Met Ala His
 610 615 620

Ile Ala Asn Gly Arg Asp Gln Gly Asn Cys Val Ser Leu Leu Arg Val
 625 630 635 640

Asn Ser Ser Asn Ser Asn Gln Ser Asn Met Leu Ile Leu Gln Glu Ser
 645 650 655

Cys Thr Asp Ala Ser Gly Ser Tyr Val Ile Tyr Ala Pro Val Asp Val
 660 665 670

Val Ala Met Asn Val Val Leu Asn Gly Gly Asp Pro Thr Arg Arg Cys
 675 680 685

Met Leu Pro Pro Ser Gly Phe Ala Ile Leu Pro Asp Xaa Xaa Arg Arg
 690 695 700

Thr Thr Val Ala Thr Ala Thr Ala Ala Ser Ala Ser Ala Xaa Gly Gly
 705 710 715 720

Ser Leu Leu Thr Val Ala Phe Gln Ile Leu Val Asp Ser Val Pro Thr
 725 730 735

Ala Lys Leu Ser Leu Gly Ser Val Ala Thr Val Asn Ser Leu Ile Ala
 740 745 750

Cys Thr Val Glu Arg Ile Lys Ala Ala Val Ser Gly Glu Ser Asn Pro
 755 760 765

Gln

<210> 43

<211> 1638

<212> DNA

<213> Oryza sativa

<400> 43

atgtgtgact gcaggtgggg cctagcgcag gggagtcgag agccacaaga tcggcagcgg 60

cagcggcagc ctctctctcc tccgccgcgc cgcctctctt tccctctccc aactcccgta 120

cgggcgtaac gccgcctcgt cgccggatct ttttttcgt tccccggcg agcttgcaact 180

ctcttttctt cgattctctt tcggctttgc ttcttcggtc gctttttttt ttggctgttc 240


```

ctggtcgatc ccggaagagg aggtgggtcgg gaaccaagat ttgaactact cctcctcatg 300
gcttccatgc cgcagctgga ggaaaaggat tcatccgatt tggcaatcaa caaaggcccg 360
tcgctggatc ttgtcaaadc tcccttgatg atgaatgatg catctgcaac tgtaactgca 420
atgcagccca acgaggggat ggaggagttc cctgtcaagc aggttcggaa gccttacacg 480
atcaccaagc agcgggagaa gtggacggaa gaagagcacg acaagttcct tgaggcattg 540
aagctgtatg gtcgctcttg gcgtcagata caagagcaca ttggcacaaa gactgctgtc 600
caaattcgga gccatgcccc gaagtttttc tcaaagggtg tgctgagacc cggttccaat 660
aatgcgattg agatccctcc acctaggcca aagagaaaac cacttcatcc gtaccctcga 720
aagtgtgcga attctggatc ggacgcaaat ccagcaacag cgcaattgaa gcttgctcct 780
ggctcatctt catctggctc tgatcaagag aatgggtctc ctatatcggg gctatctgcg 840
atgcagtcag atgcttttgg atcgtcagta tccaatccat caaccagatg tacctcccca 900
gcgtcatctg acgatggaaa taacattccc acattcacga gtggggaaga taataatgtg 960
ccttgtgaac caacagtgat agatccgtcc cagtcccaca aggaaataga ccaagataga 1020
aaagacgtga ataatatgtc tgaagaggat tcttcagaag aagagggtga agaaacaagc 1080
ttgaagctat ttgggaggac agttgtcatc ccagatccaa ggaagagaag ctctcagat 1140
ccgaagcatg aaagtgagga gcagatatca cagccttcca atgaagaaat gttgcaggct 1200
tcttcacatc ttggggagat tccagcggca tattgtgcac caaatgggtg gtttatgtca 1260
tacaattctt tccattcca attcggtgaa tcagcggcag atgctagaat tcccccttta 1320
cacgtgtggt ggccttacta cggttttgct ccattagcc atcctagagg actaagcaca 1380
gtgatgcagc agactgaagg tagtgacgag agtgacggtg tgaagagcca ctcatccgaa 1440
tcaagttcgg actccgggga aaatgtgcag atgaccgctc ccagagctc gagaatagtg 1500
gaatcacttg gagcgattta cgtccgagac tcagggttcaa gttttgagct aaaaccgagc 1560
gcaaattcag cgtttgtaag agtgaagcca agcaacagcg gagatgaaga gatttcggcc 1620
aaaattcatg agtactaa 1638

```

<210> 44

<211> 545

<212> PRT

<213> Oryza sativa

<400> 44

```

Met  Cys  Asp  Cys  Arg  Trp  Gly  Leu  Ala  Gln  Gly  Ser  Arg  Glu  Pro  Gln
1      5      10      15

Asp  Arg  Gln  Arg  Gln  Arg  Gln  Pro  Pro  Pro  Pro  Pro  Pro  Pro  Leu
      20      25      30

Leu  Phe  Pro  Leu  Pro  Thr  Pro  Val  Arg  Ala  Tyr  Gly  Arg  Leu  Val  Ala
      35      40      45

Gly  Phe  Phe  Phe  Ser  Phe  Pro  Arg  Arg  Ala  Cys  Thr  Leu  Phe  Ser  Ser
      50      55      60

Ile  Leu  Phe  Arg  Leu  Cys  Phe  Phe  Gly  Arg  Phe  Phe  Phe  Trp  Leu  Phe
65      70      75      80

Leu  Val  Asp  Pro  Gly  Lys  Gly  Gly  Gly  Arg  Glu  Pro  Arg  Phe  Glu  Leu
      85      90      95

Leu  Leu  Leu  Met  Ala  Ser  Met  Pro  Gln  Leu  Glu  Glu  Lys  Asp  Ser  Ser
      100     105     110

Asp  Leu  Ala  Ile  Asn  Lys  Gly  Pro  Ser  Leu  Asp  Leu  Val  Lys  Ser  Pro
      115     120     125

Leu  Met  Met  Asn  Asp  Ala  Ser  Ala  Thr  Val  Thr  Ala  Met  Gln  Pro  Asn
      130     135     140

Glu  Gly  Met  Glu  Glu  Phe  Pro  Val  Lys  Gln  Val  Arg  Lys  Pro  Tyr  Thr
145     150     155     160

Ile  Thr  Lys  Gln  Arg  Glu  Lys  Trp  Thr  Glu  Glu  Glu  His  Asp  Lys  Phe
      165     170     175

Leu  Glu  Ala  Leu  Lys  Leu  Tyr  Gly  Arg  Ser  Trp  Arg  Gln  Ile  Gln  Glu
      180     185     190

His  Ile  Gly  Thr  Lys  Thr  Ala  Val  Gln  Ile  Arg  Ser  His  Ala  Gln  Lys
      195     200     205

Phe  Phe  Ser  Lys  Val  Val  Arg  Glu  Pro  Gly  Ser  Asn  Asn  Ala  Ile  Glu
      210     215     220

Ile  Pro  Pro  Pro  Arg  Pro  Lys  Arg  Lys  Pro  Leu  His  Pro  Tyr  Pro  Arg
225     230     235     240

Lys  Cys  Ala  Asn  Ser  Gly  Ser  Asp  Ala  Asn  Pro  Ala  Thr  Ala  Gln  Leu
      245     250     255

Lys  Leu  Ala  Pro  Gly  Ser  Ser  Ser  Ser  Gly  Ser  Asp  Gln  Glu  Asn  Gly
      260     265     270

Ser  Pro  Ile  Ser  Val  Leu  Ser  Ala  Met  Gln  Ser  Asp  Ala  Phe  Gly  Ser

```

275		280		285
Ser Val Ser Asn Pro Ser Thr Arg Cys Thr Ser Pro Ala Ser Ser Asp				
290		295		300
Asp Gly Asn Asn Ile Pro Thr Phe Thr Ser Gly Glu Asp Asn Asn Val				
305		310		315 320
Pro Cys Glu Pro Thr Val Ile Asp Pro Ser Gln Ser His Lys Glu Ile				
		325		330 335
Asp Gln Asp Arg Lys Asp Val Asn Asn Met Ser Glu Glu Asp Ser Ser				
		340		345 350
Glu Glu Glu Val Gln Glu Thr Ser Leu Lys Leu Phe Gly Arg Thr Val				
		355		360 365
Val Ile Pro Asp Pro Arg Lys Arg Ser Ser Ser Asp Pro Lys His Glu				
		370		375 380
Ser Glu Glu Gln Ile Ser Gln Pro Ser Asn Glu Glu Met Leu Gln Ala				
385		390		395 400
Ser Ser Ser Val Gly Glu Ile Pro Ala Ala Tyr Cys Ala Pro Asn Gly				
		405		410 415
Trp Phe Met Ser Tyr Asn Ser Phe Pro Phe Gln Phe Gly Glu Ser Ala				
		420		425 430
Ala Asp Ala Arg Ile Pro Pro Leu His Val Trp Trp Pro Tyr Tyr Gly				
		435		440 445
Phe Ala Pro Ile Ser His Pro Arg Gly Leu Ser Thr Val Met Gln Gln				
		450		455 460
Thr Glu Gly Ser Asp Glu Ser Asp Gly Val Lys Ser His Ser Ser Glu				
465		470		475 480
Ser Ser Ser Asp Ser Gly Glu Asn Val Gln Met Thr Ala Pro Gln Ser				
		485		490 495
Ser Arg Ile Val Glu Ser Leu Gly Ala Ile Tyr Val Arg Asp Ser Gly				
		500		505 510
Ser Ser Phe Glu Leu Lys Pro Ser Ala Asn Ser Ala Phe Val Arg Val				
		515		520 525
Lys Pro Ser Asn Ser Gly Asp Glu Glu Ile Ser Ala Lys Ile His Glu				
		530		535 540
Tyr				
545				

<210> 45
 <211> 2196
 <212> DNA
 <213> *Oryza sativa*

<400> 45
 atggctgttt tggacacatt gcgtgtgacg acgcccctca atcatctcaa acctgatcag 60
 tactacccaa gtggtgcacg tggcatggtg cgcctgacag gtcgggcca cgggtcacga 120
 tccacatcgc ccccgaaaaa ggtcagtgga gaaaacaatg agcggttgag cgcaaacat 180
 gtggccgagg cgtgtctcgg agccctgcag cgccaacccg taagcttggt gccagggata 240
 gtccatttgt tctctcacg gtggtacgga gaagagaaca acaaacagcg ggttctctcc 300
 ctgctgtgcg cctctggttg ggcgacatgt ggcgggcgga acatgccatc gattcgtgaa 360
 gaccacgaa agcgcacggt ttactccta caccacata ctagtagtag tagcggtagc 420
 agtacgcaa ctagtggcaa ccagacacat ggacaagtag cagcgaagca ggcgaaggaa 480
 ggggcaagtg agtgccgtga agcagcgaga gccacaagaa tcgagcggct tccgggggct 540
 ctctacagga ttgaggcgaa aagtgcaaag ccggaaggag agagagagaa aaaagcgttg 600
 tctcgagaaa gattggagtt tttggtggtt tgggttgtgc ggttcttgct tttgttggcg 660
 caagcgcacg gctggttctg ggacatagaa accaaggcga ggaatgatca aggacctgtt 720
 gctgatcatg ttgggcacca aaacctcatg gaaaacctca cagatcctct ggattccagt 780
 ggcattggaca tgatggacga agcgcgaatt cctaaggcgc ggaagccata cagcataacg 840
 aagcaaaggg agaaatggac cgaggacgaa cataagctgt tcttggaagc cctgcagctc 900
 catggccgag cctggcgacg catccaagga tacctcgctc cggttgcaga gcacataggc 960
 accaagactg ccgtgcagat caggagccac gcacagaagt tcttctccaa ggtcatcaaa 1020
 gaatcatctg gggacaattg caacagcttg ggtgctgcat catcaattca gattcccccg 1080
 ccgcggccga agcgtaaagc tgttcatccg taccacgca atctaggag cacagccagc 1140
 aagaacgtcc ctgcactgaa acagctagag aagcctcagc tgcaggtgca gtctctctac 1200
 gaccaggaca atgggtcgcg gacgtcagtg ttaacagtac cacagatacg ggctgataca 1260
 ctgggaagtg agagtgggtg gtcgccaacc tcgacgattg atattgaaga gagatgtcct 1320
 acaccaagca tagcaactgc tgagttagct atggagttgc ctctacaaa tgacgaggtc 1380
 aagggcaatg gcgatcatga agaagttaca tgtgacagat caggagttcc agtccttagg 1440

ctatttggca agaggggttat ggtgaatgat ttacatcaga tgtcagcccc tgatgccggg 1500
 aacctgcaaa ctgtggcaga catggaagtg gacgcttcag ctgagacacc aactagtgga 1560
 actgggaaat tctcttccca tgggtgcagca gaagcaaata catggaaccc atggctgact 1620
 aacacacagc agtttctgta ttatcttcct aacggacaaa ttttctccgt gcattctgct 1680
 ctcccatgct tcacctacca taatgagggg gttacttgca cccagttttc aaaccacag 1740
 gtggtagcct cagatcaaca gcatcaacac caaacttctg aagctgtaga ttacaagggt 1800
 atacagaggg aaggatcttg gacggagtca aatacatcct caagcagtgt gcctgaaaca 1860
 gcaactcata attcagagac tacagaatca tatagaaacg gaaacagaaa cgaagatgaa 1920
 atggtacctt ctccagattc aagaaaatgt gtgagcccag gttccaactg caggcgaggc 1980
 tttgtgccgt acaagagatg tgttgctgat agtgaggcgc tgetgaagtc acaggcgctt 2040
 caggaggagg cagacgagag attgacgagg ctgcagattg aaaccgcgtt agtactccac 2100
 aaaacagggg gtggccgtgg ttcttgtctg gccataggtc aatggctgtc atgcaatgac 2160
 aagtattttg atagagttgt agaaggggtg tactga 2196

<210> 46
 <211> 731
 <212> PRT
 <213> *Oryza sativa*

<400> 46

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

Arg Asn Met Pro Ser Ile Arg Glu Asp Pro Arg Lys Arg Thr Val Leu
 115 120 125
 Leu Leu His Pro His Thr Ser Ser Ser Ser Gly Ser Ser Thr Pro Thr
 130 135 140
 Ser Gly Asn Gln Thr His Gly Gln Val Ala Ala Lys Gln Ala Lys Glu
 145 150 155 160
 Gly Ala Ser Glu Cys Arg Glu Ala Ala Arg Ala Thr Arg Ile Glu Arg
 165 170 175
 Leu Pro Gly Ala Leu Tyr Arg Ile Glu Ala Lys Ser Ala Lys Pro Glu
 180 185 190
 Arg Glu Arg Glu Lys Lys Ala Leu Ser Arg Glu Arg Leu Glu Phe Leu
 195 200 205
 Val Val Trp Val Val Arg Phe Leu Leu Leu Leu Ala Gln Ala His Gly
 210 215 220
 Trp Phe Trp Asp Ile Glu Thr Lys Ala Arg Asn Asp Gln Gly Pro Val
 225 230 235 240
 Ala Asp His Val Gly His Gln Asn Leu Met Glu Asn Leu Thr Asp Pro
 245 250 255
 Leu Asp Ser Ser Gly Met Asp Met Met Asp Glu Ala Arg Ile Pro Lys
 260 265 270
 Ala Arg Lys Pro Tyr Thr Ile Thr Lys Gln Arg Glu Lys Trp Thr Glu
 275 280 285
 Asp Glu His Lys Leu Phe Leu Glu Ala Leu Gln Leu His Gly Arg Ala
 290 295 300
 Trp Arg Arg Ile Gln Gly Tyr Leu Ala Pro Val Ala Glu His Ile Gly
 305 310 315 320
 Thr Lys Thr Ala Val Gln Ile Arg Ser His Ala Gln Lys Phe Phe Ser
 325 330 335
 Lys Val Ile Lys Glu Ser Ser Gly Asp Asn Cys Asn Ser Leu Gly Ala
 340 345 350
 Ala Ser Ser Ile Gln Ile Pro Pro Pro Arg Pro Lys Arg Lys Pro Val
 355 360 365
 His Pro Tyr Pro Arg Asn Leu Gly Ser Thr Ala Ser Lys Asn Val Pro
 370 375 380
 Ala Leu Lys Gln Leu Glu Lys Pro Gln Leu Gln Val Gln Ser Leu Tyr
 385 390 395 400

Asp Gln Asp Asn Gly Ser Pro Thr Ser Val Leu Thr Val Pro Gln Ile
 405 410 415
 Arg Ala Asp Thr Leu Gly Ser Glu Ser Gly Gly Ser Pro Thr Ser Thr
 420 425 430
 Ile Asp Ile Glu Glu Arg Cys Pro Thr Pro Ser Ile Ala Thr Ala Glu
 435 440 445
 Leu Ala Met Glu Leu Pro Pro Thr Asn Asp Glu Val Lys Gly Asn Gly
 450 455 460
 Asp His Glu Glu Val Thr Cys Asp Arg Ser Gly Val Pro Val Leu Arg
 465 470 475 480
 Leu Phe Gly Lys Arg Val Met Val Asn Asp Leu His Gln Met Ser Ala
 485 490 495
 Pro Asp Ala Gly Asn Leu Gln Thr Val Ala Asp Met Glu Val Asp Ala
 500 505 510
 Ser Ala Glu Thr Pro Thr Ser Gly Thr Gly Lys Phe Ser Ser His Gly
 515 520 525
 Ala Ala Glu Ala Asn Thr Trp Asn Pro Trp Leu Thr Asn Thr Gln Gln
 530 535 540
 Phe Leu Tyr Tyr Leu Pro Asn Gly Gln Ile Phe Ser Val His Ser Ala
 545 550 555 560
 Leu Pro Cys Phe Thr Tyr His Asn Glu Gly Val Thr Cys Thr Gln Phe
 565 570 575
 Ser Asn Pro Gln Val Val Ala Ser Asp Gln Gln His Gln His Gln Thr
 580 585 590
 Ser Glu Ala Val Asp Tyr Lys Gly Ile Gln Arg Glu Gly Ser Trp Thr
 595 600 605
 Glu Ser Asn Thr Ser Ser Ser Ser Val Pro Glu Thr Ala Thr His Asn
 610 615 620
 Ser Glu Thr Thr Glu Ser Tyr Arg Asn Gly Asn Arg Asn Glu Asp Glu
 625 630 635 640
 Met Val Pro Ser Pro Asp Ser Arg Lys Cys Val Ser Pro Gly Ser Asn
 645 650 655
 Cys Arg Arg Gly Phe Val Pro Tyr Lys Arg Cys Val Ala Asp Ser Glu
 660 665 670
 Ala Leu Leu Lys Ser Gln Ala Pro Gln Glu Glu Ala Asp Glu Arg Leu
 675 680 685

Thr Arg Leu Gln Ile Glu Thr Ala Leu Val Leu His Lys Thr Gly Ser
690 695 700

Gly Arg Gly Ser Cys Leu Ala Ile Gly Gln Trp Leu Ser Cys Asn Asp
705 710 715 720

Lys Tyr Phe Asp Arg Val Val Glu Gly Leu Tyr
725 730

<210> 47
<211> 764
<212> DNA
<213> Oryza sativa

<400> 47
atggggaggg gaagggttga gctgaagcgc atcgagaaca agatcaaccg gcaggtcacc 60
ttctccaagc gccgcaaccg cctcctcaag aaggcgtacg agctctccgt gctctgcgac 120
gcggagggtgg cgctcatcat cttctccagc cgtggcaagc tgtacgaatt cggtagcgcc 180
gggaataaac aagacattgg aaaagtacaa tagttgctgt tacaacgctc aagggttcaaa 240
tagtgctctt gctggtggtg aacatcagag agctggtacc aagagatgtc aaggctcaag 300
actaagcttg aatgtctcca acgctctcag agaggcacat gcttggtgaa gatcttggac 360
cattgagcat aaaggaactg cagcagctgg agaagcaact tgagtactca ctgtcacagg 420
ctcgacaacg aaagcagaca caaatcatga tggagcaggt cgacgatctt cgccggaaga 480
ggaacgccag cttggagagc tcaataagca actgaaaaac aagagctaga agctgaagcc 540
gatagcagca actgcagatc agccatccag gattcctggg tccatggcac cgtcgtcagt 600
ggcggcagag tgctgaatgc tcaaccacca ccagatattg actgtgagcc tactctgcaa 660
attggaggta ctatcaatth gtccgtcctg aggcggccaa tccaagaagc aatggaggag 720
gaggggatca gaacaacaac tttgtgatgg gatggccccct ctga 764

<210> 48
<211> 176
<212> PRT
<213> Oryza sativa

<220>
<221> X_region
<222> (87)..(87)
<223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (161)..(161)
 <223> Xaa = any naturally occurring amino acid

<400> 48

Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
 1 5 10 15

Arg Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
 20 25 30

Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
 35 40 45

Ser Ser Arg Gly Lys Leu Tyr Glu Phe Gly Ser Pro Gly Ile Asn Lys
 50 55 60

Thr Leu Glu Lys Tyr Asn Ser Cys Cys Tyr Asn Ala Gln Gly Ser Asn
 65 70 75 80

Ser Ala Ser Cys Trp Trp Xaa Thr Ser Glu Ser Trp Tyr Gln Glu Met
 85 90 95

Ser Arg Leu Lys Thr Lys Leu Glu Cys Leu Gln Arg Ser Ser Glu Arg
 100 105 110

His Met Leu Gly Glu Asp Leu Gly Pro Leu Ser Ile Lys Glu Leu Gln
 115 120 125

Gln Leu Glu Lys Gln Leu Glu Tyr Ser Leu Ser Gln Ala Arg Gln Arg
 130 135 140

Lys Gln Thr Gln Ile Met Met Glu Gln Val Asp Asp Leu Arg Arg Lys
 145 150 155 160

Xaa Glu Arg Gln Leu Gly Glu Leu Asn Lys Gln Leu Lys Asn Lys Ser
 165 170 175

<210> 49
 <211> 1228
 <212> DNA
 <213> Oryza sativa

<400> 49

atggatgctc tttgtgattt ctgcagggag caaaggtcaa tgggtctattg ccgatcagat 60

gcagcatcct tgtgcttgtc gtgtgatcgt aatgtgcatt cagccaatgc cctttcccca 120

cgccatacaa gaacccttct ttgtgatcgt tgtgttgagc agcctgcagc agtccgttgt 180

```

cttgaagaga acacctcact ttgccaaaac tgtgactgga atggacatgg ggcagcatcc 240
tcagctgctg ggcataaaag gcagaccata aactgttact cgggggtgcc atcgtcggca 300
gagctctcga gaatctggtc ctttagtatg gatatcccaa ctgtagctgc ggagccaaac 360
tgtgaggaag gaataaacat gatgagcatt aatgacaatg acgtcaataa ccattgtggt 420
gctccagaag atggccgctt gttggatata gctagcacag cactcatgag tgatttacct 480
acaggagaca agttcaaacc tttaataggt tcttcttcag gagatgggat gaatcttctg 540
ccacttaatt cagatcagcc agctgagcca gtttcaacga cgcctaagag gcaccctgtg 600
tcacagacaa agatatgttc aatgatggga gcgtatatgg agacttctgt gtggatgatg 660
ctgacctaac atttgagaat tatgaagaat tattcggtag ctctcacgtt caaacagaac 720
aactctttga tgatgctgga attgacagtt attttgaaat gaaggatgtg ccagcagatg 780
aatctaata gaagcagccca aacctgtgca gccagaatgt agcaatgtcg catctgttga 840
ttctgggatg tcaaaccag ctgcaagggc tgattccagc cattgtattc ctggtaggca 900
ggctatatcc aacatatccc tttccttctc tggtttgact ggtgaaagca gtgctggata 960
ttttcaagat tgtggggtat cgtcaatgat ccttatgggt gagccaccct ggcatcctcc 1020
tggtcctgaa agttcatctg ctggaggcag cagagataat gctcttacac ggtacaagga 1080
gaagaagaag agaagaaaat ttgacaagaa gataaggtag gcttctcgta aggctagggc 1140
agatgtgaga aaaaggggtca aaggacgggt tgtaaaggct ggcgaagcat atgactacga 1200
tccgctgagc caaacaagaa gctactaa 1228

```

<210> 50

<211> 129

<212> PRT

<213> Oryza sativa

<220>

<221> X_region

<222> (108)..(108)

<223> Xaa = any naturally occurring amino acid

<400> 50

```

Met Phe Asn Asp Gly Ser Val Tyr Gly Asp Phe Cys Val Asp Asp Ala
1           5           10           15
Asp Leu Thr Phe Glu Asn Tyr Glu Glu Leu Phe Gly Thr Ser His Val
          20           25           30

```

Gln Thr Glu Gln Leu Phe Asp Asp Ala Gly Ile Asp Ser Tyr Phe Glu
 35 40 45
 Met Lys Asp Val Pro Ala Asp Glu Ser Asn Glu Ser Ser Pro Asn Leu
 50 55 60
 Cys Ser Gln Asn Val Ala Met Ser His Leu Leu Ile Leu Gly Cys Gln
 65 70 75 80
 Thr Gln Leu Gln Gly Leu Ile Pro Ala Ile Val Phe Leu Val Gly Arg
 85 90 95
 Leu Tyr Pro Thr Tyr Pro Phe Pro Ser Leu Val Xaa Leu Val Lys Ala
 100 105 110
 Val Leu Asp Ile Phe Lys Ile Val Gly Gly Arg Asn Lys His Asp Glu
 115 120 125

His

<210> 51
 <211> 3117
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (2061)..(2061)
 <223> r = adenine or guanine

<400> 51
 atggatcatg ggcgttatgc tccacagcat ggatgggaga acaacagtgc attggatggg 60
 tatggtgtca tcaatggccc agattatagg gccggtggat cctacaatgg taggaggttt 120
 gtggatgatg gcttcccagg agattcatat ggaagggggg cattctatca ggacacacat 180
 gatagaaatg tgtatccacc tgcacctct gtcccatgt tgtctcaacc tcgaagggtt 240
 catgatgatg aatattcaac tgctagagat tacagaaggc atgacactta tcattgcaat 300
 gatggaaagc atcatgaatt tgaatcctat ggaggagttg atatactcca cgataattat 360
 gctgctgata actatgaatc tggtagttac cgtgattttt gttttgatag gaacaaaagg 420
 attgggagta gagaccgtgc ggagtttcac ggtgactttg aagaccggta ccgcagctct 480
 caccaaagta gagaggatag tagagaggat agctatgaga gagatcgtga ttatgatcgt 540
 tacagctatg attcggacta tgagaaaagc aggagagatg gtagctggag aaggcgtgat 600

ttatgtgaga	gtgaacatga	aagaagaggg	ttgagtcacg	aaagagacag	aagtccatac	660
atgcaacata	gccgttcacg	atctcgtggg	cgtagcgaac	gatcaagatc	tcggtctaga	720
tcaagatcac	cccgtggcaa	aagtcgtggg	cgaacccaaa	gagatggcct	ttatgatgac	780
aattcctttg	gcagaagaag	ggagtatgac	tgggatgaga	gaaggcatgg	agatttagtg	840
gctccatctg	ctactgttgt	tgtaagggt	ctgtcccaga	agactaatga	agatgaccta	900
aaccagattc	ttgtacaagt	tcctcttcat	aaactgttaa	actggttgca	agtgtattgt	960
ctggctgagt	ggggccctct	cogtagtggt	cgcgtaatta	aggaacgatc	ttctggcatg	1020
tctcggggat	ttgcttttat	cgacttcctt	actgtggaag	ctgcccgcag	aatgatggaa	1080
ggtgttggag	acaatggcct	tctaattgat	ggtagaaagg	tatTTTTTca	gtacagtagt	1140
aaaccaacta	gtgggatgag	tgggccttct	catggagaag	aaaattttac	aaggtataac	1200
tatgggcaca	ggactgctgc	tgcgccatgt	gattggatat	gtactatatg	tggatgcatg	1260
aactttgctc	gcagaacttc	atgttttcag	acactagcat	tttcagggtcc	aactcatggt	1320
ttagttgttc	gtggtctgga	tgaaaatgct	gacgaggaaa	tgcttcgata	tgaatttgcc	1380
aagcatgcac	ctataaagga	cattcgcctt	gttcgggata	agtttactca	tgtgtccaga	1440
ggttttgctt	tcattccattt	tcattcagtg	gaagaagcta	caaaagctct	tgaagctacg	1500
aatgggatta	cacttgagaa	gaatggtcag	gttttacgtg	taacttatgc	taaaagcaca	1560
catggaccgg	tatcaggggc	ttcacagtca	aacagccttg	ctgcagctgc	cattgaggct	1620
gcatcatttt	ctcagcagta	tgatgctata	ggttggggcc	caaaagagta	caatcctgat	1680
gacaaactga	atagcaactc	ggaacctcag	agcagtggct	ctgctccaca	gtctggattt	1740
gtttgggatg	aaaagtctgg	ttattactat	gattcagcct	caggattcta	ttatgatgga	1800
aataccggac	tttactatga	tggtaatgct	ggagtgtggt	attcatatga	tcaacaaact	1860
caacagtatg	tcccttgtag	tgaacagaac	agcagtaaag	cagctgggga	catggcaaat	1920
actagcacga	agacctcaga	aagtagtggt	aaaaatgtag	tgatttctgc	acctgcagcc	1980
acgattaaac	aaagtgagaa	aacttcatta	cctgaagctg	ttcaagctgc	tgcaagtgca	2040
gcgctggctg	cagaaaagaa	rgaaaaggag	agagcaaaag	agatcaagtt	ggcatcaaaa	2100
ggtagtcttt	tggcgaataa	gaaaaagatg	aataatgtcc	tagcaatgtg	gaaacaaaga	2160
aatcaagaag	gccaggcagg	gcgtgctatc	cttgatgaca	aggaaccatc	caactctgct	2220

gatgataaat taaacaattt acacaattca actggatttg cagtgaagc taagcctaaa 2280
 tctgatgttg gaaatgctaa ggacatgaac tcgcctgcta gctataattc tcttggacga 2340
 acagctgcac ccacagaaat gatagattct gatataaagc ctacacctgt cagtaacagt 2400
 ttaggaacta ccattatggg tgtcataaga ggctctggta gaggaatagt cagatcagac 2460
 actgcatttc atgcgccatc tgatgctggg ggtgctgatt ctttctccaa cataccaaca 2520
 agcacatgtg ggttaacagc aaatgctggg gcacctactt ctgctccctt caaaacagaa 2580
 ttatctgctc tagcatctta tactccatct ggagtttcag ggagtggtaa acgtcggttt 2640
 tctgaggctc caggacattc acagtacaga gatcgggctg cagaaaggcg aaatctatat 2700
 ggatcatctt cttcacttgg aagtgataat gatggactgg atccaactgg tgagtacct 2760
 cgaaggggtc caagtgagat ggggtcaatg ccatttcctc caggggtggg tgaacgttct 2820
 agcggtgaaa ttggcaatac tgaaaattat gaggttatca ctgctgatag agcaatagat 2880
 gagagcaatg tgggcaatcg tattctgctc aatatgggtt ggcaagaagg actgggtctc 2940
 ggaaaaactg gtagtggcat caaggaacca gtgcaggcta aatctgtcga tgtgagggt 3000
 ggacttggaa gtcagcagcg gaaatcttct gaccgctcgc tggaggcgca agctggtgat 3060
 agctacaaaa ccatcatcca gaagaaggct atggcaagat tcaggagat gtcttga 3117

<210> 52

<211> 1038

<212> PRT

<213> *Oryza sativa*

<400> 52

Met Asp His Gly Arg Tyr Ala Pro Gln His Gly Trp Glu Asn Asn Ser
 1 5 10 15

Ala Leu Asp Gly Tyr Gly Val Ile Asn Gly Pro Asp Tyr Arg Ala Gly
 20 25 30

Gly Ser Tyr Asn Gly Arg Arg Phe Val Asp Asp Gly Phe Pro Gly Asp
 35 40 45

Ser Tyr Gly Arg Gly Ala Phe Tyr Gln Asp Thr His Asp Arg Asn Val
 50 55 60

Tyr Pro Pro Ala Pro Ser Val Pro Met Leu Ser Gln Pro Arg Arg Phe
 65 70 75 80

His Asp Asp Glu Tyr Ser Thr Ala Arg Asp Tyr Arg Arg His Asp Thr

					85					90					95			
Tyr	His	Cys	Asn	Asp	Gly	Lys	His	His	Glu	Phe	Glu	Ser	Tyr	Gly	Gly			
			100					105					110					
Val	Asp	Ile	Leu	His	Asp	Asn	Tyr	Ala	Ala	Asp	Asn	Tyr	Glu	Ser	Gly			
		115					120					125						
Ser	Tyr	Arg	Asp	Phe	Cys	Phe	Asp	Arg	Asn	Lys	Arg	Ile	Gly	Ser	Arg			
	130					135					140							
Asp	Arg	Ala	Glu	Phe	His	Gly	Asp	Phe	Glu	Asp	Arg	Tyr	Arg	Ser	Ser			
145					150				155						160			
His	Gln	Ser	Arg	Glu	Asp	Ser	Arg	Glu	Asp	Ser	Tyr	Glu	Arg	Asp	Arg			
				165				170						175				
Asp	Tyr	Asp	Arg	Tyr	Ser	Tyr	Asp	Ser	Asp	Tyr	Glu	Lys	Ser	Arg	Arg			
			180					185					190					
Asp	Gly	Ser	Trp	Arg	Arg	Arg	Asp	Leu	Cys	Glu	Ser	Glu	His	Glu	Arg			
		195					200					205						
Arg	Gly	Leu	Ser	His	Glu	Arg	Asp	Arg	Ser	Pro	Tyr	Met	Gln	His	Ser			
	210					215					220							
Arg	Ser	Arg	Ser	Arg	Gly	Arg	Asp	Glu	Arg	Ser	Arg	Ser	Arg	Ser	Arg			
225					230					235					240			
Ser	Arg	Ser	Pro	Arg	Gly	Lys	Ser	Arg	Gly	Arg	Asn	Gln	Arg	Asp	Gly			
				245					250					255				
Phe	Tyr	Asp	Asp	Asn	Ser	Phe	Gly	Arg	Arg	Arg	Glu	Tyr	Asp	Trp	Asp			
			260					265					270					
Glu	Arg	Arg	His	Gly	Asp	Leu	Val	Ala	Pro	Ser	Ala	Thr	Val	Val	Val			
		275				280						285						
Lys	Gly	Leu	Ser	Gln	Lys	Thr	Asn	Glu	Asp	Asp	Leu	Asn	Gln	Ile	Leu			
	290					295					300							
Val	Gln	Val	Pro	Leu	His	Lys	Leu	Leu	Asn	Trp	Leu	Gln	Val	Tyr	Cys			
305					310				315						320			
Leu	Ala	Glu	Trp	Gly	Pro	Leu	Arg	Ser	Val	Arg	Val	Ile	Lys	Glu	Arg			
				325					330					335				
Ser	Ser	Gly	Met	Ser	Arg	Gly	Phe	Ala	Phe	Ile	Asp	Phe	Pro	Thr	Val			
			340					345					350					
Glu	Ala	Ala	Arg	Arg	Met	Met	Glu	Gly	Val	Gly	Asp	Asn	Gly	Leu	Leu			
		355					360					365						
Ile	Asp	Gly	Arg	Lys	Val	Phe	Phe	Gln	Tyr	Ser	Ser	Lys	Pro	Thr	Ser			

370		375		380
Gly Met Ser Gly Pro Ser His Gly Glu Glu Asn Phe Thr Arg Tyr Asn				
385		390		395
Tyr Gly His Arg Thr Ala Ala Ala Pro Cys Asp Trp Ile Cys Thr Ile				
	405		410	415
Cys Gly Cys Met Asn Phe Ala Arg Arg Thr Ser Cys Phe Gln Thr Leu				
	420		425	430
Ala Phe Ser Gly Pro Thr His Val Leu Val Val Arg Gly Leu Asp Glu				
	435		440	445
Asn Ala Asp Glu Glu Met Leu Arg Tyr Glu Phe Ala Lys His Ala Pro				
	450		455	460
Ile Lys Asp Ile Arg Leu Val Arg Asp Lys Phe Thr His Val Ser Arg				
465		470		475
Gly Phe Ala Phe Ile His Phe His Ser Val Glu Glu Ala Thr Lys Ala				
	485		490	495
Leu Glu Ala Thr Asn Gly Ile Thr Leu Glu Lys Asn Gly Gln Val Leu				
	500		505	510
Arg Val Thr Tyr Ala Lys Ser Thr His Gly Pro Val Ser Gly Ala Ser				
	515		520	525
Gln Ser Asn Ser Leu Ala Ala Ala Ala Ile Glu Ala Ala Ser Phe Ser				
	530		535	540
Gln Gln Tyr Asp Ala Ile Gly Trp Ala Pro Lys Glu Tyr Asn Pro Asp				
545		550		555
Asp Lys Leu Asn Ser Asn Ser Glu Pro Gln Ser Ser Gly Ser Ala Pro				
	565		570	575
Gln Ser Gly Phe Val Trp Asp Glu Lys Ser Gly Tyr Tyr Tyr Asp Ser				
	580		585	590
Ala Ser Gly Phe Tyr Tyr Asp Gly Asn Thr Gly Leu Tyr Tyr Asp Gly				
	595		600	605
Asn Ala Gly Val Trp Tyr Ser Tyr Asp Gln Gln Thr Gln Gln Tyr Val				
	610		615	620
Pro Cys Ser Glu Gln Asn Ser Ser Lys Ala Ala Gly Asp Met Ala Asn				
625		630		635
Thr Ser Thr Lys Thr Ser Glu Ser Ser Gly Lys Asn Val Val Ile Ser				
	645		650	655
Ala Pro Ala Ala Thr Ile Lys Gln Ser Glu Lys Thr Ser Leu Pro Glu				

660					665					670					
Ala	Val	Gln	Ala	Ala	Ala	Ser	Ala	Ala	Leu	Ala	Ala	Glu	Lys	Asn	Glu
	675						680					685			
Lys	Glu	Arg	Ala	Lys	Glu	Ile	Lys	Leu	Ala	Ser	Lys	Gly	Ser	Leu	Leu
	690					695					700				
Ala	Asn	Lys	Lys	Lys	Met	Asn	Asn	Val	Leu	Ala	Met	Trp	Lys	Gln	Arg
705					710					715					720
Asn	Gln	Glu	Gly	Gln	Ala	Gly	Arg	Ala	Ile	Leu	Asp	Asp	Lys	Glu	Pro
				725					730					735	
Ser	Asn	Ser	Ala	Asp	Asp	Lys	Leu	Asn	Asn	Leu	His	Asn	Ser	Thr	Gly
			740					745					750		
Phe	Ala	Val	Lys	Ala	Lys	Pro	Lys	Ser	Asp	Val	Gly	Asn	Ala	Lys	Asp
	755						760					765			
Met	Asn	Ser	Pro	Ala	Ser	Tyr	Asn	Ser	Leu	Gly	Arg	Thr	Ala	Ala	Pro
	770					775					780				
Thr	Glu	Met	Ile	Asp	Ser	Asp	Ile	Lys	Pro	Thr	Pro	Val	Ser	Asn	Ser
785					790					795					800
Leu	Gly	Thr	Thr	Ile	Met	Gly	Val	Ile	Arg	Gly	Ser	Gly	Arg	Gly	Ile
				805					810					815	
Val	Arg	Ser	Asp	Thr	Ala	Phe	His	Ala	Pro	Ser	Asp	Ala	Gly	Gly	Ala
			820					825					830		
Asp	Ser	Phe	Ser	Asn	Ile	Pro	Thr	Ser	Thr	Cys	Gly	Leu	Thr	Ala	Asn
	835						840					845			
Ala	Gly	Ala	Pro	Thr	Ser	Ala	Pro	Phe	Lys	Thr	Glu	Leu	Ser	Ala	Leu
	850					855					860				
Ala	Ser	Tyr	Thr	Pro	Ser	Gly	Val	Ser	Gly	Ser	Gly	Lys	Arg	Arg	Phe
865					870					875					880
Ser	Glu	Ala	Pro	Gly	His	Ser	Gln	Tyr	Arg	Asp	Arg	Ala	Ala	Glu	Arg
				885					890					895	
Arg	Asn	Leu	Tyr	Gly	Ser	Ser	Ser	Ser	Leu	Gly	Ser	Asp	Asn	Asp	Gly
			900						905				910		
Leu	Asp	Pro	Thr	Gly	Glu	Tyr	Pro	Arg	Arg	Gly	Pro	Ser	Glu	Met	Gly
	915						920					925			
Ser	Met	Pro	Phe	Pro	Pro	Gly	Val	Gly	Glu	Arg	Ser	Ser	Gly	Glu	Ile
	930					935					940				
Gly	Asn	Thr	Glu	Asn	Tyr	Glu	Val	Ile	Thr	Ala	Asp	Arg	Ala	Ile	Asp

945 950 955 960
 Glu Ser Asn Val Gly Asn Arg Ile Leu Arg Asn Met Gly Trp Gln Glu
 965 970 975
 Gly Leu Gly Leu Gly Lys Thr Gly Ser Gly Ile Lys Glu Pro Val Gln
 980 985 990
 Ala Lys Ser Val Asp Val Arg Ala Gly Leu Gly Ser Gln Gln Arg Lys
 995 1000 1005
 Ser Ser Asp Pro Ser Leu Glu Ala Gln Ala Gly Asp Ser Tyr Lys Thr
 1010 1015 1020
 Ile Ile Gln Lys Lys Ala Met Ala Arg Phe Arg Glu Met Ser
 1025 1030 1035

<210> 53
 <211> 2583
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (381)..(381)
 <223> m = adenine or cytosine

<220>
 <221> misc_feature
 <222> (1886)..(1886)
 <223> k = guanine or thymine

<400> 53
 atgagcttcg gtggcatgtt cgatggcgcc ggctccggcg tcttctccta cgacgccggc 60
 ggcggcgggc gcggcggcgg agtacacaac tccgcctcc tccctacgcc gcccgcccc 120
 aagcccggcg gcgggttcgc cgctcccggc ctctctctcg gcctgatgga cggtagccag 180
 ctgggcgacg tgaaccggtc gcttgccatg atgggcaacg gcgggagcgg cagcggcggc 240
 gacggcgact cgctgggacg cggccgggag gaagagaacg acagccgctc cggcagcgac 300
 aacctggacg gcgcctccgg cgacgaactc gaccgggaca acagcaaccc gcgcaagaag 360
 aagaagcggt atcaccgcca maccgcgag caaatccaag aactcgaagc gtgttcaagg 420
 agtgtccaca tcctgacgag aagcaacgga tggagctcag ccggcggctc aacctcgaga 480
 gccgccagag acgcagatcg agcggcatga gaacgcgctg ctgcggcagg agaacgacaa 540
 gctgagggcg gagaacatga cgatccggga ggcgatgagg aaccgatgt gcgcgagctg 600

cgccggcgcg gcggtgctcg gcgaggtgtc gctggaggag cagcacctgc gcatcgagaa 660
cgcgcggtc aaggacgagc tcgaccggt ctgcgcgtc gccggcaagt tcctcgccg 720
gccaatctcg agcatctcct ccccgggccc gccgtcgtg caggcgtgct cgggcctcga 780
gctcggtc ggcagcaacg gcgggttcg gctcggtcg ctcggggcct cggcggtgat 840
gcagtccatc cccgacctga tggcggtc ctccggcctg acgggggggc ccgtgggtc 900
cgccgccatg cgcctaccg cgggcattgg cggcctcgac ggcgccatgc acgccggtc 960
ggcgacggc ggcgccatcg accgcggt gtcctggag ctgccctcg cggccatgga 1020
cgagtcgtc aagggtggcg agatggacga gccgtctg ctcccagcc tcgacggcg 1080
gttcgagacg ctcaactacg acgagtacca ccgcggtc gccagggtgg tcggccagt 1140
ccccgcggc tacgtctccg aggcgacgc cgagtcggc atcgccatca tcagcagct 1200
cgacctcgtc gacagctca tggacggc ggtggtcgga gatgttccc tgcgtggtg 1260
cgagagcgag caccacggac atcatctcca gcggcatggg cggcacgcg agcgggtcaa 1320
ttcaactgga tgcacggga gctgcaggtg ctgtcgccg tggtgccgat acgggaggtg 1380
gtgttctcgtc gtttctgcaa gcagcacgc gagggcctc gggcggtggt cgacgtctc 1440
gtcgacggc tcctccgcc cgaccagaac ggcggtcg ggtcgtcgtc gtcgtcgtac 1500
atgggtgcc gcctctccc caccggtgc atcgtgcagg acatgaacaa tggctactcc 1560
aagaggtcac gtgggtggt caccggtg acgacgagc ggcggcgac cagctctacc 1620
ggcggtgct ccgtccggc caggccctc gcgccggc ctggctcgc tcgtccagc 1680
gacagtcca gtatctcgc atcctctgct ccaactccct cccgcccgc gaccagctg 1740
gcagcgataa cgccgtggg gagggaggc atgctgaagc tggcgagcg gatgacggac 1800
aacttctcgt cggggtgtg cgcgtcggc gcgcagaagt ggcgggcct ggacgagtgg 1860
cgcggaagg ttggggcg cgcgktggc ggcggtggc acggggagga caagtgcg 1920
atgatggcg ggcacagct gggcggtcg gggagccgc cggcggtcgt gctgagcgc 1980
accagtcgtg tcgggtccc cggcacgtg ccgcagcg tgttcgacta cctccgcgac 2040
gagcagggc gcggcgactg ggacatcct gccaacggc aggccatgca ggagatggac 2100
cacatcgcca agggccagca ccacggcaat gccgttccc tcctcgtcc aatcatatc 2160
ttgaaaaaac aaataactaat gtctctctc caggccacaa gcgggaacca gaacaacatg 2220

```

ctgacccctcc aggagacgtg caccgactca tcgggctccc tcgtcgtcta cgctccggtg 2280
gacgtgcagt cgatgcacgt cgtcatgaac ggcggcgact ccgcctacgt ctccctcctc 2340
ccctccgggt tcgccatcct ccccgacggc cacaacaatg gcgcctcgcc gtctccggca 2400
gaggggtggcg gtggcggtag caacaacacc gggtcgctgg tgacgggtggc gttccagata 2460
ctcgtcaaca acctgccgac ggcgaagctc accgtggagt ccgtcgacac cgtcagcaac 2520
ctcctctcct gcaccatcca gaagatcaag tcggcactgc aggccagcat catctcacca 2580
tag 2583

```

```

<210> 54
<211> 859
<212> PRT
<213> Oryza sativa

```

```

<220>
<221> X_region
<222> (127)..(127)
<223> Xaa = any naturally occurring amino acid

```

```

<220>
<221> X_region
<222> (386)..(386)
<223> Xaa = any naturally occurring amino acid

```

```

<220>
<221> X_region
<222> (521)..(521)
<223> Xaa = any naturally occurring amino acid

```

```

<220>
<221> X_region
<222> (628)..(628)
<223> Xaa = any naturally occurring amino acid

```

```

<400> 54

```

```

Met Ser Phe Gly Gly Met Phe Asp Gly Ala Gly Ser Gly Val Phe Ser
1           5           10          15

```

```

Tyr Asp Ala Gly Gly Gly Gly Gly Gly Gly Val His Asn Ser Arg
20           25           30

```

```

Leu Leu Pro Thr Pro Pro Val Pro Lys Pro Gly Gly Gly Phe Ala Ala
35           40           45

```

Pro Gly Leu Ser Leu Gly Leu Met Asp Gly Ser Gln Leu Gly Asp Val
 50 55 60
 Asn Arg Ser Leu Ala Met Met Gly Asn Gly Gly Ser Gly Ser Gly Gly
 65 70 75 80
 Asp Gly Asp Ser Leu Gly Arg Gly Arg Glu Glu Asn Asp Ser Arg
 85 90 95
 Ser Gly Ser Asp Asn Leu Asp Gly Ala Ser Gly Asp Glu Leu Asp Pro
 100 105 110
 Asp Asn Ser Asn Pro Arg Lys Lys Lys Lys Arg Tyr His Arg Xaa Thr
 115 120 125
 Pro Gln Gln Ile Gln Glu Leu Glu Ala Phe Lys Glu Cys Pro His Pro
 130 135 140
 Asp Glu Lys Gln Arg Met Glu Leu Ser Arg Arg Leu Asn Leu Glu Ser
 145 150 155 160
 Arg Gln Arg Arg Arg Ser Ser Gly Met Arg Thr Arg Cys Cys Gly Gln
 165 170 175
 Glu Asn Asp Lys Leu Arg Ala Glu Asn Met Thr Ile Arg Glu Ala Met
 180 185 190
 Arg Asn Pro Met Cys Ala Ser Cys Gly Gly Ala Ala Val Leu Gly Glu
 195 200 205
 Val Ser Leu Glu Glu Gln His Leu Arg Ile Glu Asn Ala Arg Leu Lys
 210 215 220
 Asp Glu Leu Asp Arg Val Cys Ala Leu Ala Gly Lys Phe Leu Gly Arg
 225 230 235 240
 Pro Ile Ser Ser Ile Ser Ser Pro Gly Pro Pro Ser Leu Gln Ala Cys
 245 250 255
 Ser Gly Leu Glu Leu Gly Val Gly Ser Asn Gly Gly Phe Gly Leu Gly
 260 265 270
 Ala Leu Gly Ala Ser Ala Ala Met Gln Ser Ile Pro Asp Leu Met Gly
 275 280 285
 Gly Ser Ser Gly Leu Thr Gly Gly Pro Val Gly Ser Ala Ala Met Arg
 290 295 300
 Leu Pro Ala Gly Ile Gly Gly Leu Asp Gly Ala Met His Ala Ala Ala
 305 310 315 320
 Ala Asp Gly Gly Ala Ile Asp Arg Ala Val Leu Leu Glu Leu Ala Leu
 325 330 335

Ala Ala Met Asp Glu Leu Val Lys Val Ala Gln Met Asp Glu Pro Leu
 340 345 350
 Trp Leu Pro Ser Leu Asp Gly Gly Phe Glu Thr Leu Asn Tyr Asp Glu
 355 360 365
 Tyr His Arg Ala Phe Pro Gly Trp Ser Ala Ser Ala Pro Pro Ala Thr
 370 375 380
 Ser Xaa Glu Ala Thr Arg Glu Ser Gly Ile Ala Ile Ile Ser Ser Val
 385 390 395 400
 Asp Leu Val Asp Ser Leu Met Asp Ala Ala Trp Ser Glu Met Phe Pro
 405 410 415
 Cys Val Val Ala Arg Ala Ser Thr Thr Asp Ile Ile Ser Ser Gly Met
 420 425 430
 Gly Ala Arg Ala Ala Gly Gln Phe Asn Trp Met His Ala Glu Leu Gln
 435 440 445
 Val Leu Ser Pro Leu Val Pro Ile Arg Glu Val Val Phe Leu Arg Phe
 450 455 460
 Cys Lys Gln His Ala Glu Gly Leu Trp Ala Val Val Asp Val Ser Val
 465 470 475 480
 Asp Ala Val Leu Arg Pro Asp Gln Asn Gly Gly Gly Gly Ser Ser Ser
 485 490 495
 Ser Ser Tyr Met Gly Cys Arg Leu Leu Pro Thr Gly Cys Ile Val Gln
 500 505 510
 Asp Met Asn Asn Gly Tyr Ser Lys Xaa Val Thr Trp Val Val His Ala
 515 520 525
 Glu Tyr Asp Glu Thr Ala Ala His Gln Leu Tyr Arg Pro Leu Leu Arg
 530 535 540
 Ser Gly Gln Ala Leu Gly Ala Arg Arg Trp Leu Ala Ser Leu Gln Arg
 545 550 555 560
 Gln Cys Gln Tyr Leu Ala Ile Leu Cys Ser Asn Ser Leu Pro Ala Arg
 565 570 575
 Asp Thr Leu Ala Ala Ile Thr Pro Val Gly Arg Arg Ser Met Leu Lys
 580 585 590
 Leu Ala Gln Arg Met Thr Asp Asn Phe Cys Ala Gly Val Cys Ala Ser
 595 600 605
 Ala Ala Gln Lys Trp Arg Arg Leu Asp Glu Trp Arg Gly Lys Val Gly
 610 615 620

Gly Gly Gly Xaa Gly Gly Gly Gly Asp Gly Glu Asp Lys Val Arg Met
 625 630 635 640
 Met Ala Arg His Ser Val Gly Ala Pro Gly Glu Pro Pro Gly Val Val
 645 650 655
 Leu Ser Ala Thr Thr Ser Val Arg Leu Pro Gly Thr Leu Pro Gln Arg
 660 665 670
 Val Phe Asp Tyr Leu Arg Asp Glu Gln Arg Arg Gly Asp Trp Asp Ile
 675 680 685
 Leu Ala Asn Gly Glu Ala Met Gln Glu Met Asp His Ile Ala Lys Gly
 690 695 700
 Gln His His Gly Asn Ala Val Ser Leu Leu Arg Pro Asn His Ile Leu
 705 710 715 720
 Lys Lys Gln Ile Leu Met Ser Leu Leu Gln Ala Thr Ser Gly Asn Gln
 725 730 735
 Asn Asn Met Leu Ile Leu Gln Glu Thr Cys Thr Asp Ser Ser Gly Ser
 740 745 750
 Leu Val Val Tyr Ala Pro Val Asp Val Gln Ser Met His Val Val Met
 755 760 765
 Asn Gly Gly Asp Ser Ala Tyr Val Ser Leu Leu Pro Ser Gly Phe Ala
 770 775 780
 Ile Leu Pro Asp Gly His Asn Asn Gly Ala Ser Pro Ser Pro Ala Glu
 785 790 795 800
 Gly Gly Gly Gly Gly Ser Asn Asn Thr Gly Ser Leu Val Thr Val Ala
 805 810 815
 Phe Gln Ile Leu Val Asn Asn Leu Pro Thr Ala Lys Leu Thr Val Glu
 820 825 830
 Ser Val Asp Thr Val Ser Asn Leu Leu Ser Cys Thr Ile Gln Lys Ile
 835 840 845
 Lys Ser Ala Leu Gln Ala Ser Ile Ile Ser Pro
 850 855

<210> 55

<211> 1041

<212> DNA

<213> Oryza sativa

<400> 55

atggatgtga gcaggattcg cttcttcagg ctcattgactg gagattttgc tcatggcatt

60

```

agcataccgg agaaagttgc ggacagattc agtgggtcaga tcaccaaagg gttcaacctg      120
aaagcaccca gtggcgaaac atggcgtgtc agcgtcgaaa aggttgctga tgagctgac      180
ctcatgtcag gatgggagga ttttgccaag gctcatgaat tgcaggagaa tgacctcctg      240
ttcttcactt gcaatggccg ttgcaatggc agcttctcct ttgatgtact aatctttgat      300
gcaagtggct gtgagaaagt gtcttgcttc tttattggta aaaagaacag tataggtggt      360
cagactcaga atgctgagca ataccattta agtgattctg aagatactag cacaccatca      420
acattttctag ttggatctcc tcacaaggcc tctacttcaa agaaattgaa tggcaagacc      480
aagactaate caaataaaga tgatccatth tcaacaggga aggagcctga agatccaaac      540
agcagccgct ctcatgtcaa acatgagatg attgaggaag agaagagtga tgatgatgat      600
gagcatgctg attatgaaca tgctgactac tactactcga ggtttgccaa ttacctaact      660
ggcgaggaac gtgaagagat attcagtttg gtgtcactgc aaccaggcaa tcctgtatth      720
gtgactgttc tgcaggcgcc acaaattcat cgcaagggcc tcctgatcgt cccagtgga      780
tttgcagctg atcatcttga cagcagatca caagaaatcc tgctgatgag accaaacaaa      840
aaggagaaat ggtatgtcaa atactaccat gcaagcacta ccaggggctt caattgccaa      900
cgctggatca agttcatccg cgagaacagg ctgcgagaag ggtacatctg catcttcgag      960
ctgatgaaag gcgcaaggcg agtgaccatg acagttcatg tcattgggaa ggttgatgac     1020
cggtttgctt tactgggcta a                                           1041

```

<210> 56

<211> 346

<212> PRT

<213> Oryza sativa

<400> 56

```

Met Asp Val Ser Arg Ile Arg Phe Phe Arg Leu Met Thr Gly Asp Phe
1           5           10           15

```

```

Ala His Gly Ile Ser Ile Pro Glu Lys Val Ala Asp Arg Phe Ser Gly
20           25           30

```

```

Gln Ile Thr Lys Gly Phe Asn Leu Lys Ala Pro Ser Gly Glu Thr Trp
35           40           45

```

```

Arg Val Ser Val Glu Lys Val Ala Asp Glu Leu Ile Leu Met Ser Gly
50           55           60

```

Trp Glu Asp Phe Ala Lys Ala His Glu Leu Gln Glu Asn Asp Leu Leu
 65 70 75 80
 Phe Phe Thr Cys Asn Gly Arg Cys Asn Gly Ser Phe Ser Phe Asp Val
 85 90 95
 Leu Ile Phe Asp Ala Ser Gly Cys Glu Lys Val Ser Cys Phe Phe Ile
 100 105 110
 Gly Lys Lys Asn Ser Ile Gly Gly Gln Thr Gln Asn Ala Glu Gln Tyr
 115 120 125
 His Leu Ser Asp Ser Glu Asp Thr Ser Thr Pro Ser Thr Phe Leu Val
 130 135 140
 Gly Ser Pro His Lys Ala Ser Thr Ser Lys Lys Leu Asn Gly Lys Thr
 145 150 155 160
 Lys Thr Asn Pro Asn Lys Asp Asp Pro Phe Ser Thr Gly Lys Glu Pro
 165 170 175
 Glu Asp Pro Asn Ser Ser Arg Ser His Val Lys His Glu Met Ile Glu
 180 185 190
 Glu Glu Lys Ser Asp Asp Asp Asp Glu His Ala Asp Tyr Glu His Ala
 195 200 205
 Asp Tyr Tyr Tyr Ser Arg Phe Ala Asn Tyr Leu Thr Gly Glu Glu Arg
 210 215 220
 Glu Glu Ile Phe Ser Leu Val Ser Leu Gln Pro Gly Asn Pro Val Phe
 225 230 235 240
 Val Thr Val Leu Gln Ala Pro Gln Ile His Arg Lys Gly Leu Leu Ile
 245 250 255
 Val Pro Ser Gly Phe Ala Ala Asp His Leu Asp Ser Arg Ser Gln Glu
 260 265 270
 Ile Leu Leu Met Arg Pro Asn Lys Lys Glu Lys Trp Tyr Val Lys Tyr
 275 280 285
 Tyr His Ala Ser Thr Thr Arg Gly Phe Asn Cys Gln Arg Trp Ile Lys
 290 295 300
 Phe Ile Arg Glu Asn Arg Leu Arg Glu Gly Tyr Ile Cys Ile Phe Glu
 305 310 315 320
 Leu Met Lys Gly Ala Arg Arg Val Thr Met Thr Val His Val Ile Gly
 325 330 335
 Lys Val Asp Asp Arg Phe Val Leu Leu Gly
 340 345

<210> 57
 <211> 811
 <212> DNA
 <213> Oryza sativa

<400> 57
 atggggaggg gaagagttga gctgaagcgc atcgagaaca agatcaacag gcaggtcacc 60
 ttctccaagc gccgcaacgg cctcctcaag aaggcctacg agctgtccgt tctctgcgac 120
 gccgaggtcg cgctcatcat cttctccagc cgcggaagc tctacgagtt cggcagcgcc 180
 ggggcataac aaagacttta gaaaggtacc aacattgttg ctacaatgct caagattcca 240
 acaatgcact ttctgaaacc cagagagttg gtaccatgaa atgtcaaagt tgaaagcaaa 300
 atttgaagct ttgcagcgca ctcaaagaaa ggcaattgct tggggaggat cttggaccac 360
 tcagcgtcaa agaattgcag cagctggaga aacagcttga atgtgcacta tcacaggcga 420
 gacagagaaa gagacgcaac tgatgatgga acaagtggag gaacttcgca gaaagcagga 480
 gcgtcagctg ggtgaaatta ataggcaact caagcacaag cagctcgagg ttgaaggttc 540
 caccagcaac tacagagcca tgcagcaagc ctctggggct cagggcgccg tggtaggagaa 600
 tggcgccgca tacgtgcagc cgccgccaca ctccgcggcc atggactctg aaccacactt 660
 gcaaattgga ggtatcctca tcaatttgtg cctgctgaag caaacactat tcagaggagc 720
 actgccctg caggtgcaga gaacaacttc atgctgggat gggttctttg agctaagcag 780
 ccatcgatca gctgtcagaa gttggagcta a 811

<210> 58
 <211> 257
 <212> PRT
 <213> Oryza sativa

<220>
 <221> X_region
 <222> (74)..(74)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (86)..(86)
 <223> Xaa = any naturally occurring amino acid

<220>

<221> X_region

<222> (145)..(145)

<223> Xaa = any naturally occurring amino acid

<400> 58

Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
 1 5 10 15

Arg Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
 20 25 30

Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
 35 40 45

Ser Ser Arg Gly Lys Leu Tyr Glu Phe Gly Ser Ala Arg Gly Ile Thr
 50 55 60

Lys Thr Leu Glu Arg Tyr Gln His Cys Xaa Leu Gln Cys Ser Arg Phe
 65 70 75 80

Gln Gln Cys Thr Phe Xaa Asn Pro Glu Ser Trp Tyr His Glu Met Ser
 85 90 95

Lys Leu Lys Ala Lys Phe Glu Ala Leu Gln Arg Thr Gln Glu Arg His
 100 105 110

Leu Leu Gly Glu Asp Leu Gly Pro Leu Ser Val Lys Glu Leu Gln Gln
 115 120 125

Leu Glu Lys Gln Leu Glu Cys Ala Leu Ser Gln Ala Arg Gln Arg Lys
 130 135 140

Xaa Thr Gln Leu Met Met Glu Gln Val Glu Glu Leu Arg Arg Lys Gln
 145 150 155 160

Glu Arg Gln Leu Gly Glu Ile Asn Arg Gln Leu Lys His Lys Gln Leu
 165 170 175

Glu Val Glu Gly Ser Thr Ser Asn Tyr Arg Ala Met Gln Gln Ala Ser
 180 185 190

Trp Ala Gln Gly Ala Val Val Glu Asn Gly Ala Ala Tyr Val Gln Pro
 195 200 205

Pro Pro His Ser Ala Ala Met Asp Ser Glu Pro Thr Leu Gln Ile Gly
 210 215 220

Gly Tyr Pro His Gln Phe Val Pro Ala Glu Ala Asn Thr Ile Gln Arg
 225 230 235 240

Ser Thr Ala Pro Ala Gly Ala Glu Asn Asn Phe Met Leu Gly Trp Val
 245 250 255

Leu

<210> 59

<211> 555

<212> DNA

<213> Oryza sativa

<400> 59

```

atggcacggc ttaccaaaaa cgctgccgc agagtgactc ttaggaggcg tcgtgctgga      60
ttgattaaaa aggtaacaga actatcaatc ttgtgcagtg tccaagctag tatcgtggtg      120
tacaacatcg atgaagcagg tgatccagtg gtatggccat ctattgaaga ggccaaaaac      180
atgtggagca aactcatgga catgtcagag gctacccaga agaagtggat gcaagactcc      240
aagaccctac tccagcaaca aattatgaag ctcaacaaaa agctagacaa tctgaaggct      300
gagaactaca aacgtgagat caccaacata atttctcaga taggaggtgg acaccgcaag      360
aacttaaatg atttatctcc cgagatggtc aaaaatgtca agcgggaggt ggccaagctc      420
cgggaggcca ttagaaaccg catcatcaag ctccacgcac aagggtgctt ttcttcgggtg      480
gttgtggctc cacaagtgga aatagttgca ccacatgctt ctcgatttga tttgaatgaa      540
catgcactag tttaa                                         555

```

<210> 60

<211> 184

<212> PRT

<213> Oryza sativa

<400> 60

```

Met Ala Arg Leu Thr Lys Asn Ala Ala Arg Arg Val Thr Leu Arg Arg
1           5           10           15

Arg Arg Ala Gly Leu Ile Lys Lys Val Thr Glu Leu Ser Ile Leu Cys
          20           25           30

Ser Val Gln Ala Ser Ile Val Val Tyr Asn Ile Asp Glu Ala Gly Asp
          35           40           45

Pro Val Val Trp Pro Ser Ile Glu Glu Ala Lys Asn Met Trp Ser Lys
          50           55           60

Leu Met Asp Met Ser Glu Ala Thr Gln Lys Lys Trp Met Gln Asp Ser
65           70           75           80

```

```
<210> 61
<211> 589
<212> DNA
<213> Oryza sativa
```

```
<210> 62
<211> 59
<212> PRT
<213> Oryza sativa
```

<400> 62

Ala Arg Thr Ser Ala Pro Gly Thr Ser Pro Ser Ser Thr Thr Ser Ala
 1 5 10 15

Arg Arg Ser Pro Pro Ser Thr Ser Thr Ala Ser Ala Arg Pro Ala Pro
 20 25 30

Ala Ala Gly Gly Ser Thr Gln Trp Leu Arg Ala Gln Ala Val His Gly
 35 40 45

Asp Pro Thr Ala Gln Gly Arg Gly Arg Trp Gln
 50 55

<210> 63

<211> 1696

<212> DNA

<213> Oryza sativa

<220>

<221> misc_feature

<222> (722)..(722)

<223> s = guanine or cytosine

<220>

<221> misc_feature

<222> (735)..(735)

<223> r = adenine or guanine

<220>

<221> misc_feature

<222> (756)..(756)

<223> m = adenine or cytosine

<220>

<221> misc_feature

<222> (759)..(759)

<223> y = cytosine or thymine

<220>

<221> misc_feature

<222> (786)..(786)

<223> m = adenine or cytosine

<400> 63

atgggatttcg gcgacgaacc cgagggctcc gacagccagc gccgccgcaa gcgctaccac 60

cgccacacgc ctcgccagat tcagcagctc gaggttcaagg agtgccccca cccggacgag 120

aaccagcggg	cgcagcttag	ccgggagctc	gggttgaggc	cgaggcagat	caagttctgg	180
ttccagaatc	gccggaccca	gatgaagttg	caggcgcagc	acgagcgggc	ggacaactgc	240
ttcctccgcg	ccgagaacga	caagatccgg	tgcgagaaca	tcgccatccg	cgaggccctc	300
aagaacgtca	tctgccccac	ctgcggcggc	cctcccgtcg	gcgaggacta	cttcgacgag	360
cagaagcttc	gcatggagaa	cgcccgctc	aaggaagagt	cgaagtatct	cggccggccg	420
ttcacgcagc	tggggccgtc	gctggacctg	gacctcctca	gcgggtggctc	gtccgacatg	480
gagcggccca	tgatggccga	gatggccacg	cgcgccatgg	acgagctgat	ccgcctcgcg	540
caggctcagg	ggccccgcga	gtccggcctc	gtgctcatga	gcgcgctcgc	cctcgccgac	600
gtgttcatgg	acacgaagtg	gatggagttc	ttcccaagca	tcgtgtccaa	agctcacacc	660
attgatgtgc	tcgtgaatgg	catgggaggg	agaagcgagt	ccttgattct	gcgctactgc	720
csgcagatcg	agcargggct	atgggccatc	gccgamgtyt	ccgtcgacct	gcagcgcgac	780
gcccamtctg	gcgcgtcgcc	gccgcgctcc	cgccggctcc	cttcgggggtg	cctcatcgct	840
gacatggcca	atggctactc	caagggcgcc	gcgttcgggg	cgcaccgctg	gctcgccgcg	900
ctccagcgcg	cgtgcgagcg	ctacgcctcc	ctcgtcgcg	tcggcgctcc	gcaccacatc	960
gccggtgccg	gaggggaaga	ggagcatgat	gaagctgtcg	cagcggatgg	tgaacagctt	1020
ctgctcgagc	ctgggggcgt	cgcagatgca	ccagtggacg	acgctgtcgg	gtccaacga	1080
ggtgagcgtc	cgcgtcacca	tgcaccggag	cacggacccc	ggccagccca	acggcgctcg	1140
cctcagcgcc	gccacctcca	tctggctccc	cgtcccctgc	gaccacgtct	tcgccttcgt	1200
ccgcgacgag	aacacccgct	cccagcagtg	ggacgtcctg	tcgcacggca	atcaagtcca	1260
ggaagtgtcg	cgcattccca	acggctcaaa	cccggggaac	tgcattctcg	tgctaagact	1320
gcagggcttg	aatgcgagcc	agaacagcat	gtgatactg	caggagagct	gcacggacgc	1380
gtcggggctg	ctggtggtgt	actcgccgat	cgacatcccg	gcggcgaacg	tcgtgatgag	1440
cggtgaggac	ccgtcgagca	tcccgtgct	gccgtcgggg	ttcaccatcc	tccccgatgg	1500
ccgcccgggc	tcggcggcgg	gggcgtcgac	cagcagcggg	tccgtcgtca	ccgtcgcggt	1560
ccagatcctc	gtcagcagcc	tgccgtcgtc	gaagctcaac	gccgagtctg	tcgccaccgt	1620
caacgggctg	atcaccacca	ccgtcgagca	gatcaaggcc	gcgtcgaact	gctccgcccc	1680
tggccaccac	cctga					1696

<210> 64
 <211> 565
 <212> PRT
 <213> Oryza sativa

<220>
 <221> X_region
 <222> (241)..(241)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (252)..(252)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (262)..(262)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (370)..(370)
 <223> Xaa = any naturally occurring amino acid

<400> 64

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

Arg Leu Lys Glu Glu Ser Lys Tyr Leu Gly Arg Pro Phe Thr Gln Leu
 130 135 140
 Gly Pro Ser Leu Asp Leu Asp Leu Leu Ser Gly Gly Ser Ser Asp Met
 145 150 155 160
 Glu Arg Pro Met Met Ala Glu Met Ala Thr Arg Ala Met Asp Glu Leu
 165 170 175
 Ile Arg Leu Ala Gln Val Glu Gly Ser Arg Glu Ser Gly Leu Val Leu
 180 185 190
 Met Ser Ala Val Ala Leu Ala Asp Val Phe Met Asp Thr Lys Trp Met
 195 200 205
 Glu Phe Phe Pro Ser Ile Val Ser Lys Ala His Thr Ile Asp Val Leu
 210 215 220
 Val Asn Gly Met Gly Gly Arg Ser Glu Ser Leu Ile Leu Arg Tyr Cys
 225 230 235 240
 Xaa Gln Ile Glu Gln Gly Leu Trp Ala Ile Ala Xaa Val Ser Val Asp
 245 250 255
 Leu Gln Arg Asp Ala Xaa Phe Gly Ala Ser Pro Pro Arg Ser Arg Arg
 260 265 270
 Leu Pro Ser Gly Cys Leu Ile Ala Asp Met Ala Asn Gly Tyr Ser Lys
 275 280 285
 Gly Ala Ala Phe Gly Ala His Arg Trp Leu Ala Ala Leu Gln Arg Ala
 290 295 300
 Cys Glu Arg Tyr Ala Ser Leu Val Ala Leu Gly Val Pro His His Ile
 305 310 315 320
 Ala Gly Pro Glu Gly Lys Arg Ser Met Met Lys Leu Ser Gln Arg Met
 325 330 335
 Val Asn Ser Phe Cys Ser Ser Leu Gly Ala Ser Gln Met His Gln Trp
 340 345 350
 Thr Thr Leu Ser Gly Ser Asn Glu Val Ser Val Arg Val Thr Met His
 355 360 365
 Arg Xaa Ser Thr Asp Pro Gly Gln Pro Asn Gly Val Val Leu Ser Ala
 370 375 380
 Ala Thr Ser Ile Trp Leu Pro Val Pro Cys Asp His Val Phe Ala Phe
 385 390 395 400
 Val Arg Asp Glu Asn Thr Arg Ser Gln Gln Trp Asp Val Leu Ser His
 405 410 415

Gly Asn Gln Val Gln Glu Val Ser Arg Ile Pro Asn Gly Ser Asn Pro
 420 425 430
 Gly Asn Cys Ile Ser Leu Leu Arg Leu Gln Gly Leu Asn Ala Ser Gln
 435 440 445
 Asn Ser Met Leu Ile Leu Gln Glu Ser Cys Thr Asp Ala Ser Gly Ser
 450 455 460
 Leu Val Val Tyr Ser Pro Ile Asp Ile Pro Ala Ala Asn Val Val Met
 465 470 475 480
 Ser Gly Glu Asp Pro Ser Ser Ile Pro Leu Leu Pro Ser Gly Phe Thr
 485 490 495
 Ile Leu Pro Asp Gly Arg Pro Gly Ser Ala Ala Gly Ala Ser Thr Ser
 500 505 510
 Ser Gly Ser Val Val Thr Val Ala Phe Gln Ile Leu Val Ser Ser Leu
 515 520 525
 Pro Ser Ser Lys Leu Asn Ala Glu Ser Val Ala Thr Val Asn Gly Leu
 530 535 540
 Ile Thr Thr Thr Val Glu Gln Ile Lys Ala Ala Leu Asn Cys Ser Ala
 545 550 555 560
 His Gly His His Pro
 565

<210> 65
 <211> 2363
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (681)..(681)
 <223> s = guanine or cytosine

<220>
 <221> misc_feature
 <222> (713)..(713)
 <223> y = cytosine or thymine

<400> 65
 atgggcggcc tccatgtggc caaggacgag gggatggacc tcgagatgga catggagctc 60
 agtggcggct cccgcagcgc ccacctcgac ggcctcctca gcttcgccga cgctcgacgac 120
 gaccacaagc cgcagcacag cggccatgac cagccgcccg acgccgccca gccagcggc 180

gccgcggggg gcaacgccaa gaagaagcgc taccaccgcc acacggcgca ccagatccag 240
 cagatggagg cggtctgtca aggagtcccc gcatccggac gacaagcagc ggctgaagct 300
 gagccaggaa ctgggcctca aacctcgcca ggtcaagtcc tggttccaga atcgccggac 360
 ccagatgaag aggcgcagca ggacagggcg gacaacgtga ttctgcgcgc ggagaacgag 420
 aacctcaaga gcgacaactt ccgcctccag gccgccatcc gcaacgtcgt ctgccccaac 480
 tgccggccaag ccgccgtcct ccgcgacatg tcctacgagg agcagcagct ccgcatcgag 540
 aacgccaggc tcaaagacga gagcttgacc gcttgccctg catcgccacc cgctacggcg 600
 gcggcgccgg acggcagccg gtgctgtcga cgtcgccct gagctgcatc tcggcgccgc 660
 cgccggctct catgccaccg stcgacctcg acatgaacgt ctactcgcg aayttcgccg 720
 agcaggcgcc ggtcatgggc tgccggcgacc tcatcccgca acccgctcgt ccgcagcacg 780
 atggcgccgg gccgtacatg ggccgcatga tggcgcccg gcaggagcag gacaagcagc 840
 tgggtggtcga ccttgccgcc acggcgccgg accagctcgc caggatgtgc cgcgccggcg 900
 agccgctgtg ggtgcggcag cgtggcgccg aggtgatggc cgtggaggag cacgcgcgga 960
 tggttcagctg gccggtggac ggccggaagc agggagacgg cggcgcggtg gcgagggcg 1020
 aggggacgcg tgacaacgcg gtggtcatca tgaacagcat caacttgggt gacgccttcc 1080
 tcgatgcgaa caaatggatg gagctgttcc cgtcgattgt ttgcaaggca agaacaatcc 1140
 agatcataaa ccacggagct gcgtctggtc acctcggcag tggaaactct ctcttggatg 1200
 caagcagagg tgcagttctt gtgcctctg gtggcgccgc gcgaggtggt gttcttccgt 1260
 tactgcgtgc acaacgccga cgaggggagc tggcgcatcg tggatttccc ggcggaaggg 1320
 tttgaggagg ggctcctcca agcctcggtg gtcaggtgcc ggccggcgcc gtccggctgc 1380
 atcatccagg acatgccccaa tggatactcc aggttggtgt ggggtggagca catggagatg 1440
 gtccgggagg agaagccgct gcagccggtg ttcagggact acgtcgccag cggcgccgcc 1500
 ttccggggcca cgcgtggct ctccatctc cagcgccagt gcgagcgct cgcagtgag 1560
 ctgcgccgta acatcgccga cctcgagag tgatccgcac ccctgaggca agaacaata 1620
 tgatgaagct gtgcagcgg atgatacca cttctgtgc taacatcagt gcttctggaa 1680
 cccaatcttg gacggccctc tcggattcaa ccaggacac tatcaggggt acaactcgga 1740
 agaaactga gcccgacag ccagcggtg tcacctcac cgtgtctcc acaagctggc 1800

```

ttccttttac ccatcagcag gtgtttgagc ttcttgacaga tgaacaacag cgctgccagc 1860
agcttgagat tttgtcaaac ggcggtccc ttcatgaggt ggcacatatt gcaaatggat 1920
cacatccaag gaattgcatc tctcttcttc gcattaacgc tgcaagcaac tcgtcacaga 1980
atgtggagct cctgctgcag gagagcagca cccaccctga tgggtgggagc ctggtggtgt 2040
tcgcgacggt ggacgtggac gcaattcagg tgacgatgag cggcgaggat ccttcctaca 2100
tccctctcct cccctgggc ttgccatct tccctgccac cagcccatca cctgctgcta 2160
atgcagatga gctgctgcct cccaatggct gcctctcac cgctggcatg caggtgctgg 2220
ccagcgccgt accttctgcc aagctcaacc tctcgagtgt cactgcgac aacagccatg 2280
tctgcaacgc cattcaccag atcacagctg cactcaaaag ttcagcaggt ggtgctgggg 2340
gtgagccggc gtccgatcag tag 2363

```

```

<210> 66
<211> 788
<212> PRT
<213> Oryza sativa

```

```

<220>
<221> X_region
<222> (227)..(227)
<223> Xaa = any naturally occurring amino acid

```

```

<220>
<221> X_region
<222> (525)..(526)
<223> Xaa = any naturally occurring amino acid

```

```

<220>
<221> X_region
<222> (583)..(583)
<223> Xaa = any naturally occurring amino acid

```

```

<400> 66

```

```

Met Gly Gly Leu His Val Ala Lys Asp Glu Gly Met Asp Leu Glu Met
1           5           10           15

```

```

Asp Met Glu Leu Ser Gly Gly Ser Gly Ser Ala His Leu Asp Gly Leu
20           25           30

```

```

Leu Ser Phe Ala Asp Val Asp Asp Asp His Lys Pro Gln His Ser Gly
35           40           45

```

His Asp Gln Pro Pro Asp Ala Ala Gln Pro Ser Gly Ala Ala Gly Gly
 50 55 60
 Asn Ala Lys Lys Lys Arg Tyr His Arg His Thr Ala His Gln Ile Gln
 65 70 75 80
 Gln Met Glu Ala Leu Phe Lys Glu Cys Pro His Pro Asp Asp Lys Gln
 85 90 95
 Arg Leu Lys Leu Ser Gln Glu Leu Gly Leu Lys Pro Arg Gln Val Lys
 100 105 110
 Phe Trp Phe Gln Asn Arg Arg Thr Gln Met Lys Arg Ala Gln Gln Asp
 115 120 125
 Arg Ala Asp Asn Val Ile Leu Arg Ala Glu Asn Glu Asn Leu Lys Ser
 130 135 140
 Asp Asn Phe Arg Leu Gln Ala Ala Ile Arg Asn Val Val Cys Pro Asn
 145 150 155 160
 Cys Gly Gln Ala Ala Val Leu Ala Asp Met Ser Tyr Glu Glu Gln Gln
 165 170 175
 Leu Arg Ile Glu Asn Ala Arg Leu Lys Asp Glu Leu Asp Arg Leu Ala
 180 185 190
 Cys Ile Ala Thr Arg Tyr Gly Gly Gly Gly Gly Arg Gln Pro Val Leu
 195 200 205
 Ser Thr Ser Ala Leu Ser Cys Ile Ser Ala Pro Pro Pro Val Leu Met
 210 215 220
 Pro Pro Xaa Asp Leu Asp Met Asn Val Tyr Ser Arg Asn Phe Ala Glu
 225 230 235 240
 Gln Ala Pro Val Met Gly Cys Gly Asp Leu Ile Pro Gln Pro Val Val
 245 250 255
 Pro Gln His Asp Gly Ala Ala Ala Tyr Met Gly Ala Met Met Ala Pro
 260 265 270
 Val Gln Glu Gln Asp Lys Gln Leu Val Val Asp Leu Ala Ala Thr Ala
 275 280 285
 Ala Asp Gln Leu Ala Arg Met Cys Arg Ala Gly Glu Pro Leu Trp Val
 290 295 300
 Arg Gln Arg Gly Ala Glu Val Met Ala Val Glu Glu His Ala Arg Met
 305 310 315 320
 Phe Ser Trp Pro Val Asp Gly Ala Lys Gln Gly Asp Gly Gly Ala Val
 325 330 335

Ala Arg Ala Glu Gly Thr Arg Asp Asn Ala Val Val Ile Met Asn Ser
 340 345 350
 Ile Asn Leu Val Asp Ala Phe Leu Asp Ala Asn Lys Trp Met Glu Leu
 355 360 365
 Phe Pro Ser Ile Val Cys Lys Ala Arg Thr Ile Gln Ile Ile Asn His
 370 375 380
 Gly Ala Ala Ser Gly His Leu Gly Ser Gly Thr Leu Leu Leu Met Gln
 385 390 395 400
 Ala Glu Val Gln Phe Leu Ser Pro Leu Val Ala Ala Arg Glu Val Val
 405 410 415
 Phe Phe Arg Tyr Cys Val His Asn Ala Asp Glu Gly Ser Trp Ala Ile
 420 425 430
 Val Asp Phe Pro Ala Glu Gly Phe Glu Glu Gly Leu Leu Gln Ala Ser
 435 440 445
 Val Val Arg Cys Arg Arg Arg Pro Ser Gly Cys Ile Ile Gln Asp Met
 450 455 460
 Pro Asn Gly Tyr Ser Arg Val Val Trp Val Glu His Met Glu Met Val
 465 470 475 480
 Gly Glu Glu Lys Pro Leu Gln Pro Val Phe Arg Asp Tyr Val Ala Ser
 485 490 495
 Gly Ala Ala Phe Gly Ala Thr Arg Trp Leu Ser Ile Leu Gln Arg Gln
 500 505 510
 Cys Glu Arg Leu Ala Ser Glu Leu Ala Arg Asn Ile Xaa Xaa Arg Pro
 515 520 525
 Arg Arg Val Ile Arg Thr Pro Glu Ala Arg Thr Asn Met Met Lys Leu
 530 535 540
 Ser Gln Arg Met Ile Thr Thr Phe Cys Ala Asn Ile Ser Ala Ser Gly
 545 550 555 560
 Thr Gln Ser Trp Thr Ala Leu Ser Asp Ser Thr Gln Asp Thr Ile Arg
 565 570 575
 Val Thr Thr Arg Lys Asn Xaa Thr Glu Pro Gly Gln Pro Ser Gly Val
 580 585 590
 Ile Leu Thr Ala Val Ser Thr Ser Trp Leu Pro Phe Thr His Gln Gln
 595 600 605
 Val Phe Glu Leu Leu Ala Asp Glu Gln Gln Arg Cys Gln Gln Leu Glu
 610 615 620

Ile Leu Ser Asn Gly Gly Ser Leu His Glu Val Ala His Ile Ala Asn
 625 630 635 640
 Gly Ser His Pro Arg Asn Cys Ile Ser Leu Leu Arg Ile Asn Ala Ala
 645 650 655
 Ser Asn Ser Ser Gln Asn Val Glu Leu Leu Gln Glu Ser Ser Thr
 660 665 670
 His Pro Asp Gly Gly Ser Leu Val Val Phe Ala Thr Val Asp Val Asp
 675 680 685
 Ala Ile Gln Val Thr Met Ser Gly Glu Asp Pro Ser Tyr Ile Pro Leu
 690 695 700
 Leu Pro Leu Gly Phe Ala Ile Phe Pro Ala Thr Ser Pro Ser Pro Ala
 705 710 715 720
 Ala Asn Ala Asp Glu Leu Leu Pro Pro Asn Gly Cys Leu Leu Thr Val
 725 730 735
 Gly Met Gln Val Leu Ala Ser Ala Val Pro Ser Ala Lys Leu Asn Leu
 740 745 750
 Ser Ser Val Thr Ala Ile Asn Ser His Val Cys Asn Ala Ile His Gln
 755 760 765
 Ile Thr Ala Ala Leu Lys Ser Ser Ala Gly Gly Ala Gly Gly Glu Pro
 770 775 780
 Ala Ser Asp Gln
 785

<210> 67
 <211> 485
 <212> DNA
 <213> Musa acuminata

<400> 67
 tcgtcgtcgg tgggccacgg tcgtggagac gcgattgggc gagggagggt gcagctgaag 60
 cggatcgaga acaagatcaa ccggcaggtg accttctcca agcgccgata ggggctgctc 120
 aagaaggccc acgagatctc cgtcctctgc gacgccgagg ttgccctcat cgtcttctcc 180
 accaaggcca agctctatga gtactccacc gactccagta tggaaaagat tcttgaacgt 240
 tacgagtatt actcctacgc agaaaaggct cttatgtcat ccgatcaaga ttatcagggc 300
 aactgggtgtc aagaatatgg aaaacttaag gctaagggtg aagctttaag taaaagtcaa 360
 aggcattctaa tgggagaaca acttgaggcc ttgaatctta aagaactcca gcaactggag 420

catcagcttg aaatttctct gaagcatatt aggtcaagaa agaaccaagt catgtttgac 480
tccat 485

<210> 68
<211> 499
<212> DNA
<213> Musa acuminata

<220>
<221> misc_feature
<222> (12)..(12)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (86)..(86)
<223> n = any nucleotide

<400> 68
tggaggagga gntggagatg gggaggggaa gggaggagct gaagcggatc gagaacaaga 60
tcaaccggca ggtgaccttt gcgaangcag gaatggactg ctcaagaagg cctacgagct 120
gtcggtgctc tgcgacgccg aggttgccct catcgtcttc tccaaccgcg ggaagctcta 180
cgagttctgc agcagctcca gtatgttgag gacactagag aggtaccaa agtgcaatta 240
tggtgcacca gaaacaaata tcatatcaag ggagactcag agtagtcagc aagagtatct 300
gaagcttaag gcgcgtgttg atggcttgca gcgatcaca agaaatctgc tgggtgaaga 360
tctggggcca cttaacatca aggaacttga gcagcttgag cgccaacttg atgcatcgtt 420
aagacaaata agatcaacaa ggacacaata catgcttgac cagcttggag atctccaaag 480
acgggaacaa atgctatgt 499

<210> 69
<211> 414
<212> DNA
<213> Musa acuminata

<220>
<221> misc_feature
<222> (402)..(402)
<223> n = any nucleotide

<220>

<221> misc_feature
 <222> (411)..(411)
 <223> n = any nucleotide

<400> 69
 cgaaagtgag tcatcatcat acaatctatc caatgtggca tcatagtttc ttacacaggt 60
 cctcaaaacg atgggacgac accaaacctt gaattagcta acacagcagc agtggagaac 120
 atctgatcca cttcgagctc tacatccgac ttcttcgcaa accggccttt gattcgtggc 180
 ctcgtttcag catacggcct ttctcgatgc ataccttatg gtctttctcaa acctcctcgt 240
 cttcctcttc tccctgtatc taagcaactc ggcttctctg tccatgagat tgagctgtgg 300
 ttgaacttga agaggatggg ggcccgagaa gagatccatg gttcctttgg gcgctccgat 360
 ggaagtgtct ggtacgacac ttgcatccat ggatgacaaa gnaactctgt ngct 414

<210> 70
 <211> 584
 <212> DNA
 <213> Musa acuminata

<220>
 <221> misc_feature
 <222> (579)..(579)
 <223> n = any nucleotide

<400> 70
 cctgcttctt ctaggtaacg aggaattttc aagtccagct tcgaacactg gaggctcacc 60
 cataagaagc actgatgaaa caccacaatc ttgatagtcc ccggcactgc tttcaccagt 120
 caagccagaa aaggaaaggg ataggctaaa gcaagcttga cgtgctggaa aaaactgact 180
 tgaatctgca ttttttactg gttttgacat cacagaatca gcagagactg cattgctaca 240
 tgttgcctgc ataggcttaa cctgccccgc agatgatgcc tctccaacaa gttcaccctg 300
 gcagttggaa ttgccagcag atgttgccct tgtatcaaaa aaaatatcta tttcatcatc 360
 atcaaaaagg tgccctgttt ggttatgaga tacaccaaaa agttcttcat agttttgaaa 420
 tgtcaaatca acatcatcca tattgagatc ttcgtaaata tcatctttgc agaattcaaa 480
 atctttgggc ccaggacaag atagcttagg tgtagttaaa tccactgcac cagctgtcag 540
 atcagcagta cagggcatag gacacattgg cagaccctna agat 584

<210> 71
 <211> 531

<212> DNA
 <213> Musa acuminata

<220>
 <221> misc_feature
 <222> (386)..(386)
 <223> n = any nucleotide

<400> 71
 ttgtccttat catattcaaa agtggttaagt gtttttaagt aactgacata cttctaaagc 60
 atgcatcaga tgcataattg catctccgtc aactgcaata caccaatggc agacaaccac 120
 acaccatgac aaacagtaaa attgcaccta ctgacttgct caacatggct tataaaatgc 180
 actgctgcac caacagatac tagagagtca ttttatttga agctaaagtc aataaccatt 240
 gatggtcttc aagtattctt cttctcgatg tagagaaaaa gcagaatgga aacaatgcc 300
 agcaaacc aa taagaagagg caaccgttcc aaaaccaaac tgagaggaga gatgaccata 360
 ttgttctctg gttaccactc aggaancctg gaaaaaaaca gacaaaagat gacagtgaaa 420
 agttaaaaag gatcagagaa aaccttggtg ctgccagtt ttcctatcct ttatgaaagc 480
 cacttcaatg acatctccat gactcaaca atgggcgata tcttcttcag t 531

<210> 72
 <211> 744
 <212> DNA
 <213> Musa acuminata

<220>
 <221> misc_feature
 <222> (55)..(55)
 <223> n = any nucleotide

<400> 72
 agattaaaat tccccttaga tcaaacagca aaaatttcag aactgcac gaagnaccaa 60
 atcaaacctt ggcgatcaaa ctggatacta gaaatatatg tcaatcagtc gatagcatat 120
 aaacacataa gcttcaactg ttgaacatgg tgatacactt ccaacataat aaacagacta 180
 atcttctggc tacaaattta gttaacatat attaggtaca ctaaggacta aaagctactt 240
 gttagaagat atcatgattc ctgcagcttt ctccattaat aaagataggc ggtctggtac 300
 tggacctggt ggacattacc atagaacgtg gaatgtgtgc gggggatggt atatggcgca 360
 gaaccattg gcgtaggatg atgagtttca gcagcataga cgtactgcct atcatatgcg 420
 tatgggtacc tctcaaaggg tgatcgatgg aagcttttgt caggcatgat gttaatacga 480

gggtaactgc tgccattggc acgggggtctc ttggtttgcg gctttgcagc ttcagcagcc 540
ctcctcttgt cggcctttgc cttctctagc agaaggatgc gcttctgaag cggccaacag 600
ggatttgctc ctcaagtttg tgctcttcaa tgcatttgat gaccgcttta agggcagata 660
gttccgttca ttcatttcat tctgcgcacc aggaaaaagg gcggccgctc taaggatcaa 720
gcttacgtac cgtcatggac gtca 744

<210> 73
<211> 898
<212> DNA
<213> Musa acuminata

<220>
<221> misc_feature
<222> (570)..(570)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (674)..(674)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (709)..(709)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (839)..(839)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (847)..(847)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (871)..(871)
<223> n = any nucleotide

<220>

<221> misc_feature
 <222> (879)..(879)
 <223> n = any nucleotide

<400> 73
 aacatgtctc ggtctaaagg attaacacat aagtttccac caatcaaggt tgagaaacac 60
 ttaatacaat tacaggacta atcttccggt tatggagatt cagttaatta acatactggt 120
 taaaagctaa gtcaaaagag aacatgatcg attcagcctt ctccattagt gtaaataggc 180
 agtctggtag tggacctggt agctgtttcc atagtaattg gtctgagtag gtgagatgga 240
 gtatggagca gaacccatcc gcagcgggta atgattttcg gcagcatagg catactgcat 300
 gtcatatgga tatggagatc tctcagggag tgcacggtaa aagcttttgt caggcatgct 360
 ggtaatacga gcagtgtaaa tgcagccagt ggcacggggc cgcttggatt gcggctttgc 420
 aacttcagca accctcttct tgtcagcctt ggctttctca agcagaacga tacgcttctg 480
 aagcggattc aaccggatat tgctcctcaa gtttgtgctc ctcgatgcac tttataacaa 540
 ttttaagggg cagctagttc gcgttcattn tattcattct gtgcaccagg agacatgctt 600
 ccagctttga cctgtgcaaa cctttcttgc gtcctccaga tatgctttaa gcaatggcac 660
 tgggtggaaat tgtncggtaa gtttgaatgc acaagcaaga ttaacagcnt caatttgcct 720
 cccgctgttc agaagatatt caatcaaacc tggcattcta tttgacaaac caagttagca 780
 acacaagttc aacagtttga cggcgctgac aacagctgga ataagtttgc agatttcant 840
 tgatcanatt ctgaaacat gtcaaagtag ncaacagtng aagatatgca tgtgttcc 898

<210> 74
 <211> 476
 <212> DNA
 <213> Musa acuminata

<220>
 <221> misc_feature
 <222> (273)..(273)
 <223> n = any nucleotide

<400> 74
 taatcactag gatttcgata aacattattg actcttatat acatattcac attggctttt 60
 cttgatagat agattgttcc ctattgtaag aaaagagagg tgagaagaaa gaggaaggct 120
 cccaacctca ccagagcctg tattgtatac tgtgtacatg tatgcatatc caagctacat 180
 cctcctccca ccggagcctg actctctctg gcagttgaag tagacggcgg cgaccggcga 240

tccgaggtta tagagctccg cgaagtcctt gnggttgaag ttctgccgcc acccgggagc 300
atccacagct tgccggctgg ggagctgcct gaagagcacg aagacgtagc ggtggatgcc 360
gagctgaggg cgccgggtct cgtagcacac tacctcctga ccgtatgttg cttcgggttc 420
gcgggggatgt cggtgatcaa ccagtgcata tactctctga gggttggatc acttgg 476

<210> 75
<211> 413
<212> DNA
<213> Musa acuminata

<220>
<221> misc_feature
<222> (11)..(11)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (23)..(23)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (38)..(38)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (85)..(85)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (88)..(88)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (312)..(312)
<223> n = any nucleotide

<220>
<221> misc_feature
<222> (326)..(327)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (387)..(387)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (392)..(392)

<223> n = any nucleotide

<400> 75

atatatattc nacatcatgt ttnttgcaaa tggattgnac acacatgaca tatgcatctc	60
ctcctgcaga actatggagc actgnaangg gtaaacacaa acccatttac ttcaaacatc	120
atgtgttctt ctggccatga cttgcatcaa gagtttctgg gatgtctgtc accaaccagt	180
gcagatactc tctatcagtt ggattgtttg gacttgggtgc atcaggatcc accatcacct	240
tccctgaatg gcgcccggcc aggtagagac acaggccgac gttgctcatc tgcccgcg	300
aatcggaggt gncaagtcac ggccannccg ggctggcgag cactatgtgg gacgaagatt	360
cacgtcatc aacatcacgg caaagangtg tngaggagag cgatcagtat gca	413

<210> 76

<211> 564

<212> DNA

<213> Musa acuminata

<220>

<221> misc_feature

<222> (489)..(489)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (563)..(563)

<223> n = any nucleotide

<400> 76

agattatact gcagcacaga gactaagagc ggtacatctg atcacggaga ggagggagga	60
gggagctgca tggatagtgg aggtacagta gattgcacca gagacagtag acaacagaga	120
cggagcgggtg gacgcatgca cgatggctct ctatcgccgc ctggcagcgg tctcccgtg	180

```

agaattgaag tagacggcgg cgacgggggag gtcgagggcc gttctcctcg gcgaaacggc      240
gcgtgctgaa gcggtccctc gaggcggcg ggctgacggc ttgccgtcgc ttctgcttga      300
agagcacgaa gacgaacctg tggatgccaa tggtcggccg ggggtctctg tagctcacca      360
cctcccttcc aaaagaagca tctgttgtgc ctggaatgtc ggtcacaatc caatgcaggt      420
gctccctgag atatggatcg cttggcccct gggacgtctg gatctgtcat aaccagtgtg      480
aagaagganc ctcatgtcac caacctgcac ctcgactctg ggccttgatg tgactgtaga      540
agggtagaac tcatggccat tгна                                     564

```

```

<210> 77
<211> 555
<212> DNA
<213> Musa acuminata

```

```

<220>
<221> misc_feature
<222> (13)..(13)
<223> n = any nucleotide

```

```

<400> 77
ttcaggttcc aangcaaaca agattctgaa agcaaggaa accaaagcta aaccaaagct      60
caaaacgaac acttcaacga acacttcaag gccagcaaca caaaaggatc gataggagag      120
agagacacag acagagcgag agaactgtgg ttgtagtcat gaaatggggg gaagtcaata      180
cccgaaccag ccctccttgc atggcggcgg cgggtggccga gagccaaggt tagcgtgagg      240
tgcgttcttg acttccaca gatgatgtcc taatctccac cctcccccaa cttcctctcc      300
tcctatccat gccgcgtgct tgctattctc agcgagccac ttgcctgca actgaagcct      360
tgatccgttc gacggtgcaa gcaatgaggc tgttcaccgt cgcgaccgat ccgagggaga      420
gcttggcagt ggggacggag tcgaccagga tctggaacgc cacagtcaac agggagccac      480
cggagcctac ttctccacc ttctccgcc cgatggccgt cgggaagaat ggcgaacctg      540
acggaagagg gcaac                                             555

```

```

<210> 78
<211> 467
<212> DNA
<213> Musa acuminata

```

```

<400> 78
gtgttctggt tgtgctggat gggcatcacc aggttgggag gggcctgccc ttgcaagctc      60

```

gaagccattg tggccatcag gggttggcta agcttgttgt tctcatgcgc caacgcgttg 120
 cagaaggctc ggtgcgtcat gaagctgtct ctctggaga agatgggtgcc acagtgcgat 180
 ttgtactgcc tcgtcccgca gatcttggag tgagccttcc aatctgactg cacggcgat 240
 ttcttcgagc acttgtcgca cttccacttc ttctcgccat gcttgcgga gtagtgcttc 300
 ttgattccgg taagatcgcc gagggcgcg ctggggttgt ggtgcacgca tgtgggctcc 360
 ggacacacgt agaccttctt cctggcctcg ctgccgcgt ttcgtggcct cagcttcac 420
 ggaaggttgt ggcctcgcg gtgcagttgc aggttttggg ccctctg 467

<210> 79
 <211> 759
 <212> DNA
 <213> Musa acuminata

<220>
 <221> misc_feature
 <222> (9)..(9)
 <223> n = any nucleotide

<220>
 <221> misc_feature
 <222> (11)..(11)
 <223> n = any nucleotide

<400> 79
 acaattttnc ntcatcaga tttaatactg ccaacaatat ccttgtaag ttgggtacaa 60
 ataccaatat tgatatcata gtgttgccat atggattggg cgaaactcca gcagaccctg 120
 atatcctgtc ctagactact ggtctcactg gcgtaacagc ataaatgaaa ttacagtggc 180
 acccaaactt ccaagtcaga ttagcttata attattgata gcacatttag atcttcttaa 240
 cagtgaaca agcatgattc ttttcaagga agagaagtgc aaaggacaag tattggaaga 300
 gaatctccct tacgatgggc caggtggccc ttgaggggat tgctgcattg caggtgggtg 360
 tgggtggagggt ggtcgggaca tcatggactg gtaaggggga tactgctgct gatacgggtg 420
 tgggggtggg gccatatacc cgtatgggtg ataataatgt ggatactggc cctgaagatg 480
 tggaggaacc ataggataag catgcccctg aacatttggg gcttgctgcc ctttctgctt 540
 ttctgatcca gttttgcat cgctacctcc ctgcgattgg attacagcac ccctcctctg 600
 aggatccatg gatgggtaga atgtccgtgc agcttgtggc ggtgcaggga tgttgaaata 660

gtaggttgt tgctgctgat gagtctcagt gcctggtggt agaggttgct ggttgctctg 720
 ctgctgagat attacagccc ttgggagcat ccctccatg 759

<210> 80
 <211> 1064
 <212> DNA
 <213> Musa acuminata

<220>
 <221> misc_feature
 <222> (12)..(12)
 <223> n = any nucleotide

<220>
 <221> misc_feature
 <222> (30)..(30)
 <223> n = any nucleotide

<400> 80
 attgttcac tncctcttat tagctaagan gttacttttt gatgccaaact taatctcttt 60
 caacttctcc ttttcctttt tctctgcagc taatgctgct gttgcagcag catgaaccgc 120
 atcaggcaat gatgcttttt catttgattt taccgtgggt gcagctgcag atatcactac 180
 ttttttgctt gcagttccat cggatccttt tgaagtttca tttgccatgt ctccagctgc 240
 cttattgttg ctctgctcgt tgcaagggac atattgctga gtttgatgat cataagaata 300
 ccaaacacca ttgttcccat catagtaaag accagtattt ccatcataat aaaatcctga 360
 ggcagcatca taataatagc cagatttctc atcccataca aaccagatt gtggagcaga 420
 accacctctc tgaacttctg ccttgctatt ttgctcaaac ccagttgtag attgtttttg 480
 atctggattg gattcctttg gcgtccaacc aacagcatca tactgttgag caaatgttgc 540
 tgctcaata gcagctgctg caagactggc tcgattgtgg agtcccagat ccaggcccat 600
 gaatactttt tgcataatgct acacgtagga ctggccattt ttttctaggg ttgttccatt 660
 gggtgcttca agagcttttg ttgcatcttc caccgagtgg aaatgcacaa aagcaaatcc 720
 ccgagaaaca tgagtaaatt tatctctgac aaggcgaaga tccttaattg cagcatgctt 780
 agcaaattca taacgaagca tttcttcac tgcattttca tccaaccctc gtacaactaa 840
 gacatgagta ggacctaaat ctgatcccct tttccctaag ggtggtggat tgggtgcttgc 900
 tacatcagct ggaggagcac ctcagtcagg ggctcattgc actggaaaca ggatgtccgc 960

cgtgcaaaat ttaaacaacc acacatggta caaatccagt cacatgggtgc agtaatactt 1020
 ctaccgtgtc ccatggctat cttatgtatg cttgtcttga tgta 1064

<210> 81
 <211> 763
 <212> DNA
 <213> Triticum aestivum

<400> 81
 cttgtggctc aacctctggg gtacctctca tagaagtcaa tgctcagcaa tactaccagc 60
 aagaagctgc aaagctgcgc caccagatcc agatgtgtgca aagcactaac aagcacttgg 120
 ttgggtgattc tgttgggaac ctgtcactca aggagctgaa gcaactcgaa agccggcttg 180
 agaaaggcat tgcaagatc agggccagga agaatgagct cctatcttct gagatcaact 240
 acatgggtcaa aaggagata gagcttcaga gtgacagcat tgatctcaga accaagattg 300
 cggaggagga gcagcgggtg cagcaggtga ctatcgccag gccctcgggtg gcgccggagc 360
 tgaacctttt caccgactg gacatgaagt gcttcttccc tgccaacctc ttcgaggcgg 420
 cggtgcatgc gcacgcgag gcacaggctc aggctcaggc tcaggcctcg cttcagctga 480
 acctaggcta ccagcagctt gctccaccgg gcgccggcga tgtgggtcat cagttctagt 540
 caacctgcag gagctcttgc ttgactacac aacttgcaat aaaactggaa agtcgtcggc 600
 agtgttatca gtactcaact gttccgggtca aaccgtctgg ctaagatata tgtgtgctaa 660
 gttttattgt ctcaagtacc tctaaatttg ccgctggtgt agaaccaaat ttctgagata 720
 aaacctattg attaataatt atttttttcc ttttcttttt tga 763

<210> 82
 <211> 471
 <212> DNA
 <213> Triticum aestivum

<220>
 <221> misc_feature
 <222> (110)..(132)
 <223> n = any nucleotide

<400> 82
 tgtgccccctc cccctgtccc ctccagcgcg cgcgagagag ctgagctgag cttctcgggtt 60
 agctgagctg agctgctttg gtggtcggga cggtcgaacg gcgcgtagcn nnnnnnnnnn 120

nnnnnnnnnn nntcgatctg gagctagcgg tcccggacgc ggcgggccggc ggcgacggca 180
 tggggcggtgg ccgggtcgag ctcaagcggg tcgagaacaa gatcaaccga caggtcacct 240
 tcgccaagcg ccgcaacggc ctgctcaaga aggcctacga gctctccgtc ctctgcgatg 300
 ccgagggtgc gctcatcatc ttctccaacc gcggcaagct ctacgagttc tgcagcggcc 360
 agagcatgcc caaaacactt gagagatacc agaaatgcag ttatgggtggg ccagatacag 420
 cagtacaaaa taaggagaat gagttagtgc agagcagtcg aaatgagtac c 471

<210> 83
 <211> 734
 <212> DNA
 <213> Triticum aestivum

<400> 83
 acgcgtcgag caaggagtac cgcgccacg gtgtcgagag ccacgagggc agccacgagc 60
 tcgtccgtgc ccggcgagcc ggtggcgag ctgcatgagc ggcagggctt cacggcggag 120
 atgtcgtacg atgcacaaaa ttctaaccat gggtagcgat ttggcgcaac attccaacgc 180
 agcctttcaa tgtcatcatc accggacaac agtagcaccg ttcaggatgt gaccagctcg 240
 tacatgagga gcagcgagag cagcgtcgac ctcttctcag cggcgggcgca catgtcgccg 300
 cagttcatgg gcatggccat ggacagggag gccaggggtgc accggtacag ggagaagagg 360
 aagatgagggc ggttcgagaa gaccatcagg tacgcgtcgc ggaaggccta cgccgagacc 420
 aggccacgga tcaaggggcy gttcgccaag cgcgccgacg ccgacctga ggtggaccag 480
 tactttctcc cagccgcact gtccgactcc agctgcggcg ttgtgcccac tttctaggcc 540
 ggctcgggtgg agagaccatg agcgttcctc cttccatggt cgatcgacgc ctacatgttg 600
 gttggctctc cttaatctgc caaatattca gaaatttatt ggaggcgtca aagcttcacc 660
 tgtaaataata tggaaagatg ccgatatagat ctttcttctc ttctacctat caaaggctaa 720
 tgaactgacg tcag 734

<210> 84
 <211> 1416
 <212> DNA
 <213> Triticum aestivum

<400> 84
 ggggtgcccc tcatctgcag agctttcgag aatgtgggtca ttcatttttg atattcctaa 60

tgtggctcct gagctgaatt gtgagcaagt aataagcatg atgagcatca gtgacagtgc 120
 tgtcagtaat ggagacaatg ctcaagggga caacagtttg ttagatatgg ctgctgcaac 180
 acttgatgaa gaagacaaac aaaagtctgt gataggatca tcttctgaag ctgctctgaa 240
 ccttcttcca cctgctaata atcagacagc tgtatctgtg gattcgacga cagctaagta 300
 tacgccagac aaacatatgt tcagcaagga cagcatatat gaagatttct tctatggatg 360
 atattgatct gagttacgaa aattatgaag aactatttgg taactctcat attcagagtg 420
 aggaactatt tgatgatgct ggcattgaca gttactttga acctaaagaa gcgactgctg 480
 tgaattctga tgagcagccc aagccaaaag agccggctgc tagcaatgag ttatctgctg 540
 actctgggat gtcaaatacca gcagtaaaag aagatgattc cagtctttgt attcctgtga 600
 ggcaagctat ttcctattct ggctttactg gtgagagcat tcctgaagaa taccaagact 660
 gtggtgtatc accaatgctc cttatgggag agcctccttg gcttcctcct ggccctgatg 720
 ggtcatttgc tggaattaga gacagtgcga tcacacggta caaggagaag aagaaaagaa 780
 gaaagtttga ccacaagatc aggtatgaat ctgcaaggc tagggcgatg gtgagaaaga 840
 gagtcaaggg acggttttgt aaggccggtg aagcgatga ctatgatccg ctcgacacaa 900
 ggagctactg aatgcagaaa agcattgtgc ttagattagc aatcttaaac cagtgcctat 960
 tgcagctggg gtaaatacat aaatcagacg ctgaagcaat catggcatac ccacctctgg 1020
 aaagaaaaga atggaaagca accagcgagg caaaaaagat ttgaaagttg tgcaagtttt 1080
 tgacgctgtt tctcccgttc cttctgtaca gcgatggat ctgggtctt ctattgcttt 1140
 cagagcagtt ttggtgaacg aagacaacgc aactagatgg actttctgag tatttctccc 1200
 ttctactgac tctggcatcc acctaccctt gttgtggtgt ttacggcggt atgcttccag 1260
 attgtgttct cttaggcgaa agctgtagtt gtgaagtatc ctgctgcgtt atatttttat 1320
 gtgaattctg ctgcaatccg gagttttgtg caattatgca tgtgtaaata ccaggtcgga 1380
 aataagccgg tgggtggtgt aaagtgtgtt gccgcc 1416

<210> 85

<211> 835

<212> DNA

<213> Triticum aestivum

<400> 85

```

cataatgagg ggtccgagca accaattaaa tagttcgatt tgctgatcc gaagaggcta      60
agacctggag aatcaagggg gtgcccctgc atttcggagg tcctggtgtc agttctcgat    120
ctgatgcagc actagttatc agccgactgc caatcttgac gagcaaatac cgcacacatg    180
cctcctgaca gttggcgccc ttcaagccca agctcaatgg ctctcatca gtttaataac    240
tttgggtctg acaattctat gggcctgatg ggtggccctg ttacatcagc agcagataat    300
gttacttttc ggcctcagat gtttcatggg aatggttctt tgtcaagtca gacagctgtg    360
ccgacatcat ctcatatggg cataaatcct tccttgtcac aagggcatca tctaggtggg    420
ccacagatct cgcccttgca aaagccaact ggcccaccgc agaatttccc tgtacagttg    480
cagaatgctc agcaagggca gcttcatgcc tcacaaccct tggggcctgg ttcttttggc    540
cagaatatac caactatgca attacctggc cagctccctg tgtcacaacc gttgacgcag    600
caaaatgctt ctgcaggcgc tctacaggcg ccttcagctg tacagtcaa tcccatgcaa    660
gctgttcctg gacaacagca acttccttcc aatgtaacac cgcaaagct gcagcagcct    720
gtccagcaga tgctgtcaca agcaccacag ttgctacttc aacagcagca ggcagctata    780
cagtccagtt atcaatcttc acagcagacg atttttcagc ttcagcaaca gctgc      835

```

```

<210> 86
<211> 928
<212> DNA
<213> Triticum aestivum

```

```

<220>
<221> misc_feature
<222> (92)..(113)
<223> n = any nucleotide

```

```

<220>
<221> misc_feature
<222> (856)..(856)
<223> n = any nucleotide

```

```

<400> 86
gcaccattta ctttctccaa tcgaccggcc agccggcagg atagattaga catctagaga      60
gcaaagaagc aacaattagt gtcggtagct gnnnnnnnnn nnnnnnnnnn nnnngggaggg    120
ggaagattgc gaatgagagg atcgacaacg cgacgaaccg tcaggtgacc ttctccaagc    180
gacgcgggagg gctgatgaag aaggcccggg agctcgccat cctctgtgac gccgatcttg    240

```

cactcatcgt cttctccagc accggtcgcc tctacgactt cgcgagctcc agaggcatgg 300
 aggcaatact agagcgctac caggaggcca aacaggagca tggcggagtg ttgaatccaa 360
 catcagaggc aaagttatgg cagagggagg ttacaacctt gaggcaacaa gtgcacaact 420
 tacaacacaa taataggcaa ttattgggag aggagctatc tggttccact gttcgagatt 480
 tgcagtttct cgtgaaccaa ttggagacga gtttgcattc tgtacgaaag aggaaggagc 540
 aagttatggc tgaggagatc cacgaactta accaaaaggg atttcttatt caaaaggaaa 600
 atatagaact tggcaagaag ctttctgggg ccatgagatc gagtgagcag caagccagcg 660
 gaagcagctc aaaagctgct gccggattat tattgtggac tagagcaggt gaaccaagga 720
 tttttattga ccttgagttg cggcagcagg agcatgagga tgaataacac ctactacctc 780
 gtccactaca tgctactgca gcttcaatct atattttcag tgcattgcgtg gacgatggaa 840
 gctcgatcat caaatnagtt atatggttcc tgtgtaattt gcataaggta ttaacaaaca 900
 tgtatgcatg gagtgcagaa ctacaaat 928

<210> 87

<211> 947

<212> DNA

<213> Triticum aestivum

<400> 87

ggaaactgct tggaagcaca tgcatttctt caactgcttg ctacctttgg tattttttct 60
 gaatttaatg aagatgaact ttgcaaactt cttccttctg tcagtcgacg tcgtcaaaca 120
 cctgagcttt gccgactgct tgggttatca cagaagatgc cagggtgcat tggagttctg 180
 gtggatagtg caagaccaat tgatgcaatt aacttggcct acgcgtttgg gctcactgaa 240
 cagtttgagc cagtacaatt gctgaaagca tatctgaggg aggtcaagaa ggtgtcacat 300
 gccagaatg gaaaaatgtc tcctggagca cagaatgaga tgaatgagcg tgagctgtct 360
 gcaactgaaag ctgtcatcaa gtgcatcgag gagcacaac tggaggagca atatcccgtg 420
 gatccacttc agaaaagggg gattcagctg gagaaggcca aggcagacaa gaggagggct 480
 gtogaagccg ccaagccaca gtogaagagg ccacgtgcc aacggatcgct ctatgcaccc 540
 cgtgtcagca gcttcctga gaagagcttc taccaggcaa caccagagag gcaccatac 600
 ccttacgaaa ggcagtttgt gtacggcgct gaggcccacc accaccctac gatgatgaac 660

gcggctccct acaccatctc acccgccac acgccgtact acggtaatgg ctacccggtg 720
 cagtaccaga taccttatat ccactaatgg gggagagcta caatggctct ttcctaacta 780
 tccctagctt ttaaccttcc gctgttgttg tagttgtacc aagtcattcta gctagtttgc 840
 ctagctgaaa atagtggac cactcgcttg tgtgtgtgct atgagtggct attgtgagaa 900
 tctgttaatt ttatgtcacg acttatgaca taatttactt ctaatgt 947

<210> 88
 <211> 1077
 <212> DNA
 <213> Triticum aestivum

<400> 88
 atagaagaag gttatgcgta ggatagcacc cctgtgttgt caggtttggg acgtgtggac 60
 gcttcgtccg accaccacct caggtgacct tcgggacaga ccgtgtccaa cgctgcgagc 120
 tcaaaccgtc catgtcgccc agcagcccag ggttgagggtg gcggcaatga gatgaggacc 180
 ttcttacaca ctctgatgg tagaccaga tgctccaagt ccaagcgatc ccaaccttag 240
 ggagtatctc cactggcttg tgacagatat ccccggtaca actggtgcgt cgttcgggca 300
 ggagggtgatg tgctacgaga gccctcgtec gaccatgggg atccaccgct tcgtgctcgt 360
 actcttccag cagctcgggc gccagacggg gtacgcccc ggggtggcgcc agaacttcaa 420
 caccagggac ttcgccgagc tctacaacct cgccccgct gtcgccgccc tctacttcaa 480
 ctgccagcgt gaggccggct ccggcggcag gaggatgtac aattgatcta cccacggccc 540
 tcgtacgcca ccagcccgcc gccaaagtcag caaattatcc aacgtggcta gtttactagt 600
 atatagtttg tgataagaag ccagccacga attaagcatt acctatata tggcaacaca 660
 tacactacat atatgcatac tatgatcgat gtataactag ccgcatgcat atatgcaatc 720
 aacggctaata taaggggggg tgaaccctag atcaatggct tggtagtgca ctatataaat 780
 atagtctgca ataaactgat gccaatagta tacagcacac aaatattgga ggagctacac 840
 gccatgtgca acttagtgct acctggtaca tatctgcagg ttggtcttgt gcgttcactt 900
 atgcgtgcat gaacatcagt caatcatata gacatagtta tgcatgggag acaaacatgt 960
 aactgacagc gactgctaca agacacagt atagtacgc acttagcgca gtagcaaagc 1020
 acatgcatgt gttgggtctt gtacaacttc ctcatgatta tggcaatata atcaagt 1077

<210> 89
 <211> 426
 <212> DNA
 <213> Triticum aestivum

<400> 89
 ggaagagcag agcagcaggc atgtcccgt gctgggatcc gctggtggtg gggcgggtga 60
 tcggcgaggt ggtggacatg ttcgtgccgt cgtctacttc aactcccaga aggagccgtc 120
 cgggcaccgc cggcgctgaa gaaagcgcgc gccggccggc ctagctactc cggcaccgcc 180
 cgctgcatgc gcaccgagct acaataactca atcccggatga agaataagta actaaacttc 240
 cataaactag cctgaactgc tacatgtatc cgctgtctac cgttgtgtaa tcgcactgtg 300
 cttgcgtatg catgcatgag atgcatggac catatatgtc ttgggggtttt gcagagcaac 360
 tctgttcttg taataatagt actactagca tgcagctggc gcttaatttc atgttgtttg 420
 gtcttc 426

<210> 90
 <211> 623
 <212> DNA
 <213> Triticum aestivum

<400> 90
 gcgttctctc tgaccgtggc attccagatt cttgtgacag ccagcccccc gccaaagtca 60
 ccgtggaatc ggtcgagacc gtgaacaatc ttatctcctg cactatcaag aagatcaaga 120
 cggcgctgca atgtgcaacg cctgagtgtt aggggtggatt ttgccttggg gatgaccggc 180
 gacgtcgagc ctgagtcaag aacgaaccgc gcaatgaaaa aaaaaatcct ttgctaggtg 240
 ataaggatct tacttgaaag gtccttagca cagagcaatg gtggtggttc gggatttgac 300
 tcggcctctt ccgtcggtg tattccccgg tgcccatgca ttttagacat ttcttgaagg 360
 acatgctata gttccatcgt tttagggatga actttcggtt ggctcggttc ctctgttttg 420
 tccgtctctc tctttctgtt gtaaggatga gctggtgtgc aagtggagac tcaggaatgc 480
 tgggttgatc tactagtgcc gtccataggt tctggtagag gtctagaatc aaacatgaac 540
 gtttgaaagc tatcagccaa ataatgtctg gcgttgctgt ggatcaactg ggatttattt 600
 ttgcaatcaa tggccggcca agt 623

<210> 91
 <211> 272

<212> DNA

<213> Triticum aestivum

<400> 91

```

cgcgtcttcg acttcctccg cgacgagacc tctcgcagcg agtgggacat cctctccaat      60
ggcggcatcg tccaagaaat ggctcacatc gccaatggcc gcgaccatgg caactgcgtc      120
tcactcctcc gtgtcaatag caccaactcg aaccaaagca acatgctgat cctgcaagag      180
agctgcacgg acgcgtcggg ctctacgtc atctacgcgc cggaggacgt ggtggccatg      240
aatgtggtgc tcaacggcgg cgacccagac ta                                     272

```

<210> 92

<211> 703

<212> DNA

<213> Triticum aestivum

<400> 92

```

tccggctttc catcctcccc gacgcccacc cgcgccgccc cggtactga cgtttcttc      60
cagggctcgt cgccacgct catgcgcagc agcacacagc cccggctcgc tcgtcacctg      120
ggcgttccag atactggtca ataacctgcc gacggcgaag ctacacctgg agtcggtcga      180
caccgtcagc aacctgctct cctgcacat ccagaagatc aagtccgcc tgcaggccag      240
catcatctcg ccctagctcc atggctagcg ggcatttgac atggattccg gccaacacac      300
cgatacagca ccggccggat gccatttcaa atgtctaact agctcgatca ggtcgtgtag      360
gatcaciaaag ctagaggaag aagaacaaga agaagatttc aggcgagact tcggattatc      420
atagaggctt cttgggatac aaataaatta tagtgtggag atcgaattag tatttcatag      480
aggaaatgaa cggggcgaga tgcgaattag ggattaagtc aagaacgaac ctgcgcctag      540
gcatgcctgc attttaattt gccgctgctt aaactagctg atgtggtggt tcgggtattg      600
acttaaaactc tggtagaggac tgaagagaac cccctagggc tttgttttat tttatattta      660
tataatatat ctagtgcctt gggtttattaa tgcaccttc tgt                                     703

```

<210> 93

<211> 473

<212> DNA

<213> Triticum aestivum

<400> 93

```

ccccggcagg tcaagttctg gttccagaac cggcgcacgc agatgaaggc gcagcaggac      60

```


cgcgcccagaca acgtcatcct ccgcgccgag aacgagtccc tcaagacaga caactaccgc 120
 ctccaggccg ccatccgcaa cgtcgtctgc cctagctgcg gccacgccgc cgtcattggt 180
 gacatgtcct acgaggagca gaacctccgc atcgagaacg caccgctcaa agacgagcta 240
 gaccgcttgg cgtgcatcgc cagcaggtag ggtggcgggc gccagcctgg catgtcgtcg 300
 gcttcagcgt tgagctgcat gtcggcgccg ccgccggtgc tcatgccgcc ctggacctcg 360
 acatgaacgt atactcccgg cacttcgccg agcagcagca cggcgccatg gacctcatgg 420
 ccttgccgca gcagatggcc gccggcgggc acggcgctcg ggcgacggcg tca 473

<210> 94
 <211> 577
 <212> DNA
 <213> Triticum aestivum

<400> 94
 cacgcgtgcc tcgtgacgcc gccgcacagg gacatcgccg gtgtgacgtt ggagggcaag 60
 agtgagcatg atgaggctgt cacagcggat ggtgggcagc ttctgcgcca gcctgagcgc 120
 gtcgcagcag caccggtgga cgacgtgtc ggggccgggc gcgggcgtgg acgaagcggc 180
 gggccgtgcg cgtcatggtg caccgcagca cggaccccg gcagcccagc ggcgtggtgc 240
 tccgcgccgc cacctccatc tggctccccg tgttctgcga cggggtgttc gccttccgtt 300
 gccgcgaaga ggaacacgcg ctcccagtgg gacgtggctg tcgcattggg cgaacccggg 360
 tggcggcgag ggtggctcgc cattccccaa acgggcttcg cgattccggg ggccaactcc 420
 gaatcctcct tggcttcagg agggccttgg aaccgccgaa gccaggaaac aggcattggc 480
 tggatccctt ggccacggga ggaagcttgg cgcccgggga ccggcccgat cccggggggg 540
 tcccgccatg gggatggggg atggggtacc ccgggcg 577

<210> 95
 <211> 510
 <212> DNA
 <213> Triticum aestivum

<400> 95
 agccggatct gctgaattcc ggcagcgaca cgcgatgat ggccgaagcc cgcctcccca 60
 aggcgcggaa gccgtacatg ataaccaagc agcgggagaa atggaccgag gaggagcaca 120
 agctcttctt ggaagccatg cagctccacg gccgcgcctg gcgccgaatc caagagcaca 180

taggcaccaa gacggccgtg caaatcagga gccacgcgca gaagttcttc tccaaggtca 240
 cccgggactc gtccggggac agcggcggca gcagctccgg cgacggggccg ccgatccaga 300
 tcccgcgcgc gcgggccaag aggaagtcgg tgcacccgta cccgtgcaac cagaggagcg 360
 cgcagccggg gaagcacgcc catgccgcgc tcctgcgggt cgacaagcct cagctgtcca 420
 tgtgtgagca ggggaatggc tcgccgacat ccgccgtgac cacctcgcag ataggcttgg 480
 gctccgagag cttcgacagc gacacctcga 510

<210> 96
 <211> 891
 <212> DNA
 <213> Triticum aestivum

<400> 96
 gctatactac ctcattcctc tcgtgcttct actagctacc tgcaggctgt agatatttgt 60
 catcgatcat caccaacttt tcagttcaag ctaggtagct agccatggct aggggtgctgg 120
 agcctctcat cgttggaaag gtcacgcggg aggtcatcga caacttcaac ccaacagtga 180
 agatgacggt gacctacagc tccaacaagc aggtgttcaa cggccacgag ttcttcccgt 240
 cggcggctgt gtccaagcca cgcattgaag tccagggtgg ggacatgaga tccttcttca 300
 ctctggctcat gacggacca gatgtgccag ggctagtga tccatacctt agggagcatc 360
 tccactggat cgtcagtgat attcctggca ccacagatgc ttcttttggg agggaggtgg 420
 tgagctacga gagccccaag cccaacattg gcattcacag gttcaccttc gtgctgttcc 480
 aacagaagaa gcggcaggcc atgaaccctc cctccaccag ggactacttc aacacgcgcc 540
 gcttcgcaa cgagaacgac cttggcctcc cggtggctgc cgtctacttc aacgcgcagc 600
 gggagacagc ggcacgcgc cgctgacgga aacaatcagc ccaacccttc ttctcttgca 660
 tgcattggcg atgcaacaag cctatggatc aacctatagc tactattagc taccatctac 720
 atatattctc actcaaataa gagtttgctt gtgagcgtct tgtgtgtgct gtttccagta 780
 gtctttctac tgttcagttt tttcttcggt caatttagtt actagttgat ctgtgaacca 840
 gtgtacgtgc aatctacatc taccactagt gtgaatttga ttcagctcct a 891

<210> 97
 <211> 1361
 <212> DNA
 <213> Triticum aestivum

<400> 97
 tctgcactta ggcccaagtt tgagcatatg ttagatggtc tagatgacat ggatgggtgct 60
 gaggatgtgc gttttgatga gctcataggg atgcaaggcg tagtatttca cttgggtggg 120
 aacgcaatta cacgcatacc ggacaagttt gcgagcagtt tcatcagaca gatgcagacc 180
 tcagagggggc tcgacctgaa actgaaagca ccgagcgggtg agacatggca tgtgggtgct 240
 agcaagggttg ccaatgagct gttcttgagg tcaggatggg gaggtttcgc taaggctcat 300
 gaactgcagg agaatgacct cttgctcttc acatgcactg gcaatgcctc cttcgaggctc 360
 ctaatcttcg acccgagtgg ctgcgagaaa ctgtctcttc tcttcgccgg cagaatgcgc 420
 aaacactttg acaatatggg ggggtcaaggc ggcgttgagc aataccctga ttctgatgat 480
 actaggggtgc cgtctcagtt tgttggatcc cctcaaaagg cctctacttc aagaaaatgc 540
 agtggcaaaa ctaaatacaag caaagagctt cctgaatcac caagcagcac cagctgtcat 600
 gttaagcttg aggcaactga ggaacagaag agtgatgatg acacatgcac tgagcccggt 660
 tactactctt cgaggactgc aagtcagcta aatgaggatg aaaaacagga aatcatggag 720
 cgtgcttcca tccgaccggg gaatcccgca ttcgtgggtg ttctgctgat gacctctc 780
 caacgcaaga acaactttct gaccatcccc agtaaatttg cagctgatca tcttcagagc 840
 aagccgcgtg aggtcctgct gcttaggctg agcagagaag acaagtggca tgtcagggtgc 900
 tactattcca gcaggaccag atgcttcaac tgccaacggt ggggtcaagtt tgtgaggggac 960
 aatgggctgc gcgagggcga cgtctgcgtc tttgagctga tcaaaggcgc aaggaagatg 1020
 acaacgatga ccgtccatgt tgccaggagg aagaaagatg gccggtttgt cgccggacaa 1080
 tgggctgcgc gagggcgacg tctgcgtctt tgagctgatc aaaggcgcaa ggaagatgac 1140
 aacgatgacc gtccatgtcg ccaggaggaa gaaagatggc cggtttgtca ccgtgggcta 1200
 attataagtt cctgccacc ctttctccta cattttggta tgcaatatgt tgtaccgacc 1260
 tgccactcca aactgaacct gctgtacctg gtgatgcatt tgcataaaaa tgcaatggca 1320
 catttttttag ccttgtaaat gataatcatg ctaacaagca c 1361

<210> 98
 <211> 1001
 <212> DNA
 <213> Triticum aestivum

<400> 98
 aaaaccctcg cttccccact ctggactgcc ctctctctcc ctggcggtt tgttcatcgc 60
 ctctgccgaa aaatcaaacc caaaaccaag aaagggggcg aaaatggctg aggcgaactc 120
 gtacgaggag cagcgcagga ggcaggtgga ggagaacaag cggaagctgg aggagctgcg 180
 cctgcaccac ctctccgccg ccgtccgcga ggccgccgtc aagcccaagc tgaagaggaa 240
 ggccccgaag ccgcgggacg ccgccgatga cgccccgccc cggcggtccg gccgtatcgc 300
 caccctcccc gagcagcccg actatcgcca caatctgggc accgggaaac cacggcagca 360
 gaaggaagta aagcctgatc atgcgtacgc catcgccaag gccgaagagc tacaggatga 420
 gctaggctcc gactaccgca cctttgtcaa acccatgcc aagtctctta cctcgctgca 480
 tatcccgga cagttcagca tggagcatct cccggacaat ggcattgagaa tggttttggt 540
 ggatgaggag gcggaggagt ctaagggtgag ctatcgcca cacagtagca gtcttgcgc 600
 cgggtggagc gagtttgcca tcgacaatga gctagttgaa ggtgattgct tgggtttcca 660
 gctgatcaag agggcattgt tcaagggtcta catattcagg gcaagttcgt actatgaaaa 720
 tgagcattaa tggagcagcc ggcaggctcc tggttgcttc atgtcaactt gtgttgaact 780
 agcacagttc agaaatctgc actatctcag cggttagtta tgtatgtatg gctactgtca 840
 ggcattatta gaatctctcg tgtgattgcc aagtcgtatg gtaactgtcg gcatcttgta 900
 aagtttgtca tgcgacggca agtgtgaaac ttgacatgtg ataccatca tggttgatat 960
 cctattgatt cgttatgtgg acgaaccatt taccgcgcgt g 1001

<210> 99
 <211> 706
 <212> DNA
 <213> Triticum aestivum

<400> 99
 cgcaccaccg ccaaaagaag agacttacca acccgacgcg gaggtgatcg cgctgtcgcc 60
 caagacgctg atgcgacgaa ccggttcgtg tgcgaggtgt gcaacaagg gttccagcgg 120
 gagcagaacc tcgccagctg caccgccgcg gccacaacct gccgtggaag ctgaagcaga 180
 agaacccgaa ccaggtgcag cgccggcgcg tgtacctgtg cccggagccg acgtgcgtcc 240
 accacgaccc gtcccgcgcc tgggcgacct caccggcatc aagaagcact tctgccgcaa 300
 gcacggcgag aagaagtgga agtgcgacaa gtgctccaag cgctacgccg tgcagtccga 360

ctggaaggcg cactccaaga tctgcggcac ccgcgagtac cgctgcgact gcggcaccct 420
 cttctcacgg gagggacagc ttcatcaccc acagggcctt ctgcgacgcc ctggcgagg 480
 agagcgccag gctgccgccc acgagcctga gcagcctcac cagcacctct acggggctac 540
 taacgctggc aacatggcgc tgagcctgtc ccaggtggga tcccacctca actccacccc 600
 tgcagcatga cggtcaccac cacacccgtc cccggagctc ctgcgcctcg gcggagcggc 660
 tggcgaggag agcagatcgc cgcgcgcctc aacacctcct gtcgcc 706

<210> 100

<211> 1077

<212> DNA

<213> Triticum aestivum

<400> 100

atagaagaag gttatgcgta ggatagcacc cctgtgttgt cagggttggg acgtgtggac 60
 gcttcgtccg accaccacct caggtgacct tcgggacaga ccgtgtccaa cgctgcgagc 120
 tcaaaccgtc catgtcgccc agcagcccag ggttgagggt gcggcaatga gatgaggacc 180
 ttcttacaca ctctgatgg tagaccaga tgctccaagt ccaagcgatc ccaaccttag 240
 ggagtatctc cactggcttg tgacagatat ccccggtaca actggtgcgt cgttcgggca 300
 ggaggatgat tgctacgaga gccctcgtcc gaccatgggg atccaccgct tcgtgctcgt 360
 actcttccag cagctcgggc ggagacgggt gtacgcccc gggtagcgcc agaacttcaa 420
 caccagggac ttcgcccagc tctacaacct cggcccgctc gtcgcccgcg tctacttcaa 480
 ctgccagcgt gagggccgct ccggcggcag gaggatgtac aattgatcta cccacggccc 540
 tcgtacgcca ccagcccgcc gccaaagtcag caaattatcc aacgtggcta gtttactagt 600
 atatagtttg tgataagaag ccagccacga attaaagcatt acctatatat tggcaacaca 660
 tacactacat atatgcatac tatgatcgat gtataactag ccgcatgcat atatgcaatc 720
 aacggctaata taaggggggg tgaaccctag atcaatggct tggtagtga ctatataaat 780
 atagtctgca ataaactgat gccaatagta tacagcacac aaatattgga ggagctacac 840
 gccatgtgca acttagtgct acctggtaca tatctgcagg ttggtcttgt gcgttcactt 900
 atgcgtgcat gaacatcagt caatcatata gacatagtta tgcattggag acaaacatgt 960
 aactgacagc gactgctaca agacacagtg atatgtacgc acttagcgca gtagcaaagc 1020
 acatgcatgt gttgggtctt gtacaacttc ctcatgatta tggcaatata atcaagt 1077

<210> 101
 <211> 864
 <212> DNA
 <213> Triticum aestivum

<400> 101
 cggaaggatg gggcgcgga aggtggagct gaagcggatc gacaacaaga tcagccggca 60
 ggtgacgttc gccaaagcgcc gcaacgggct gctcaagaag gcgtacgagc tgtcgggtgct 120
 ctgcgacgcc gaggtcgcg ccatcatctt ctccaccgc ggccgctct tggagttctc 180
 cacatcctca tgcattgtaca agacactaga gcggtaccgc agctgcaact tcaactccga 240
 ggcaactgca actccggaga ctgaacaagc aattaccagg agtacttgaa gctgaagaca 300
 agagttgagt tcctacagac aactcagaga aatcttcttg gcgaggactt gggcccactt 360
 aacatgaagg aacttgagca gcttgagaac caaattgaga tatctctcaa acatattcagg 420
 gcgacaaaga gccaacagtc acttgatcag ctctttgagc tcaagcgcaa ggaacaacaa 480
 ctgcaagatg ttaataaaga cttacggaag aagatacaag aaactagtgt ggaaaatgtg 540
 ctgcaaatgt cttgccagga tgttggaact agtgggtcta gtggccatgc taatcaagct 600
 aatcaacagg agtatcttca tcatgattgt gacccttccc tgcattatagg gtatcaacga 660
 aatttcctgg atcagttgaa caaagaatga atgggtgtat gagctgagcc tgagcgcaga 720
 tgtcctacgt tgggtcgatg tgataaaaat ttaatgcaat gctttctttt atcaatgggt 780
 tctaacttaa aacaacctat tatgtactat tcaatcatat tgtgtgttat tttaaacttg 840
 agagaagaat tatctgaatg ttgg 864

<210> 102
 <211> 588
 <212> DNA
 <213> Triticum aestivum

<400> 102
 cattcgaatg agccattcca aggagtgatg tcaaccgtga gtactttcag aagagcatga 60
 tcgtaaggct gcagctggca tagattatga ctcttcatat ggaaaggccc gtccaaatga 120
 taccattctg aagcttcaga ggacttcacc atattacaag aggaaccgag ctcatgtttg 180
 cagtttctat gtgcggggtg aatgtacaag aggcgctgag tgcccatatc gccacgagat 240
 gcctgagaca ggggagttat cccagcagaa catcaaagat cgctactatg gagttaatga 300

cccagtggct ctcaagcttt tgggtaaggc aggtgagatg ccatctctgc aaccaccaga 360
 tgatgagact ataaggaccc tctacattgg tggacttgat aaccgagtca gtgagcagga 420
 tttgagggat cagttttatg cacatgggtga gattgaatcc atcaggatgg taatccagcg 480
 tgcttgtgca tttgtgacat acacaactag agaaggcgct gagcgtgctg cagaggagct 540
 tgctaacaag ctagtcatca aggggtgtgcg cctgaagctc atgtgggg 588

<210> 103
 <211> 1254
 <212> DNA
 <213> Triticum aestivum

<400> 103
 attttcagct tcggaatttg ttatgggcaa cttcaaagca tgatgtttat ttcattgtcga 60
 attccacagt aggccactgg tcgtcattgt ctcacaaaat gacagatgtt cttgatattct 120
 cagggcatgt tgctccagca aagaaacacc ctggctgtgc actagaaggg tttactggcg 180
 ttcaagttag cacacttgca gtaaatgagg gtttatttgg cgctgggtgg tttcaaggag 240
 aactagtttg caagagtcta ggagaacgtg atgttaagtt ctgcacaagg actactttga 300
 gcgacaatgc tatcacaaat gctatggata tacacagatc tacaagtgga agcttgcgca 360
 ttacgggtatc gaataatgat tctgggtgtc gtgaattcga catggaaaga tttcagctct 420
 tgaatcactt ccgttttaac tggccagtga atcacacatc cgtgagccc gacaagaaac 480
 ttttggcagt agttggagat gaccgggacg ctcttcttgt tgattcacga aatggcaagg 540
 taacttctac cttagtggc catctggact actcatttgc ctcagcgtgg caccgggatg 600
 gtgtcacctt cgcaaccggg aaccaggaca agacctgccg ggtctgggac atccggaacc 660
 catcgacctc cctcgcggtc ctgaggggca acatcggcgc gatccggtgc atccgctact 720
 cctcggacgg gcggttcctt ctcttctccg agcccgcga ctttgtgcac gtctacagcg 780
 ccgccgagtg ctaccgtaag cggcaggaga tcgacttctt cggcgagatc tcggggatct 840
 cgctcagccc ggacgacgag tccctgtttg tgggggtgtg cgaccgcgtc tacgccagcc 900
 tgctgaatta caggatggta cacgccaacg ggtacctgga ctcgtacatg tagaggaata 960
 aaacctttgt gtttgatgtg aaagaaagag cgccagaact gctgaaaatt tgaagatctc 1020
 gctcggcccc gccgacgagc ccctgtcgtt cgagggggcg agatatgtca gcgtgctcca 1080

ttacaggata gtccagggtg gccttcgggc acctggactc tgtacatgta gagaaagaaa 1140
gcttggtgct aagtttagtg tgttcgtgta aaggaagagc ggaattgctt aaaatctgcc 1200
gaacagtaac ttttaaggag atcagtatgg aatttatacc aacttcactt tccg 1254

<210> 104
<211> 724
<212> DNA
<213> Triticum aestivum

<400> 104
taagcgacgt ttctccgagg caccagggca atcacagtac agagatcggg ctgcagaaaag 60
acgaagtcta tatggatcaa cttcatcgct tcccaatgat ggactggata caactggtga 120
atattcgtct cgaaaggggt caagtgagat gggatcaatg ccatttcctc caggggtggg 180
tgaacgttct gttggtgaaa gtgacaacac tgaaaattat gaggtgatca ctgctgacag 240
agcaatagac gaaaacaatg tgggcaatcg aattctgcgc aacatgggct ggcaagaagg 300
actgggcctt ggaaaggatg gcagtggcat caaggagccg gtgcaggcca agcctggtga 360
cgtgagagcc gggcttggtg gtcagcagaa gagagctgct gacctatcac tggaggcccc 420
agctggcgat agctataaaa ccataatcca gaagaaggcc attgagagat tcagggagat 480
gtcctaagtg ctttctttat ttcatcgctg tgtataatta gttcggatac tcgtttttat 540
cgggtgtaat gtcaaaaacc ccaatgctgt gtgcatttgt actgttaaatt ttatcttatg 600
tgtctggctc cataaatctt ggtgtaattt aagcaccgag tacagattta atttgtgtac 660
tcacattggt ctgaaactag ctctgtacct gtttttcttt tcttgacctac ttgaattcaa 720
tggt 724

<210> 105
<211> 568
<212> DNA
<213> Triticum aestivum

<400> 105
ggtctgtgta acccgaaaca aaaaccatgt tctttcaagt gggagagagg gtgtgccaca 60
agcttttttg atgttagaat gccgacggag gtagtggaga tttgtggaac attcagggcc 120
ccgagtaaca gattggttgg cagctggggc agagcttgta taagcccga tgagaactgc 180
atagcgtctg gtcctctga tgggtctgtt tgcatatggt cgagatcgaa gaacgagggg 240


```

ccgaccatcc tagagggcca ctcatggccg gttgttacaa gtgcatggtc tgagtttgga      300
cctcttgcca ccgccgacaa gaatcatatt catatatggg cttgagaccc gtgttttcctt      360
cggtcattca aaccaggatg tggaggggtt cgcttcaata tggatttggt ccattcttga      420
ggacactcgt ttgcatgtaa atgacttagt tagtgtctaa ttcaattgta attatgaaaa      480
ctgacagcac ggtacggcca tctgctttcc tctctgtatg ttttagggg atttctgtca      540
tctgttgtct ggtaaaatga tgggtttg      568

```

```

<210> 106
<211> 795
<212> DNA
<213> Zea mays

```

```

<220>
<221> misc_feature
<222> (698)..(698)
<223> n = any nucleotide

```

```

<220>
<221> misc_feature
<222> (721)..(721)
<223> n = any nucleotide

```

```

<220>
<221> misc_feature
<222> (770)..(770)
<223> n = any nucleotide

```

```

<400> 106
atggggaggg gaaggattga gatcaagagg atcgagaaca acacgagccg gcaggtcacc      60
ttctgcaagc gccgcaatgg gtcctcaag aaggcgtacg agctctccgt cctctgcgac      120
gccgaggtgg cgctcgtcgt cttctccagc cgcgagcgcc tctacgagta cgccaacaac      180
agtgtgaagg ctacgatcga gaggtacaag aaggcacacg ccgttggttc ttctcttggc      240
ccccactct tagagcacia tgcccagcaa ttctaccagc aagaatccgc aaaactgcgc      300
aaccagatcc aaatgctgca aaacactaac aggcacttgg ttggtgactc tgtgggaaac      360
ttgtcactca aagagctgaa gcagctggag agccgccttg agaaaggcat ctctaagatc      420
agggccagga agagtgagct gttggctgcc gagatcaatt acatggccaa aaggagagact      480
gagcttcaga atgaccacat gaacctcagg accaagattg aggagggaga gcaacagctg      540

```

cagcaggtga ctgtggccca gtctgtcgca gcagcagcgg ccaccgacgt ggagctgaat 600
 ccgttcttgg agatggatac caaatgcttc ttccctggcg gccccttcgc gacactggac 660
 atgaagtgct tcttccccgg cagcttgcag atgctggnag cacagcagcg ccagatgctt 720
 ngcaccgagc ctgacctccg ctaccaactg gcgccgcctg acactgacgn tgccaacaat 780
 aaccctcage agttc 795

<210> 107
 <211> 720
 <212> DNA
 <213> Zea mays

<400> 107
 atgggggaggg gccgggtcga gctcaagcgg atcgagaaca agatcaaccg ccaggtcacc 60
 ttccccaagc gccgcaacgg cctgctcaag aaggcgtacg agctctcggg gctctgcgac 120
 gccgaggtcg cgctcatcat cttctccaac cgcggcaagc tctacgagtt ctccagcacg 180
 cagagcatgc caaaaacact tgaaaagtac caaaaatgca gttttgcagg gcctgaaaca 240
 gcactccaga atagagagaa tgagcaactg aaaagtagcc gcaatgagta cctcaaactg 300
 aaggcacgtg ttgataatth acagcggact caaaggaact tgcttggtga agatcttgag 360
 tcattaggca taaaagagct ggagcacctg gagaagcagc tcgattcgtc cttgaagcac 420
 ataagatcta caaggacaca acacatggtt gatcaactga cagaacttca gaaaaaagaa 480
 caaatgtttt gtgaagcaaa taagtgtctt cgaagaagac tggaggagag caaccagggt 540
 atatggcagc atgctgtgga gcaaagcgag cggcattctg aagtgcagcc gcagcagctc 600
 aatggcaata acttcttcca tccctcgat ggtgctggtg aaccaccct tcagataggg 660
 tatccttcag aggcactgac tagctcatgc atgaccacct tcctgcccc gtggttgcca 720

<210> 108
 <211> 492
 <212> DNA
 <213> Zea mays

<400> 108
 atggtcggga ccgggaagag ggagcggata gcgatacgga ggatcgacaa cctggcggcc 60
 aggcaggtga ccttctccaa gaggcggcgg ggctgttca agaaggccga ggagctctcc 120
 atcctctgcg acgccgaggt cggccttgct gtcttctccg ccaccggcaa actcttccac 180

ttccgccagct ccagcatgaa gcaggtaatc gatcggtagc actctcattc caagactctc 240
 cagagggtccg aaccgcagtc gtctcaactg cagtcacata tggatgacgg cacttgtgca 300
 aggctaaagg aggaacttgc tgaaacaagc cttaagctca ggcagatgag aggagaggag 360
 ctccagagggc tgagcgtcga acagctgcag gagctcgaga agaccctcga atccggcctc 420
 ggctcttact caaaaccaag agccaaaaaa tccttgacga gatcagcggg ctggaaagaa 480
 agagaacgca gc 492

<210> 109
 <211> 1602
 <212> DNA
 <213> Zea mays

<400> 109
 atgtctgaaa tggagtctgt cgcggctctt atggactcga caagctccaa gatacaacag 60
 cttcaggagg cattcgtgga gcttgagagc cagagcgctt tctgcatgaa cttcaagtgg 120
 aagcagcttg aagaccactt tcgtggtctt gaacaatccc tcaagaaaaa gtttgacgag 180
 ctaaaagagc aggagaagga gttccaggag actgtcgcaa aatcagagca gatgctggag 240
 cagcgggagg cagctgttgt ggccaaggaa ctgacctctt tggagagggt gcaggagaag 300
 agggaagctg cattagctat gatcttttagc aagtccaggc tttccttacc tgctccttcc 360
 atcaacccaa tgaacaaggc actggataat cttggtgtca agtggcctaa accggcttct 420
 gaggagagtg tgcacttgca agttgataat gctgcagtga ggcctcgctt tgagcttggt 480
 gcactctgtg aggaaatgaa tgtgaacggg cttcataagt tcatatccga caacaggaag 540
 aacttggcag ccatccgga ggaaattcct agtgactga agaaaacatc tcatccttat 600
 ggcctggtgt tggattccct ggaagacttt tattctggag ataatcttgt attagatgga 660
 aaaaaggatg gtgaccttct ggggtgttaga aggacatgtt tgatgttgat ggagtcactt 720
 ggacaactgc actctaata tataacttgt tcttctttgg agaggcacat gcttacaact 780
 aatatcatag agcgagcaaa gacgattgca ttaagtggga agtccaagtt agataatctt 840
 gacattgatg ctagcaatgg gaactgtctt gaagctcatg cgtttcttca actcctggca 900
 acctttggta tttctgctga attcaatgaa gatgacttgt gcaaattgct tccatatgtc 960
 agtcgtcgtc gtcagacacc tgaactttgt cgattgcttg ggttgctgca gaaaatgcc 1020

ggtgtcattg aagttctggt ggaaagtgga agacctattg atgcaattaa tttggcttat 1080
 gtgtttgagc ttactgaaca atttgaacca gtacaccttc ttaaagcata tctaagggat 1140
 gttaagaaaa tgtcacacgc caggaatgtc aaaacttctc ctggagcaca gaatgagatg 1200
 aatgaacgtg agttatctgc gctgaaatct gtcattaagt gcattgaaga gcacaaactt 1260
 gagcagcagt accctgtgga tccacttcag aaaagggttt tgcagctaga gaaagcaaag 1320
 gcagacaaga ggatggctgt tgaagcagca aagccgcagt ccaagaggcc tcgtgccaat 1380
 ggatcggcct ttgcaaccgc tgctacaggc tttgctgaca agagcttcta ctgagcaacc 1440
 ccagagaggc atccattcaa tccttgcgag aggcagtttg tgtatggcgc tgaggttcat 1500
 cttcctocca tgatgacctc agcgtcttac acgatgcaac ctgcccattg accttattac 1560
 ggtaatggct acccagttca gtaccaggtc tcgtatatcc ac 1602

<210> 110

<211> 600

<212> DNA

<213> Zea mays

<400> 110

atggagctgc cagaggagat tgggcggccg aggaacttgc cggcgagggc gcagacgcgg 60
 tccagctcgt ccttgagccg cgcgttctct atgcggaggt gctgctctc cagcgacacc 120
 tccccgagca cggccgcgcc gccacagttg gcgcagatgg gggtgcgcat cgcctcccgg 180
 atcgtcatgt tctccgccct cagcttgctg ttctctgcc tcagcagcgc gttctcgtgc 240
 cgctcgatct gcgtcttcat ctgcgtgcgg cggttctgga accagaactt gacctggcgg 300
 ctctctaggt tgagccgctt gctgagctcc atcctttgct tctcgtcggg gtgggggcac 360
 tccttgaaca cagcttcgag ttcttggatt tgctgcggtg tgtggcgggtg gtatcgcttc 420
 ttcttctctgc gagggttact gttgtccggg tcgatctcgt caccggaggc gccgtcgacg 480
 ttgtcgtgc ctgaccggct gtcgatctcg tcctcacggc cgcggcccag cgagtccccg 540
 tcgcctccgc tcccgtccc gccgcgccca atcaggccca tacgacccaa gtcgcccagc 600

<210> 111

<211> 444

<212> DNA

<213> Zea mays

<400> 111

atgttgcctt ggttcgagtt ggcgttatcg acgcggagca gcgagacagc gttgccgtgg 60
 tcgcggccgt tggcaatgtg ggccatctcc tgcacgtcgc cgccgttgga gagcatgtcc 120
 cactcgtcgc ggggtggcgtc gtcccggagg aagccgaaga cgcgcgcggg ggccacgggg 180
 agccagaagg aggtggcggc gttgaggatg atgcccgggc gccggcccgg gtcgtccacg 240
 ctcttgccgc tcatcacgcg cacgtcctcc gcgcgcgtgc ccgacagcgt ggtccactgg 300
 tgcgtcgtcg acgccgtcac gccgccgag aagctcgcca ccatgcgctc cgccagccgc 360
 atcatgctgc ggcgcccctc tggcgacgtc accacgcccg cgactgtccc tgctgcttcc 420
 gcccaatcta acgacacgct ggcc 444

<210> 112
 <211> 588
 <212> DNA
 <213> Zea mays

<400> 112
 atggaggagc tcatgaaagt ggcgcagatg gacgagccgc tctggctccg cagtcccagc 60
 ggcagcggct tggaggcgct caacctcgac gagtaccacc gagcgttcgc caggggtgttc 120
 gggccgagcc ccgcgggcta cgtcaccgag gctaccgcgc aggcggcggt cgccataacc 180
 agcagcgtag acctcgtaga cagcctcatg gacgcggctc ggtggtcgga gatgttcccg 240
 tgcacgtcgg cgaggcgag caccacggac atcatctcga gcggcatggg cggcacgcgc 300
 agcgggtcaa tccagctgat gcacgcggag ctgcagggtc tgcgcgcgt ggtgccgac 360
 cgtgaggtgg tgttccttcg gttctgcaag cagcacgcgc aggggctgtg ggccgtggtc 420
 gacgtctcgg tggacgcat cctccggccc gacggcggcc accaccacgc gcagaacggc 480
 ggcggcgccg gggacatggg ctgccgcctg ctccccaccc gctgcatcgt gcaagacatg 540
 aacaacggct actccagggt cacgtgcgtt gtccaccggg agaacaac 588

<210> 113
 <211> 375
 <212> DNA
 <213> Zea mays

<400> 113
 atggtggaga tgcggtcgac ctgcgtcgcg aggcgcgcgt tctcgacgcg gaggtggtgc 60
 tcgtcgaagg acatctcgcc gatgacggcc gtggccgcgc ggccgccgca gctcgggcac 120

gaggcggttgg cgagcgcgtc cttgtaccgc gcgttctcgg cgcgcagctt ctcgttctcc 180
 gccgcgagct gcatgttctc gtgccgttcc tgctgcgtct tcatctgggt acgcttggtc 240
 tggaaccaga acttgacctg gagaggctcc aggcccagct cccggctcag ctccttcctc 300
 tgcttgctgt caggggtgcgg gtactccttg aagaacgctt cgagctcctg gatctgggtg 360
 tgggtgtggc ggtgg 375

<210> 114
 <211> 444
 <212> DNA
 <213> Zea mays

<400> 114
 atgcgcgaca cctcctgcac cgggttgccg tgcgacagca cgtccactg ggagcgcggtg 60
 tgctcgtcgc ggacaaaggc aaacgcccgg tgcgacggga ccggcagcca gatggacgtg 120
 gccgcgctga gcaccacgcc gctgggctgc cccgggtccg tgctgcgggtg caccatgacg 180
 cggacgcccc cgtcgttggg ccccgacagc gtcgtccacc ggtggagctg cgacgcgcta 240
 aggctcgcgc agaagctgtt caccatccgc tgcgacagct tcatcatgct ccgcttcctt 300
 tccggcgta ctctgccgc tgcaatgtcc cggtcgggca taatgccggc cgtggcgagg 360
 cacgcgcacc gctcgcacgc cctctgcagc gcggcgagcc aacggtgcgc gccgaacgcc 420
 gccccgctga ggacgaggtc gcgg 444

<210> 115
 <211> 999
 <212> DNA
 <213> Zea mays

<400> 115
 atggttttca tgagcaccac gccgaagccg cttgactccg ccgtcgtcag cgagccggtt 60
 gacggcgggg atgaggggtg tgggtggaggc ggtggcgggga agaagcaggc gcacggggcg 120
 gtgatggcgc cgccaccgat ggtgggtcccj gccgcggcgc ccgcggccga ggaggaggtt 180
 cggaaggta ggaagccgta caccatcacc aagtcgcggg agagctggac ggagccggag 240
 cacgacaagt tcctcgaggc cctgcagctt tttgatcgtg attggaagaa aattgaagca 300
 tatgttggct caaagactgt tatacagata aggagccatg cacagaagta ctctctgaag 360
 gttcagaaga atggtacggg cgagcatctg ccccccacca ggccaaaacg aaaggcagct 420

```

catccatatac cacagaaggc atctaaaaat gtttctcaag ctattttatc acaagaacaa      480
cctacccaga gggagcaagg ttctgtcatg cctatggata ctgctactgt gagaaacaca      540
aatgcgaatg tggcggtgcc ttctggggac aatactcttg ctcaaccttt cagcgctggt      600
catgttcaag gtgcagctgc aacaaataac tgttcaagta gcatggagag tccatctggt      660
acttggccaa cttctgaagc agtcgagcaa gaaaatatgg ttccaccact ccgtgccatg      720
ccagattttg cccaagtata caacttcttg ggaagcatat ttgatccgga tacaagtggg      780
catttgcaga ggttaaaagc gatggatcca gttgatattg aaacggcact actacttatg      840
aggaatcttt ccatgaacct taggagtccc gactttgagc aacatcggag gttgctatca      900
tcatatagct atggtggcgc ggaccatgtc aaatctgaag ccatggaaaa tcttgatct      960
cctcagagtg gccatcttcc atccctggta acaagcgaa                               999

```

<210> 116

<211> 636

<212> DNA

<213> Zea mays

<400> 116

```

atggcggaag cgacgccgta cgaggagcag cggcggaggc aggtggaggc gaacaaacgg      60
aagctggagg agctgcagct ccaccacctc tctgctgcc aacagagagg cgccgccgcc      120
gccaagccct cgctggtcaa gaagcggacg gcgcgggtgc ctctcgacgt cgccgacgcc      180
gtgatggagc cagtacggcg gtccggcccg ctcgccaacc tcccggataa gcccgtgtac      240
cgcgagaagg ttcccgattt tgggatgaag attaggagga cgtacggatc ggtgagaaga      300
gatctgacca accgggcgta cgcgaccgac gatgcgagga gctaccccat ctctaaggcg      360
gaggatctgg agcaagagct ggactccagc ttcccatct tcatcaagcc catgaccag      420
tcccatgtca ccgggggctt ctggctgggc cttccgattc atttctgccg gaagtatctg      480
cccaagcgtg acgagatgat caccctggtg gatgaggatg atgacgagtc cgacacgctg      540
tacctcgcaa tgaagcgggg gctcagtgcc tggatggaga gggtttgccg tccagcagaa      600
gctggtggat ggagattgcc tggctctcca gctgat                               636

```

<210> 117

<211> 1710

<212> DNA

<213> Zea mays

<400> 117

atgttttgaa ttagggatgc cgaccaccaa gaccagatga agccgctgct tgcgcagcag	60
caccagcacc agcaccagca ccagatgcc a cctgcgccgc tgetcaacgc cgcttcgctc	120
tcggcggggc ccggccaggc ggcaggggca accccgccgg cagtgaagaa gaagagaaac	180
ctaccagacc ctgacgcgga ggtgatcgcg ctatcgcca agacgctgct ggcgacgaac	240
cggttcgtgt gcgagggtgt caacaagggg ttccagcggg agcagaacct gcagctgcac	300
cggcgcgggc acaacctgcc gtggaagctg aagcagaagg acccgctcgca ggcgcagcgg	360
cggcgcgtgt acctgtgcc ggagccgacg tgcgcgcacc acgaccctc ccgcgcgctc	420
ggcgacctga cggggatcaa gaagcacttc tgccgcaagc acggcgagaa gaagtggaag	480
tgcgacaagt gctccaagcg ctacgcgctg cagtcggact ggaaggcgca ctccaaggtc	540
tgcggcacgc gcgagtaccg ctgcgactgc ggcacactct tctcgcgag ggatagcttc	600
atcaccacaa gggctttctg cgacgccctc gcgcaggaga gcgccggct gccgccccca	660
ggcctcaccg ccagccacct ctacggcgcc accaacgcgg ccaacatggg gctcagcctc	720
tcgcaggctg gttccacct cgcatccaca ctcggtgcgg acgtcacgg ccaccatcag	780
gacctctec ggctcgggtg cagcaacgcc gcgagccgct ttgagcatct tctggggccg	840
tccaacgcct ccgcgttcg gccctgccc cctccaccat cgtcggcatt tctcatgggc	900
gctccgcagg agtttgggc aggcgacggc agcgggtccc acgggttctt gcagggcaag	960
ccgttcacaa gtctcatgca tctgccgat ctccaaggca acggcgccgg ggggtgcctc	1020
gcgtcgtcgg ctacgggtct cttaacctt ggctacatcg caaacagtgc aaacagctca	1080
ggtacttcca gccatggcca tgcaagccag ggacacctga caagtgatca gttcagcgaa	1140
ggaggtggcg gggcggtc cgagtcacg gccgcgatgc tttttagcgg cggcggaac	1200
ttcgccggtg gtgaccacca agttgtcct ggcgggatgt acaacaatga tccggccgtg	1260
atgctgccgc agatgtccgc gacggcgctt ctgcagaagg cgtcccagat gggctccagc	1320
gcgagcgcgc acggcggtgt gtccgtgtt ggggggctc tgggctcgtc ggcgcctcc	1380
gctaggcacg cccgggcacc catgctcgac cagagccaga tgcacctaca gagcttgatg	1440
agctcgtcgg ctgccggcgg catgttcggc ggcgccaaca gcggcagcat gattgacctg	1500
aggatgtatg acatgaacca agatgtgaag tttagccagg gccgtggtgg cgctgagatg	1560

acgcgcgact tccttggcgt cggcggcgat ggtgtcatga gaggcacatgac agtggcgaga 1620
 gcggggcacc aagacggtgc cgatgacatg aactccttgg aagccgagat gaagtccggcc 1680
 tcgtcgccct tcaactggaga caggatgcaa 1710

<210> 118
 <211> 153
 <212> DNA
 <213> Zea mays

<400> 118
 atgatgagcg cgacctcggc gtcgcacagc agcgacagct cgtacgcctt cttgagcagc 60
 ccgttctctgc gcttggcgaa cgtcacctgc cggctgatct tgttctcgat ccgcttcagc 120
 tccaccttgc cgcgacccat cttccctgct agc 153

<210> 119
 <211> 678
 <212> DNA
 <213> Zea mays

<400> 119
 atggcgcacc ggctgctgcg ggacgcgcag gcggacgggt gggagcggtc ggacttcccc 60
 atcatctgcg agtcatgcct cggcgacaac ccctacgtcc ggatgttaag agctgaatat 120
 gacaaagaat gcaaaatttg tgcacgtcct tttactgtat tccggtggcg tcctgggcga 180
 gatgcaaggt acaagaagac tgagatctgc cagacatggt gcaagctaaa gaatgtctgc 240
 caggtctgcc tgctggatct tgaatatggc ctgcctgttc aggtccgaga tacagcactc 300
 tctatcaatt ccaatgatgc aattccaagg agtgatgtca atcggaata ttttgcagaa 360
 gagcatgata ggagggccag agctggcata gactatgatt catcctatgg gaaggctcgc 420
 ccaaatgata ccattctgaa gcttcaaagg acagcaccat attacaagag gaacagagct 480
 catgtttgca gtttctttgt gcgaggtgaa tgtacaagag gagctgagtg cccttatcgt 540
 catgagatgc ctgaaactgg ggagttatct cagcagaaca ttaaagatcg ctactatgga 600
 gttaatgata cagttgcact gaaacttttg agcaaggcag gtgagatgcc atccctgaca 660
 ccacagatga tgagacta 678

<210> 120
 <211> 1305
 <212> DNA

<213> Zea mays

<400> 120

```

atgtcgcagc gcgacggcga cctcgccgcc ggagctcatc cggaagaagg agaccgccac      60
gccgttcgaa tggatggctc cgacgatgag ctggaccaac ctacgagtaa agcgactgag      120
cacatcacgg ctatggacgt gaggaagggg aaggatatgc aagggatccc ctggaacgtg      180
atgccgacta caagggacaa gtacaggcag tcacggctgc aggggtacgc caactttgag      240
aacgtgccta attccggaag gctttctgag aaggaatgca tgcctgcaga gaaaggccaa      300
ctctactacg agtttcagca cagcacgagg tcggtgaaac cgaccatccg tcatttccag      360
ttgaggaatt tagtgtgggc aacaacaagg cacgatgtgt atcttatgtc aaacttttct      420
gtgcttcatt ggtcagcaat gacttataag aagcaagaag tcattaatct tcaaggacat      480
gtgacaccgt gcgagaagca ccatcggaat cactacgagg gtttttatca gactccagtc      540
agtactttgg cggatgaagca taacttgctt gttactggtg ggtttcatgg ggagataata      600
tgcaagtttc tggatcgta aggagtaagc tactgttgca agacaacaaa tgatgacgac      660
ggatattacta attctctgga gatatttgag aagcctagtg gttctctgca ctctcttgcg      720
tcaaacaatg attgtggact tagagacttt gacatggaaa agtttcgaat atgcaaccac      780
ttccattttg cttggcctgt gaatcataca tcgttgagcc cagatggtaa acttgctgct      840
attgtggggg acagtcctga tggacttctg attgacgcta attcaggaaa gacggttcac      900
aaactccatg gccatttgga ttactccttt gcgtcggcat ggaatccaga cggccgaaca      960
ttcgccacgg ggaaccaaga caagacgtgc agggctctggg acgtccgcaa tctctccaga     1020
tcagttgttg tattgggtgg caacatcgga gccgtcaggt cgattcgcta cagccagac     1080
gggaagttca tggcgaatgac tgaaccgcg gacttcatcc acatcttcga tgttgagagt     1140
gggtacagca ggaagcagga gctggacttc tttggtgaga tcgcaggcat atcgttcagt     1200
cctgacactg aagctctctt tgtcgggtgtg catgatcgct tcttcaactg ctgcctctct     1260
cagttcagtc ggcgacgggt ctactcgta cttgattcaa cactg                        1305

```

<210> 121

<211> 792

<212> DNA

<213> Zea mays

<400> 121

```

atgggcaatt ctcttggcca aggaatgaca tcgactcaaa cattggattc tgttgtaaag      60
cctagacctg tcagtaacag cttaaggact acagttatgg gtgtcataag aggttctgct      120
agaggagtga tcaagtcaga tactacgttt cacgcattat ctgatgctgg tagcactgat      180
tcccgcacca ctataaccac aagagcaagt gggttaatga caagttctga ggcacttata      240
actcctgcac ccttcaaaac tgatgtgtct gccttagtct ctaatactcc atcgggagtt      300
tctgggagtg gtaaactgct gttctctgag gcaccgggac aatcccagta cagggatcga      360
gctgcagaaa gacgaaatct gtatggatca tctcatggca ctgatactgt tggattggat      420
tcaactgggtg actatccgtc ccgaaagggt tcaagtgaga tagggtcgat gccgtttcct      480
cctggtgtgg gtgaacgttc cagtggtgaa attggcaaca gcgaaaacta tgaagttatt      540
accgctgata gagcaatcga tgaaagcaat gtgggcaacc gcattctacg caacatgggg      600
tggcaggagg gactgggtct cggaaaggat gccagcggta tcaaggaacc agtcctggca      660
aaatccgtcg acgccagggc tggactggga agtcagcagc ggaagcctga tccatccctt      720
gaagcccaag cgggggacag ctacaaaacc atcatccaga agaaggccat ggcaagattc      780
aaggagatgt ct                                                    792

```

```

<210> 122
<211> 759
<212> DNA
<213> Musa acuminata

<220>
<221> misc_feature
<222> (9)..(9)
<223> n = any nucleotide

```

```

<220>
<221> misc_feature
<222> (11)..(11)
<223> n = any nucleotide

```

```

<400> 122
acaattttnc ntcattcaga tttaatactg ccaacaatat ccttggttaag ttgggtacaa      60
ataccaatat tgatatcata gtgttgccat atggattggg cgaaactcca gcagaccctg      120
atatcctgtc ctagactact ggtctcactg gcgtaacagc ataaatgaaa ttacagtggc      180
acccaaactt ccaagtcaga ttagcttata attattgata gcacatttag atcttcttaa      240

```

```

cagtgaaca agcatgattc ttttcaagga agagaagtgc aaaggacaag tattggaaga      300
gaatctccct tacgatggtc caggtggccc ttgagggatg tgctgcattg caggtgggtgc      360
tggtggaggt ggtcgggaca tcatggactg gtaaggggga tactgctgct gatacgggtg      420
tggggggtggg gccatatacc cgtatgggtgg ataataatgt ggatactggc cctgaagatg      480
tggaggaacc ataggataag catgcccctg aacatttggt gcttgctgcc ctttctgctt      540
ttctgatcca gttttgtcat cgctacctcc ctgcgattgg attacagcac ccatcctctg      600
aggatccatg gatgggtaga atgtccgtgc agcttgtggc ggtgcaggga tgttgaaata      660
gttaggttgt tgctgctgat gagtctcagt gcctgggtgg agaggttgct ggttgctctg      720
ctgctgagat attacagccc ttgggagcat ccctccatg                               759

```

```

<210> 123
<211> 522
<212> DNA
<213> Oryza sativa

```

```

<400> 123
atgtctaggg tgctggagcc tctcgtcgtc gggaagggtga tcggagaggt catcgacaac      60
ttcaacccca cgggtgaagat gacggcgacc tacagctcca acaagcaggt gttcaacggc      120
cacgagttat tcccgtcggc ggtcgtgtcc aagccgcgag tcgaggttca gggcggcgac      180
ctgaggtctt tcttcacact gggtatgaca gatccagacg tgccagggcc tagtgatccg      240
tacctgaggg agcacctcca ctggatcgtc actgatattc ctggcaccac tgatgcttcc      300
tttgggaggg aggtggtgag ctacgagagc ccgaagccca acattggcat ccacaggttc      360
gtcctcgtgc tgttcaagca gaagcgccgt caggcgggtga ccccgccatc ctccagggac      420
tacttcagca cccgcccgtt cgccgccgac aacgacctcg gctccccgt cgccgccgct      480
tacttcaacg cgagcgaga gacggccgct cgccgccgct aa                          522

```

```

<210> 124
<211> 173
<212> PRT
<213> Oryza sativa

```

```

<400> 124

```

```

Met Ser Arg Val Leu Glu Pro Leu Val Val Gly Lys Val Ile Gly Glu
1           5           10          15

```

Val Ile Asp Asn Phe Asn Pro Thr Val Lys Met Thr Ala Thr Tyr Ser
 20 25 30
 Ser Asn Lys Gln Val Phe Asn Gly His Glu Leu Phe Pro Ser Ala Val
 35 40 45
 Val Ser Lys Pro Arg Val Glu Val Gln Gly Gly Asp Leu Arg Ser Phe
 50 55 60
 Phe Thr Leu Val Met Thr Asp Pro Asp Val Pro Gly Pro Ser Asp Pro
 65 70 75 80
 Tyr Leu Arg Glu His Leu His Trp Ile Val Thr Asp Ile Pro Gly Thr
 85 90 95
 Thr Asp Ala Ser Phe Gly Arg Glu Val Val Ser Tyr Glu Ser Pro Lys
 100 105 110
 Pro Asn Ile Gly Ile His Arg Phe Val Leu Val Leu Phe Lys Gln Lys
 115 120 125
 Arg Arg Gln Ala Val Thr Pro Pro Ser Ser Arg Asp Tyr Phe Ser Thr
 130 135 140
 Arg Arg Phe Ala Ala Asp Asn Asp Leu Gly Leu Pro Val Ala Ala Val
 145 150 155 160
 Tyr Phe Asn Ala Gln Arg Glu Thr Ala Ala Arg Arg Arg
 165 170

<210> 125
 <211> 2601
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (735)..(735)
 <223> n = any nucleotide

<220>
 <221> misc_feature
 <222> (829)..(829)
 <223> n = any nucleotide

<220>
 <221> misc_feature
 <222> (861)..(861)
 <223> n = any nucleotide

<400> 125

atgggcgggg cgcctcctc cgccgacgcg ctcttcgcct cgccgctcat ccccaacccc 60
gctctcgccg gcttcatgtc ctcttcgcc gccatgccct tccaccactt ctccaacgcc 120
gccgccacct taatacctaa ggaggagggg ttgatgggcg gcctccatgt ggccaaggac 180
gaggggatgg acctcgagat ggacatggag ctcatgtggc gctccggcag cgcccacctc 240
gacggcctcc tcagcttcgc cgacgtcgac gacgaccaca agccgcagca cagcggccat 300
gaccagccgc ccgacgccgc ccagcccagc ggcgccgcgg ggggcaacgc caagaagaag 360
cgctaccacc gccacacggc gcaccagatc cagcagatgg aggcgctgtt caaggagtgc 420
ccgcatccgg acgacaagca gcggtgaag ctgagccagg aactgggcct caaacctcgc 480
caggtcaagt tctggttcca gaatcgccg acccagatga aggcgcagca ggacagggcg 540
gacaacgtga ttctgcgcgc ggagaacgag aacctcaaga gcgacaactt ccgcctccag 600
gccgccatcc gcaacgtcgt ctgccccaac tgcggccacg ccgccgtcct cgccgacatg 660
tcctacgagg agcagcagct ccgcatcgag aacgccaggc tcaaagacga gcttgaccgc 720
ttggcctgca tcgcnaccgc ctacggcggc ggcgggcgac ggcagccggt gctgtcgacg 780
tcggccctga gctgcatctc ggcgccgcgc ccggtcctca tgccaccgnt cgacctcgac 840
atgaacgtct actcgcgga nttcgccgag caggcgccgc tcatgggctg cggcgacctc 900
atcccgaac ccgtcgtgcc gcagcacgat ggcgcgggcg cgtacatggg cgccatgatg 960
gcgcccgtgc aggagcagga caagcagctg gtggtcgacc ttgccgccac ggcgggcgac 1020
cagctcgcca ggatgtgccg cgccggcgag ccgctgtggg tgcggcagcg tggcgcgag 1080
gtgatggccg tggaggagca cgcgcggtatg ttcagctggc cgggtggacg cgcggaagcag 1140
ggagacggcg gcgcggtggc gagggcgag gggacgcgtg acaacgcggt ggtcatcatg 1200
aacagcatca acttggtgga cgcttcctc gatgcgaaca aatggatgga gctgttcccg 1260
tcgattgttt gcaaggcaag aacaatccag atcataaacc acggagctgc gtctgggtcac 1320
ctcggcagtg gaactctct cttgatgcaa gcagaggtgc agttcttgc gcctctgggtg 1380
gcggcgcgcg aggtggtgtt cttccgttac tgcgtgcaca acgccgacga ggggagctgg 1440
gcgatcgtgg atttcccggc ggaagggtt gaggaggggc tcctccaagc ctcggtggtc 1500
aggtgccggc ggcgccgctc cggctgcac atccaggaca tgccaatgg atactccagg 1560
gtggtgtggg tggagcacat ggagatggtc ggggaggaga agccgctgca gccggtgttc 1620

```

agggactacg tcgccagcgg cgccgccttc ggggccacgc gctgggtctc catcctccag 1680
cgccagtgcg agcgcctcgc cagtgcgttc gcccgtaaca tcgcgcacct cggagtgatc 1740
cgcacccctg aggcaagaac aaatatgatg aagctgtcgc agcggatgat caccactttc 1800
tgtgctaaca tcagtgcctc tggaacccaa tcttggtacg cctctcggga ttcaaccagg 1860
gacactatca gggttacaac tcggaagaac actgagcccg gacagcccag cgggtgtcatc 1920
ctcaccgctg tctccacaag ctggcttctt tttacccatc agcagggtgt tgagcttctt 1980
gcagatgaac aacagcgtg ccagcttgag attttgtcaa acggcggctc cttcatgag 2040
gtggcacata ttgcaaattg atcacatcca aggaattgca tctctcttct tcgcattaac 2100
gctgcaagca actcgtcaca gaatgtggag ctctgctgc aggagagcag caccaccct 2160
gatggtggga gcctggtggt gttcgcgacg gtggacgtgg acgcaattca ggtgacgatg 2220
agcggcgagg atccttctta catccctctc ctccccctgg gcttcgcat cttccctgcc 2280
accagcccat cacctgctgc tgcaccacc atcagctcaa gcaccaccac caccactggc 2340
aatggcaatg gcgaaaccag tagcactcct ccaagaaaca gcagcagcaa caacaacaat 2400
gcagatgagc tgctgcctcc caatggtgc ctctcaccg tcggcatgca ggtgctggcc 2460
agcgcgttac cttctgcaa gctcaacctc tcgagtgtca ctgcgatcaa cagccatgtc 2520
tgcaacgcca ttcaccagat cacagctgca ctcaaaagtt cagcagggtg tgctgggggt 2580
gagccggcgt ccgatcagta g 2601

```

<210> 126

<211> 866

<212> PRT

<213> *Oryza sativa*

<400> 126

```

Met Ala Gly Ala Ala Ser Ser Ala Asp Ala Leu Phe Ala Ser Pro Leu
1           5           10           15

Ile Pro Asn Pro Ala Leu Ala Gly Phe Met Ser Ser Ser Ala Ala Met
          20           25           30

Pro Phe His His Phe Ser Asn Ala Ala Ala Thr Leu Ile Pro Lys Glu
          35           40           45

Glu Gly Leu Met Gly Gly Leu His Val Ala Lys Asp Glu Gly Met Asp
50           55           60

```

Leu Glu Met Asp Met Glu Leu Ser Gly Gly Ser Gly Ser Ala His Leu
 65 70 75 80
 Asp Gly Leu Leu Ser Phe Ala Asp Val Asp Asp Asp His Lys Pro Gln
 85 90 95
 His Ser Gly His Asp Gln Pro Pro Asp Ala Ala Gln Pro Ser Gly Ala
 100 105 110
 Ala Gly Gly Asn Ala Lys Lys Lys Arg Tyr His Arg His Thr Ala His
 115 120 125
 Gln Ile Gln Gln Met Glu Ala Leu Phe Lys Glu Cys Pro His Pro Asp
 130 135 140
 Asp Lys Gln Arg Leu Lys Leu Ser Gln Glu Leu Gly Leu Lys Pro Arg
 145 150 155 160
 Gln Val Lys Phe Trp Phe Gln Asn Arg Arg Thr Gln Met Lys Ala Gln
 165 170 175
 Gln Asp Arg Ala Asp Asn Val Ile Leu Arg Ala Glu Asn Glu Asn Leu
 180 185 190
 Lys Ser Asp Asn Phe Arg Leu Gln Ala Ala Ile Arg Asn Val Val Cys
 195 200 205
 Pro Asn Cys Gly His Ala Ala Val Leu Ala Asp Met Ser Tyr Glu Glu
 210 215 220
 Gln Gln Leu Arg Ile Glu Asn Ala Arg Leu Lys Asp Glu Leu Asp Arg
 225 230 235 240
 Leu Ala Cys Ile Ala Thr Arg Tyr Gly Gly Gly Gly Gly Arg Gln Pro
 245 250 255
 Val Leu Ser Thr Ser Ala Leu Ser Cys Ile Ser Ala Pro Pro Pro Val
 260 265 270
 Leu Met Pro Pro Leu Asp Leu Asp Met Asn Val Tyr Ser Arg Asn Phe
 275 280 285
 Ala Glu Gln Ala Pro Val Met Gly Cys Gly Asp Leu Ile Pro Gln Pro
 290 295 300
 Val Val Pro Gln His Asp Gly Ala Ala Ala Tyr Met Gly Ala Met Met
 305 310 315 320
 Ala Pro Val Gln Glu Gln Asp Lys Gln Leu Val Val Asp Leu Ala Ala
 325 330 335
 Thr Ala Ala Asp Gln Leu Ala Arg Met Cys Arg Ala Gly Glu Pro Leu
 340 345 350

Trp Val Arg Gln Arg Gly Ala Glu Val Met Ala Val Glu Glu His Ala
 355 360 365
 Arg Met Phe Ser Trp Pro Val Asp Gly Ala Lys Gln Gly Asp Gly Gly
 370 375 380
 Ala Val Ala Arg Ala Glu Gly Thr Arg Asp Asn Ala Val Val Ile Met
 385 390 395 400
 Asn Ser Ile Asn Leu Val Asp Ala Phe Leu Asp Ala Asn Lys Trp Met
 405 410 415
 Glu Leu Phe Pro Ser Ile Val Cys Lys Ala Arg Thr Ile Gln Ile Ile
 420 425 430
 Asn His Gly Ala Ala Ser Gly His Leu Gly Ser Gly Thr Leu Leu Leu
 435 440 445
 Met Gln Ala Glu Val Gln Phe Leu Ser Pro Leu Val Ala Ala Arg Glu
 450 455 460
 Val Val Phe Phe Arg Tyr Cys Val His Asn Ala Asp Glu Gly Ser Trp
 465 470 475 480
 Ala Ile Val Asp Phe Pro Ala Glu Gly Phe Glu Glu Gly Leu Leu Gln
 485 490 495
 Ala Ser Val Val Arg Cys Arg Arg Arg Pro Ser Gly Cys Ile Ile Gln
 500 505 510
 Asp Met Pro Asn Gly Tyr Ser Arg Val Val Trp Val Glu His Met Glu
 515 520 525
 Met Val Gly Glu Glu Lys Pro Leu Gln Pro Val Phe Arg Asp Tyr Val
 530 535 540
 Ala Ser Gly Ala Ala Phe Gly Ala Thr Arg Trp Leu Ser Ile Leu Gln
 545 550 555 560
 Arg Gln Cys Glu Arg Leu Ala Ser Glu Leu Ala Arg Asn Ile Ala Asp
 565 570 575
 Leu Gly Val Ile Arg Thr Pro Glu Ala Arg Thr Asn Met Met Lys Leu
 580 585 590
 Ser Gln Arg Met Ile Thr Thr Phe Cys Ala Asn Ile Ser Ala Ser Gly
 595 600 605
 Thr Gln Ser Trp Thr Ala Leu Ser Asp Ser Thr Gln Asp Thr Ile Arg
 610 615 620
 Val Thr Thr Arg Lys Asn Thr Glu Pro Gly Gln Pro Ser Gly Val Ile
 625 630 635 640

Leu Thr Ala Val Ser Thr Ser Trp Leu Pro Phe Thr His Gln Gln Val
 645 650 655
 Phe Glu Leu Leu Ala Asp Glu Gln Gln Arg Cys Gln Leu Glu Ile Leu
 660 665 670
 Ser Asn Gly Gly Ser Leu His Glu Val Ala His Ile Ala Asn Gly Ser
 675 680 685
 His Pro Arg Asn Cys Ile Ser Leu Leu Arg Ile Asn Ala Ala Ser Asn
 690 695 700
 Ser Ser Gln Asn Val Glu Leu Leu Leu Gln Glu Ser Ser Thr His Pro
 705 710 715 720
 Asp Gly Gly Ser Leu Val Val Phe Ala Thr Val Asp Val Asp Ala Ile
 725 730 735
 Gln Val Thr Met Ser Gly Glu Asp Pro Ser Tyr Ile Pro Leu Leu Pro
 740 745 750
 Leu Gly Phe Ala Ile Phe Pro Ala Thr Ser Pro Ser Pro Ala Ala Ala
 755 760 765
 Pro Thr Ile Ser Ser Ser Thr Thr Thr Thr Thr Gly Asn Gly Asn Gly
 770 775 780
 Glu Thr Ser Ser Thr Pro Pro Arg Asn Ser Ser Ser Asn Asn Asn Asn
 785 790 795 800
 Ala Asp Glu Leu Leu Pro Pro Asn Gly Cys Leu Leu Thr Val Gly Met
 805 810 815
 Gln Val Leu Ala Ser Ala Val Pro Ser Ala Lys Leu Asn Leu Ser Ser
 820 825 830
 Val Thr Ala Ile Asn Ser His Val Cys Asn Ala Ile His Gln Ile Thr
 835 840 845
 Ala Ala Leu Lys Ser Ser Ala Gly Gly Ala Gly Gly Glu Pro Ala Ser
 850 855 860
 Asp Gln
 865

<210> 127
 <211> 2451
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (49)..(49)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (51)..(51)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (699)..(699)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (708)..(708)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (1858)..(1858)

<223> n = any nucleotide

<220>

<221> misc_feature

<222> (1885)..(1885)

<223> n = any nucleotide

<400> 127

atgcagttcc ctttcgccag tggcttcgcg tcctcgccgg cgttttccnt ngcgctggta 60

tgcgttcggc ggctgtcgcg gaccgggggt gcgtcagctg ttcgtagaca cgagagtttt 120

ctcctcgatc gagaggacaa cgccggcggg ggcacggcg ggcggatgct cggcggcggc 180

gctggcgctg gttctagcgc gggcgggcgc atgacgaggg aactgaggc ggagaacgac 240

agccggtcgg gaagcgacca tctcgacgcc atctcgcccg ccggcgagga cgacgtcgag 300

gacgccgagc ccagcaactc ccgcaagagg aagaagcgat accaccgcca cacgccgcag 360

cagatccaag agctcgaagc gctgttcaag gagtgcctcc atccggacga gaagcagcgc 420

gccgagctca gcaggaggct ctccctcgat gcgcggcagg tcaagttctg gttccagaat 480

cgccgcacgc agatgaagac gcaactggag cggcacgaga acgcgcttct caagcaggag 540

aacgacaagc tgcgcgccga gaacatgacg atccgggagg cgatgcgcag cccaatgtgc 600

ggcagctgcg ggagccccgc catgctcggg gaggtgtccc tagaggagca gcacctgcgc 660

atcgagaatg	cgcgccctcaa	ggacgagctc	aaccgcgtnt	gcgccctngc	caccaagttc	720
ctaggcaagc	ccatctccct	cctgtcgccg	ccgcccttgc	tgcagccgca	cctctcgctg	780
cctatgccta	actcgtoget	ggagctagcg	atcgagggca	tcggtggctt	agggctctctc	840
ggtacactac	ccggctgtat	gaatgagttc	gccggaggcg	tgtcgagccc	gatgggcacg	900
gtgatcacgc	ctgcgcgcgc	aaccggggct	gctataccat	cgctgggtggg	taacatcgac	960
aggtcctgtg	tcttagagct	tgcaatcagc	gcgatggacg	aacttgtcaa	gatggcacag	1020
atggacgata	cattgtgggt	tccggcactg	cctggttctc	ccagcaagga	ggtgctcaac	1080
tttgaggagt	atcttactc	cttctaccg	tgcatggca	tgaagcctgc	cggtacgctc	1140
tctgaggcct	cgagggaatc	cggctctagc	atcatcgaca	acagccttgc	ccttgtagag	1200
accctcatgg	atgagagacg	gtggtctgac	atgttctcgt	gcatgattgc	taaggcaaca	1260
gtgcttgagg	agggtctctac	cggcattgca	ggaagcagaa	atggcgcggt	gctgctgatg	1320
aaggctgagc	tacaggtgct	ctcacctcta	gtaccaatta	gggagggttac	atctctccgg	1380
ttttgtaagc	agctggctga	gggtgcatgg	gcagtagttg	atgtgtccat	tgatgggttta	1440
gtgagagata	ataattctgg	aacggcaccc	acaggtggaa	atgtgaagtg	taggagggta	1500
ccttccgggt	gtgtgatgca	agacactccc	aatgggtatt	gcaagggtcac	atgggttgag	1560
catacggaat	acgatgaggc	atcagtgac	cagctctacc	gtccactcct	tcggtctggt	1620
ctcgcccttg	gcgccaggcg	ttggcttgcg	acgctgcagc	gccaatgcga	atgccttgcc	1680
atcctcatgt	cctctgctac	agtgcagca	aatgactcga	ctgctatata	gcaagagggc	1740
aagcgaagca	tgctgaagct	ggcacgccgg	atgacggaga	acttctgtgc	tggggtgagc	1800
gcatcgtctg	cacgtgaatg	gagcaagctg	gatggtgcga	ccggtagcat	cggggagnac	1860
gtgcgcgtga	tggcacggaa	gagcntgagc	gagcctggag	agccaccggg	cgtgggtgctg	1920
agcgcggcca	cctcggtgtg	ggtgcccgctg	gctccggaga	agctcttcaa	cttctcgcc	1980
gatgagcagc	tgcgtgcgga	gtgggacatc	ctcagcaatg	gaggcccat	gcaggagatg	2040
accagatcg	ccaaggggca	gcgggatggg	aactcggttt	cactcctcag	ggccagtgt	2100
gtgagtgcc	accagagcag	catgctgata	cttcaggaga	cgtgcaccga	cgcttctggc	2160
tccattgttg	tgtacgctcc	ggtagacatc	ccggcaatgc	agcttgtcat	gaatgggtggc	2220
gactccacct	atgttgctct	gcttccatcc	ggattcgcca	tcctgcccga	tgggcctcgc	2280

atcggcgcca cggggtacga gacgggcggc tctctgctca ccgtggcggt ccagattctt 2340
 gtcaacaacc agcccaccgc caagctcacc gtggaatcgg tcgagaccgt gaacaatctc 2400
 atctcctgca ccatcaagaa gatcaagacg gcgctgcagt gcgacgcctg a 2451

<210> 128
 <211> 816
 <212> PRT
 <213> Oryza sativa

<400> 128

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

Leu Gly Glu Val Ser Leu Glu Glu Gln His Leu Arg Ile Glu Asn Ala
 210 215 220
 Arg Leu Lys Asp Glu Leu Asn Arg Val Cys Ala Leu Ala Thr Lys Phe
 225 230 235 240
 Leu Gly Lys Pro Ile Ser Leu Leu Ser Pro Pro Pro Leu Leu Gln Pro
 245 250 255
 His Leu Ser Leu Pro Met Pro Asn Ser Ser Leu Glu Leu Ala Ile Gly
 260 265 270
 Gly Ile Gly Gly Leu Gly Ser Leu Gly Thr Leu Pro Gly Cys Met Asn
 275 280 285
 Glu Phe Ala Gly Gly Val Ser Ser Pro Met Gly Thr Val Ile Thr Pro
 290 295 300
 Ala Arg Ala Thr Gly Ala Ala Ile Pro Ser Leu Val Gly Asn Ile Asp
 305 310 315 320
 Arg Ser Val Phe Leu Glu Leu Ala Ile Ser Ala Met Asp Glu Leu Val
 325 330 335
 Lys Met Ala Gln Met Asp Asp Pro Leu Trp Val Pro Ala Leu Pro Gly
 340 345 350
 Ser Pro Ser Lys Glu Val Leu Asn Phe Glu Glu Tyr Leu His Ser Phe
 355 360 365
 Leu Pro Cys Ile Gly Met Lys Pro Ala Gly Tyr Val Ser Glu Ala Ser
 370 375 380
 Arg Glu Ser Gly Leu Val Ile Ile Asp Asn Ser Leu Ala Leu Val Glu
 385 390 395 400
 Thr Leu Met Asp Glu Arg Arg Trp Ser Asp Met Phe Ser Cys Met Ile
 405 410 415
 Ala Lys Ala Thr Val Leu Glu Glu Val Ser Thr Gly Ile Ala Gly Ser
 420 425 430
 Arg Asn Gly Ala Leu Leu Leu Met Lys Ala Glu Leu Gln Val Leu Ser
 435 440 445
 Pro Leu Val Pro Ile Arg Glu Val Thr Phe Leu Arg Phe Cys Lys Gln
 450 455 460
 Leu Ala Glu Gly Ala Trp Ala Val Val Asp Val Ser Ile Asp Gly Leu
 465 470 475 480
 Val Arg Asp His Asn Ser Gly Thr Ala Pro Thr Gly Gly Asn Val Lys
 485 490 495

Cys Arg Arg Val Pro Ser Gly Cys Val Met Gln Asp Thr Pro Asn Gly
 500 505 510
 Tyr Cys Lys Val Thr Trp Val Glu His Thr Glu Tyr Asp Glu Ala Ser
 515 520 525
 Val His Gln Leu Tyr Arg Pro Leu Leu Arg Ser Gly Leu Ala Phe Gly
 530 535 540
 Ala Arg Arg Trp Leu Ala Thr Leu Gln Arg Gln Cys Glu Cys Leu Ala
 545 550 555 560
 Ile Leu Met Ser Ser Ala Thr Val Thr Ala Asn Asp Ser Thr Ala Ile
 565 570 575
 Ser Gln Glu Gly Lys Arg Ser Met Leu Lys Leu Ala Arg Arg Met Thr
 580 585 590
 Glu Asn Phe Cys Ala Gly Val Ser Ala Ser Ser Ala Arg Glu Trp Ser
 595 600 605
 Lys Leu Asp Gly Ala Thr Gly Ser Ile Gly Glu His Val Arg Val Met
 610 615 620
 Ala Arg Lys Ser Leu Ser Glu Pro Gly Glu Pro Pro Gly Val Val Leu
 625 630 635 640
 Ser Ala Ala Thr Ser Val Trp Val Pro Val Ala Pro Glu Lys Leu Phe
 645 650 655
 Asn Phe Leu Arg Asp Glu Gln Leu Arg Ala Glu Trp Asp Ile Leu Ser
 660 665 670
 Asn Gly Gly Pro Met Gln Glu Met Thr Gln Ile Ala Lys Gly Gln Arg
 675 680 685
 Asp Gly Asn Ser Val Ser Leu Leu Arg Ala Ser Ala Val Ser Ala Asn
 690 695 700
 Gln Ser Ser Met Leu Ile Leu Gln Glu Thr Cys Thr Asp Ala Ser Gly
 705 710 715 720
 Ser Ile Val Val Tyr Ala Pro Val Asp Ile Pro Ala Met Gln Leu Val
 725 730 735
 Met Asn Gly Gly Asp Ser Thr Tyr Val Ala Leu Leu Pro Ser Gly Phe
 740 745 750
 Ala Ile Leu Pro Asp Gly Pro Arg Ile Gly Ala Thr Gly Tyr Glu Thr
 755 760 765
 Gly Gly Ser Leu Leu Thr Val Ala Phe Gln Ile Leu Val Asn Asn Gln
 770 775 780

Pro Thr Ala Lys Leu Thr Val Glu Ser Val Glu Thr Val Asn Asn Leu
 785 790 795 800

Ile Ser Cys Thr Ile Lys Lys Ile Lys Thr Ala Leu Gln Cys Asp Ala
 805 810 815

<210> 129
 <211> 1977
 <212> DNA
 <213> Oryza sativa

<220>
 <221> misc_feature
 <222> (18)..(18)
 <223> n = any nucleotide

<220>
 <221> misc_feature
 <222> (26)..(26)
 <223> n = any nucleotide

<220>
 <221> misc_feature
 <222> (1864)..(1864)
 <223> n = any nucleotide

<400> 129
 atgcacgcgg cggcgacngt tccacngacc ctctccggc cccgcgcccg gttctcgcg 60
 cggcgggggac ggccgattcg gccgcggccc ttctcgctgg tcgtccggcg aagcggcggc 120
 ggcggcgagt acggggagcc gggtagcggg ccgagacacc gctacggcag tggccgtggt 180
 gaccactcag atcatgataa cagaaacaac tatgttaaac tttttattgg gtcagttcca 240
 aggacagcga ctgaggatga tgtacgtcct ttatttgagg agcatggaga tgttggtgaa 300
 gttgctttga tcaaggatag gaagactggt gaacagcaag gttggtgttt tgttaaatat 360
 gctacttctg aagaggctga gcgagccatc agagctcttc acaaccagta cactttacct 420
 ggggcaatgg gccctattca agtcagatat gctgatggtg aaaggggaacg ccatgggggt 480
 attgagcaca agctgtttgt cgcacactc aataagcaag ctactgcaa ggagattgaa 540
 gagatTTTTG cgccgatggt tcatgtggaa gatgtctaca ttatgaaaga tggcatgagg 600
 cagagccgag gttgtggctt tgtcaaattc tcatcacgag aaccggcgct ggcagccatg 660
 agtgctctga gtgggaatta tgtaatgagg ggggtgcgagc aaccattaat aattcggttc 720

gctgatccta agagacctag acctggagaa tcaaggggtg gtccctgcatt tggagggtcct 780
 ggtttcagtc cacgatctga tgcagctcct gttatcaggc cgaccgcaa ccttgatgaa 840
 ccaaggggtc ggcatatgcc tctgactct tggcatccat caagcccaag gtcagcacct 900
 catcagttta ataactttgg atctgacaat cctatggctc cgaaagggag tactgttaca 960
 tcaacaacag atacgatgtt cagcggaaat ggctctttgt caagccaaac agctgtgcca 1020
 agttcgtcac atatgggcat gaaccctcct cccatggcac aaggccatca tctgggaggc 1080
 cagcagatcc cacctttgca aaagctacca gggttgcctc agaatttccc tgttcaattg 1140
 cagaataatc agctgggaca acccttgagc ggtcctgctc aacagattgg tcagcttcaa 1200
 gttccacaat ctatgggtcc tggttctttt ggccagaata gggtatctgg tcagcttctc 1260
 gtgtcacagc cgttgatgca acaaaatgct tccgtcagcg ctgtgcaggc gccttcagcc 1320
 gtgtccaatt ccatgcaagc tattcccggg caacagcatc ttccctcaa tgttgcacca 1380
 caaatgcttc agcagccagt gcagcagatg ccatgcaag caccacaatt gctactcaa 1440
 cagcaagctg ctttgcaatc cagttatcaa tcttcacagc aggcaattta tcagcttcaa 1500
 caacagctgc aactaatgca gcagcagcag caatcgaatc taaaccatca acaaccaca 1560
 cagggtcagc cagtgaatc tagtaacct ggtgccccaa atgccattat tccatcaaac 1620
 attaatacaa ttccccagca ggctacttca cctgcagttc ctttaacctg caattggact 1680
 gaacatacgt ctctgaagg ttttaaatac tactacaata gcataactcg agagagtaag 1740
 tgggataagc ctgaagagta tgtgctgtat gagcaacagc aacagcagca gcaacagcag 1800
 aaactacttt tgcttcaaca gcaccaacaa aaacttgcca tgcagcaact tcagtcacca 1860
 cctnaggctc agacacatcc agcaatgcaa cctgtccaac aaattcctca ggcacaacaa 1920
 ggacagcaac aaatgcagat gaaacagcag gtgcacatgc tgttcctga gctataa 1977

<210> 130

<211> 658

<212> PRT

<213> *Oryza sativa*

<220>

<221> X_region

<222> (6)..(6)

<223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (9)..(9)
 <223> Xaa = any naturally occurring amino acid

<220>
 <221> X_region
 <222> (622)..(622)
 <223> Xaa = any naturally occurring amino acid

<400> 130

Met His Ala Ala Ala Xaa Val Pro Xaa Thr Leu Leu Arg Pro Arg Ala
 1 5 10 15

Arg Phe Ser Arg Arg Arg Gly Arg Pro Ile Arg Pro Arg Pro Phe Ser
 20 25 30

Leu Val Val Arg Arg Ser Gly Gly Gly Gly Glu Tyr Gly Glu Pro Gly
 35 40 45

Ser Gly Pro Arg His Arg Tyr Gly Ser Gly Arg Gly Asp His Ser Asp
 50 55 60

His Asp Asn Arg Asn Asn Tyr Val Lys Leu Phe Ile Gly Ser Val Pro
 65 70 75 80

Arg Thr Ala Thr Glu Asp Asp Val Arg Pro Leu Phe Glu Glu His Gly
 85 90 95

Asp Val Val Glu Val Ala Leu Ile Lys Asp Arg Lys Thr Gly Glu Gln
 100 105 110

Gln Gly Cys Cys Phe Val Lys Tyr Ala Thr Ser Glu Glu Ala Glu Arg
 115 120 125

Ala Ile Arg Ala Leu His Asn Gln Tyr Thr Leu Pro Gly Ala Met Gly
 130 135 140

Pro Ile Gln Val Arg Tyr Ala Asp Gly Glu Arg Glu Arg His Gly Ala
 145 150 155 160

Ile Glu His Lys Leu Phe Val Ala Ser Leu Asn Lys Gln Ala Thr Ala
 165 170 175

Lys Glu Ile Glu Glu Ile Phe Ala Pro Tyr Gly His Val Glu Asp Val
 180 185 190

Tyr Ile Met Lys Asp Gly Met Arg Gln Ser Arg Gly Cys Gly Phe Val
 195 200 205

Lys Phe Ser Ser Arg Glu Pro Ala Leu Ala Ala Met Ser Ala Leu Ser
 210 215 220

Gly Asn Tyr Val Met Arg Gly Cys Glu Gln Pro Leu Ile Ile Arg Phe
 225 230 235 240
 Ala Asp Pro Lys Arg Pro Arg Pro Gly Glu Ser Arg Gly Gly Pro Ala
 245 250 255
 Phe Gly Gly Pro Gly Phe Ser Pro Arg Ser Asp Ala Ala Leu Val Ile
 260 265 270
 Arg Pro Thr Ala Asn Leu Asp Glu Pro Arg Gly Arg His Met Pro Pro
 275 280 285
 Asp Ser Trp His Pro Ser Ser Pro Arg Ser Ala Pro His Gln Phe Asn
 290 295 300
 Asn Phe Gly Ser Asp Asn Pro Met Ala Pro Lys Gly Ser Thr Val Thr
 305 310 315 320
 Ser Thr Thr Asp Thr Met Phe Ser Gly Asn Gly Ser Leu Ser Ser Gln
 325 330 335
 Thr Ala Val Pro Ser Ser Ser His Met Gly Met Asn Pro Pro Pro Met
 340 345 350
 Ala Gln Gly His His Leu Gly Gly Gln Gln Ile Pro Pro Leu Gln Lys
 355 360 365
 Leu Pro Gly Leu Pro Gln Asn Phe Pro Val Gln Leu Gln Asn Asn Gln
 370 375 380
 Leu Gly Gln Pro Leu Gln Gly Pro Ala Gln Gln Ile Gly Gln Leu Gln
 385 390 395 400
 Val Pro Gln Ser Met Gly Pro Gly Ser Phe Gly Gln Asn Arg Leu Ser
 405 410 415
 Gly Gln Leu Pro Val Ser Gln Pro Leu Met Gln Gln Asn Ala Ser Val
 420 425 430
 Ser Ala Val Gln Val Pro Ser Ala Val Ser Asn Ser Met Gln Ala Ile
 435 440 445
 Pro Gly Gln Gln His Leu Pro Ser Asn Val Ala Pro Gln Met Leu Gln
 450 455 460
 Gln Pro Val Gln Gln Met Pro Ser Gln Ala Pro Gln Leu Leu Leu Gln
 465 470 475 480
 Gln Gln Ala Ala Leu Gln Ser Ser Tyr Gln Ser Ser Gln Gln Ala Ile
 485 490 495
 Tyr Gln Leu Gln Gln Gln Leu Gln Leu Met Gln Gln Gln Gln Gln Ser
 500 505 510

Asn Leu Asn His Gln Gln Pro Thr Gln Gly Gln Pro Val Gln Ser Ser
 515 520 525
 Asn Pro Gly Ala Pro Asn Ala Ile Ile Pro Ser Asn Ile Asn Thr Ile
 530 535 540
 Pro Gln Gln Ala Thr Ser Pro Ala Val Pro Leu Thr Cys Asn Trp Thr
 545 550 555 560
 Glu His Thr Ser Pro Glu Gly Phe Lys Tyr Tyr Tyr Asn Ser Ile Thr
 565 570 575
 Arg Glu Ser Lys Trp Asp Lys Pro Glu Glu Tyr Val Leu Tyr Glu Gln
 580 585 590
 Gln Gln Gln Gln Gln Gln Gln Gln Lys Leu Leu Leu Leu Gln Gln His
 595 600 605
 Gln Gln Lys Leu Ala Met Gln Gln Leu Gln Ser Pro Pro Xaa Ala Gln
 610 615 620
 Thr His Pro Ala Met Gln Pro Val Gln Gln Ile Pro Gln Ala Gln Gln
 625 630 635 640
 Gly Gln Gln Gln Met Gln Met Lys Gln Gln Val His Met Leu Phe Pro
 645 650 655
 Glu Leu

<210> 131

<211> 984

<212> DNA

<213> *Oryza sativa*

<400> 131

atggggaggg gcaggattga gatcaagagg atcgagaaca cgacaagccg ccagggtgacc	60
ttctgcaagc gccgcaacgg acttctcaag aaggcgtatg agctctccgt cctctgcgat	120
gccgaggtgg ctctcatcgt cttctccagc cgtggccgcc tctacgagta ctccaacaac	180
aacaactcgt gtataactcat ttcatctcgt tcttggttgc acttcaacac agttcttgca	240
gctagagaca gcatgcattc caatggctac ctttcttctg cacaatttta tgctgaagta	300
attaaacgct tgctcctcac atgcattatt tctgcattca tgacatgtat agtcattgtg	360
aaggctacaa ttgacaggta caagaaggcg catgcttgtg gctcaacttc tgggtgcacct	420
ctcatagagg tcaatgctca gcaatactac cagcaggagt ctgccaaact gcgccaccag	480

attcagatgc tgcaaaacac caacaagcac ctggttggcg ataatgtgag caacctgtca 540
 ctgaaggagc tgaagcaact tgaaagccgc ctggagaaag gcatttcaaa gatcagagcc 600
 aggaagaatg aactgctggc ttcagagatc aattacatgg ccaaagggga gattgagctt 660
 cagaacgaca acatggacct cagaaccaag attgctgagg aggagcagca gctgcagcag 720
 gtgacggtgg cccggtcggc cgccatggag ctgcaggctg cggcggcggc gcagcagcag 780
 cagcagaatc cgttcgcggt ggcggcggcg cagctggaca tgaagtgctt cttcccgttg 840
 aacctgttcg aggcggcggc gcaggtgcag gccgtggcgg cgcagcgcca gcagatcatc 900
 cccaccgagc tcaacctcgg ctaccaccac caccttgcca ttcccggcgc cgccgccgcc 960
 gacgcgccgc ctccctcactt ctga 984

<210> 132

<211> 327

<212> PRT

<213> *Oryza sativa*

<400> 132

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

145 150 155 160
 Ile Gln Met Leu Gln Asn Thr Asn Lys His Leu Val Gly Asp Asn Val
 165 170 175
 Ser Asn Leu Ser Leu Lys Glu Leu Lys Gln Leu Glu Ser Arg Leu Glu
 180 185 190
 Lys Gly Ile Ser Lys Ile Arg Ala Arg Lys Asn Glu Leu Leu Ala Ser
 195 200 205
 Glu Ile Asn Tyr Met Ala Lys Arg Glu Ile Glu Leu Gln Asn Asp Asn
 210 215 220
 Met Asp Leu Arg Thr Lys Ile Ala Glu Glu Glu Gln Gln Leu Gln Gln
 225 230 235 240
 Val Thr Val Ala Arg Ser Ala Ala Met Glu Leu Gln Ala Ala Ala Ala
 245 250 255
 Ala Gln Gln Gln Gln Gln Asn Pro Phe Ala Val Ala Ala Ala Gln Leu
 260 265 270
 Asp Met Lys Cys Phe Phe Pro Leu Asn Leu Phe Glu Ala Ala Ala Gln
 275 280 285
 Val Gln Ala Val Ala Ala Gln Arg Gln Gln Ile Ile Pro Thr Glu Leu
 290 295 300
 Asn Leu Gly Tyr His His His Leu Ala Ile Pro Gly Ala Ala Ala Ala
 305 310 315 320
 Asp Ala Pro Pro Pro His Phe
 325

<210> 1723

<211> 765

<212> DNA

<213> Oryza sativa

<400> 133

atggggaggg gaagagttga gctgaagcgc atcgagaaca agatcaacag gcaggtcacc 60
 ttctccaagc gccgcaacgg cctcctcaag aaggcctacg agctgtccgt tctctgcgac 120
 gccgaggtcg cgctcatcat cttctccagc cgcggaacg tctacgagtt cggcagcgcc 180
 gggtgtttcg tactagctca acttgagat cttttcatta tttgcctagt tgggtggcata 240
 acaaagactt tagaaaggta ccaacattgt tgctacaatg ctcaagattc caacaatgca 300
 ctttctgaaa ccagagttg gtaccatgaa atgtcaaagt tgaaagcaaa atttgaagct 360
 ttgcagcgca ctcaaaggca cttgcttggg gaggatcttg gaccactcag cgtcaaagaa 420

```

ttgcagcagc tggagaaaca gcttgaatgt gcactatcac aggcgagaca gagaaagacg      480
caactgatga tggaacaagt ggaggaactt cgcagaaagc tcgaggttga aggttccacc      540
agcaactaca gagccatgca gcaagcctcc tgggctcagg gcgccgtggt ggagaatggc      600
gccgcatacg tgcagccgcc gccacactcc gcggccatgg actctgaacc caccttgcaa      660
attgggtatc ctcatcaatt tgtgcctgct gaagcaaaca ctattcagag gagcactgcc      720
cctgcaggtg cagagaacaa cttcatgctg ggatgggttc ttgga                        765

```

<210> 134
 <211> 254
 <212> PRT
 <213> *Oryza sativa*

<400> 1724

```

Met Gly Arg Gly Arg Val Glu Leu Lys Arg Ile Glu Asn Lys Ile Asn
1              5              10              15

Arg Gln Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Lys Ala
                20              25              30

Tyr Glu Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe
        35              40              45

Ser Ser Arg Gly Lys Leu Tyr Glu Phe Gly Ser Ala Gly Cys Phe Val
        50              55              60

Leu Ala Gln Leu Gly Asp Leu Phe Ile Ile Cys Leu Val Gly Gly Ile
65              70              75              80

Thr Lys Thr Leu Glu Arg Tyr Gln His Cys Cys Tyr Asn Ala Gln Asp
                85              90              95

Ser Asn Asn Ala Leu Ser Glu Thr Gln Ser Trp Tyr His Glu Met Ser
                100             105             110

Lys Leu Lys Ala Lys Phe Glu Ala Leu Gln Arg Thr Gln Arg His Leu
        115              120              125

Leu Gly Glu Asp Leu Gly Pro Leu Ser Val Lys Glu Leu Gln Gln Leu
        130              135              140

Glu Lys Gln Leu Glu Cys Ala Leu Ser Gln Ala Arg Gln Arg Lys Thr
145              150              155              160

Gln Leu Met Met Glu Gln Val Glu Glu Leu Arg Arg Lys Leu Glu Val
                165              170              175

```

Glu Gly Ser Thr Ser Asn Tyr Arg Ala Met Gln Gln Ala Ser Trp Ala
 180 185 190

Gln Gly Ala Val Val Glu Asn Gly Ala Ala Tyr Val Gln Pro Pro Pro
 195 200 205

His Ser Ala Ala Met Asp Ser Glu Pro Thr Leu Gln Ile Gly Tyr Pro
 210 215 220

His Gln Phe Val Pro Ala Glu Ala Asn Thr Ile Gln Arg Ser Thr Ala
 225 230 235 240

Pro Ala Gly Ala Glu Asn Asn Phe Met Leu Gly Trp Val Leu
 245 250

<210> 135

<211> 1224

<212> DNA

<213> Oryza sativa

<400> 135

atggatgctc tttgtgattt ctgcagggag caaagggtcaa tgggtctattg ccgatcagat	60
gcagcatcct tgtgcttgct gtgtgatcgt aatgtgcatt cagccaatgc cctttcccga	120
cgccatacaa gaacccttct ttgtgatcgt tgtgttgga agcctgcagc agtccgttgt	180
cttgaagaga acacctcact ttgccaaaac tgtgactgga atggacatgg ggcagcatcc	240
tcagctgctg ggcataaaag gcagaccata aactgttact cggggtgccc atcgctcgga	300
gagctctcga gaatctggct ctttagtatg gatatoccaa ctgtagctgc ggagccaaac	360
tgtgaggaag gaataaacat gatgagcatt aatgacaatg acgtcaataa ccatttgtgt	420
gctccagaag atggccgctt gttggatata gctagcacag cactcatgag tgatttacct	480
acaggagaca agttcaaacc tttaataggt tcttcttcag gagatgggat gaatcttctg	540
ccacttaatt cagatcagcc agctgagcca gtttcaacga cgcctaaggc accctgtgtc	600
acagacaaag atatgttcaa tgatgggagc gtatatggag acttctgtgt ggatgatgct	660
gacctaacat ttgagaatta tgaagaatta ttcggtacct ctcacgttca aacagaacaa	720
ctctttgatg atgctggaat tgacagttat tttgaaatga aggatgtgcc agcagatgaa	780
tctaatgagc agcccaaacc tgtgcagcca gaatgtagca atgtcgcatc tgttgattct	840
gggatgtcaa acccagctgc aagggtgat tccagccatt gtattcctgg taggcaggct	900
atatccaaca tatccctttc cttctctggt ttgactggtg aaagcagtgc tggatatttt	960


```

caagattgtg gggatcgtc aatgatacctt atgggtgagc caccctggca tcctcctggt 1020
cctgaaagtt catctgctgg aggcagcaga gataatgctc ttacacggta caaggagaag 1080
aagaagagaa gaaaatttga caagaagata aggtatgctt ctcgtaaggc tagggcagat 1140
gtgagaaaaa ggggtcaaagg acggtttgta aaggctggcg aagcatatga ctacgatccg 1200
ctgagccaaa caagaagcta ctaa 1224

```

```

<210> 136
<211> 407
<212> PRT
<213> Oryza sativa

```

```

<400> 136

```

```

Met Asp Ala Leu Cys Asp Phe Cys Arg Glu Gln Arg Ser Met Val Tyr
1          5          10          15

Cys Arg Ser Asp Ala Ala Ser Leu Cys Leu Ser Cys Asp Arg Asn Val
          20          25          30

His Ser Ala Asn Ala Leu Ser Arg Arg His Thr Arg Thr Leu Leu Cys
          35          40          45

Asp Arg Cys Val Gly Gln Pro Ala Ala Val Arg Cys Leu Glu Glu Asn
50          55          60

Thr Ser Leu Cys Gln Asn Cys Asp Trp Asn Gly His Gly Ala Ala Ser
65          70          75          80

Ser Ala Ala Gly His Lys Arg Gln Thr Ile Asn Cys Tyr Ser Gly Cys
          85          90          95

Pro Ser Ser Ala Glu Leu Ser Arg Ile Trp Ser Phe Ser Met Asp Ile
          100          105          110

Pro Thr Val Ala Ala Glu Pro Asn Cys Glu Glu Gly Ile Asn Met Met
          115          120          125

Ser Ile Asn Asp Asn Asp Val Asn Asn His Cys Gly Ala Pro Glu Asp
          130          135          140

Gly Arg Leu Leu Asp Ile Ala Ser Thr Ala Leu Met Ser Asp Leu Pro
145          150          155          160

Thr Gly Asp Lys Phe Lys Pro Leu Ile Gly Ser Ser Ser Gly Asp Gly
          165          170          175

Met Asn Leu Leu Pro Leu Asn Ser Asp Gln Pro Ala Glu Pro Val Ser
          180          185          190

```

Thr Thr Pro Lys Ala Pro Cys Val Thr Asp Lys Asp Met Phe Asn Asp
195 200 205

Gly Ser Val Tyr Gly Asp Phe Cys Val Asp Asp Ala Asp Leu Thr Phe
210 215 220

Glu Asn Tyr Glu Glu Leu Phe Gly Thr Ser His Val Gln Thr Glu Gln
225 230 235 240

Leu Phe Asp Asp Ala Gly Ile Asp Ser Tyr Phe Glu Met Lys Asp Val
245 250 255

Pro Ala Asp Glu Ser Asn Glu Gln Pro Lys Pro Val Gln Pro Glu Cys
260 265 270

Ser Asn Val Ala Ser Val Asp Ser Gly Met Ser Asn Pro Ala Ala Arg
275 280 285

Ala Asp Ser Ser His Cys Ile Pro Gly Arg Gln Ala Ile Ser Asn Ile
290 295 300

Ser Leu Ser Phe Ser Gly Leu Thr Gly Glu Ser Ser Ala Gly Tyr Phe
305 310 315 320

Gln Asp Cys Gly Val Ser Ser Met Ile Leu Met Gly Glu Pro Pro Trp
325 330 335

His Pro Pro Gly Pro Glu Ser Ser Ser Ala Gly Gly Ser Arg Asp Asn
340 345 350

Ala Leu Thr Arg Tyr Lys Glu Lys Lys Lys Arg Arg Lys Phe Asp Lys
355 360 365

Lys Ile Arg Tyr Ala Ser Arg Lys Ala Arg Ala Asp Val Arg Lys Arg
370 375 380

Val Lys Gly Arg Phe Val Lys Ala Gly Glu Ala Tyr Asp Tyr Asp Pro
385 390 395 400

Leu Ser Gln Thr Arg Ser Tyr
405

<210> 137

<211> 555

<212> DNA

<213> Oryza sativa

<400> 137

atgtctaggg acccattggt tgtcgggtcat gtcgtcggcg atatcgtgga cccgttcgtc 60

accaccgctt cgcttagggt cttctacaac agcaaggaga tgacaaatgg gtctgagctc 120

aagccatctc aggtgttgaa ccaaccaagg atttatatcg aaggtcgcga catgaggacg 180

ctctacacgc ttgtaatggg ggaccctgat gcaccaagcc ccagcaaccc tactaaaaga 240
gagtaccttc attggatggg gacagacatt ccagagacca ctgatgccag atttggtaat 300
gagattgtcc cctatgagag cccacgccc actgcaggca tccatcgctt cgtgttcac 360
ctattcaggg agtcagtcag gcagaccacc tatgcaccag ggtggcgcca aaacttcaat 420
acaagggact ttgctgagct ctacaacctc ggttcgccgg tcgccgcgct cttcttcaac 480
tgccagaggg agaacggctg tggaggaaga agaaagcttg agctgcgcat tgttccaatt 540
gactacatgg tttga 555

<210> 138

<211> 184

<212> PRT

<213> Oryza sativa

<400> 138

Met Ser Arg Asp Pro Leu Val Val Gly His Val Val Gly Asp Ile Val
1 5 10 15

Asp Pro Phe Val Thr Thr Ala Ser Leu Arg Val Phe Tyr Asn Ser Lys
20 25 30

Glu Met Thr Asn Gly Ser Glu Leu Lys Pro Ser Gln Val Leu Asn Gln
35 40 45

Pro Arg Ile Tyr Ile Glu Gly Arg Asp Met Arg Thr Leu Tyr Thr Leu
50 55 60

Val Met Val Asp Pro Asp Ala Pro Ser Pro Ser Asn Pro Thr Lys Arg
65 70 75 80

Glu Tyr Leu His Trp Met Val Thr Asp Ile Pro Glu Thr Thr Asp Ala
85 90 95

Arg Phe Gly Asn Glu Ile Val Pro Tyr Glu Ser Pro Arg Pro Thr Ala
100 105 110

Gly Ile His Arg Phe Val Phe Ile Leu Phe Arg Gln Ser Val Arg Gln
115 120 125

Thr Thr Tyr Ala Pro Gly Trp Arg Gln Asn Phe Asn Thr Arg Asp Phe
130 135 140

Ala Glu Leu Tyr Asn Leu Gly Ser Pro Val Ala Ala Leu Phe Phe Asn
145 150 155 160

Cys Gln Arg Glu Asn Gly Cys Gly Gly Arg Arg Lys Leu Glu Leu Arg

165 170 175

Ile Val Pro Ile Asp Tyr Met Val
180

<210> 139
<211> 1821
<212> DNA
<213> Oryza sativa

<220>
<221> misc_feature
<222> (1262)..(1262)
<223> n = any nucleotide

<400> 139
atggagaagt atatgagggg atctagaatg actcttgagg agaaggttac tttcatgcct 60
tttcagttga ctgggtcaggc ggacatttgg tgggatggag ttgttgaaga gtggaaccct 120
gaccgtgggtg ctattacttg ggaagttttt cttcagcagt tcagagctaa gtattatcca 180
gaatcattta gggacaagat gagtaatgag ttaaacaata tccaacaggg aagcaagact 240
gttgatgagt atgaacggga gttcaacagt attgtacgct ttgtgcctaa cgtggcaagt 300
gatgagcgtg agaaggctag aaagttttct cgagggctca ttgcccgata tagagaagtg 360
atgggaagaa acccacctac tatatacctc aatgcagttg aggaggctcg agggatggag 420
gttgagctcc aaattactac tgcacagaag aatcgttctg gcaatgtcac tgggaaaggg 480
ctattttgc ttaagtgtgg aaagaatcac cattcagcag agtgtggggt tactgaggca 540
tgtcgtgggt gtggcaaggc tggacacatg gagttgatga tgacagetta tccacttct 600
taccagcagt tcacacataa ttaccaccca ccaccattac cgggacagta ttggtccatg 660
cctcaatctc ttgctccacg tgcttcactc cagttgcctg catgcctat ggtgcaacct 720
tctgcttcag ctccacctca ccagcaagcc tcagggtcta ctacccagc tatggctgga 780
gtctacaaca tgctgcagc ttcattccctt ggtcgcccgg atgtcgttgc agatacactt 840
tcagttgatt tgcatgatgc tctagttcta tttgattctg gtgctacatt ttcctttgtt 900
tcaattgagt ttgcgaaaag agctaattcta tttagtcaag aggtctctcg atctgtccaa 960
gtcagctctc cagggggcga ggagatgggt tttcaaggga gcgctcctcc gcattctatt 1020
ttcgtgatgg caaggttatt tcctgggcag cgtgctgtaa agacaggggtt attatgggcc 1080
ttagccgaga aggcattcaa gcctttattg attgaagatg ttcctgtagt gcaagattat 1140

```

cctgatgttt ttctgcaga attgccaggg atgccacctg agcatgatgt agagtttcgt 1200
attgatcttg ttctggtac acgacctatc tcagagcccc gtaccgaaga gttgaagaaa 1260
cnattagatg atctattatc caagaagtta atccgacgga gtgtgtcacc atggggtgca 1320
cctgttattt ttactgagaa gaatgatgga agttggcgta tgtgtgtgga ctatcgggcc 1380
ttaaatgcta cgacaatcaa gaacaagcac ccattgcctc gtattgaaga tttgtttgat 1440
cagttaaagg gagtgaagggt gttttctaaa attgatttgc aatctgggta taaccaaatt 1500
cgagtgcgag aagaggatat tgagaagacc gcattttcca cgaggtagtg gcattatgag 1560
tatagagtaa tgtcttttgg gctgaccaat gctccggcct attttatgga ggcgatgaat 1620
agcatgttgc atagctttct agataatttt gtgggtggtgt ttattgatga tattctgatt 1680
tactccaaag atgtagagga gcatgagcgt catctgagga tggtagtagg gaccctacga 1740
gaaaacaagt tctatgctaa actgaagaag tgtgagttct ggcttcctga agtggctttc 1800
ctagggcatg taattaatta g 1821

```

```

<210> 140
<211> 606
<212> PRT
<213> Oryza sativa

```

```

<400> 140

```

```

Met Glu Lys Tyr Met Arg Gly Ser Arg Met Thr Leu Glu Glu Lys Val
1           5           10           15

Thr Phe Met Pro Phe Gln Leu Thr Gly Gln Ala Asp Ile Trp Trp Asp
          20           25           30

Gly Val Val Glu Glu Trp Asn Pro Asp Arg Gly Ala Ile Thr Trp Glu
          35           40           45

Val Phe Leu Gln Gln Phe Arg Ala Lys Tyr Tyr Pro Glu Ser Phe Arg
          50           55           60

Asp Lys Met Ser Asn Glu Leu Asn Asn Ile Gln Gln Gly Ser Lys Thr
65           70           75           80

Val Asp Glu Tyr Glu Arg Glu Phe Asn Ser Ile Val Arg Phe Val Pro
          85           90           95

Asn Val Ala Ser Asp Glu Arg Glu Lys Ala Arg Lys Phe Ser Arg Gly
          100          105          110

```

Leu Ile Ala Arg Tyr Arg Glu Val Met Gly Arg Asn Pro Pro Thr Ile
 115 120 125
 Tyr Leu Asn Ala Val Glu Glu Ala Arg Gly Met Glu Val Glu Leu Gln
 130 135 140
 Ile Thr Thr Ala Gln Lys Asn Arg Ser Gly Asn Val Thr Gly Lys Gly
 145 150 155 160
 Leu Phe Cys Phe Lys Cys Gly Lys Asn His His Ser Ala Glu Cys Gly
 165 170 175
 Phe Thr Glu Ala Cys Arg Gly Cys Gly Lys Ala Gly His Met Glu Leu
 180 185 190
 Met Met Thr Ala Tyr Pro Leu Pro Tyr Gln Gln Phe Thr His Asn Tyr
 195 200 205
 Pro Pro Pro Pro Leu Pro Gly Gln Tyr Trp Ser Met Pro Gln Ser Leu
 210 215 220
 Ala Pro Arg Ala Ser Leu Gln Leu Pro Ala Ser Pro Met Val Gln Pro
 225 230 235 240
 Ser Ala Ser Ala Pro Pro His Gln Gln Ala Ser Gly Ser Thr Thr Pro
 245 250 255
 Ala Met Ala Gly Val Tyr Asn Met Pro Ala Ala Ser Ser Leu Gly Arg
 260 265 270
 Pro Asp Val Val Ala Asp Thr Leu Ser Val Asp Leu His Asp Ala Leu
 275 280 285
 Val Leu Phe Asp Ser Gly Ala Thr Phe Ser Phe Val Ser Ile Glu Phe
 290 295 300
 Ala Lys Arg Ala Asn Leu Phe Ser Gln Glu Val Ser Arg Ser Val Gln
 305 310 315 320
 Val Ser Ser Pro Gly Gly Glu Glu Met Val Phe Gln Gly Ser Ala Pro
 325 330 335
 Pro His Ser Ile Phe Val Met Ala Arg Leu Phe Pro Gly Gln Arg Ala
 340 345 350
 Val Lys Thr Gly Leu Leu Trp Ala Leu Ala Glu Lys Ala Ser Lys Pro
 355 360 365
 Leu Leu Ile Glu Asp Val Pro Val Val Gln Asp Tyr Pro Asp Val Phe
 370 375 380
 Pro Ala Glu Leu Pro Gly Met Pro Pro Glu His Asp Val Glu Phe Arg
 385 390 395 400

Ile Asp Leu Val Pro Gly Thr Arg Pro Ile Ser Glu Pro Arg Thr Glu
 405 410 415
 Glu Leu Lys Lys Gln Leu Asp Asp Leu Leu Ser Lys Lys Leu Ile Arg
 420 425 430
 Arg Ser Val Ser Pro Trp Gly Ala Pro Val Ile Phe Thr Glu Lys Asn
 435 440 445
 Asp Gly Ser Trp Arg Met Cys Val Asp Tyr Arg Ala Leu Asn Ala Thr
 450 455 460
 Thr Ile Lys Asn Lys His Pro Leu Pro Arg Ile Glu Asp Leu Phe Asp
 465 470 475 480
 Gln Leu Lys Gly Val Lys Val Phe Ser Lys Ile Asp Leu Gln Ser Gly
 485 490 495
 Tyr Asn Gln Ile Arg Val Arg Glu Glu Asp Ile Glu Lys Thr Ala Phe
 500 505 510
 Ser Thr Arg Tyr Gly His Tyr Glu Tyr Arg Val Met Ser Phe Gly Leu
 515 520 525
 Thr Asn Ala Pro Ala Tyr Phe Met Glu Ala Met Asn Ser Met Leu His
 530 535 540
 Ser Phe Leu Asp Asn Phe Val Val Val Phe Ile Asp Asp Ile Leu Ile
 545 550 555 560
 Tyr Ser Lys Asp Val Glu Glu His Glu Arg His Leu Arg Met Val Leu
 565 570 575
 Gly Thr Leu Arg Glu Asn Lys Phe Tyr Ala Lys Leu Lys Lys Cys Glu
 580 585 590
 Phe Trp Leu Pro Glu Val Ala Phe Leu Gly His Val Ile Asn
 595 600 605

<210> 141

<211> 1317

<212> DNA

<213> Oryza sativa

<400> 141

atggcttcca tgccgcagct ggaggaaaag gattcatccg atttggcaat caacaaaggc 60

ccgtcgtctgg atcttgtaaa atctcccttg atgatgaatg atgcatctgc aactgtaact 120

gcaatgcagc ccaacgaggg gatggaggag ttccctgtca aggttcggaa gccttacacg 180

atcaccaagc agcgggagaa gtggacggaa gaagagcacg acaagttcct tgaggcattg 240

```

aagctgtatg gtcgctcttg gcgtcagata caagagcaca ttggcacaaa gactgctgtc   300
caaattcgga gccatgcccga gaagtttttc tcaaagggtg tgctgagacc cggttccaat   360
aatgcgattg agatccctcc acctaggcca aagagaaaac cacttcatcc gtaccctcga   420
aagtgtgcga attctggatc ggacgcaaat ccagcaacag cgcaattgaa gcttgctcct   480
ggctcatctt catctggctc tgatcaagag aatggttctc ctatatcggg gctatctgcg   540
atgcagtcag atgcttttgg atcgtcagta tccaatccat caaccagatg tacctcccca   600
gcgtcatctg acgatggaaa taacattccc acattcacga gtggggaaga taataatgtg   660
ccttgtgaac caacagtgat agatccgtcc cagtcccaca aggaaataga ccaagataga   720
aaagacgtga ataatatgtc tgaagaggat tcttcagaag aagaggtgca agaaacaagc   780
ttgaagctat ttgggaggac agttgtcatc ccagatccaa ggaagagaag ctctcagat   840
ccgaagcatg aaagtgagga gcagatatca cagccttcca atgaagaaat gttgcaggct   900
tcttcatcgg ttggggagat tccagcggca tattgtgcac caaatgggtg gtttatgtca   960
tacaattctt tccattcca attcggtgaa tcagcggcag atgctagaat tcccccttta 1020
cacgtgtggg ggccttacta cggttttgct ccattagcc atcctagagg actaagcaca 1080
gtgatgcagc agactgaagg tagtgacgag agtgacgggt tgaagagcca ctcatccgaa 1140
tcaagttcgg actccgggga aaatgtgcag atgaccgctc ccagagctc gagaatagtg 1200
gaatcacttg gagcgattta cgtccgagac tcaggttcaa gttttgagct aaaaccgagc 1260
gcaaattcag cgtttgtaag agtgaagcca agcaacagcg gagatgaaga gatttga   1317

```

<210> 142

<211> 438

<212> PRT

<213> *Oryza sativa*

<400> 142

```

Met Ala Ser Met Pro Gln Leu Glu Glu Lys Asp Ser Ser Asp Leu Ala
1           5           10           15

```

```

Ile Asn Lys Gly Pro Ser Leu Asp Leu Val Lys Ser Pro Leu Met Met
20           25           30

```

```

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

```

```

Glu Glu Phe Pro Val Lys Val Arg Lys Pro Tyr Thr Ile Thr Lys Gln

```


50		55		60	
Arg Glu Lys Trp Thr	Glu Glu Glu His Asp	Lys Phe Leu Glu Ala Leu			
65	70	75			80
Lys Leu Tyr Gly Arg	Ser Trp Arg Gln Ile	Gln Glu His Ile Gly Thr			
	85	90			95
Lys Thr Ala Val Gln	Ile Arg Ser His Ala	Gln Lys Phe Phe Ser Lys			
	100	105			110
Val Val Arg Glu Pro	Gly Ser Asn Asn Ala	Ile Glu Ile Pro Pro Pro			
	115	120			125
Arg Pro Lys Arg Lys	Pro Leu His Pro Tyr	Pro Arg Lys Cys Ala Asn			
	130	135			140
Ser Gly Ser Asp Ala	Asn Pro Ala Thr Ala	Gln Leu Lys Leu Ala Pro			
145	150	155			160
Gly Ser Ser Ser Ser	Gly Ser Asp Gln Glu	Asn Gly Ser Pro Ile Ser			
	165	170			175
Val Leu Ser Ala Met	Gln Ser Asp Ala Phe	Gly Ser Ser Val Ser Asn			
	180	185			190
Pro Ser Thr Arg Cys	Thr Ser Pro Ala Ser	Ser Ser Asp Asp Gly Asn Asn			
	195	200			205
Ile Pro Thr Phe Thr	Ser Gly Glu Asp Asn	Asn Val Pro Cys Glu Pro			
	210	215			220
Thr Val Ile Asp Pro	Ser Gln Ser His Lys	Glu Ile Asp Gln Asp Arg			
225	230	235			240
Lys Asp Val Asn Asn	Met Ser Glu Glu Asp	Ser Ser Glu Glu Glu Val			
	245	250			255
Gln Glu Thr Ser Leu	Lys Leu Phe Gly Arg	Thr Val Val Ile Pro Asp			
	260	265			270
Pro Arg Lys Arg Ser	Ser Ser Ser Asp Pro	Lys His Glu Ser Glu Glu Gln			
	275	280			285
Ile Ser Gln Pro Ser	Asn Glu Glu Met Leu	Gln Ala Ser Ser Ser Val			
	290	295			300
Gly Glu Ile Pro Ala	Ala Tyr Cys Ala Pro	Asn Gly Trp Phe Met Ser			
305	310	315			320
Tyr Asn Ser Phe Pro	Phe Gln Phe Gly Glu	Ser Ala Ala Asp Ala Arg			
	325	330			335
Ile Pro Pro Leu His	Val Trp Trp Pro Tyr	Tyr Gly Phe Ala Pro Ile			

[illegible]

caccttcggt tatgggacat ccgaagtgca aaatgcacta gtcagacatt cgctcatctt 900
 gacgtaagct ctgtcagtgt gtcacggaat agaaatttca ttcttacaag tggaaaggat 960
 aacgtgcaca acctttttga tcctagaaca atggaagttt gtggaaaatt caaggccatg 1020
 ggtaacagag tggtttagcag ctggggcagg ccttgcataa gccctgatga aaactccata 1080
 gctgctggtg ctaatgatgg gtctgtttac atatggtcaa gattgaaaaa ggatggggta 1140
 ccgacgatcc tacagggcca ttcttcgtcg gtagtttcaa gttcatggtg tgggttagga 1200
 cctcttgcta ccgctgacaa gcatcatatc tataatctgga cttga 1245

<210> 144

<211> 414

<212> PRT

<213> *Oryza sativa*

<400> 144

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

	165		170		175
Ser Val Asn Asp Leu Ala Val Thr Asn Asp Asn Lys Phe Val Ile Ala	180		185		190
Ala Cys Ser Ser Asn Lys Leu Phe Val Trp Glu Val Asn Gly Gly Arg	195		200		205
Pro Arg His Thr Leu Thr Gly His Thr Lys Asn Val Ser Ser Val Asp	210		215		220
Ala Ser Trp Val Lys Ser Cys Val Leu Ala Ser Ser Ser Asn Asp His	225		230		235
Thr Ile Lys Ile Trp Asp Leu Gln Ser Gly Phe Cys Lys Ser Thr Ile	245		250		255
Met Ser Gly Ser Asn Ala Asn Ser Leu Ala Phe Ile Asp Gly Val Thr	260		265		270
Leu Cys Ser Gly His Arg Asp Gly His Leu Arg Leu Trp Asp Ile Arg	275		280		285
Ser Ala Lys Cys Thr Ser Gln Thr Phe Ala His Leu Asp Val Ser Ser	290		295		300
Val Ser Val Ser Arg Asn Arg Asn Phe Ile Leu Thr Ser Gly Lys Asp	305		310		315
Asn Val His Asn Leu Phe Asp Pro Arg Thr Met Glu Val Cys Gly Lys	325		330		335
Phe Lys Ala Met Gly Asn Arg Val Val Ser Ser Trp Gly Arg Pro Cys	340		345		350
Ile Ser Pro Asp Glu Asn Ser Ile Ala Ala Gly Ala Asn Asp Gly Ser	355		360		365
Val Tyr Ile Trp Ser Arg Leu Lys Lys Asp Gly Val Pro Thr Ile Leu	370		375		380
Gln Gly His Ser Ser Ser Val Val Ser Ser Ser Trp Cys Gly Leu Gly	385		390		395
Pro Leu Ala Thr Ala Asp Lys His His Ile Tyr Ile Trp Thr	405		410		

<210> 145

<211> 1449

<212> DNA

<213> Oryza sativa

<220>

<221> misc_feature
 <222> (1313)..(1313)
 <223> n = any nucleotide

<400> 145
 atggctcacc ggctgctgcg agacgcgcag gcggacgggt gggagcggtc cgacttcccc 60
 atcatctgcg agtcctgcct cggcgacaac ccctacgtcc gcatgttgag agcagaatat 120
 gataaagagt gcaaaatatg tgcacgtccg ttactgtct tccgttgcg acctggccgg 180
 gatgcaagat acaagaagac agagatctgc cagacatgtt gcaaactgaa aaatgtctgc 240
 caggtctgcc tgctagatct tgaatatggc ctaccgggtcc aggttaggga tacagcactc 300
 gctattaatt caaatgatgc aatcccaagg agtgatgtga accgtgagta ctttgcagaa 360
 gagcatgatc gcaaggccag ggctggaata gattatgatt cttcccatgg gaaggcacgt 420
 cccaatgaca ccattctgaa gcttcagagg acggcaccat attacaagag gaatcgagct 480
 catgtctgca gtttctatgt gcgtgggtgaa tgtacaagag gtgctgaatg cccttatcga 540
 catgagatgc ctgaaactgg agagttatct cagcagaaca ttaaagatcg ttactatgga 600
 gttaatgatc cagttgcttt gaaacttttg ggtaaggccg gtgagatgcc atccctgaca 660
 cctccagacg atgagagtat aaggacactt tacattgggtg gccttaataa tcgaatcacc 720
 gagcaggatt tgagggatca attttacgca catgggtgaaa ttgagtccat aaggatgggt 780
 ctgcaacgtg cctgtgcatt tgtgacgtac actacaagag aaggtgctga gaaagctgca 840
 gaggagcttg ccaacaagtt agtcatcaag ggtatacgcc tgaagctcat gtggggtaag 900
 cctcaagcac caaagccaga agacgatgaa gctggaaggc aaggcatgt tgacatgga 960
 ggaatgctcc ccagggtgt tatatctcag cagcagagtg gtgaccagcc tcaaccacct 1020
 ggaatggagg gacaacagca agcaccatcg ggatcatact acttcaacat tccagcaccg 1080
 ccaggagcag agcggacact gtacccttca atggatcccc agaggatggg tgccttggtc 1140
 aagtcacagg agggatgatg caaacccggc ccacagcagg cagcccaagc tcaggcatca 1200
 agcagctcag gacagagtta ccctatgcca ccacaatatt accatgggtc gtaccacccc 1260
 tactaccac cgtatggcgg ttacatgccc ccacctggca tgccataccc acngccaccg 1320
 cagtatccgc cttatcagcc gatgttggca acaccagcac agtcgcaagc tagctcatca 1380
 caacaacctg caccagcaac gctgcaccag gctcaagtc ctcctccgca gcagacgacc 1440
 cagaactga 1449

<210> 146
 <211> 482
 <212> PRT
 <213> Oryza sativa

<400> 146

```

Met Ala His Arg Leu Leu Arg Asp Ala Gln Ala Asp Gly Trp Glu Arg
1          5          10          15

Ser Asp Phe Pro Ile Ile Cys Glu Ser Cys Leu Gly Asp Asn Pro Tyr
          20          25          30

Val Arg Met Leu Arg Ala Glu Tyr Asp Lys Glu Cys Lys Ile Cys Ala
          35          40          45

Arg Pro Phe Thr Val Phe Arg Trp Arg Pro Gly Arg Asp Ala Arg Tyr
          50          55          60

Lys Lys Thr Glu Ile Cys Gln Thr Cys Cys Lys Leu Lys Asn Val Cys
65          70          75          80

Gln Val Cys Leu Leu Asp Leu Glu Tyr Gly Leu Pro Val Gln Val Arg
          85          90          95

Asp Thr Ala Leu Ala Ile Asn Ser Asn Asp Ala Ile Pro Arg Ser Asp
          100          105          110

Val Asn Arg Glu Tyr Phe Ala Glu Glu His Asp Arg Lys Ala Arg Ala
          115          120          125

Gly Ile Asp Tyr Asp Ser Ser His Gly Lys Ala Arg Pro Asn Asp Thr
          130          135          140

Ile Leu Lys Leu Gln Arg Thr Ala Pro Tyr Tyr Lys Arg Asn Arg Ala
145          150          155          160

His Val Cys Ser Phe Tyr Val Arg Gly Glu Cys Thr Arg Gly Ala Glu
          165          170          175

Cys Pro Tyr Arg His Glu Met Pro Glu Thr Gly Glu Leu Ser Gln Gln
          180          185          190

Asn Ile Lys Asp Arg Tyr Tyr Gly Val Asn Asp Pro Val Ala Leu Lys
          195          200          205

Leu Leu Gly Lys Ala Gly Glu Met Pro Ser Leu Thr Pro Pro Asp Asp
          210          215          220

Glu Ser Ile Arg Thr Leu Tyr Ile Gly Gly Leu Asn Asn Arg Ile Thr
225          230          235          240

```

[illegible]