

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
2 December 2004 (02.12.2004)

PCT

(10) International Publication Number
WO 2004/104174 A2

(51) International Patent Classification⁷: **C12N**

(21) International Application Number:
PCT/US2004/015286

(22) International Filing Date: 14 May 2004 (14.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/471,007 16 May 2003 (16.05.2003) US

(71) Applicant (for all designated States except US): **NORTH CAROLINA STATE UNIVERSITY** [US/US]; 2401 Research Drive, Campus Box 8210, Raleigh, NC 27695-8210 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ELUMALAI, Sivamani** [IN/US]; 2108 Star Sapphire Drive, Raleigh, NC 27610 (US). **QU, Rongda** [US/US]; 103 Hounslow Ct., Cary, NC 27511 (US).

(74) Agent: **MYERS BIGEL SIBLEY & SAJOVEC, P.A.**; P.O. Box 37428, Raleigh, NC 27627 (US).

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

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

Published:

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

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: POLYUBIQUITIN RUBI3 PROMOTER AND 5' REGULATORY SEQUENCES

(57) Abstract: The present invention provides novel expression control elements and methods for expressing a heterologous nucleotide sequence of interest in plants. Isolated nucleic acids, expression cassettes and vectors comprising the expression control elements are provided. Also provided are transformed plants, plant tissues, plant cells, plant seed, and the like, comprising the isolated nucleic acids, expression cassettes and vectors of the invention. Further provided are methods of expressing a heterologous nucleotide sequence of interest in a plant, plant tissue, plant cell, and the like.

WO 2004/104174 A2

POLYUBIQUITIN *rubi3* PROMOTER AND 5' REGULATORY SEQUENCES

5

RELATED APPLICATION INFORMATION

This application claims the benefit of United States provisional
Application Serial No. 60/471,007; Filed May 16, 2003; the disclosure of which
10 is incorporated by reference herein in its entirety.

FIELD OF THE INVENTION

The present invention relates to the field of plant molecular biology,
15 more particularly to regulation of nucleic acid expression in plants.

BACKGROUND OF THE INVENTION

Ubiquitin is a highly conserved 76-amino acid protein present in all
20 eukaryotic cells (Callis and Vierstra (1989) *Oxford Surv. Plant Mol. Cell Biol.*
6:1-30). Ubiquitin is encoded by small multigenic families containing two
types of genes, polyubiquitin and ubiquitin extension/fusion genes. The
polyubiquitin gene consists of tandem repeats of a 228 bp ubiquitin coding
region (Burke, *et al.* (1988) *Mol. Gen. Genet.* 213:435-443; Binet, *et al.* (1991)
25 *Plant Mol. Biol.* 17:395-407; Christensen, *et al.* (1992) *Plant Mol. Biol.* 18:675-
689). The ubiquitin genes are translated as a polyprotein precursor, then
proteolytically processed to functional ubiquitin monomers. The polyubiquitin
genes are constitutively expressed in all plant organs that have been tested,
with increased levels in young tissues (Binet, *et al.* (1991) *Plant Sci.* 79:87-94;
30 Kawalleck, *et al.* (1993) *Plant Mol. Biol.* 21:673-684; Burke, *et al.* (1988)
supra; Cornejo, *et al.* (1993) *Plant Mol Biol Rep* 14:19-21).

Promoters of polyubiquitin genes have been tested and reported to
have high constitutive expression of foreign genes placed under their control.
Such promoters have been isolated from Arabidopsis (Callis, *et al.* (1990) *J.*
35 *Biol. Chem.* 265:12486-12493; Norris, *et al.* (1993) *Plant Mol. Biol.* 21:895-

906), sunflower (Binet, *et al.* (1991) *supra*), tobacco (Genschik, *et al.* (1994) *Gene* 148:195-202), potato (Garbino and Belknap (1994) *Plant Mol. Biol.* 24:119-127), maize (Christensen, *et al.* (1992) *supra*), sugarcane (Wei, *et al.* (1999) *J. Plant Physiol.* 155:513-519) and rice (Wang, *et al.* (2000) *Plant Sci.* 156(2):201-211). Among these, the maize polyubiquitin Ubi1 promoter has been used in maize and other monocot transformation systems (Christensen, *et al.* (1992) *supra*; Cornejo, *et al.* (1993) *supra*; Christensen and Quail (1996) *Transgenic Res.* 5:213-218).

10

SUMMARY OF THE INVENTION

The ability to manipulate gene expression in plants facilitates the production of plants with improved and/or new characteristics. There are some situations in which it is desirable to express a foreign (*i.e.*, exogenous) nucleic acid sequence or an additional or modified copy of an endogenous sequence in a plant, plant tissue culture or plant cell, *etc.*, for example, for the production of polypeptides of agronomic, nutritional or commercial value, or the regulation of endogenous gene expression with antisense RNA or interfering RNA (RNAi).

The present invention provides compositions and methods for regulating expression of a heterologous nucleotide sequence(s) of interest in plants. In particular, the invention provides novel regulatory elements and methods for expressing a nucleotide sequence of interest in a plant, plant tissue, plant cell, and the like.

Accordingly, as a first aspect, the present invention provides an isolated nucleic acid comprising an expression control element comprising a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20; (b) a nucleotide sequence consisting essentially of a biologically active fragment of at least 50 consecutive nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20; (c) a nucleotide sequence that hybridizes to the complement of the nucleotide sequences of (a) or (b) under stringent hybridization conditions,

wherein the nucleotide sequence is biologically active as an expression control element; and (d) a nucleotide sequence having at least 95% sequence identity to the nucleotide sequences of (a) or (b), wherein the nucleotide sequence is biologically active as an expression control element.

5 A further aspect of the invention provides expression cassettes comprising the isolated nucleic acids of the invention operably linked to a heterologous nucleotide sequence of interest.

 As still further aspects, the present invention provides vectors and cells comprising the isolated nucleic acids and expression cassettes of the
10 invention.

 As yet another aspect, the present invention provides stably transformed plants and seeds comprising the isolated nucleic acids, expression cassettes and vectors of the invention.

 Also provided by the present invention are methods of introducing a
15 nucleotide sequence into a plant, plant tissue, plant cell or plant seed comprising transforming a plant, plant tissue, plant cell or plant seed with an isolated nucleic acid, expression cassette or vector of the invention.

 Further provided are methods of expressing a heterologous nucleotide sequence in a plant, plant tissue, plant cell or plant seed comprising
20 transforming a plant, plant tissue, plant cell or plant seed with an isolated nucleic acid, expression cassette or vector of the invention, and expressing the heterologous nucleotide sequence of interest in the plant, plant tissue, plant cell or plant seed.

 As yet another aspect, the invention provides methods of expressing a
25 nucleotide sequence in a plant comprising: stably transforming a plant cell or plant tissue with an isolated nucleic acid, expression cassette or vector of the invention, and regenerating a stably transformed plant, and expressing the heterologous nucleotide sequence of interest in the plant.

 Also provided are methods of producing a polypeptide of interest in a
30 plant, plant cell, plant tissue or plant seed of the invention.

 These and other aspects of the invention are set forth in more detail in the description of the invention below.

BRIEF DESCRIPTION OF THE DRAWINGS

- Figures 1A–1H** show regulatory sequences of the *rubi3* gene. **Figure 1A**, *rubi3* promoter and 5' UTR (untranslated region) intron (SEQ ID NO:1);
- 5 **Figure 1B**, *rubi3* promoter only (SEQ ID NO:2); **Figure 1C**, 5' UTR intron only (SEQ ID NO:3); **Figures 1D-1G**, *rubi3* promoter, 5' UTR intron, and first nine nucleotides of the coding sequence (double-underlined), wherein the third position of said coding sequence contains a G (**Figure 1D**; SEQ ID NO:17), A (**Figure 1E**; SEQ ID NO:18), T (**Figure 1F**; SEQ ID NO:19), or C (**Figure 1G**; SEQ ID NO:20). **Figure 1H** depicts the sequence of an 780 bp deletion mutant of the *rubi3* intron (SEQ ID NO:27). The TATA box is underlined. The transcription initiation site is double-underlined. CAAT boxes are in lower case letters.
- 15 **Figures 2A and 2B** show the nucleotide sequence (SEQ ID NO:8) and amino acid sequence (SEQ ID NO:9) of rice *rubi3*. The 5'-upstream region of rice *rubi3* is depicted in lowercase letters. The TATA box and transcription initiation site (26 nucleotides downstream of the TATA box) are double-underlined. Downstream is the 5' UTR intron shown by single underline. The
- 20 5' and 3' intron splice sites are boxed. The 918 nucleotide polyubiquitin coding region of *rubi3* and deduced amino acid sequence are indicated.
- Figures 3A-C** show a CLUSTAL W multiple sequence alignment of RUBQ1 (R1; accession number AF184279; **SEQ ID NO:14**), RUBQ2 (R2; accession number AF184280; **SEQ ID NO:15**) and *rubi3* (R3; **SEQ ID NO:3**) ubiquitin introns of rice and the maize (M; **SEQ ID NO:16**) ubiquitin intron.
- Figures 4A-E** illustrate constructs containing the *rubi3* promoter, GUS coding sequence and NOS terminator.
- 30 **Figure 5A** shows transient GUS activity in rice suspension cells transformed with pRESQ4 which contains the *rubi3* promoter and *rubi3* intron driving GUS gene expression.

Figure 5B shows transient GUS activity in rice coleoptiles transformed with pRESQ4 which contains the *rubi3* promoter and *rubi3* intron driving GUS gene expression (left), and with pRESQ17 which contains the rice actin promoter and *rubi3* intron driving GUS gene expression (right), respectively.

Figure 6 shows transient GUS activity in rice suspension cells transformed with the rice *rubi3* promoter-driving GUS gene expression in comparison with other constructs.

10

Figure 7 is a schematic representation of the pRESQ42 construct and its derivatives showing the nucleotide fusion between the *rubi3* 5' UTR intron and GUS coding sequence and the resulting N-terminal sequence of the translated products. Mutated nucleotides are underlined.

15

Figure 8 shows transient GUS activity in rice suspension cells transformed with *rubi3*::GUS fusion constructs. pRESQ29, maize *ubi*; pRESQ38, translational fusion of the first monomer of *rubi3* with GUS; pRESQ36, translational fusion of the first 9 amino acids of *rubi3* with GUS; pRESQ42, translational fusion of the first 3 amino acids of *rubi3* with GUS; and pRESQ46, pRESQ47 and pRESQ48, non-translated fusion of the first 9 nucleotides of *rubi3* with GUS, wherein the G of the initiation codon of *rubi3* is mutated from G to A (pRESQ 46), G to T (pRESQ 47), or G to C (pRESQ48).

Figure 9 demonstrates transient GUS expression in rice callus lines transformed with plasmids containing modified *rubi3* promoter. The plasmid construct pRESQ48 was made by fusing 9 nucleotides encoding the first 3 amino acids of the *rubi3* gene, where the third nucleotide (G) is replaced with a C. The plasmid construct pRESQ60 was made by deleting nucleotide sequences between *Bam*HI and *Bgl*II restriction enzyme sites in the 5' UTR of the *rubi3* intron sequence in pRESQ48. The plasmid pRESQ4 contains the un-altered *rubi3* promoter, and plasmid pRESQ29 contains maize *ubi1* promoter sequence driving the GUS gene.

30

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based, in part, on the discovery of new 5' expression control elements from the rice *rubi3* polyubiquitin gene. These regulatory elements include the *rubi3* promoter, an intron sequence that is located immediately 5' to the translational start site in the rice gene, and a downstream enhancing element (*i.e.*, downstream of the regulatory intron) located within the first 9 nucleotides of the *rubi3* coding sequence. These regulatory elements can be used independently, together, or in operable association with other, heterologous (*e.g.*, from a foreign gene) regulatory elements. For example, the inventors disclose herein expression constructs containing a rice actin or H3 promoter operably associated with the *rubi3* intron sequence.

The present invention will now be described in more detail with reference to the accompanying drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The terminology used in the description of the invention herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

Nucleotide sequences are presented herein by single strand only, in the 5' to 3' direction, from left to right, unless specifically indicated otherwise. Nucleotides and amino acids are represented herein in the manner recommended by the IUPAC-IUB Biochemical Nomenclature Commission, or

(for amino acids) by either the one-letter code, or the three letter code, both in accordance with 37 CFR §1.822 and established usage.

Except as otherwise indicated, standard methods known to those skilled in the art may be used for cloning genes, amplifying and detecting
5 nucleic acids, and the like. Such techniques are known to those skilled in the art. See, e.g., SAMBROOK *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed. (Cold Spring Harbor, NY, 1989); F. M. AUSUBEL *et al.* CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Green Publishing Associates, Inc. and John Wiley & Sons, Inc., New York).

10

I. Definitions.

As used in the description of the invention and the appended claims, the singular forms "a," "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise.

15 As used herein a "chimeric nucleic acid" or "chimeric nucleotide sequence" comprises an expression control element operably linked to a nucleotide sequence of interest that is heterologous (*i.e.*, foreign) to the expression control element. In particular embodiments, the "chimeric nucleic acid" or "chimeric nucleotide sequence" comprises a *rubi3* expression control
20 element operably associated with a heterologous nucleotide sequence of interest to be transcribed.

An "expression control element" is a nucleotide sequence that controls or regulates the expression of a coding sequence. The expression control element may act at the transcriptional level, the RNA processing level, or the
25 translational level. A promoter is one type of expression control element, as are other 5' regulatory sequences such as the *rubi3* and Act1 intron sequences described herein. Also disclosed herein is an enhancing element at the 5' end of the *rubi3* coding sequence.

A "promoter" refers to the nucleotide sequence that contains a binding
30 site for RNA polymerase II and directs the initiation of transcription. In general, promoters are found 5', or upstream, relative to the start of the coding region of the corresponding gene. A promoter sequence is necessary, but not always sufficient, to drive transcription of a downstream sequence. The

promoter region may also comprise other elements that act as regulators of gene expression. These include a TATA box consensus sequence, and often a CAAT box consensus sequence (Breathnach and Chambon, (1981) *Annu. Rev. Biochem.* **50**:349). In plants, the CAAT box may be substituted by the AGGA box (Messing *et al.*, (1983) in Genetic Engineering of Plants, T. Kosuge, C. Meredith and A. Hollaender (eds.), Plenum Press, pp. 211-227). The promoter region, including all the ancillary regulatory elements, typically contain between 100 and 1000 nucleotides, but can be as long as 2 kb, 3 kb, 4 kb or longer in length. Promoters according to the present invention can function as constitutive or inducible regulatory elements.

By "operably linked" or "operably associated" as used herein, it is meant that the indicated elements are functionally related to each other, and are also generally physically related. For example, a promoter is operatively linked to a coding sequence if it controls the transcription of the sequence. Thus, the term "operatively linked" or "operably associated" as used herein, refers to nucleotide sequences on a single nucleic acid molecule that are functionally associated.

A "heterologous nucleotide sequence" or "heterologous nucleotide sequence of interest" as used herein is a coding sequence that is heterologous to (*i.e.*, foreign, exogenous or non-native to) the rubi3 expression control elements of the invention (*i.e.*, is not the native rubi3 coding sequence). The heterologous nucleotide sequence can encode a polypeptide or a nontranslated RNA.

By the term "express" or "expression" of a nucleic acid coding sequence, it is meant that the sequence is transcribed. In particular embodiments, the terms "express" and "expression" can refer to both transcription and translation to produce an encoded polypeptide.

As used herein, the term "polypeptide" encompasses both peptides and proteins, unless indicated otherwise.

A "biologically active" or "functional" nucleotide sequence is one that substantially retains at least one biological activity normally associated with that nucleotide sequence, for example, ability to drive transcription, translation and/or RNA processing, and the like of an operatively associated coding

sequence. In particular embodiments, the "biologically active" or "functional" nucleotide sequence substantially retains all of the biological activities possessed by the unmodified sequence. By "substantially retains" biological activity, it is meant that the nucleotide sequence retains at least about 50%,
5 60%, 75%, 85%, 90%, 95%, 97%, 98%, 99%, or more, of the biological activity of the native nucleotide sequence (and can even have a higher level of activity than the native nucleotide sequence). For example, a biologically active expression control element is able to promote or enhance the expression of a nucleic acid operably associated with the expression control
10 element. Methods of measuring expression of a nucleotide sequence are well known in the art and include Northern blots, RNA run-on assays and methods of measuring the presence of the encoded polypeptide (e.g., antibody based methods or visual inspection in the case of a reporter polypeptide).

As used herein, an "isolated" nucleic acid (e.g., an "isolated DNA")
15 means a nucleic acid separated or substantially free from at least some of the other components of the naturally occurring organism or virus, such as for example, the cell or viral structural components or other polypeptides or nucleic acids commonly found associated with the nucleic acid.

Likewise, an "isolated" polypeptide means a polypeptide that is
20 separated or substantially free from at least some of the other components of the naturally occurring organism or virus, for example, the cell or viral structural components or other polypeptides or nucleic acids commonly found associated with the polypeptide. In particular embodiments, the "isolated" polypeptide is at least about 1%, 5%, 10%, 25%, 50%, 60%, 70%, 75%, 80%,
25 85%, 90%, 95%, 97%, 98%, 99% or more pure (w/w). In other embodiments, an "isolated" polypeptide indicates that at least about a 5-fold, 10-fold, 25-fold, 100-fold, 1000-fold, 10,000-fold, or more enrichment of the protein (w/w) is achieved as compared with the starting material.

30 II. ***rubi3* Expression Control Elements.**

As one aspect, the invention provides isolated nucleic acids comprising one or more *rubi3* expression control elements (or biologically active fragments or variations thereof). The *rubi3* expression control elements of the

invention encompass homologs from any species (e.g., plants) as well as biologically active fragments of naturally occurring *rub13* expression control elements or biologically active variants thereof that contain insertions, deletions and/or substitutions in the naturally occurring sequences and fragments. The expression control elements of the invention can further be isolated from natural sources or may be wholly or partially synthetic. Those skilled in the art will appreciate that in addition to the promoter, 5' UTR intron and downstream enhancing element (at the 5' end of the *rub13* coding sequence) described herein, the isolated nucleic acid can comprise other sequences from the *rub13* gene, e.g., all or a portion of the 5' UTR sequence that lies between the promoter and the intron sequences (see **Figure 1A, 2A and 2B**) or some of the flanking sequences (e.g., up to about 20, 50, 100, 200, 300, 500, 1000, 2000, 3000 nucleotides or more) that are positioned upstream of the *rub13* promoter. Further, as described below, the isolated nucleic acid can comprise some or all of the *rub13* coding sequence. In particular embodiments, described in more detail below, the first codon of the *rub13* coding sequence is modified so that it no longer acts as a translational start site (i.e., the modified coding sequence is transcribed but not translated).

Homologous sequences from other organisms, in particular other plants, can be routinely identified using methods known in the art. For example, PCR and other amplification techniques and hybridization techniques can be used to identify such homologs based on their sequence similarity to the sequences set forth herein.

Thus, in exemplary embodiments, the isolated nucleic acid comprises, consists essentially of, or consists of the nucleotide sequence of **SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20**. With respect to the intron sequence of **SEQ ID NO:3** (as well as **SEQ ID NO:1, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 and SEQ ID NO:20**), it will be understood by those skilled in the art that efficient intron splicing generally requires the complete 5' and 3' splice recognition sites. It has been documented that the splice recognition site regions are well conserved and have the following consensus sequences: NNCAAG/gtragu . . . y(n)ncag/RNN (Dibb, (1993) *FEBS* 325:135-139).

Thus, in particular embodiments, the intron sequences of the invention comprise the complete 5' and 3' splice recognition sites (see, e.g., Figure 2A).

In other representative embodiments, the isolated nucleic acid comprises, consists essentially of or consists of a biologically active fragment or biologically active variant of **SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20**. In particular embodiments, the biologically active fragment or variant of the *rubi3* promoter sequence comprises the TATA box sequence and/or one or more CAAT box sequences (**Figures 1A, 1B and 2A**).

The length of the biologically active fragment is not critical as long as it substantially retains the biological activity of the expression control element. Illustrative fragments comprise at least about 20, 30, 40, 50, 60, 100, 150, 200, 250, 300, 400, 500, 600 or 700 contiguous nucleotides or more of the full-length sequence. In other embodiments, a biologically active fragment of the *rubi3* promoter comprises the nucleotide sequence of **SEQ ID NO:2** having a deletion of up to 5, 10, 20, 30, 50, 100, 200, 250, 300, 400 or 500 nucleotides at the 5' and/or 3' end. In still further embodiments, the biologically active fragment of the *rubi3* promoter sequence comprises a fragment from about nucleotide 1, 5, 10, 20, 30, 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 550, 600 or 650 of **SEQ ID NO:2** to about nucleotide 300, 350, 400, 450, 500, 550, 600, 700, 750, 800, 810, 820, 830, 840 or 851 (inclusive) of **SEQ ID NO:2**, including any combination thereof.

Further, in other representative embodiments, a biologically active variant of the *rubi3* promoter of **SEQ ID NO:2** comprises internal deletions of up to about 5, 10, 20, 30, 50, 100, 200, 250, 300, 400 or 500 nucleotides. There may be a single or multiple internal deletions.

Those skilled in the art will appreciate that biologically active fragments or internal deletion variants of the intron regulatory sequence will generally require appropriate 5' and 3' splice signals for proper processing. Thus, in representative embodiments, a biologically active variant of the *rubi3* intron sequence comprises 5' and 3' splice site recognition sequence motifs and optionally further comprises an internal nucleotide sequence positioned between the splice recognition sites which is a deleted form of the sequence

from nucleotide 5 through nucleotide 1137 inclusive) of **SEQ ID NO:3** having deletion of up to about 5, 10, 20, 30, 50, 100, 200, 250, 300, 400, 500, 600, 700, 800, 900 or 1000 nucleotides, where the deletion(s) is internal and/or is at the 5' and/or 3' ends of the indicated sequence (see, e.g., **Figure 1H**). In
5 another illustrative embodiment, the entire internal sequence of the intron (*i.e.*, other than the splice recognition sites), or essentially the entire internal sequence of the intron (e.g., all but about 5, 10, 25, 50 nucleotides or less), is deleted.

In one representative embodiment, the variant 5' UTR intron comprises
10 5' and 3' splice recognition signals. The 5' UTR intron can optionally further comprise an internal nucleotide sequence positioned between the 5' and 3' splice recognition signals. In particular embodiments, the internal nucleotide sequence comprises at least about 5, 10, 20, 30, 50, 100, 200, 250, 300, 400, 500, 600, 700, 800, 900 or 1000 consecutive nucleotides from the nucleotide
15 sequence of nucleotide 5 through nucleotide 1137 (inclusive) of **SEQ ID NO:3**, wherein the 5' UTR intron is biologically active as an expression control element. The 5' UTR intron can further comprise two or more segments of the intron sequence from nucleotide 5 through nucleotide 1137 (inclusive) of **SEQ ID NO:3**. For example, the 5' UTR can include separate segments of the
20 intron sequence adjacent to each of the splice recognition sites, which can optionally be contiguous to each other as a result of the deletion of the intervening region of the intron sequence.

Alternatively, the internal intron sequence can be modified by insertion or substitution of one or more nucleotides.

25 The 5' and 3' splice site motifs of the 5' UTR intron elements of the invention can be those shown in **Figure 2A** or, alternatively, can be any other functional splice site recognition region known in the art (see consensus sequence above).

In other embodiments of the invention, one or both of the intron splice
30 recognition sites is altered so that the intron is not efficiently spliced.

In representative embodiments, the expression control element of the invention is a biologically active variant of the *rubi3* sequences or fragments described above and comprises, consists essentially of, or consists of a

nucleotide sequence that will hybridize to the *rubi3* nucleotide sequences disclosed herein (*i.e.*, **SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, or SEQ ID NO:20** or biologically active fragments thereof) under standard conditions as known by those skilled in the art.

For example, hybridization of such sequences may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions (*e.g.*, conditions represented by a wash stringency of 35-40% Formamide with 5x Denhardt's solution, 0.5% SDS and 1x SSPE at 37°C; conditions represented by a wash stringency of 40-45% Formamide with 5x Denhardt's solution, 0.5% SDS, and 1x SSPE at 42°C; and conditions represented by a wash stringency of 50% Formamide with 5x Denhardt's solution, 0.5% SDS and 1x SSPE at 42°C, respectively) to the *rubi3* expression control elements or biologically active fragments thereof disclosed herein. See, *e.g.*, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2d Ed. 1989) (Cold Spring Harbor Laboratory).

In other embodiments, the nucleotide sequence of an expression control element of the invention has at least about 60%, 70%, 80%, 85%, 90%, 95%, 97% or higher nucleotide sequence similarity to a nucleotide sequence specifically disclosed herein (*e.g.*, **SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20** or biologically active fragments thereof).

As is known in the art, a number of different programs can be used to identify whether a nucleic acid or amino acid has sequence identity or similarity to a known sequence. Sequence identity or similarity may be determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, *Adv. Appl. Math.* **2**, 482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* **48**,443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* **85**,2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), the Best Fit sequence

program described by Devereux *et al.*, *Nucl. Acid Res.* **12**, 387-395 (1984), preferably using the default settings, or by inspection.

5 An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* **35**, 351-360 (1987); the method is similar to that described by Higgins & Sharp, *CABIOS* **5**, 151-153 (1989).

10 Another example of a useful algorithm is the BLAST algorithm, described in Altschul *et al.*, *J. Mol. Biol.* **215**, 403-410, (1990) and Karlin *et al.*, *Proc. Natl. Acad. Sci. USA* **90**, 5873-5787 (1993). A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul *et al.*, *Methods in Enzymology*, **266**, 460-480 (1996);
15 [http://blast.wustl.edu/blast/ README.html](http://blast.wustl.edu/blast/README.html). WU-BLAST-2 uses several search parameters, which are preferably set to the default values. The parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however,
20 the values may be adjusted to increase sensitivity.

An additional useful algorithm is gapped BLAST as reported by Altschul *et al.* *Nucleic Acids Res.* **25**, 3389-3402 (1997).

The CLUSTAL program can also be used to determine sequence similarity. This algorithm is described by Higgins *et al.* (1988) *Gene* **73**:237;
25 Higgins *et al.* (1989) *CABIOS* **5**:151-153; Corpet *et al.* (1988) *Nucleic Acids Res.* **16**: 10881-90; Huang *et al.* (1992) *CABIOS* **8**: 155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol.* **24**: 307-331.

The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer
30 nucleotides than the nucleic acids disclosed herein, it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical nucleotides acids in relation to the total number of nucleotide bases. Thus, for example, sequence identity of

sequences shorter than a sequence specifically disclosed herein, will be determined using the number of nucleotide bases in the shorter sequence, in one embodiment. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, *etc.*

In other representative embodiments, the isolated nucleic acid further comprises a downstream enhancing element comprising all or part of the *rubi3* coding sequence, or a non-translated modification thereof, that enhances expression (transcription and/or translation) of an operably associated heterologous nucleotide sequence. In particular embodiments, the *rubi3* downstream enhancing element comprises, consists essentially of, or consists of at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 20, 30, 40, 50, 60, 75, 100, 125, 150 or 200 nucleotides of the *rubi3* coding sequence (counting from the 5' end; the *rubi3* coding sequence is shown as nucleotides 2106 to 2933 of **SEQ ID NO:8 in Figures 2A and 2B**) or even the entire *rubi3* coding sequence. In representative embodiments, the downstream enhancing element comprises, consists essentially of or consists of the first 9 nucleotides of the *rubi3* coding sequence. Accordingly, in some embodiments, the downstream enhancing element encodes the N-terminal 1, 2, 3, 4, 5, 6, 9, 12, 15, 20, 25, 30, 40, 50, 60, 70 amino acids or all 76 amino acids of a ubiquitin monomer (see, e.g., **Figures 2A and 2B**; amino acids 1-76 of **SEQ ID NO:9**).

The inventors have discovered that the first codon (*i.e.*, nucleotides 1-3, ATG) of the *rubi3* coding sequence can be modified so that it no longer functions as a translational start codon, *i.e.*, so that the modified *rubi3* coding sequence is transcribed but not translated without impairing the enhancing activity of the 5' *rubi3* downstream enhancing element. One advantage of this embodiment of the invention is that it avoids expression of a fusion protein comprising part or all of the ubiquitin protein fused to the polypeptide encoded by the heterologous nucleotide sequence. Further, in particular embodiments, modification of the initiation codon to ablate translation can result in increased enhancing activity from the downstream enhancing element.

The translational start codon (ATG) can be modified so that it no longer functional by substitution, deletion and/or insertion at any position within the codon. For example, a C, T or G can be substituted for the A at position 1 and/or an A, C or G can be substituted for the T at position 2 and/or an A, T or C can be substituted for the G at position 3. In illustrative embodiments, the first codon of the *rubi3* coding sequence is modified from ATG → ATA, ATC or ATT.

In other representative embodiments, the *rubi3* downstream enhancing element comprises, consists essentially of, or consists of the first 9 nucleotides of the *rubi3* coding sequence (ATGCAGATA) or the modified sequences ATACAGATA, ATCCAGATA or ATTCAGATA. Those skilled in the art will appreciate that the invention further encompasses variants of these downstream enhancing element having one, two, three, four, five or more nucleotide substitutions as long as the downstream enhancing element still provides enhanced expression of an operably associated heterologous nucleotide sequence of interest.

In further exemplary embodiments, the isolated nucleic acid comprises, consists essentially of, or consists of the nucleotide sequence of **SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, or SEQ ID NO:20** or biologically active fragments or variants thereof (as described above).

In other embodiments of the invention, the isolated nucleic acid does not include any of the *rubi3* coding sequence.

The invention further provides an expression cassette comprising an expression control element(s) of the invention. The expression cassette generally comprises a promoter, which in particular embodiments is a *rubi3* promoter of the invention or a biologically active fragment or variant thereof (as described above), operably linked to a heterologous nucleotide sequence of interest. In particular embodiments, the expression cassette comprises more than one (e.g., two, three, four or more) heterologous nucleotide sequences. The expression cassette can further have a plurality of restriction sites for insertion of a heterologous nucleotide sequences to be operably linked to the regulatory regions.

The expression cassette may further comprise a transcriptional termination sequence. Any suitable termination sequence known in the art may be used in accordance with the present invention. The termination region may be native with the transcriptional initiation region, may be native
5 with the nucleotide 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 synthetase and nopaline synthetase termination regions. See also, Guerineau *et al.*, *Mol. Gen. Genet.* **262**, 141 (1991); Proudfoot, *Cell* **64**, 671 (1991); Sanfacon *et al.*, *Genes Dev.* **5**, 141
10 (1991); Mogen *et al.*, *Plant Cell* **2**, 1261 (1990); Munroe *et al.*, *Gene* **91**, 151 (1990); Ballas *et al.*, *Nucleic Acids Res.* **17**, 7891 (1989); and Joshi *et al.*, *Nucleic Acids Res.* **15**, 9627 (1987). Additional exemplary termination sequences are the pea RubP carboxylase small subunit termination sequence and the Cauliflower Mosaic Virus 35S termination sequence. Other suitable
15 termination sequences will be apparent to those skilled in the art.

Further, in particular embodiments, the heterologous nucleotide sequence is operably associated with a translational start site. The translational start site can be derived from the *rubi3* coding sequence or, alternatively, can be the endogenous translational start site associated with
20 the heterologous nucleotide sequence or any other suitable translational start codon. Thus, in some embodiments, the heterologous nucleotide sequence is expressed as a fusion protein. In other embodiments, the heterologous nucleotide sequence is not expressed as a fusion protein.

In illustrative embodiments, the expression cassette includes in the 5'
25 to 3' direction of transcription, a promoter, a heterologous nucleotide sequence of interest, and a transcriptional and translational termination region functional in plants.

The expression cassette can further comprise a regulatory intron, for example, the *rubi3* 5' UTR intron (or biologically active variants thereof, as
30 described above) or another regulatory intron from a monocot plant gene, which is generally positioned in the 5' UTR between the promoter and the heterologous nucleotide sequence. In particular embodiments, the expression cassette further comprises a downstream enhancing element that

is positioned between the intron and the heterologous nucleotide sequence. The intron can optionally be positioned adjacent to (with only an insignificant number of intervening nucleotide bases, e.g., less than about 25, 20, 15, 10, or 5 intervening nucleotide bases or even no intervening nucleotides) the downstream enhancing element (if present) or adjacent to the heterologous nucleotide sequence of interest (i.e., if the downstream enhancing element is not present).

The expression cassette can additionally comprise a *rubi3* downstream enhancing element or a modified *rubi3* non-translated downstream enhancing element or variants thereof as described above. In representative embodiments, the downstream enhancing element is positioned between a 5' UTR regulatory intron and the heterologous sequence of interest, i.e., downstream of the intron and upstream of the heterologous sequence). The downstream enhancing element may optionally be positioned adjacent to of the heterologous nucleotide sequence of interest and/or adjacent to the regulatory intron.

In one particular embodiment, the orientation of these elements within the expression cassette from the 5' to 3' direction is the *rubi3* promoter (or biologically active variant or fragment thereof), optionally the *rubi3* 5' UTR intron (or biologically active variant or fragment thereof), and a *rubi3* downstream enhancing element (or biologically active variant thereof) and the heterologous nucleotide sequence of interest.

In other particular embodiments of the invention, the expression cassette comprises the *rubi3* 5' UTR intron sequence (or biologically active variants or fragments thereof) operably linked to a heterologous (i.e., foreign) promoter and a heterologous nucleotide sequence of interest such that the intron enhances expression of a heterologous nucleotide sequence from the promoter. According to representative embodiments, these elements are configured in the 5' to 3' direction as: promoter, 5' UTR *rubi3* intron (or biologically active variant or fragment thereof) and a heterologous nucleotide sequence of interest. The expression cassette can further optionally comprise a downstream enhancing element of the invention positioned between the 5' UTR intron and the heterologous nucleotide sequence. The heterologous

promoter can be any suitable promoter known in the art (including bacterial, yeast, fungal, insect, mammalian, and plant promoters). In particular embodiments, the promoter is a promoter for expression in plants. Exemplary promoters include, but are not limited to, the Cauliflower Mosaic Virus 35S promoter, the opine synthetase promoters (e.g., nos, mas, ocs, etc.), another ubiquitin promoter, an actin promoter (e.g., Act1 promoter), H3 promoter, the ribulose biphosphate (RubP) carboxylase small subunit promoter, and the alcohol dehydrogenase promoter. Other suitable promoters include promoters from viruses that infect the host plant including, but not limited to, promoters isolated from Dasheen mosaic virus, Chlorella virus (e.g., the Chlorella virus adenine methyltransferase promoter; Mitra *et al.*, (1994) *Plant Molecular Biology* **26**:85), tomato spotted wilt virus, tobacco rattle virus, tobacco necrosis virus, tobacco ring spot virus, tomato ring spot virus, cucumber mosaic virus, peanut stump virus, alfalfa mosaic virus, and the like.

In illustrative embodiments, the expression cassette comprises in the 5' to 3' direction: a *rubi3* promoter, a 5' UTR *rubi3* intron, and a *rubi3* downstream enhancing element (or biologically active variants or fragments of the foregoing), and a heterologous nucleotide sequence(s) of interest, wherein the 5' UTR *rubi3* intron is adjacent to the *rubi3* downstream enhancing element, which is in turn adjacent to the heterologous nucleotide sequence of interest.

Those skilled in the art will understand that the expression cassettes of the invention can further comprise enhancer elements and/or tissue preferred elements in combination with the expression control sequences of the invention.

Further, in some embodiments, it is advantageous for the expression cassette to comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (*NEO*) and hygromycin phosphotransferase (*HPT*), as well as genes conferring resistance to herbicidal compounds. Herbicide resistance genes generally code for a modified target protein insensitive to the herbicide or for an enzyme that degrades or detoxifies the herbicide in the plant before it can act. See,

DeBlock *et al.*, *EMBO J.* **6**, 2513 (1987); DeBlock *et al.*, *Plant Physiol.* **91**, 691 (1989); Fromm *et al.*, *BioTechnology* **8**, 833 (1990); Gordon-Kamm *et al.*, *Plant Cell* **2**, 603 (1990). For example, resistance to glyphosphate or sulfonylurea herbicides has been obtained using genes coding for the mutant
5 target enzymes, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) and acetolactate synthase (ALS). Resistance to glufosinate ammonium, boromoxynil, and 2,4-dichlorophenoxyacetate (2,4-D) have been obtained by using bacterial genes encoding phosphinothricin acetyltransferase, a nitrilase, or a 2,4-dichlorophenoxyacetate monooxygenase, which detoxify the
10 respective herbicides.

Selectable marker genes that can be used according to the present invention further include, but are not limited to, genes encoding: neomycin phosphotransferase II (Fraley *et al.*, *CRC Critical Reviews in Plant Science* **4**, 1 (1986)); cyanamide hydratase (Maier-Greiner *et al.*, *Proc. Natl. Acad. Sci. USA* **88**, 4250 (1991)); aspartate kinase; dihydrodipicolinate synthase (Perl *et al.*, *BioTechnology* **11**, 715 (1993)); the *bar* gene (Toki *et al.*, *Plant Physiol.* **100**, 1503 (1992); Meagher *et al.*, *Crop Sci.* **36**, 1367 (1996)); tryptophane decarboxylase (Goddijn *et al.*, *Plant Mol. Biol.* **22**, 907 (1993)); neomycin phosphotransferase (*NEO*; Southern *et al.*, *J. Mol. Appl. Gen.* **1**, 327 (1982));
20 hygromycin phosphotransferase (*HPT* or *HYG*; Shimizu *et al.*, *Mol. Cell. Biol.* **6**, 1074 (1986)); dihydrofolate reductase (DHFR; Kwok *et al.*, *Proc. Natl. Acad. Sci. USA* **83**, 4552 (1986)); phosphinothricin acetyltransferase (DeBlock *et al.*, *EMBO J.* **6**, 2513 (1987)); 2,2- dichloropropionic acid dehalogenase (Buchanan-Wollatron *et al.*, *J. Cell. Biochem.* **13D**, 330 (1989));
25 acetohydroxyacid synthase (United States Patent No. 4,761,373 to Anderson *et al.*; Haughn *et al.*, *Mol. Gen. Genet.* **221**, 266 (1988)); 5-enolpyruvylshikimate-phosphate synthase (*aroA*; Comai *et al.*, *Nature* **317**, 741 (1985)); haloarylnitrilase (WO 87/04181 to Stalker *et al.*); acetyl-coenzyme A carboxylase (Parker *et al.*, *Plant Physiol.* **92**, 1220 (1990)); dihydropteroate
30 synthase (*sulI*; Guerineau *et al.*, *Plant Mol. Biol.* **15**, 127 (1990)); and 32 kDa photosystem II polypeptide (*psbA*; Hirschberg *et al.*, *Science* **222**, 1346 (1983)).

Also included are genes encoding resistance to: chloramphenicol (Herrera-Estrella *et al.*, *EMBO J.* **2**, 987 (1983)); methotrexate (Herrera-Estrella *et al.*, *Nature* **303**, 209 (1983); Meijer *et al.*, *Plant Mol. Biol.* **16**, 807 (1991)); hygromycin (Waldron *et al.*, *Plant Mol. Biol.* **5**, 103 (1985); Zhijian *et al.*, *Plant Science* **108**, 219 (1995); Meijer *et al.*, *Plant Mol. Bio.* **16**, 807 (1991)); streptomycin (Jones *et al.*, *Mol. Gen. Genet.* **210**, 86 (1987)); and spectinomycin (Bretagne- Sagnard *et al.*, *Transgenic Res.* **5**, 131 (1996)); bleomycin (Hille *et al.*, *Plant Mol. Biol.* **7**, 171 (1986)); sulfonamide (Guerineau *et al.*, *Plant Mol. Bio.* **15**, 127 (1990); bromoxynil (Stalker *et al.*, *Science* **242**, 419 (1988)); 2,4-D (Streber *et al.*, *Bio/Technology* **7**, 811 (1989)); phosphinothricin (DeBlock *et al.*, *EMBO J.* **6**, 2513 (1987)); spectinomycin (Bretagne-Sagnard and Chupeau, *Transgenic Research* **5**, 131 (1996)).

Other selectable marker genes include the *pat* gene (for bialaphos and phosphinothricin resistance), the *ALS* gene for imidazolinone resistance, the *HPH* or *HYG* gene for hygromycin resistance, the *Hm1* gene for resistance to the Hc-toxin, and other selective agents used routinely and known to one of ordinary skill in the art. See generally, Yarranton, *Curr. Opin. Biotech.* **3**, 506 (1992); Christopherson *et al.*, *Proc. Natl. Acad. Sci. USA* **89**, 6314 (1992); Yao *et al.*, *Cell* **71**, 63 (1992); Reznikoff, *Mol. Microbiol.* **6**, 2419 (1992); BARKLEY ET AL., THE OPERON 177-220 (1980); Hu *et al.*, *Cell* **48**, 555 (1987); Brown *et al.*, *Cell* **49**, 603 (1987); Figge *et al.*, *Cell* **52**, 713 (1988); Deuschle *et al.*, *Proc. Natl. Acad. Sci. USA* **86**, 5400 (1989); Fuerst *et al.*, *Proc. Natl. Acad. Sci. USA* **86**, 2549 (1989); Deuschle *et al.*, *Science* **248**, 480 (1990); Labow *et al.*, *Mol. Cell. Biol.* **10**, 3343 (1990); Zambretti *et al.*, *Proc. Natl. Acad. Sci. USA* **89**, 3952 (1992); Baim *et al.*, *Proc. Natl. Acad. Sci. USA* **88**, 5072 (1991); Wyborski *et al.*, *Nuc. Acids Res.* **19**, 4647 (1991); Hillenand-Wissman, *Topics in Mol. And Struc. Biol.* **10**, 143 (1989); Degenkolb *et al.*, *Antimicrob. Agents Chemother.* **35**, 1591 (1991); Kleinschmidt *et al.*, *Biochemistry* **27**, 1094 (1988); Gatz *et al.*, *Plant J.* **2**, 397 (1992); Gossen *et al.*, *Proc. Natl. Acad. Sci. USA* **89**, 5547 (1992); Oliva *et al.*, *Antimicrob. Agents Chemother.* **36**, 913 (1992); HLAVKA ET AL., HANDBOOK OF EXPERIMENTAL PHARMACOLOGY 78 (1985); and Gill *et al.*, *Nature* **334**, 721 (1988). Such disclosures are herein incorporated by reference in their entireties.

The heterologous nucleotide sequence of interest can additionally be operably linked to a sequence that encodes a transit peptide that directs expression of an encoded polypeptide of interest to a particular cellular compartment. Transit peptides that target protein accumulation in higher
5 plant cells to the chloroplast, mitochondrion, vacuole, nucleus, and the endoplasmic reticulum (for secretion outside of the cell) are known in the art. Transit peptides that target proteins to the endoplasmic reticulum are desirable for correct processing of secreted proteins. Targeting protein expression to the chloroplast (for example, using the transit peptide from the
10 RubP carboxylase small subunit gene) has been shown to result in the accumulation of very high concentrations of recombinant protein in this organelle. The pea RubP carboxylase small subunit transit peptide sequence has been used to express and target mammalian genes in plants (U.S. Patent Nos. 5,717,084 and 5,728,925 to Herrera-Estrella *et al.*). Alternatively,
15 mammalian transit peptides can be used to target recombinant protein expression, for example, to the mitochondrion and endoplasmic reticulum. It has been demonstrated that plant cells recognize mammalian transit peptides that target endoplasmic reticulum (U.S. Patent Nos. 5,202,422 and 5,639,947 to Hiatt *et al.*).

20 Further, the expression cassette can comprise a 5' leader sequence that acts to enhance expression (transcription, post-transcriptional processing and/or translation) of an operably associated heterologous sequence. Leader sequences are known in the art and include sequences from: picornavirus leaders, e.g., EMCV leader (Encephalomyocarditis 5' noncoding region;
25 Elroy-Stein *et al.*, *Proc. Natl. Acad. Sci USA*, **86**, 6126 (1989));; potyvirus leaders, e.g., TEV leader (Tobacco Etch Virus; Allison *et al.*, *Virology*, **154**, 9 (1986));; human immunoglobulin heavy-chain binding protein (BiP; Macajak and Sarnow, *Nature* **353**, 90 (1991));; untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4; Jobling and Gehrke, *Nature* **325**,
30 622 (1987));; tobacco mosaic virus leader (TMV; Gallie, *MOLECULAR BIOLOGY OF RNA*, 237-56 (1989));; and maize chlorotic mottle virus leader (MCMV; Lommel *et al.*, *Virology* **81**, 382 (1991)). See also, Della-Cioppa *et al.*, *Plant Physiology* **84**, 965 (1987).

The heterologous nucleotide sequence(s) in the expression cassette can be any nucleotide sequence(s) of interest and can be obtained from prokaryotes or eukaryotes (e.g., bacteria, fungi, yeast, viruses, plants, mammals) or the nucleotide sequence can be synthesized in whole or in part.

5 Further, the heterologous nucleotide sequence can encode a polypeptide or can be transcribed to produce a non-translated RNA (e.g., an antisense RNA, an RNAi molecule, and the like). In particular embodiments, the non-translated RNA can be expressed to improve an agronomic trait in the plant (e.g., drought resistance, heat resistance, salt resistance, disease resistance,

10 insect and other pest resistance [e.g., a *Bacillus thuringiensis* endotoxin], herbicide resistance, and the like), to confer male sterility, to improve fertility and/or enhance nutritional quality (e.g., enzymes that enhance nutritional quality). The nucleotide sequence may further be used in the sense orientation to achieve suppression of endogenous plant genes, as is known

15 by those skilled in the art (see, e.g., U.S. Patent Nos. 5,283,184; 5,034,323; the disclosures of which are incorporated by reference herein in their entireties).

The heterologous nucleotide sequence can encode a polypeptide that imparts a desirable agronomic trait to the plant (as described above), confers

20 male sterility, improves fertility and/or improves nutritional quality. Other suitable polypeptides include enzymes that can degrade organic pollutants or remove heavy metals. Such plants, and the enzymes that can be isolated therefrom, are useful in methods of environmental protection and remediation. Alternatively, the heterologous nucleotide sequence can encode a

25 therapeutically or pharmaceutically useful polypeptide or an industrial polypeptide (e.g., an industrial enzyme). Such polypeptides include, but are not limited to antibodies and antibody fragments, cytokines, hormones, growth factors, receptors, enzymes and the like. In general, the heterologous nucleotide sequence can be any sequence other than the endogenous *rubi3*

30 coding sequence associated with the *rubi3* promoter/intron in the naturally occurring *rubi3* gene.

Heterologous nucleotide sequences suitable to confer tolerance to the herbicide glyphosate include, but are not limited to the *Agrobacterium* strain

CP4 glyphosate resistant EPSPS gene (aroA:CP4) as described in U.S. Patent No. 5,633,435 or the glyphosate oxidoreductase gene (GOX) as described in U.S. Patent No. 5,463,175. Other heterologous nucleotide sequences include genes conferring resistance to herbicides that act to inhibit
5 the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., mutant forms of the acetolactate synthase (ALS) gene that lead to such resistance, in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides that act to inhibit the action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene). The bar
10 gene encodes resistance to the herbicide basta, the nptII gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Suitable heterologous nucleotide sequences that confer insect tolerance include those which provide resistance to pests such as rootworm,
15 cutworm, European Corn Borer, and the like. Exemplary nucleotide sequences include, but are not limited to, a *Bacillus* insect control protein gene (see, e.g., WO 99/31248; U.S. Patent Nos. 5,689,052; 5,500,365; 5,880,275); *Bacillus thuringiensis* toxic protein genes (see, e.g., U.S. Patent Nos. 5,366,892; 5,747,450; 5,737,514; 5,723,756; 5,593,881; 6,555,655;
20 6,541,448; 6,538,109; Geiser, et al. (1986) *Gene* 48:109); and lectins (Van Damme et al. (1994) *Plant Mol. Biol.* 24:825).

Alternatively, the heterologous nucleotide sequence can encode a reporter polypeptide (e.g., an enzyme), including but not limited to Green Fluorescent Protein, β -galactosidase, luciferase, alkaline phosphatase, the
25 GUS gene encoding β -glucuronidase, and chloramphenicol acetyltransferase.

Where appropriate, the heterologous nucleic acids may be optimized for increased expression in a transformed plant, e.g., by using plant preferred codons. Methods for synthetic optimization of nucleic acid sequences are available in the art. The nucleotide sequence can be optimized for expression
30 in a particular host plant or alternatively can be modified for optimal expression in monocots. See, e.g., EP 0 359 472, EP 0 385 962, WO 91/16432; Perlak et al., *Proc. Natl. Acad. Sci. USA* 88, 3324 (1991), and Murray et al., *Nuc. Acids Res.* 17, 477 (1989), and the like, herein

incorporated by reference. Plant preferred codons can be determined from the codons of highest frequency in the proteins expressed in that plant.

Additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences
5 encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other such well-characterized sequences which may be deleterious to gene expression. The G-C content of the sequence may be adjusted to levels average for a given cellular host, as calculated by reference to known genes expressed in the host cell. When possible, the
10 sequence is modified to avoid predicted hairpin secondary mRNA structures.

The invention further provides vectors comprising the nucleic acids and expression cassettes of the invention, including expression vectors, transformation vectors and vectors for replicating and/or manipulating the nucleotide sequences in the laboratory. The vector can be a plant vector,
15 animal (e.g., insect or mammalian) vector, bacterial vector, yeast vector or fungal vector. Generally, according to the present invention, the vector is a plant vector, a bacterial vector, or a shuttle vector that can replicate in either host under appropriate conditions. Bacterial and plant vectors are well-known in the art. Exemplary plant vectors include plasmids (e.g., pUC or the Ti
20 plasmid), cosmids, phage, bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs) and plant viruses.

III. Methods of Expressing a Nucleotide Sequence of Interest.

The invention also provides methods of delivering a nucleic acid,
25 expression cassette or vector of the invention to a target plant or plant cell (including callus cells or protoplasts), plant parts, seed, plant tissue (including callus), or the like. The invention further comprises host plants, cells, plant parts, seeds, tissue culture (including callus) transiently or stably transformed with the nucleic acids, expression cassettes or vectors of the invention.

30 As used herein, unless indicated otherwise, a "plant cell" includes callus cells and protoplasts. The term "plant tissue" includes callus tissue as well as tissue culture. Further, the term "tissue culture" encompasses cultures of tissue, cells, protoplasts and callus.

Plant parts, organs and tissues in which the *rubi3* promoter may be useful in expressing a heterologous nucleotide sequence of interest include, but are not limited to reproductive tissues (e.g., petals, sepals, stamens, pistils, receptacles, anthers, pollen, flowers, fruits, flower bud, ovules, seeds, 5 embryos, and nuts); vegetative tissues (e.g., petioles, stems, roots, root hairs, root tips, pith, coleoptiles, stalks, shoots, bark, apical meristem, axillary bud, cotyledon, hypocotyls, and leaves); vascular tissues (e.g., phloem and xylem); specialized cells such as epidermal cells, parenchyma cells, chlorenchyma cells, schlerenchyma cells, stomates, guard cells, cuticle, mesophyll cells; 10 callus tissue; and cuttings.

Any plant can be employed in practicing the present invention including angiosperms or gymnosperms, monocots or dicots. In particular embodiments, the plant is a monocot.

Exemplary plants include, but are not limited to corn (*Zea mays*), 15 canola (*Brassica napus*, *Brassica rapa* ssp.), alfalfa (*Medicago saliva*), rice (*Oryza sativa*), rape (*Brassica napus*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), sunflower (*Helianthus annus*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton 20 (*Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), 25 macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), apple (*Malus pumila*), blackberry (*Rubus*), strawberry (*Fragaria*), walnut (*Juglans regia*), grape (*Vitis vinifera*), apricot (*Prunus armeniaca*), cherry (*Prunus*), peach (*Prunus persica*), plum (*Prunus domestica*), pear (*Pyrus communis*), watermelon (*Citrullus vulgaris*). 30 duckweed (*Lemna*), oats, barley, vegetables, ornamentals, conifers, and turfgrasses (e.g., for ornamental, recreational or forage purposes).

Vegetables include Solanaceous species (e.g., tomatoes; *Lycopersicon esculentum*), lettuce (e.g., *Lactuea sativa*), carrots (*Caucus carota*), cauliflower (*Brassica oleracea*), celery (*apium graveolens*), eggplant (*Solanum melongena*), asparagus (*Asparagus officinalis*), ochra (5 *Abelmoschus esculentus*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), members of the genus *Cucurbita* such as Hubbard squash (*C. Hubbard*), Butternut squash (*C. moschata*), Zucchini (*C. pepo*), Crookneck squash (*C. crookneck*), *C. argyrosperma*, *C. argyrosperma* ssp *sororia*, *C. digitata*, *C. ecuadorensis*, *C. foetidissima*, *C.* 10 *lundelliana*, and *C. martinezii*, and members of the genus *Cucumis* such as cucumber (*Cucumis sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*).

Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* 15 spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*dianthus caryophyllus*), poinsettia (*Euphorbia pulcherima*), and chrysanthemum.

Conifers, which may be employed in practicing the present invention, include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine 20 (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); 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 25 cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*).

Turfgrass include but are not limited to zoysiagrasses, bentgrasses, fescue grasses, bluegrasses, St. Augustinegrasses, bermudagrasses, bufallograsses, ryegrasses, and orchardgrasses.

Also included are plants that serve primarily as laboratory models, e.g., 30 *Arabidopsis*.

In particular embodiments, the invention is used to transform a monocotyledonous plant (or plant part, plant tissue, plant cell, callus, seed, etc.). Exemplary monocotyledonous species include corn (*Zea mays*); millet

such as pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), or finger millet (*Eleusine coracana*); rice (*Oryza sativa*); rye (*Secale cereale*); sorghum (*Sorghum bicolor*, *Sorghum vulgare*); wheat (*Triticum aestivum*); banana and plantain (*Musa* spp.); sugarcane
5 (*Saccharum* spp.), duckweed (*Lemna*); oats; barley; turfgrasses as well as forage and biomass grasses.

In particular embodiments, the invention provides a method of producing a polypeptide of interest from a cell or tissue culture, wherein the cultured cells or tissue comprise an isolated nucleic acid, expression construct
10 or vector (all as described above) according to the present invention. In one particular embodiment, the method comprises culturing a plant cell or tissue comprising an isolated nucleic acid, expression construct or vector according to the invention comprising a heterologous nucleotide sequence encoding a polypeptide of interest under conditions sufficient for the polypeptide of
15 interest to be expressed by the cultured cells or tissue and, optionally, collecting the polypeptide. According to this embodiment, the polypeptide of interest can be a secreted polypeptide. The method can further comprise introducing the isolated nucleic acid, expression construct or vector of the invention into the cultured plant cell or tissue.

20 Methods of introducing nucleic acids, transiently or stably, into plants, plant tissues, cells, protoplasts, seed, callus and the like are known in the art. Stably transformed nucleic acids can be incorporated into the genome or stably maintained as episomes. Exemplary transformation methods include biological methods using viruses and *Agrobacterium*, physicochemical
25 methods such as electroporation, floral dip methods, polyethylene glycol, ballistic bombardment, microinjection, and the like. Other transformation technology includes the whiskers technology that is based on mineral fibers (see e.g., U.S. Patent No. 5,302,523 and 5,464,765) and pollen tube transformation.

30 In one form of direct transformation, the vector is microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA (Crossway, *Mol. Gen. Genetics* **202**: 179 (1985)).

In another protocol, the genetic material is transferred into the plant cell using polyethylene glycol (Krens, *et al. Nature* **296**, 72 (1982)).

In still another method, protoplasts are fused with minicells, cells, lysosomes, or other fusible lipid-surfaced bodies that contain the nucleotide
5 sequence to be transferred to the plant (Fraley, *et al., Proc. Natl. Acad. Sci. USA* **79**, 1859 (1982)).

Nucleic acids may also be introduced into the plant cells by electroporation (Fromm *et al., Proc. Natl. Acad. Sci. USA* **82**, 5824 (1985)). In this technique, plant protoplasts are electroporated in the presence of nucleic
10 acids comprising the expression cassette. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the nucleic acid. Electroporated plant protoplasts reform the cell wall, divide and regenerate. One advantage of electroporation is that large pieces of DNA, including artificial chromosomes, can be transformed by this method.

15 Ballistic transformation typically comprises the steps of: (a) providing a plant material as a target; (b) propelling a microprojectile carrying the heterologous nucleotide sequence at the plant target at a velocity sufficient to pierce the walls of the cells within the target and to deposit the nucleotide sequence within a cell of the target to thereby provide a transformed target.
20 The method can further include the step of culturing the transformed target with a selection agent and, optionally, regeneration of a transformed plant. As noted below, the technique may be carried out with the nucleotide sequence as a precipitate (wet or freeze-dried) alone, in place of the aqueous solution containing the nucleotide sequence.

25 Any ballistic cell transformation apparatus can be used in practicing the present invention. Exemplary apparatus are disclosed by Sandford *et al. (Particulate Science and Technology* **5**, 27 (1988)), Klein *et al. (Nature* **327**, 70 (1987)), and in EP 0 270 356. Such apparatus have been used to transform maize cells (Klein *et al., Proc. Natl. Acad. Sci. USA* **85**, 4305
30 (1988)), soybean callus (Christou *et al., Plant Physiol.* **87**, 671 (1988)), McCabe *et al., BioTechnology* **6**, 923 (1988), yeast mitochondria (Johnston *et al., Science* **240**, 1538 (1988)), and *Chlamydomonas* chloroplasts (Boynton *et al., Science* **240**, 1534 (1988)).

Alternately, an apparatus configured as described by Klein *et al.* (*Nature* **70**, 327 (1987)) may be utilized. This apparatus comprises a bombardment chamber, which is divided into two separate compartments by an adjustable-height stopping plate. An acceleration tube is mounted on top of the bombardment chamber. A macroprojectile is propelled down the acceleration tube at the stopping plate by a gunpowder charge. The stopping plate has a borehole formed therein, which is smaller in diameter than the microprojectile. The macroprojectile carries the microprojectile(s), and the macroprojectile is aimed and fired at the borehole. When the macroprojectile is stopped by the stopping plate, the microprojectile(s) is propelled through the borehole. The target is positioned in the bombardment chamber so that a microprojectile(s) propelled through the bore hole penetrates the cell walls of the cells in the target and deposit the nucleotide sequence of interest carried thereon in the cells of the target. The bombardment chamber is partially evacuated prior to use to prevent atmospheric drag from unduly slowing the microprojectiles. The chamber is only partially evacuated so that the target tissue is not desiccated during bombardment. A vacuum of between about 400 to about 800 millimeters of mercury is suitable.

In alternate embodiments, ballistic transformation is achieved without use of microprojectiles. For example, an aqueous solution containing the nucleotide sequence of interest as a precipitate may be carried by the macroprojectile (e.g., by placing the aqueous solution directly on the plate-contact end of the macroprojectile without a microprojectile, where it is held by surface tension), and the solution alone propelled at the plant tissue target (e.g., by propelling the macroprojectile down the acceleration tube in the same manner as described above). Other approaches include placing the nucleic acid precipitate itself ("wet" precipitate) or a freeze-dried nucleotide precipitate directly on the plate-contact end of the macroprojectile without a microprojectile. In the absence of a microprojectile, it is believed that the nucleotide sequence must either be propelled at the tissue target at a greater velocity than that needed if carried by a microprojectile, or the nucleotide sequenced caused to travel a shorter distance to the target (or both).

It particular embodiments, the nucleotide sequence is delivered by a microprojectile. The microprojectile can be formed from any material having sufficient density and cohesiveness to be propelled through the cell wall, given the particle's velocity and the distance the particle must travel. Non-limiting examples of materials for making microprojectiles include metal, glass, silica, ice, polyethylene, polypropylene, polycarbonate, and carbon compounds (e.g., graphite, diamond). Non-limiting examples of suitable metals include tungsten, gold, and iridium. The particles should be of a size sufficiently small to avoid excessive disruption of the cells they contact in the target tissue, and sufficiently large to provide the inertia required to penetrate to the cell of interest in the target tissue. Particles ranging in diameter from about one-half micrometer to about three micrometers are suitable. Particles need not be spherical, as surface irregularities on the particles may enhance their carrying capacity.

The nucleotide sequence may be immobilized on the particle by precipitation. The precise precipitation parameters employed will vary depending upon factors such as the particle acceleration procedure employed, as is known in the art. The carrier particles may optionally be coated with an encapsulating agents such as polylysine to improve the stability of nucleotide sequences immobilized thereon, as discussed in EP 0 270 356 (column 8).

Alternatively, plants may be transformed using *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. *Agrobacterium*-mediated nucleic acid transfer exploits the natural ability of *A. tumefaciens* and *A. rhizogenes* to transfer DNA into plant chromosomes. *Agrobacterium* is a plant pathogen that transfers a set of genes encoded in a region called T-DNA of the Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, into plant cells. The typical result of transfer of the Ti plasmid is a tumorous growth called a crown gall in which the T-DNA is stably integrated into a host chromosome. Integration of the Ri plasmid into the host chromosomal DNA results in a condition known as "hairy root disease". The ability to cause disease in the host plant can be removed by deletion of the genes in the T-DNA without loss

of DNA transfer and integration. The DNA to be transferred is attached to border sequences that define the end points of an integrated T-DNA.

Transfer by means of engineered *Agrobacterium* strains has become routine for many dicotyledonous plants. Some difficulty has been

5 experienced, however, in using *Agrobacterium* to transform monocotyledonous plants, in particular, cereal plants. However, *Agrobacterium* mediated transformation has been achieved in several monocot species, including cereal species such as rye, maize (Rhodes *et al.*, *Science* **240**, 204 (1988)), and rice (Hiei *et al.*, (1994) *Plant J.* **6**:271).

10 While the following discussion will focus on using *A. tumefaciens* to achieve gene transfer in plants, those skilled in the art will appreciate that this discussion also applies to *A. rhizogenes*. Transformation using *A. rhizogenes* has developed analogously to that of *A. tumefaciens* and has been successfully utilized to transform, for example, alfalfa, *Solanum nigrum* L., and
15 poplar (U.S. Patent No. 5,777,200 to Ryals *et al.*). As described by U.S. Patent No. 5, 773,693 to Burgess *et al.*, it is preferable to use a disarmed *A. tumefaciens* strain (as described below), however, the wild-type *A. rhizogenes* may be employed. An illustrative strain of *A. rhizogenes* is strain 15834.

In particular protocols, the *Agrobacterium* strain is modified to contain
20 the nucleotide sequences to be transferred to the plant. The nucleotide sequence to be transferred is incorporated into the T-region and is typically flanked by at least one T-DNA border sequence, optionally two T-DNA border sequences. A variety of *Agrobacterium* strains are known in the art particularly, and can be used in the methods of the invention. See, e.g.,
25 Hooykaas, *Plant Mol. Biol.* **13**, 327 (1989); Smith *et al.*, *Crop Science* **35**, 301 (1995); Chilton, *Proc. Natl. Acad. Sci. USA* **90**, 3119 (1993); Mollony *et al.*, *Monograph Theor. Appl. Genet NY* **19**, 148 (1993); Ishida *et al.*, *Nature Biotechnol.* **14**, 745 (1996); and Komari *et al.*, *The Plant Journal* **10**, 165 (1996), the disclosures of which are incorporated herein by reference.

30 In addition to the T-region, the Ti (or Ri) plasmid contains a *vir* region. The *vir* region is important for efficient transformation, and appears to be species-specific.

Two exemplary classes of recombinant Ti and Ri plasmid vector systems are commonly used in the art. In one class, called "cointegrate," the shuttle vector containing the gene of interest is inserted by genetic recombination into a non-oncogenic Ti plasmid that contains both the cis-
5 acting and trans-acting elements required for plant transformation as, for example, in the PMLJ1 shuttle vector of DeBlock *et al.*, *EMBO J* **3**, 1681 (1984), and the non-oncogenic Ti plasmid pGV2850 described by Zambryski *et al.*, *EMBO J* **2**, 2143 (1983). In the second class or "binary" system, the gene of interest is inserted into a shuttle vector containing the cis-acting
10 elements required for plant transformation. The other necessary functions are provided in trans by the non-oncogenic Ti plasmid as exemplified by the pBIN19 shuttle vector described by Bevan, *Nucleic Acids Research* **12**, 8711 (1984), and the non-oncogenic Ti plasmid PAL4404 described by Hoekma, *et al.*, *Nature* **303**, 179 (1983).

15 Binary vector systems have been developed where the manipulated disarmed T-DNA carrying the heterologous nucleotide sequence of interest and the *vir* functions are present on separate plasmids. In this manner, a modified T-DNA region comprising foreign DNA (the nucleic acid to be transferred) is constructed in a small plasmid that replicates in *E. coli*. This
20 plasmid is transferred conjugatively in a tri-parental mating or via electroporation into *A. tumefaciens* that contains a compatible plasmid with virulence gene sequences. The *vir* functions are supplied in *trans* to transfer the T-DNA into the plant genome. Such binary vectors are useful in the practice of the present invention.

25 In particular embodiments of the invention, super-binary vectors are employed. See, e.g., United States Patent No. 5,591,615 and EP 0 604 662, herein incorporated by reference. Such a super-binary vector has been constructed containing a DNA region originating from the hypervirulence region of the Ti plasmid pTiBo542 (Jin *et al.*, *J. Bacteriol.* **169**, 4417 (1987))
30 contained in a super-virulent *A. tumefaciens* A281 exhibiting extremely high transformation efficiency (Hood *et al.*, *Biotechnol.* **2**, 702 (1984); Hood *et al.*, *J. Bacteriol.* **168**, 1283 (1986); Komari *et al.*, *J. Bacteriol.* **166**, 88 (1986); Jin

et al., *J. Bacteriol.* **169**, 4417 (1987); Komari, *Plant Science* **60**, 223 (1987); ATCC Accession No. 37394.

Exemplary super-binary vectors known to those skilled in the art include pTOK162 (Japanese patent Appl. (Kokai) No. 4-222527, EP 504,869, EP 604,662, and United States Patent No. 5,591,616, herein incorporated by reference) and pTOK233 (Komari, *Plant Cell Reports* **9**, 303 (1990); Ishida *et al.*, *Nature Biotechnology* **14**, 745 (1996); herein incorporated by reference). Other super-binary vectors may be constructed by the methods set forth in the above references. Super-binary vector pTOK162 is capable of replication in both *E. coli* and in *A. tumefaciens*. Additionally, the vector contains the *virB*, *virC* and *virG* genes from the virulence region of pTiBo542. The plasmid also contains an antibiotic resistance gene, a selectable marker gene, and the nucleic acid of interest to be transformed into the plant. The nucleic acid to be inserted into the plant genome is typically located between the two border sequences of the T region. Super-binary vectors of the invention can be constructed having the features described above for pTOK162. The T-region of the super-binary vectors and other vectors for use in the invention are constructed to have restriction sites for the insertion of the genes to be delivered. Alternatively, the DNA to be transformed can be inserted in the T-DNA region of the vector by utilizing *in vivo* homologous recombination. See, Herrera-Esterella *et al.*, *EMBO J.* **2**, 987 (1983); Horsch *et al.*, *Science* **223**, 496 (1984). Such homologous recombination relies on the fact that the super-binary vector has a region homologous with a region of pBR322 or other similar plasmids. Thus, when the two plasmids are brought together, a desired gene is inserted into the super-binary vector by genetic recombination via the homologous regions.

In plants stably transformed by *Agrobacteria*-mediated transformation, the nucleotide sequence of interest is incorporated into the plant nuclear genome, typically flanked by at least one T-DNA border sequence and generally two T-DNA border sequences.

Plant cells may be transformed with *Agrobacteria* by any means known in the art, e.g., by co-cultivation with cultured isolated protoplasts, or transformation of intact cells or tissues. The first uses an established culture

system that allows for culturing protoplasts and subsequent plant regeneration from cultured protoplasts. Identification of transformed cells or plants is generally accomplished by including a selectable marker in the transforming vector, or by obtaining evidence of successful bacterial infection.

5 Protoplasts, which have been transformed by any method known in the art, can also be regenerated to produce intact plants using known techniques.

Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Handbook of Plant Cell Cultures*, Vol. 1: (MacMilan Publishing Co. New York, 1983); and Vasil I. R. (ed.), *Cell Culture and Somatic Cell Genetics of*
10 *Plants*, Acad. Press, Orlando, Vol. I, 1984, and Vol. II, 1986). Essentially all plant species can be regenerated from cultured cells or tissues, including but not limited to, all major species of sugar-cane, sugar beet, cotton, fruit trees, and legumes.

Means for regeneration vary from species to species of plants, but
15 generally a suspension of transformed protoplasts or a petri plate containing transformed explants is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently root. Alternatively, somatic embryo formation can be induced in the callus tissue. These somatic embryos germinate as natural embryos to form plants. The culture media will generally
20 contain various amino acids and plant hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is usually
25 reproducible and repeatable.

The regenerated plants are transferred to standard soil conditions and cultivated in a conventional manner. The plants are grown and harvested using conventional procedures.

Alternatively, transgenic plants may be produced using the floral dip
30 method (See, *e.g.*, Clough and Bent (1998) *Plant Journal* 16:735-743, which avoids the need for plant tissue culture or regeneration. In one representative protocol, plants are grown in soil until the primary inflorescence is about 10 cm tall. The primary inflorescence is cut to induce the emergence of multiple

secondary inflorescences. The inflorescences of these plants are typically dipped in a suspension of *Agrobacterium* containing the vector of interest, a simple sugar (e.g., sucrose) and surfactant. After the dipping process, the plants are grown to maturity and the seeds are harvested. Transgenic seeds
5 from these treated plants can be selected by germination under selective pressure (e.g., using the chemical bialaphos). Transgenic plants containing the selectable marker survive treatment and can be transplanted to individual pots for subsequent analysis. See Bechtold, N. and Pelletier, G. Methods Mol Biol 82, 259-266 (1998); Chung, M.H. et al. Transgenic Res 9, 471-476
10 (2000); Clough, S.J. and Bent, A.F. Plant J 16, 735-743 (1998); Mysore, K.S. et al. Plant J 21, 9-16 (2000); Tague, B.W. Transgenic Res 10, 259-267 (2001); Wang, W.C. et al. Plant Cell Rep 22, 274-281 (2003); Ye, G.N. et al. Plant J., 19:249-257 (1999).

The particular conditions for transformation, selection and regeneration
15 can be optimized by those of skill in the art. Factors that affect the efficiency of transformation include the species of plant, the target tissue or cell, composition of the culture media, selectable marker genes, kinds of vectors, and light/dark conditions. Therefore, these and other factors may be varied to determine what is an optimal transformation protocol for any particular plant
20 species. It is recognized that not every species will react in the same manner to the transformation conditions and may require a slightly different modification of the protocols disclosed herein. However, by altering each of the variables, an optimum protocol can be derived for any plant species.

Having described the present invention, the same will be explained in
25 greater detail in the following examples, which are included herein for illustration purposes only, and which are not intended to be limiting to the invention.

EXAMPLE 1

30

Plant Material

Rice seeds (cv. Nipponbare) were obtained from the National Small Grains Collection (USDA-ARS, Aberdeen, ID). Eighteen-month-old fine rice

suspension cells derived from primary calli raised from mature seeds were maintained in 125 ml flasks containing 50 ml of AA medium (Müller and Grafe (1978) *Mol. Gen. Genet.* 161:67-76) on a gyratory shaker (125 rpm). Cultures were grown in the dark at 25°C and subcultured once a week.

5

EXAMPLE 2

Isolation of *rubi3* Sequences

A rice genome sequence database (<http://www.rice-research.org>) was used in the BLAST search for polyubiquitin gene sequences. The genomic
10 sequence for *rubi3* was identified in BAC clone OSM14960/14965 (Monsanto, St. Louis, MO).

Genomic DNA from rice leaves was isolated using well-established methods (Dellaporta, *et al.* (1983) *Plant Mol. Biol. Rep.* 14:19-21). The 5' flanking region of *rubi3* containing the promoter region plus *rubi3* intron (SEQ
15 ID NO:1; Figure 1A), *rubi3* promoter alone (SEQ ID NO:2; Figure 1B) and *rubi3* intron alone (SEQ ID NO:3; Figure 1C) were amplified using the following primers: 4 - 5'-GTC GAC CAC CCA ACC CCA TAT CGA CAG AGG-3' (SEQ ID NO:4); UR7sma - 5'-CCC GGG CTG GAA GAG GCA AGA AAG GAT TGG AAT TAA C-3' (SEQ ID NO:5); 2sma - 5'-CCC GGG TAA CTA ATC
20 AAT CAC CTC GT-3' (SEQ ID NO:6); UR8 - 5'-CTT GAC GAG GCG ATT AGA GAA CGC-3' (SEQ ID NO:7).

The *rubi3* gene (SEQ ID NO:8) contained an open reading frame of 1146-bp, arranged as five tandem, head-to-tail repeats of 228-bp, encoding a pentameric precursor ubiquitin (Figures 2A and 2B). The deduced amino acid
25 sequence of the five ubiquitin monomers of *rubi3* (SEQ ID NO:9) was identical to ubiquitin sequences of maize and *A. thaliana*.

EXAMPLE 3

Characterization of the *rubi3* Gene

30 The coding sequence of the *rubi3* ubiquitin gene shared a high degree of homology with ubiquitin genes from *Arabidopsis* (Burke, *et al.* (1988) *supra*; Callis, *et al.* (1995) *Genetics* 139:921-939), sunflower (Binet, *et al.* (1991) *supra*), potato (Garbarino and Belknap (1994) *supra*), tomato (Hoffman, *et al.*

(1991) *Plant Mol. Biol.* 17:1189-1201), *N. tabacum* (Genschik, *et al.* (1994) *supra*), sugarcane (Albert, *et al.* (1995) *Plant Physiol.* 109:337) and yeast (Ozkaynak, *et al.* (1987) *EMBO J.* 6:1427-1439). For example, the coding sequence of the *rubi3* gene showed 84% identity to the maize polyubiquitin gene (Genbank accession number S94464), 79% to the *A. thaliana* ubiquitin gene (accession number U84968), 80% to the sunflower ubiquitin gene (accession number X57004), 86% to the rice RUBQ1 (accession number AF184279), 86% to the rice RUBQ2 (accession number AF184280) and 85% to the rice RUBQ1 (accession number U37687).

10 An intron was identified at the 5' untranslated region (5' UTR) immediately upstream of the ATG translation initiation codon for the *rubi3* gene (**Figure 2A**). the presence of the intron and the precise splice sites for *rubi3* were confirmed by comparison of *rubi3* genomic and cDNA sequences. The *rubi3* intron was 1140 bp long and did not have significant homology with
15 known maize or rice ubiquitin introns (**Figures 3A-3C**) or any other sequences. Intron splice sites of *rubi3* are in good agreement with the consensus sequences CAAG/gta at the 5' end and cag/ATG at the 3' end, which are common to plant polyubiquitin genes (Binet, *et al.* (1991) *supra*; Christensen, *et al.* (1992) *supra*; Kawalleck, *et al.* (1993) *supra*; Norris, *et al.* (1993) *Plant Mol. Biol.* 21:895-906).

A heat shock sequence similar to the *Drosophila* consensus sequence 5'-CTGGAATnTTCTAGA-3' (**SEQ ID NO:10**) (Pelham (1982) *Cell* 30:517-528), is generally present in the isolated ubiquitin promoters (Christensen, *et al.* (1992) *supra*; Wang, *et al.* (2000) *Plant Sci.* 156(2):201-211). Such a
25 consensus sequence was not observed in the *rubi3* promoter sequence. An enhancer core consensus sequence 5'-GGTGTGG(AAA/TTT)G-3' (**SEQ ID NO:11**) (Weiher, *et al.* (1983) *Science* 219:626-631; Khoury and Gruss (1983) *Cell* 33:313-314) is observed in RUBQ1 and RUBQ2 promoter sequences (Wang *et al.* (2000) *supra*); however, this sequence was not
30 observed in the *rubi3* promoter sequence (**Figure 2A**).

EXAMPLE 4

RT-PCR

Reverse transcription (RT)-coupled PCR reactions were performed to obtain the cDNA encoding *rubi3* and to confirm the predicted intron and splice sites. Two primers were synthesized to encompass the *rubi3* coding sequence and 5'- and 3'-UTR sequences. These primers were: Intron1, 5'-CGA ATC GAC CGA AGG GGA GG-3' (**SEQ ID NO:12**) and 5'Race3, 5'-ACA CGA TGA TAT GAC AGA CGA GC-3' (**SEQ ID NO:13**).

Total RNA was prepared using the method of Haffner, *et al.* ((1978) *Can. J. Biochem.* 56:729-733). RT reactions were performed using SUPERScript™ II reverse transcriptase (INVITROGEN™ Corp., Carlsbad, CA) followed by standard PCR reactions using Taq DNA polymease (EPPENDORF®, Westbury, NY). Conditions for RT-PCR reactions were in accordance with manufacturer's recommendations.

EXAMPLE 5

Chimeric gene constructs

A 2-kb PCR product encompassing the *rubi3* promoter and *rubi3* 5' UTR intron was amplified with primers 4 and UR7sma and cloned into the pCR®2.1 to generate pRESQ21. The sequence and orientation of the 2-kb fragment were confirmed and the 2-kb *SpeI/SmaI* fragment was cloned into *XbaI/SmaI*-digested pRESQ8, a pUC119-based vector containing a β -glucuronidase gene (GUS) and a NOS terminator. The resulting plasmid, pRESQ4 (**Figure 4A**), harbored the *rubi3* promoter, intron, GUS and NOS terminator.

The *rubi3* promoter, a 875-bp PCR product obtained using primers 4 and UR8, was cloned into pCR®2.1 to generate pRESQ32. A 0.9-kb *BamHI/EcoRV* fragment of pRESQ32 encompassing the *rubi3* promoter was liberated and ligated into *BamHI/SmaI*-digested pRESQ8 to create pRESQ33 (**Figure 4B**) which harbored the *rubi3* promoter, GUS and NOS.

The *rubi3* intron, a 1.1-kb PCR product obtained using primers 2sma and UR7sma, was cloned into pCR®2.1 to generate pRESQ13. A 1.1-kb *SmaI* fragment of pRESQ13 encompassing the *rubi3* intron was liberated and

ligated into *Sma*I-digested pRESQ8 to create pRESQ18 (**Figure 4C**) which harbored the *rubi3* intron, GUS and NOS.

The 1.1-kb *Sma*I fragment of pRESQ18 was liberated and ligated in the *Sma*I site of pRQ202 which encompasses an intronless rice Actin1D promoter, GUS and NOS. The resulting plasmid, designated pRESQ17 (**Figure 4D**), contained an intronless rice Actin1D promoter, *rubi3* intron, GUS and NOS.

A 3.8-kb *Hind*III cassette containing the maize ubi promoter, GUS and NOS was liberated from pAHC25 (Christensen and Quail (1996) *Transgenic Res.* 5:213–218) and ligated into *Hind*III-digested pUC119 to create pRESQ29. This construct along with pRQ6, which contains the cauliflower mosaic virus CaMV 35S promoter, GUS and NOS, and pAct1D, which contains the rice actin promoter, GUS and NOS, were used as controls.

15

EXAMPLE 6

Transient GUS Assays

The chimeric gene constructs provided herein were used in the transformation of rice suspension cells via particle bombardment, with appropriate controls.

20 Approximately 200 mg of rice suspension cells (cv. Nipponbare) were evenly distributed over the surface of a 15 mm Whatman no. 2 filter paper in a 100 x 15 mm² petri dish for particle bombardment. Plasmids were coated onto gold particles (1.0 μ m in diameter; BIO-RAD[®], Hercules, CA) and introduced into suspension cells via particle bombardment according to the method
25 described by Christou, *et al.* (1991) *BioTechnology* 9:957–962). Equal molar amounts of each construct containing the chimeric GUS gene were transformed into suspension cells with three replications. After a 24-hour incubation at 25°C, cells were transferred into a microfuge tube and homogenized with 400 μ L GUS extraction buffer (Jefferson (1987) *Plant Mol. Biol. Rep.* 5:387-405). Cell homogenates were centrifuged at 10,000 x *g* for 5
30 minutes at 4°C and the supernatant was collected and assayed.

GUS enzyme assays were performed using 20 μ L of supernatant in accordance with well-established methods (Jefferson (1987) *supra*; Gallagher

(1992) In: GUS protocols: using the GUS gene as a reporter of gene expression. S.R. Gallagher ed. Academic Press Inc. pp 47-59) in a FLUOstar BMG fluorometer (BMG Lab Technologies, Durham, NC). GUS activity was expressed as $\mu\text{mol MU}$ (4 methylumbelliferone)/min/mg total protein.

- 5 Simultaneously, histochemical GUS assays (Jefferson (1987) *supra*) were also performed.

Chimeric gene constructs containing the *rubi3* promoter sequence driving expression of GUS were introduced into rice suspension cells (**Figure 5A**) and coleoptiles (**Figure 5B**). A significant number of blue spots
10 were observed as compared with control samples (**Figures 5A and 5B**).

The activity of the *rubi3* promoter and the effect of the *rubi3* intron on gene expression were analyzed. **Figures 4A-4C** show schematic features of chimeric genes containing the *rubi3* promoter. These constructs were introduced into rice suspension cells via particle bombardment. GUS activity
15 (pmol MU/min/mg total protein) was determined and the relative values across multiple independent experiments were consistent. The quantitative MUG assays and the histochemical assays provided similar results.

When compared with other commonly used promoters, the *rubi3* promoter with the *rubi3* intron exhibited more than a 10-fold higher level of
20 activity than the CaMV 35S promoter, 30% lower than the maize *ubi1* promoter, and nearly a 50% lower level of expression than the rice Act1 promoter (**Table 1; Figure 6**). However, the expression of rice Act1D construct may have been positively affected by the first six amino acid residues of the Act1 coding region which was fused to the GUS gene.

25 The *rubi3* intron had a significant, positive effect on GUS reporter gene expression. A 20-fold higher level of expression was observed from constructs containing the *rubi3* intron (pRESQ4) as compared to constructs lacking the *rubi3* intron (pRESQ33) (**Table 1; Figure 6**). Moreover, when the *rubi3* intron was placed downstream of the rice actin promoter which lacked its own intron
30 (pRESQ17), the enhancement of transient GUS activity was nearly 18-fold more than that of an intronless Act1 promoter construct (pRQ202). These results indicate that the *rubi3* intron enhances the activity of other unrelated promoters.

In other studies, expression from the rice H3 promoter isolated by the inventors' laboratory was enhanced by 21-fold by the *rub13* intron (data not shown).

5

TABLE 1

Plasmid construct	Promoter	Transient GUS activity μMoles MU/min/mg total protein*
pRESQ4	<i>rub13</i> with the <i>rub13</i> intron	14.74
pRESQ33	<i>rub13</i> without an intron	0.74
pRQ202	Rice Act1 without an intron	1.18
	Rice Act1 with the <i>rub13</i> intron	
pRESQ17		20.88
pAct1D	Rice Act1 with the Act1 intron	27.25
	Maize ubi1 with the ubi1 intron	
pRESQ29		21.35
pRQ6	CaMV 35S	1.35
Control		0.63

* Values represent the mean of three independent bombardments of rice suspension cells

10

EXAMPLE 7

Rub13 Promoter Activity Enhancement

To further enhance the levels of expression of proteins fused to *rub13* regulatory regions, three constructs were generated which fused the coding sequence of a full monomer, first nine amino acids, or first three amino acids of the *rub13* gene in-frame with nucleic acid sequences encoding GUS. Using standard cloning and three-primer PCR techniques (Yon and Fried (1989) *Nucl. Acid Res.* 17(12):4895), precise translational fusions were generated. All fusions were verified by DNA sequence analysis. The resulting constructs were designated pRESQ 38 encoding a translational fusion between GUS and the first monomer of RUBI3, pRESQ36 encoding a translational fusion between GUS and the first nine amino acids of RUBI3, and pRESQ 42 encoding a translational fusion between GUS and the first three amino acids of RUBI3.

The constructs were tested in rice callus lines via transient GUS assays. Rice suspension cells (cv. Nipponbare) were bombarded in triplicates

and repeated at least once with respective plasmid constructs.

Transformation through biolistic bombardment and transient MUG assays were performed as described herein.

The results of these experiments indicated that fusions proteins
5 between *rubi3* and GUS enhanced the activity of GUS by 4- to 4.5-fold, with the three amino acid fusion construct being the highest (about 1.5-fold of a maize *ubi1*-GUS construct)(**Figure 8**). Thus, the first nine nucleotides of the *rubi3* coding sequence (ATGCAGATA) are sufficient for enhancing the expression of transgenes from *rubi3* regulatory sequences.

10 To maintain a high level of expression and eliminate the additional amino acids at the N-terminus of a transgene of interest, the third nucleotide of the initiation codon of *rubi3* was mutated to disrupt translation initiation from the *rubi3* coding sequence, *i.e.*, so that translation starts from the ATG codon of the transgene of interest. Accordingly, the G of the *rubi3* initiation codon
15 was mutated to A, T, or C (**Figure 7**). Using standard methods, site-directed mutagenesis of the third nucleotide of the initiation codon in the first 9 nucleotides of the *rubi3* coding sequence (ATGCAGATA) and fusion with GUS was carried out by PCR using vector pRESQ42 as a template (**Figure 7**). The resulting vectors were designated pRESQ46 which contained a G to
20 A substitution, pRESQ47 which contained a G to T substitution, and pRESQ48 which contained a G to C substitution (**Figure 7**). All the mutations were verified by DNA sequence analysis.

Unexpectedly, these mutations further enhanced GUS activity by 20-50% over the three amino acid fusion construct (pRESQ42)(**Figure 8**). The G
25 to C mutation generated the highest level of expression; about 8-fold higher than the original *rubi3* promoter containing the 5' UTR intron (pRESQ4), and approximately 2.5-fold of the maize *ubi1* promoter (pRESQ29)(**Figure 8**). These results also indicate that the first nine nucleotides of the *rubi3* coding sequence (ATGCAGATA) function as a downstream factor (with its 5' UTR or
30 other regulatory 5' UTR) to enhance *rubi3* promoter activity.

EXAMPLE 8

Rubi3 Intron Deletion Mutant

To further define sequences involved in enhanced gene expression from the *rubi3* intron, approximately 800 bp from the middle of the *rubi3* intron
5 were deleted. To generate this deletion mutant, plasmid pRESQ48 was restricted with *Bam*HI, which cuts the intron at nucleotide 969 of **Figure 1A**, and *Bgl*II, which cuts the intron at nucleotide 1755 of **Figure 1A**; treated with Klenow enzyme to blunt the ends and ligated to obtain pRESQ60 which lacked about 780 bases in the *rubi3* intron (**Figure 4E**). The resulting
10 promoter sequence of pRESQ60 is depicted in **Figure 1H**. Unexpectedly, transient GUS activity of pRESQ60 was comparable, if not higher, than that of pRESQ48 in rice suspension cells (**Figure 9**).

The foregoing is illustrative of the present invention, and is not to
15 be construed as limiting thereof. The invention is defined by the following claims, with equivalents of the claims to be included therein.

That which is claimed is:

1. An isolated nucleic acid comprising an expression control element comprising a nucleotide sequence selected from the group consisting of:
- 5 of:
- (a) the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20;
 - 10 (b) a nucleotide sequence consisting essentially of a biologically active fragment of at least 50 consecutive nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20;
 - 15 (c) a nucleotide sequence that hybridizes to the complement of the nucleotide sequences of (a) or (b) under stringent hybridization conditions, wherein said nucleotide sequence is biologically active as an expression control element; and
 - 20 (d) a nucleotide sequence having at least 95% sequence identity to the nucleotide sequences of (a) or (b), wherein said nucleotide sequence is biologically active as an expression control element.
2. The isolated nucleic acid of Claim 1, wherein said expression control element comprises the nucleotide sequence of SEQ ID NO:1 or a
- 25 biologically active fragment thereof.
3. The isolated nucleic acid of Claim 1, wherein said expression control element comprises the nucleotide sequence of SEQ ID NO:2 or a biologically active fragment thereof.
- 30
4. The isolated nucleic acid of Claim 3, wherein said expression control element further comprises the nucleotide sequence of SEQ ID NO:3

which is operably associated with said nucleotide sequence of SEQ ID NO:2 or a biologically active fragment thereof.

5 5. The isolated nucleic acid of Claim 3, wherein said expression control element further comprises an intron which is an internal deletion construct of the nucleotide sequence of SEQ ID NO:3, wherein said intron is operably associated with said nucleotide sequence of SEQ ID NO:2 or a biologically active fragment thereof.

10 6. The isolated nucleic acid of Claim 1, wherein said expression control element comprises the nucleotide sequence of SEQ ID NO:3.

 7. The isolated nucleic acid of Claim 6, wherein said expression control element further comprises the nucleotide sequence of SEQ ID NO:2 or a biologically active fragment thereof operably associated with the nucleotide sequence of SEQ ID NO:3.

 8. The isolated nucleic acid of Claim 6, wherein said expression control element further comprises a heterologous promoter operably associated with the nucleotide sequence of SEQ ID NO:3.

 9. The isolated nucleic acid of Claim 1, wherein said expression control element further comprises in the 5' direction a segment comprising at least the first nucleotide of the polyubiquitin coding sequence of Figure 2A.

25

 10. The isolated nucleic acid of Claim 9, wherein said segment of the polyubiquitin coding sequence consists of the first 9 nucleotides of the polyubiquitin coding sequence (ATGCAGATA).

30 11. The isolated nucleic acid of Claim 9, wherein said expression control element comprises at least the first three nucleotides of the polyubiquitin coding sequence, wherein said first three nucleotides are modified so that they no longer function as a translational start codon.

12. The isolated nucleic acid of Claim 9, wherein the first three nucleotides of the polyubiquitin coding sequence are modified from ATG to ATA, ATC or ATT.

5

13. The isolated nucleic acid of Claim 12, wherein the modified segment of the polyubiquitin coding sequence comprises ATACAGATA, ATCCAGATA or ATTCAGATA.

10

14. The isolated nucleic acid of Claim 12, wherein the modified segment of the polyubiquitin coding sequence consists of ATACAGATA, ATCCAGATA or ATTCAGATA.

15. The isolated nucleic acid of Claim 1, wherein said expression control element comprises the nucleotide sequence of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20 or a biologically active fragment thereof.

16. The isolated nucleic acid of Claim 15, wherein said expression control element comprises the nucleotide sequence of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 or SEQ ID NO:20.

17. The isolated nucleic acid of Claim 1, wherein said expression control element does not comprise any of the polyubiquitin coding sequence of Figure 2A.

18. An isolated nucleic acid comprising an expression control element comprising an intron comprising 5' and 3' splice recognition signals and a nucleotide sequence positioned therebetween, said nucleotide sequence selected from the group consisting of:

30

- (a) a nucleotide sequence comprising at least 10 consecutive nucleotides from the nucleotide sequence of nucleotide 5 through nucleotide 1137 of SEQ ID NO:3, wherein said

nucleotide sequence is biologically active as an expression control element;

(b) a nucleotide sequence that hybridizes to the complement of the nucleotide sequence of (a) under stringent hybridization conditions, wherein said nucleotide sequence is biologically active as an expression control element; and

(c) a nucleotide sequence having at least 95% sequence identity to the nucleotide sequences of (a) or (b), wherein said nucleotide sequence is biologically active as an expression control element.

19. The isolated nucleic acid of Claim 18, wherein said expression control element further comprises the nucleotide sequence of SEQ ID NO:2 or a biologically active fragment thereof operably associated with said intron.

20. The isolated nucleic acid of Claim 18, wherein said expression control element further comprises a heterologous promoter operably associated with said intron.

21. An expression cassette comprising the isolated nucleic acid of any of Claims 1-20 operably associated with a heterologous nucleotide sequence of interest.

22. The expression cassette of Claim 21, wherein said heterologous nucleotide sequence of interest encodes a polypeptide.

23. The expression cassette of Claim 21, wherein said heterologous nucleotide sequence of interest can be transcribed to produce an antisense RNA or RNAi.

24. The expression cassette of Claim 21, wherein said expression cassette further comprises a sequence that encodes a selectable marker.

25. A vector comprising the expression cassette of Claim 21.
26. A cell comprising the isolated nucleic acid of any of Claims 1-20.
- 5 27. The cell of Claim 26, wherein the cell is a plant cell.
28. The cell of Claim 27, wherein the isolated nucleic acid is stably incorporated into the genome of the cell.
- 10 29. A cell comprising the expression cassette of Claim 21.
30. The cell of Claim 29, wherein the cell is a plant cell.
31. The cell of Claim 30, wherein the expression cassette is stably
15 incorporated into the genome of the cell.
32. A cell comprising the vector of Claim 25.
33. The cell of Claim 32, wherein the cell is a plant cell.
- 20 34. A plant comprising the cell of Claim 27.
35. A plant comprising the cell of Claim 30.
- 25 36. A plant comprising the cell of Claim 33.
37. A plant comprising the expression cassette of any of Claims 21-
24 stably incorporated into its genome.
- 30 38. The plant of Claim 37, wherein the plant is a monocotyledonous plant.
39. A crop comprising a plurality of the plant of Claim 37.

40. Seed comprising the expression cassette of any of Claims 21-24 stably incorporated in its genome.

5 41. A method of introducing a nucleic acid into a plant or plant cell comprising transforming said plant or plant cell with the isolated nucleic acid of any of Claims 1-20, the expression cassette of any of Claims 21-24 or the vector of Claim 25.

10 42. The method of Claim 41, wherein the plant or plant cell is a monocotyledonous plant or plant cell.

 43. The method of Claim 41, wherein the heterologous nucleotide sequence of interest encodes a polypeptide.

15 44. The method of Claim 43, further comprising the step of collecting the polypeptide.

 45. The method of Claim 41, wherein the heterologous nucleotide
20 sequence of interest is transcribed to produce an antisense RNA or RNAi.

 46. The method of Claim 41, wherein the method comprises stably transforming the plant or plant cell with the isolated nucleic acid, expression cassette or vector.

25 47. A method of expressing a nucleotide sequence in a plant or plant cell comprising transforming said plant or plant cell with the expression cassette of any of Claims 21-24 or the vector of Claim 25, and expressing the heterologous nucleotide sequence of interest in the plant or plant cell.

30 48. The method of Claim 47, wherein the plant or plant cell is a monocotyledonous plant or plant cell.

49. The method of Claim 47, wherein the heterologous nucleotide sequence of interest encodes a polypeptide.

50. The method of Claim 49, further comprising the step of
5 collecting the polypeptide.

51. The method of Claim 47, wherein the heterologous nucleotide sequence of interest can be transcribed to produce an antisense RNA or RNAi.
10

52. The method of Claim 47, wherein the method comprises stably transforming the plant or plant cell with the expression cassette or vector.

53. A method of expressing a nucleotide sequence in a plant
15 comprising: stably transforming a plant cell with the expression cassette of any of Claims 21-24 or the vector of Claim 25, regenerating a stably transformed plant, and expressing the heterologous nucleotide sequence of interest in the plant.

20 54. The method of Claim 53, wherein the plant is a monocotyledonous plant.

55. The method of Claim 53, wherein the heterologous nucleotide sequence of interest encodes a polypeptide.
25

56. The method of Claim 55, further comprising the step of collecting the polypeptide.

57. The method of Claim 53, wherein the heterologous nucleotide sequence of interest can be transcribed to produce an antisense RNA or RNAi.
30

58. A method of producing a polypeptide in a plant cell comprising culturing the plant cell of Claim 30, wherein said heterologous nucleotide sequence of interest encodes a polypeptide, under conditions sufficient for the production of the polypeptide by the plant cell.

5

59. A method of producing a polypeptide in a plant comprising growing the plant of Claim 35, wherein said heterologous nucleotide sequence of interest encodes a polypeptide, under conditions sufficient for the production of the polypeptide by the plant.

10

1 CCACCCAACCCCATATCGACAGAGGATGTGAAGAACAGGTAAATCACGCA
51 GAAGAACCCATCTCTGATAGCAGCTATCGATTAGAACAACGAATCCATAT
101 TGGGTCCGTGGGAAATACTTACTGCACAGGAAGGGGGGCGATCTGACGAG
151 GCCCCGCCACCGGCCTCGACCCGAGGCCGAGGCCGACGAAGCGCCGGCGA
201 GTACGGCGCCGCGGCGGCCTCTGCCCGTGCCCTCTGCGCGTGGGAGGGAG
251 AGGCCGCGGTGGTGGGGGCGCGCGCGCGCGCGCGCAGCTGGTGCGGCG
301 GCGCGGGGGTCAGCCGCCGAGCCGGCGGCGACGGAGGAGCAGGGCGGCGT
351 GGACGCGAACTTCCGATCGGTTGGTCAGAGTGCGCGAGTTGGGCTTAGCc
401 aatTAGGTCTCAAc^aatCTATTGGGCCGTAAAATTCATGGGCCCTGGTTT
451 GTCTAGGCC^aaatATCCCGTTCATTTAGCC^aCAAATATTTCCCCAGAG
501 GATTATTAAGGCC^aCACACGCAGCTTATAGCAGATCAAGTACGATGTTTCC
551 TGATCGTTGGATCGGAAACGTACGGTCTTGATCAGGCATGCCGACTTCGT
601 CAAAGAGAGGCGGCATGACCTGACGCGGAGTTGGTTCCGGGCACCGTCTG
651 GATGGTCGTACCGGGACCGGAC^aCGTGTCTGCGCCTCCA^aCTACATGGACA
701 CGTGTGGTGCTGCCATTGGGCCGTACGCGTGGCGGTGACCGCACCGGATG
751 CTGCCTCGCACCGCCTTGCCACGCTTTATATAGAGAGGTTTTCTCTCCA
801 TTAATCGC^aATAGCGAGTCGAATCGACCGAAGGGGAGGGGGAGCGAAGCTT
851 TCGGTTCTCTAATCGCCTCGTCAAGGTA^aCTAATCAATCACCTCGTCCTA
901 ATCCTCGAATCTCTCGTGGTGCCCGTCTAATCTCGCGATTTTGATGCTCG
951 TGGTGGAAAGCGTAGGAGGATCCCGTGCGAGTTAGTCTCAATCTCTCAGG
1001 GTTTCGTGCGATTTTAGGGTGATCCACCTCTTAATCGAGTTACGGTTTCG
1051 TGCGATTTTAGGGTAATCCTCTTAATCTCTCATTGATTTAGGGTTTCGTG
1101 AGAATCGAGGTAGGGATCTGTGTTATTTATATCGATCTAATAGATGGATT
1151 GGTTTTGAGATTGTTCTGTCTAGATGGGGATTGTTTCGATATATTACCCTA
1201 ATGATGTGTCTAGATGGGGATTGTTTCGATATATTACCCTAATGATGTGTC
1251 AGATGGGGATTGTTTCGATATATTACCCTAATGATGGATAATAAGAGTAG
1301 TTCACAGTTATGTTTTGATCCTGCCACATAGTTTGAGTTTTGTGATCAGA
1351 TTTAGTTTTACTTATTTGTGCTTAGTTCGGATGGGATTGTTCTGATATTG
1401 TTCCAATAGATGAATAGCTCGTTAGGTTAAAATCTTTAGGTTGAGTTAGG
1451 CGAC^aCATAGTTTATTTCTCTGGATTTGGATTGGAATTGTGTTCTTAGT
1501 TTTTTTCCCCTGGATTTGGATTGGAATTGTGTGGAGCTGGGTTAGAGAAT
1551 TACATCTGTATCGTGTACACCTACTTGA^aCTGTAGAGCTTGGGTTCTAAG
1601 GTCAATTTAATCTGTATTGTATCTGGCTCTTTGCCTAGTTGA^aCTGTAGT
1651 GCTGATGTTGTACTGTGTTTTTTTACCCGTTTTTATTTGCTTTACTCGTGC
1701 AAATCAAATCTGTCAGATGCTAGAACTAGGTGGCTTTATTTCTGTGTTCTT
1751 ACATAGATCTGTTGTCTGTAGTTACTTATGTCAGTTTTGTTATTATCTG
1801 AAGATATTTTTGGTTGTTGCTTGTTGATGTGGTGTGAGCTGTGAGCAGCG
1851 CTCTTATGATTAATGATGCTGTCCAATTGTAGTGTAGTATGATGTGATTG
1901 ATATGTTTCATCTATTTTGAGCTGACAGTACCGATATCGTAGGATCTGGTG
1951 CCAACTTATTCTCCAGCTGCTTTTTTTTACCTATGTTAATCCAATCCTT
2001 TCTTGCCCTCTCCAG (SEQ ID NO:1)

FIG. 1A

1 CCACCCAACCCCATATCGACAGAGGATGTGAAGAACAGGTAAATCACGCA
51 GAAGAACCCATCTCTGATAGCAGCTATCGATTAGAACAACGAATCCATAT
101 TGGGTCCGTGGGAAATACTTACTGCACAGGAAGGGGGGCGATCTGACGAG
151 GCCCCGCCACCGGCCTCGACCCGAGGCCGAGGCCGACGAAGCGCCGGCGA
201 GTACGGCGCCGCGGCGGCCTCTGCCCGTGCCCTCTGCGCGTGGGAGGGAG
251 AGGCCGCGGTGGTGGGGGCGCGCGCGCGCGCGCGCAGCTGGTGCGGCG
301 GCGCGGGGGTCAGCCGCCGAGCCGGCGGCGACGGAGGAGCAGGGCGGCGT
351 GGACGCGAACTTCCGATCGGTTGGTCAGAGTGCGCGAGTTGGGCTTAGCc
401 aatTAGGTCTCAAcAatCTATTGGGCCGTAAAATTCATGGGCCCTGGTTT
451 GTCTAGGCCcaatATCCCGTTCATTTAGCCACAAATATTTCCCCAGAG
501 GATTATTAAGGCCACACGCAGCTTATAGCAGATCAAGTACGATGTTTCC
551 TGATCGTTGGATCGGAAACGTACGGTCTTGATCAGGCATGCCGACTTCGT
601 CAAAGAGAGGCGGCATGACCTGACGCGGAGTTGGTTCGGGGCACCGTCTG
651 GATGTCGTACCGGGACCGGACACGTGTGCGGCCTCCAACTACATGGACA
701 CGTGTGGTGCTGCCATTGGGCCGTACGCGTGGCGGTGACCGCACCGGATG
751 CTGCCTCGCACCGCCTTGCCCACGCTTTATATAGAGAGGTTTTCTCTCCA
801 TTAATCGCATAGCGAGTCGAATCGACCGAAGGGGAGGGGGAGCGAAGCTT
851 TGCGTTCTCTAATCGCCTCGTCAAG (SEQ ID NO:2)

FIG. 1B

1 GTAACATAATCAATCACCTCGTCCTAATCCTCGAATCTCTCGTGGTGCCCG
51 TCTAATCTCGCGATTTTGATGCTCGTGGTGGAAGCGTAGGAGGATCCCG
101 TGCGAGTTAGTCTCAATCTCTCAGGGTTTCGTGCGATTTTAGGGTGATCC
151 ACCTCTTAATCGAGTTACGGTTTCGTGCGATTTTAGGGTAATCCTCTTAA
201 TCTCTCATTGATTTAGGGTTTCGTGAGAATCGAGGTAGGGATCTGTGTTA
251 TTTATATCGATCTAATAGATGGATTGGTTTTGAGATTGTTCTGTGATG
301 GGGATTGTTTCGATATATTACCCTAATGATGTGTGATGAGGATTGTTT
351 CGATATATTACCCTAATGATGTGTGATGAGGATTGTTTCGATATATTA
401 CCCTAATGATGGATAATAAGAGTAGTTCACAGTTATGTTTTGATCCTGCC
451 ACATAGTTTGAGTTTTGTGATCAGATTTAGTTTTACTTATTTGTGCTTAG
501 TTCGGATGGGATTGTTCTGATATTGTTCCAATAGATGAATAGCTCGTTAG
551 GTTAAAATCTTTAGGTTGAGTTAGGCGACACATAGTTTATTTCTCTGGA
601 TTTGGATTGGAATTGTGTTCTTAGTTTTTTTCCCCTGGATTTGGATTGGA
651 ATTGTGTGGAGCTGGGTTAGAGAATTACATCTGTATCGTGTACACCTACT
701 TGAAGTGTAGAGCTTGGGTTCTAAGGTCAATTTAATCTGTATTGTATCTG
751 GCTCTTTGCCTAGTTGAAGTGTAGTGCTGATGTTGTACTGTGTTTTTTTA
801 CCCGTTTTATTTGCTTTACTCGTGCAAATCAAATCTGTCAGATGCTAGAA
851 CTAGGTGGCTTTATTCTGTGTTCTTACATAGATCTGTTGTCTGTAGTTA
901 CTTATGTCAGTTTTGTTATTATCTGAAGATATTTTTGGTTGTTGCTTGTT
951 GATGTGGTGTGAGCTGTGAGCAGCGCTCTTATGATTAATGATGCTGTCCA
1001 ATTGTAGTGTAGTATGATGTGATTGATATGTTTCATCTATTTTGAGCTGAC
1051 AGTACCGATATCGTAGGATCTGGTGCCAACTTATTTCTCCAGCTGCTTTTT
1101 TTTACCTATGTTAATTCGAATCCTTTCTTGCTCTTCCAG
(SEQ ID NO:3)

FIG. 1C

1 CCACCCAACCCCATATCGACAGAGGATGTGAAGAACAGGTAAATCACGCA
51 GAAGAACCCATCTCTGATAGCAGCTATCGATTAGAACAACGAATCCATAT
101 TGGGTCCGTGGGAAATACTTACTGCACAGGAAGGGGGGCGATCTGACGAG
151 GCCCCGCCACCGGCCTCGACCCGAGGCCGAGGCCGACGAAGCGCCGGCGA
201 GTACGGCGCCGCGGCGGCCTCTGCCCGTGCCCTCTGCGCGTGGGAGGGAG
251 AGGCCGCGGTGGTGGGGGCGCGCGCGCGCGCGCGCAGCTGGTGCGGCG
301 GCGCGGGGGTCAGCCGCCGAGCCGGCGGCGACGGAGGAGCAGGGCGGCGT
351 GGACGCGAACTTCCGATCGGTTGGTCAGAGTGC GCGAGTTGGGCTTAGCc
401 aatTAGGTCTCAAc aatCTATTGGGCCGTAAAATT CATGGGCCCTGGTTT
451 GTCTAGGCCcaatATCCCGTTCATTT CAGCCCACAAATATTTCCCCAGAG
501 GATTATTAAGGCCACACGCAGCTTATAGCAGATCAAGTACGATGTTTCC
551 TGATCGTTGGATCGGAAACGTACGGTCTTGATCAGGCATGCCGACTTCGT
601 CAAAGAGAGGCGGCATGACCTGACGCGGAGTTGGTTCCGGGCACCGTCTG
651 GATGGTCGTACCGGGACCGGACACGTGTCGCGCCTCCAACTACATGGACA
701 CGTGTGGTGCTGCCATTGGGCCGTACGCGTGGCGGTGACCGCACCGGATG
751 CTGCCCTCGCACCGCCTTGCCACGCTTTATATAGAGAGGTTTTCTCTCCA
801 TTAATCGCATAGCGAGTCGAATCGACCGAAGGGGAGGGGGAGCGAAGCTT
851 TCGGTTCTCTAATCGCCTCGTCAAGGTAAC TAATCAATCACCTCGTCCTA
901 ATCCTCGAATCTCTCGTGGTGCCCGTCTAATCTCGCGATTTTGATGCTCG
951 TGGTGGAAGCGTAGGAGGATCCCGTGCGAGTTAGTCTCAATCTCTCAGG
1001 GTTTCGTGCGATTTTAGGGTGATCCACCTCTTAATCGAGTTACGGTTTCG
1051 TGCGATTTTAGGGTAATCCTCTTAATCTCTCATTGATTTAGGGTTTCGTG
1101 AGAATCGAGGTAGGGATCTGTGTTATTTATATCGATCTAATAGATGGATT
1151 GGTTTTGAGATTGTTCTGT CAGATGGGGATTGTTTCGATATATTACCCTA
1201 ATGATGTGTCAGATGGGGATTGTTTCGATATATTACCCTAATGATGTGTC
1251 AGATGGGGATTGTTTCGATATATTACCCTAATGATGGATAATAAGAGTAG
1301 TTCACAGTTATGTTTTGATCCTGCCACATAGTTTGAGTTTTGTGATCAGA
1351 TTTAGTTTTACTTATTTGTGCTTAGTTCGGATGGGATTGTTCTGATATTG
1401 TTCCAATAGATGAATAGCTCGTTAGGTTAAAACTTTAGGTTGAGTTAGG
1451 CGACACATAGTTTATTTCTCTGGATTTGGATTGGAATTGTGTTCTTAGT
1501 TTTTTTCCCCTGGATTTGGATTGGAATTGTGTGGAGCTGGGTTAGAGAAT
1551 TACATCTGTATCGTGTACACCTACTTGAAGTGTAGAGCTTGGGTTCTAAG
1601 GTCAATTTAATCTGTATTGTATCTGGCTCTTTGCCTAGTTGAAGTGTAGT
1651 GCTGATGTTGTACTGTGTTTTTTTACCCGTTTTATTTGCTTTACTCGTGC
1701 AAATCAAATCTGTCAGATGCTAGAACTAGGTGGCTTTATTCTGTGTTCTT
1751 ACATAGATCTGTTGTCTGTAGTTACTTATGTCAGTTTTGTTATTATCTG
1801 AAGATATTTTTGGTTGTTGCTTGTTGATGTGGTGTGAGCTGTGAGCAGCG
1851 CTCTTATGATTAATGATGCTGTCCAATTGTAGTGTAGTATGATGTGATTG
1901 ATATGTT CATCTATTTTGAGCTGACAGTACCGATATCGTAGGATCTGGTG
1951 CCAACTATTCTCCAGCTGCTTTTTTTTACCTATGTTAATCCAATCCTT
2001 TCTTGCCCTCTTCCAGATGCAGATA (SEQ ID NO:17)

FIG. 1D

1 CCACCCAACCCCATATCGACAGAGGATGTGAAGAACAGGTAAATCACGCA
51 GAAGAACCCATCTCTGATAGCAGCTATCGATTAGAACAACGAATCCATAT
101 TGGGTCCGTGGGAAATACTTACTGCACAGGAAGGGGGGCGATCTGACGAG
151 GCCCCGCCACCGGCCTCGACCCGAGGCCGAGGCCGACGAAGCGCCGGCGA
201 GTACGGCGCCGCGGCGGCCTCTGCCCGTGCCCTCTGCGCGTGGGAGGGAG
251 AGGCCGCGGTGGTGGGGGCGCGCGCGCGCGCGCAGCTGGTGC GGCG
301 GCGCGGGGGTCAGCCGCCGAGCCGGCGGCGACGGAGGAGCAGGGCGGCGT
351 GGACGCGAACTTCCGATCGGTTGGTCAGAGTGCGCGAGTTGGGCTTAGCc
401 aatTAGGTCTCAAc aatCTATTGGGCCGTAAAATTTCATGGGCCCTGGTTT
451 GTCTAGGCCcaatATCCCGTTTCATTTTCAGCCCACAAATATTTCCCCAGAG
501 GATTATTAAGGCCCCACACGCAGCTTATAGCAGATCAAGTACGATGTTTCC
551 TGATCGTTGGATCGGAAACGTACGGTCTTGATCAGGCATGCCGACTTCGT
601 CAAAGAGAGGCGGCATGACCTGACGCGGAGTTGGTTCCGGGACCGTCTG
651 GATGGTCGTACCGGGACCGGACACGTGTCGCGCCTCCA ACTACATGGACA
701 CGTGTGGTGCTGCCATTGGGCCGTACGCGTGGCGGTGACCGCACCGGATG
751 CTGCCTCGCACCGCCTTGCCACGCTTTATATAGAGAGGTTTTCTCTCCA
801 TTAATCGCATAGCGAGTCGAATCGACCGAAGGGGAGGGGGAGCGAAGCTT
851 TGC GTTCTCTAATCGCCTCGTCAAGGTA ACTAATCAATCACCTCGTCCTA
901 ATCCTCGAATCTCTCGTGGTGCCCGTCTAATCTCGCGATTTTGATGCTCG
951 TGGTGGAAAGCGTAGGAGGATCCCGTGCGAGTTAGTCTCAATCTCTCAGG
1001 GTTTCGTGCGATTTTAGGGTGATCCACCTCTTAATCGAGTTACGGTTTCG
1051 TGCGATTTTAGGGTAATCCTCTTAATCTCTCATTGATTTAGGGTTTCGTG
1101 AGAATCGAGGTAGGGATCTGTGTTATTTATATCGATCTAATAGATGGATT
1151 GGTTTTGAGATTGTTCTGT CAGATGGGGATTGTTTCGATATATTACCCTA
1201 ATGATGTGTCAGATGGGGATTGTTTCGATATATTACCCTAATGATGTGTC
1251 AGATGGGGATTGTTTCGATATATTACCCTAATGATGGATAATAAGAGTAG
1301 TTCACAGTTATGTTTTGATCCTGCCACATAGTTTGAGTTTTGTGATCAGA
1351 TTTAGTTTTACTTATTTGTGCTTAGTTCGGATGGGATTGTTCTGATATTG
1401 TTCCAATAGATGAATAGCTCGTTAGGTTAAAATCTTTAGGTTGAGTTAGG
1451 CGACACATAGTTTATTTCTCTGGATTTGGATTGGAATTGTGTTCTTAGT
1501 TTTTTTCCCCTGGATTTGGATTGGAATTGTGTGGAGCTGGGTTAGAGAAT
1551 TACATCTGTATCGTGTACACCTACTTGA ACTGTAGAGCTTGGGTTCTAAG
1601 GTCAATTTAATCTGTATTGTATCTGGCTCTTTGCCTAGTTGAACTGTAGT
1651 GCTGATGTTGTACTGTGTTTTTTTACCCGTTTTATTTGCTTTACTCGTGC
1701 AAATCAAATCTGTCAGATGCTAGAACTAGGTGGCTTTATCTGTGTTCTT
1751 ACATAGATCTGTTGTCTGTAGTTACTTATGTCAGTTTTGTTATTATCTG
1801 AAGATATTTTTGGTTGTTGCTTGTTGATGTGGTGTGAGCTGTGAGCAGCG
1851 CTCTTATGATTAATGATGCTGTCCAATTGTAGTGTAGTATGATGTGATTG
1901 ATATGTTTCATCTATTTTGAGCTGACAGTACCGATATCGTAGGATCTGGTG
1951 CCAACTTATTCTCCAGCTGCTTTTTTTTACCTATGTTAATCCAATCCTT
2001 TCTTGCCCTCTCCAGATACAGATA (SEQ ID NO:18)

FIG. 1E

1 CCACCCAACCCCATATCGACAGAGGATGTGAAGAACAGGTAAATCACGCA
51 GAAGAACCCATCTCTGATAGCAGCTATCGATTAGAACAACGAATCCATAT
101 TGGGTCCGTGGGAAATACTTACTGCACAGGAAGGGGGGCGATCTGACGAG
151 GCCCCGCCACCGGCCTCGACCCGAGGCCGAGGCCGACGAAGCGCCGGCGA
201 GTACGGCGCCGCGGCGGCCTCTGCCCGTGCCCTCTGCGCGTGGGAGGGAG
251 AGGCCGCGGTGGTGGGGGCGCGCGCGCGCGCGCGCAGCTGGTGCGGCG
301 GCGCGGGGGTCAGCCGCCGAGCCGGCGGGCGACGGAGGAGCAGGGCGGCGT
351 GGACGCGAACTTCCGATCGGTTGGTCAGAGTGCGCGAGTTGGGCTTAGCc
401 aatTAGGTCTCAAcAatCTATTGGGCCGTAAAATTCATGGGCCCTGGTTT
451 GTCTAGGCCcaatATCCCGTTCATTTAGCCCAAAATATTTCCCCAGAG
501 GATTATTAAGGCCCCACACGCAGCTTATAGCAGATCAAGTACGATGTTTCC
551 TGATCGTTGGATCGGAAACGTACGGTCTTGATCAGGCATGCCGACTTCGT
601 CAAAGAGAGGCGGCATGACCTGACGCGGAGTTGGTTCCGGGCACCGTCTG
651 GATGGTCGTACCGGGACCGGACACGTGTGCGGCCTCCAACTACATGGACA
701 CGTGTGGTGCTGCCATTGGGCCGTACGCGTGCGGGTGACCGCACCGGATG
751 CTGCCCTCGCACCGCCTTGCCCACGCTTTATATAGAGAGGTTTTCTCTCCA
801 TTAATCGCATAGCGAGTCGAATCGACCGAAGGGGAGGGGGAGCGAAGCTT
851 TCGCTTCTCTAATCGCCTCGTCAAGGTAACATAATCAATCACCTCGTCCTA
901 ATCCTCGAATCTCTCGTGGTGCCCGTCTAATCTCGCGATTTTGATGCTCG
951 TGGTGGAAGCGTAGGAGGATCCCGTGCGAGTTAGTCTCAATCTCTCAGG
1001 GTTTCGTGCGATTTTAGGGTGATCCACCTCTTAATCGAGTTACGGTTTCG
1051 TGCGATTTTAGGGTAATCCTCTTAATCTCTCATTGATTTAGGGTTTCGTG
1101 AGAATCGAGGTAGGGATCTGTGTTATTTATATCGATCTAATAGATGGATT
1151 GGTTTTGAGATTGTTCTGTGATGGGGATTGTTTCGATATATTACCCTA
1201 ATGATGTGTCAGATGGGGATTGTTTCGATATATTACCCTAATGATGTGTC
1251 AGATGGGGATTGTTTCGATATATTACCCTAATGATGGATAATAAGAGTAG
1301 TTCACAGTTATGTTTTGATCCTGCCACATAGTTTGAGTTTTGTGATCAGA
1351 TTTAGTTTTACTTATTTGTGCTTAGTTTCGGATGGGATTGTTCTGATATTG
1401 TTCCAATAGATGAATAGCTCGTTAGGTTAAATCTTTAGGTTGAGTTAGG
1451 CGACACATAGTTTATTTCCCTCTGGATTGGGATTGGAATTGTGTTCTTAGT
1501 TTTTTTCCCCTGGATTGGATTGGAATTGTGTGGAGCTGGGTTAGAGAAT
1551 TACATCTGTATCGTGTACACCTACTTGAAGTGTAGAGCTTGGGTTCTAAG
1601 GTCAATTTAATCTGTATTGTATCTGGCTCTTTGCCTAGTTGAAGTGTAGT
1651 GCTGATGTTGTACTGTGTTTTTTTACCCGTTTTATTTGCTTTACTCGTGC
1701 AAATCAAATCTGTCAGATGCTAGAACTAGGTGGCTTTATTCTGTGTTCTT
1751 ACATAGATCTGTTGTCCTGTAGTTACTTATGTCAGTTTTGTTATTATCTG
1801 AAGATATTTTTGGTTGTTGCTTGTTGATGTGGTGTGAGCTGTGAGCAGCG
1851 CTCTTATGATTAATGATGCTGTCCAATTGTAGTGTAGTATGATGTGATTG
1901 ATATGTTTATCTATTTTGTAGCTGACAGTACCGATATCGTAGGATCTGGTG
1951 CCAACTTATTCTCCAGCTGCTTTTTTTTACCTATGTTAATTCCAATCCTT
2001 TCTTGCCTCTTCCAGATTCAGATA (SEQ ID NO:19)

FIG. 1F

1 CCACCCAACCCCATATCGACAGAGGATGTGAAGAACAGGTAAATCACGCA
51 GAAGAACCCATCTCTGATAGCAGCTATCGATTAGAACAACGAATCCATAT
101 TGGGTCCGTGGGAAATACTTACTGCACAGGAAGGGGGGCGATCTGACGAG
151 GCCCCGCCACCGGCCTCGACCCGAGGCCGAGGCCGACGAAGCGCCGGCGA
201 GTACGGCGCCGCGGCGGCCTCTGCCCGTGCCCTCTGCGCGTGGGAGGGAG
251 AGGCCGCGGTGGTGGGGGCGCGCGCGCGCGCGCGCAGCTGGTGCGGCG
301 GCGCGGGGGTCAGCCGCCGAGCCGGCGGCGACGGAGGAGCAGGGCGGCGT
351 GGACGCGAACTTCCGATCGGTTGGTCAGAGTGCGCGAGTTGGGCTTAGCc
401 aatTAGGTCTCAAc aatCTATTGGGCCGTAAAATTCATGGGCCCTGGTTT
451 GTCTAGGCCcaatATCCCGTTCATTTAGCCCAAAATATTTCCCCAGAG
501 GATTATTAAGGCCCCACACGCAGCTTATAGCAGATCAAGTACGATGTTTCC
551 TGATCGTTGGATCGGAAACGTACGGTCTTGATCAGGCATGCCGACTTCGT
601 CAAAGAGAGGCGGCATGACCTGACGCGGAGTTGGTTCCGGGCACCGTCTG
651 GATGGTCGTACCGGGACCGGACACGTGTCGCGCCTCCAACTACATGGACA
701 CGTGTGGTGCTGCCATTGGGCCGTACGCGTGCGGGTGACCGCACCGGATG
751 CTGCCCTCGCACCGCCTTGCCCACGCTTTATATAGAGAGGTTTTCTCTCCA
801 TTAATCGCATAGCGAGTCGAATCGACCGAAGGGGAGGGGGAGCGAAGCTT
851 TGCGTTCTCTAATCGCCTCGTCAAGGTAAC TAATCAATCACCTCGTCCTA
901 ATCCTCGAATCTCTCGTGGTGCCCGTCTAATCTCGCGATTTTGATGCTCG
951 TGGTGGAAGCGTAGGAGGATCCCGTGCGAGTTAGTCTCAATCTCTCAGG
1001 GTTTCGTGCGATTTTAGGGTGATCCACCTCTTAATCGAGTTACGGTTTCG
1051 TGCGATTTTAGGGTAATCCTCTTAATCTCTCATTGATTTAGGGTTTCGTG
1101 AGAATCGAGGTAGGGATCTGTGTTATTTATATCGATCTAATAGATGGATT
1151 GGTTTTGAGATTGTTCTGTGATGAGGATTGTTTCGATATATTACCCTA
1201 ATGATGTGTCAGATGGGGATTGTTTCGATATATTACCCTAATGATGTGTC
1251 AGATGGGGATTGTTTCGATATATTACCCTAATGATGGATAATAAGAGTAG
1301 TTCACAGTTATGTTTTGATCCTGCCACATAGTTTGAGTTTTGTGATCAGA
1351 TTTAGTTTTACTTATTTGTGCTTAGTTTCGGATGGGATTGTTCTGATATTG
1401 TTCCAATAGATGAATAGCTCGTTAGGTTAAATCTTTAGGTTGAGTTAGG
1451 CGACACATAGTTTATTTCTCTGGATTGGATTGGAATTGTGTTCTTAGT
1501 TTTTTTCCCCTGGATTGATTGGAATTGTGTGGAGCTGGGTTAGAGAAT
1551 TACATCTGTATCGTGTACACCTACTTGAAGTGTAGAGCTTGGGTTCTAAG
1601 GTCAATTTAATCTGTATTGTATCTGGCTCTTTGCCTAGTTGAAGTGTAGT
1651 GCTGATGTTGTACTGTGTTTTTTTACCCGTTTTATTTGCTTTACTCGTGC
1701 AAATCAAATCTGTGATGCTAGAACTAGGTGGCTTTATTCTGTGTTCTT
1751 ACATAGATCTGTTGTCCTGTAGTTACTTATGTCAGTTTTGTTATTATCTG
1801 AAGATATTTTTGGTTGTTGCTTGTTGATGTGGTGTGAGCTGTGAGCAGCG
1851 CTCTTATGATTAATGATGCTGTCCAATTGTAGTGTAGTATGATGTGATTG
1901 ATATGTTTCTATTTTGGAGCTGACAGTACCGATATCGTAGGATCTGGTG
1951 CCAACTATTCTCCAGCTGCTTTTTTTTACCTATGTTAATCCAATCCTT
2001 TCTTGCTCTTCCAGATCCAGATA (SEQ ID NO:20)

FIG. 1G

1 CCACCCAACCCCATATCGACAGAGGATGTGAAGAACAGGTAAATCACGCA
51 GAAGAACCCATCTCTGATAGCAGCTATCGATTAGAACAACGAATCCATAT
101 TGGGTCCGTGGGAAATACTTACTGCACAGGAAGGGGGGCGATCTGACGAG
151 GCCCCGCCACCGGCCTCGACCCGAGGCCGAGGCCGACGAAGCGCCGGCGA
201 GTACGGCGCCGCGGCGGCCTCTGCCCGTGCCCTCTGCGCGTGGGAGGGAG
251 AGGCCGCGGTGGTGGGGGCGCGCGCGCGCGCGCGCAGCTGGTGCGGCG
301 GCGCGGGGGTCAGCCGCCGAGCCGGCGGCGACGGAGGAGCAGGGCGGCGT
351 GGACGCGAACTTCCGATCGGTTGGTCAGAGTGCGCGAGTTGGGCTTAGCc
401 aatTAGGTCTCAAc^aatCTATTGGGCCGTAAAATTCATGGGCCCTGGTTT
451 GTCTAGGCC^caatATCCCGTTCA^tTTTCAGCCCACAAATATTTCCCAGAG
501 GATTATTAAGGCCCACACGCAGCTTATAGCAGATCAAGTACGATGTTTCC
551 TGATCGTTGGATCGGAAACGTACGGTCTTGATCAGGCATGCCGACTTCGT
601 CAAAGAGAGGCGGCATGACCTGACGCGGAGTTGGTTCCGGGCACCGTCTG
651 GATGGTCGTACCGGGACCGGACACGTGTCGCGCCTCCA^aCTACATGGACA
701 CGTGTGGTGCTGCCATTGGGCCGTACGCGTGGCGGTGACCGCACCGGATG
751 CTGCCTCGCACCGCCTTGCCACGCTTTATATAGAGAGGTTTTCTCTCCA
801 TTAATCGC^aTAGCGAGTCGAATCGACCGAAGGGGAGGGGGAGCGAAGCTT
851 TGCGTTCTCTAATCGCCTCGTCAAGGTA^aCTAATCAATCACCTCGTCCTA
901 ATCCTCGAATCTCTCGTGGTGCCCGTCTAATCTCGCGATTTTGATGCTCG
951 TGGTGGAAAGCGTAGGAGGATCGATCTGTTGTCCTGTAGTTACTTATGTC
1001 AGTTTTGTTATTATCTGAAGATATTTTTGGTTGTTGCTTGTTGATGTGGT
1051 GTGAGCTGTGAGCAGCGCTCTTATGATTAATGATGCTGTCCAATTGTAGT
1101 GTAGTATGATGTGATTGATATGTTCA^tCTATTTTGAGCTGACAGTACCGA
1151 TATCGTAGGATCTGGTGCCAACTTATTCTCCAGCTGCTTTTTTTTACCTA
1201 TGTTAATTCCAATCCTTTCTTGCCCTCTTCCAGATCCAGATA

(SEQ ID NO:27)

FIG. 1H

1 acacccaaccccatatcgacagaggatgtgaagaacaggtaaatacacgca
51 gaagaacccatctctgatagcagctatcgattagaacaacgaatccatat
101 tgggtccgtgggaaatacttactgcacaggaagggggcgatctgacgag
151 gccccgccaccggcctcgacccgaggccgagggccgacgaagcgccggcga
201 gtacggcgccgcggcgccctctgcccgtgccctctgcgcgtgggaggag
251 aggcgcggtggtggggcgcgcgcgcgcgcgcgcgagctggtgcggcg
301 gcgcgggggtcagccgagcggcgggcgacggaggagcagggcgggcgt
351 ggacgcgaacttccgatcggttggtcagagtgcgcgagttgggcttagcc
401 aattaggtctcaacaatctattgggcccgtaaaattcatgggccctggttt
451 gtctaggcccaatatccggttcatttcagcccacaaatatttcccagag
501 gattattaaggcccacacgcagcttatagcagatcaagtacgatgtttcc
551 tgatcgttggatcggaaacgtacggtcttgatcaggcatgccgacttcgt
601 caaagagaggcgccatgacctgacgcggagttggttccgggcaccgtctg
651 gatggtcgtaccgggaccggacacgtgtcgcgcctccaactacatggaca
701 cgtgtggtgctgccattgggcccgtacgcgtggcggtgaccgcaccggatg
751 ctgcctcgcaccgccttgcccacgctttatatagagaggttttctctcca
801 ttaatcgcatagcgagtcgaatcgaccgaagggggagggggagcgaagctt
851 tgcgttctctaatacgccctcgtcaaggtaactaatcaatcacctcgtccta
901 atcctcgaatctctcgtggtgccctcctaatactcgcgatttttgatgctcg
951 tgggtggaaagcgtaggaggatcccgtgcgagttagtctcaatctctcagg
1001 gtttcgtgcgatttttagggatccacctcttaatcgagttacgggtttcgtg
1051 tgcgatttttagggtaatcctcttaatctctcattgatttaggggtttcgtg
1101 agaatcgaggtagggatctgtgttatattatcgatctaatagatggatt
1151 ggttttgagattgttctgtcagatggggattgtttcgatatattacccta
1201 atgatgtgtcagatggggattgtttcgatatattaccctaataatgatgtgtc
1251 agatggggattgtttcgatatattaccctaataatgatggataataagagtag
1301 ttcacagttatgttttgatcctgccacatagtttgagttttgtgatcaga
1351 tttagttttacttattttgtgcttagttcggatgggattgttctgatattg
1401 tccaatagatgaatagctcgttaggttaaaatctttagggttgagttagg
1451 cgacacatagttttatttccctctggattttggattggaattgtgttcttagt
1501 ttttttcccctggattttggattggaattgtgtggagctgggttagagaat
1551 tacatctgtatcgtgtacacctacttgaactgtagagcttgggttctaag
1601 gtcaatttaaatctgtattgtatctggcctctttgcctagttgaactgtagt
1651 gctgatgttgactgtgtttttttacccggttttattttgctttactcgtgc
1701 aaatcaaactctgtcagatgctagaactaggtggctttattctgtgttctt
1751 acatagatctgttgctcctgtagttacttatgtcagttttgttattatctg
1801 aagatatttttggttggttgcttggtgatgtggtgtgagctgtgagcagcg
1851 ctcttatgattaatgatgctgtccaattgtagtgtagtatgatgtgattg
1901 atatgttcatctattttgagctgacagtaccgatatcgtaggatctgggtg
1951 ccaacttattctccagctgcttttttttacctatgttaattccaatcctt
2001 tcttgccctcttcagATGCAGATATTCGTTAAGACCCCTCACTGGCAAGAC

1 M Q I F V K T L T G K T

2051 CATCACCTTGGAGGTTGAGTTCTCCGATACGATTGACAATGTGAAGGCTA

13 I T L E V E F S D T I D N V K A

2101 AGATTCAGGACAAGGAGGGCATCCCTCCGACCAGCAACGCCTTATCTTC

29 K I Q D K E G I P P D Q Q R L I F

FIG. 2A
8/16

2151 GCTGGCAAGCAGCTTGAGGATGGGCGTACTCTCGCGGATTATAACATCCA
46 A G K Q L E D G R T L A D Y N I Q
2201 GAAGGAGTCCACCTTGACCTTGTCTCCGCCTTCGTGGAGGCATGCAAA
63 K E S T L H L V L R L R G G M Q
2251 TATTCGTGAAGACCCTCACCGGCAAGACCATTACCCTGGAGGTCGAGTCC
79 I F V K T L T G K T I T L E V E S
2301 TCCGACACGATCGATAATGTGAAGGCCAAGATCCAGGACAAGGAGGGAAT
96 S D T I D N V K A K I Q D K E G I
2351 CCCACCAGACCAGCAGCGTCTCATCTTTGCTGGGAAGCAGCTCGAGGATG
113 P P D Q Q R L I F A G K Q L E D
2401 GCCGCACCCTTGACAGACTACAACATCCAAAAGGAATCCACCCTGCACCTT
129 G R T L A D Y N I Q K E S T L H L
2451 GTCCTGCGCCTCCGGGGCGGTATGCAGATCTTTGTGAAGACCCTTACTGG
146 V L R L R G G M Q I F V K T L T G
2501 CAAGACGATCACCTTGGAGGTTGAGTCCTCTGACACGATCGACAATGTGA
163 K T I T L E V E S S D T I D N V
2551 AGGCCAAGATCCAGGACAAGGAGGGTATTCCACCAGACCAGCAGCGGCTC
179 K A K I Q D K E G I P P D Q Q R L
2601 ATCTTCGCTGGCAAGCAGCTTGAGGACGGCCGCACCCTTGCTGACTATAA
196 I F A G K Q L E D G R T L A D Y N
2651 CATCCAGAAGGAGTCCACCTTGCACTTGGTCCTCCGCCTTCGTGGAGGTA
213 I Q K E S T L H L V L R L R G G
2701 TGCAGATCTTCGTGAAGACCCTCACCGGCAAGACCATCACCTTGGAGGTT
229 M Q I F V K T L T G K T I T L E V
2751 GAGTCGTCTGACACGATCGACAATGTGAAGGCCAAGATCCAGGACAAGGA
246 E S S D T I D N V K A K I Q D K E
2801 GGGTATTCCACCAGACCAGCAGCGCCTCGTCTTCGCTGGCAAGCAGCTTG
263 G I P P D Q Q R L V F A G K Q L
2851 AGGATGGTCGCACCCTTGACAGATTACAACATCCAGAAGGAGTCCACCCTG
279 E D G R T L A D Y N I Q K E S T L
2901 CACCTTGTCTGCGCCTTCGTGGAGGTATGCAGATCTTCGTGAAGACCCT
296 H L V L R L R G G M Q I F V K T L
2951 CACTGGCAAGACCATCACTCTGGAGGTTGAGTCCTCTGACACCATCGACA
313 T G K T I T L E V E S S D T I D
3001 ACGTGAAGGCCAAGATCCAGGATAAGGAGGGCATCCCTCCAGACCAGCAG
329 N V K A K I Q D K E G I P P D Q Q
3051 CGCCTCATCTTCGCCGGCAAGCAGCTCGAGGATGGCCGCACCCTTGCCGA
346 R L I F A G K Q L E D G R T L A D
3101 CTATAACATCCAGAAGGAGTCCACCCTGCACCTTGTCTGCGCCTCCGTG
363 Y N I Q K E S T L H L V L R L R
3151 GTGGTCTCTGA (SEQ ID NO:8)
379 G G L * (SEQ ID NO:9)

FIG. 2B
9/16

```

R1 -----
R2 -----
R3 GTAAC TAATCAATCACCTCGTCCTAATCCTCGAATCTCTCGTGGTGCCCG 50
M -----

R1 -----
R2 -----
R3 TCTAATCTCGCGATTTTGATGCTCGTGGTGGAAAGCGTAGGAGGATCCCG 100
M -----

R1 -----
R2 -----GTACGCTGCTTCTCCTCTCCTCGCT 25
R3 TGCGAGTTAGTCTCAATCTCTCAGGGTTTCGTGCGATTTTAGGGTGATCC 150
M -----GTACGCCGCTCGTCCTCCCCCCCCC 25

R1 -----GTAT---GCGATTTTCTGAT 17
R2 TCGTTTCGATTTCGATTTCGGACGGGTGAGGTTGT---TTTGTGCTAGAT 72
R3 ACCTCTTAATCGAGTTACGGTTTCGTGCGATTTTAGGGTAATCCTCTTAA 200
M CCCCTCTCTACCTTCTCTAGATCGGCGTTCCGGT---CCATGGTTAGGG 71
          *                *

R1 CCTCTCCGTTCCCTCGCGTT--TGATTTGATT-TCCC-GGCCTGTTCTGTA 63
R2 CCGATTGGTGGTTAGGGTTGTCGATGTGATTATCGT-GAGATGTTTAGGG 121
R3 TCTCTCATTGATTTAGGGTTTCGTGAGAATCGAGGTAGGGATCTGTGTTA 250
M CC---CGGTAGTTCTACTTCTGTTTCATGTTTGTGTTAGATCCGTGTTTGT 118
  *      *      *      *      *      *      *

R1 TTGTGA-----GATGTTGTGGTTAGTCTCCGTTTTGCGATCTGT--GGTA 106
R2 GTTGTA-----GATCTGATGGTTGTGATTTGGGC-ACGGTTGGTTCGATA 165
R3 TTTATATC---GATCTAATAGATGG-ATTGGTTTTGAGATTGTTCTGTCA 296
M GTTAGATCCGTGCTGCTAGCGTTTCGTACACGGAT-GCGACCTGTACGTCA 167
  *      *      *      *      *      *      *      *      *

R1 GATTTGAACAGGGT--TAG-ATGGGGTTCGCGTGGTATGCTGGATCT--G 151
R2 GGTGGAATCGTGGT--TAG-GTT---TTGGGATTGGATGTTGGTTCT--G 207
R3 GATGGGGATTGTTT--CGATATATTACCCTAATGATGTGTCAGATGG--G 342
M GACACGTTCTGATTGCTAACTTGCCAGTGTTTCTCTTTGGGGAATCCTGG 217
  *                *      *      **      *      *

R1 --TGATTATGAGCGATG--CTGAACGT-GGTCCAAGTATTGATTGGTTTCG 196
R2 A-TGATTGGGGGGAATT--TTTACGGTTAGATGAATTGTTGGATGATTTCG 254
R3 GATTGTTTCGATATATTACCCTAATGATGTGTCAGATGGGGATTGTTTCG 392
M GATGGCTCTAGCCGTTCCGCAGACGG--GATCGATTTTCATGATTTTTTTT 265
  *      *      *      *      *      *      *      *      *

R1 GA-----TCTAGAAGTAGA-AGTAGAACTGTGCTAGGGTTGTG-ATTT 237
R2 AT-----TGGGGAAATCGG-TGTAGATCTGTTGGGGAATTGTGGAAC 296
R3 ATATATTACCCTAATGATGGATAATAAGAGTAGTTCACAGTTATG-TTTT 441
M GT-----TTCGTTGCATAGG-GTTTGGTTTGCCCTTTTCCTTTATTTC 308
          * *      *      *      *      *      *

```

FIG. 3A
10/16

```

R1 GTTCCGATCT----GTTCAATCAGTAGGATTTAGTCTCTGTTTTTCTCGT 283
R2 AGTCATGCCT----GAGTGATTGGTGCGATTTGTAGCGTGTTCCATCTAG 342
R3 GATCCTGCCACATAGTTTGAGTTTTGTGATCAGATTTAGTTTTACTTATT 491
M ATATATGCCG--TGCACTTGTTTGTCTGGGTCATCTTTTCATGCTTTTTTT 356
      *                * * *

R1 TGAGCCAAGT-----AGCAGCGTCAGGTATATT-----TTGCTT 358
R2 TAGGCCTTGTT---GCGAGCATGTTTCAGATCTACTGTTCCGCTCTTGATT 389
R3 TGTGCTTAGTTTCGGATGGGATTGTTCTGATATTGT-TCCAATAGATGAAT 540
M TGT-CTTGTTGTGAT-GATGTGGTCTGGTTGGGCGGTC-GTTCTAGATC 403
      * * **                * * * *

R1 AGGTTGTT-----TTTGATTTCAGTCC---CTCTAGTTGCA-TAGATTCTA 383
R2 GAGTTATTGGTGCCATGGGTGGGTGCAAACACAGGCTGCAATATGTTATA 439
R3 AGCTCGTTAGGTTAAATCTTTAGGTTGAGTTAGGCGACA-CATAGTTTA 589
M GGAGTAGAA---TTCTGTTTCAAACCTACCTGGTGGAATTTAATTTTGG 450
      *                * * *

R1 CTCTGTTTCATGTTTAATCTA-----A-----GGGCTGCGTCTTGTTGA 396
R2 TCTGTTTTGTGTTTGATGTAGATCTGTA-----GGGTAGTTCTTCTTAGA 484
R3 TTTCCTCTGGATTTGGATTGGAATTGTGTTCTTAGTTTTTTTCCCCTGGA 639
M ATCTGTATGTGTGTGCCATACATATTCA-----TAGTTACGAATTGAAGA 495
      * * * * **

R1 T-TAG--T--GATTACATAGCATAGCTGTCAGGATATTTTACTTGCTT-A 446
R2 CATGG--TTCAATTATGTAGC-TTGTCTGTTTCGATTTGATGCTCATAT-G 530
R3 TTTGGATTGGAATTGTGTGGAGCTGG-GTTAGAGAATTACATCTGTATCG 688
M TGATGGATGGAAATATCGATCTAGGATAGGTATACATGTTGAT-GCGGGT 544
      * * * * *

R1 TGCCTATCT-TATCAACTGTTGCACC-----TGTA AATTCTAGCCTATGT 484
R2 TTCACAGAT-TAGATAATGATGAACTC---TTTTAATTAATTGTCAATGG 576
R3 TGTACACCTACTTGAACTGTAGAGCTTGGGTTCTAAGGTCAATTTAATCT 738
M TTTACTGATGCATATACAGAGATGCTT---TTTGTTTCGCTTGGTTGTGA 590
      * * * * *

R1 TAATT--AACCTGCCTTATG--TGCT--CTCGGGATAG-TGCTAGTA-GT 526
R2 TAAAT--AGGAAGTCTTATCGCTATA--TCTGTCATAA-TGATCTCATGT 621
R3 GTATTGTATCTGGCTCTTTGCCTAGT--TGAAGTGTAG-TGCTGATGTTG 785
M TGATGTGGTGTGGTTGGGCGGTCGTTTCATTCGTTCTAGATCGGAGTAGAA 640
      * * ** *

R1 TATTG-----AATCAGTT-TGCCGA-----TGGAATTCTA--- 555
R2 TACTAT--CTGCCAGTAATTTATGCTAAG-----CACTATATTA--- 658
R3 TACTGTGTTTTTTTACCCGTTTTATTTGCTTTACTCGTGCAAATCAA--- 832
M TACTGTTTCAAACCTACCTGGTGTATTTAT--TAATTTTGGAAGTGTATGT 688
      ** * * *

```

R1	GTAGTTCATA----GACCTGCAG-ATTAT--TTTTGTGAACTCGAGCACG	598
R2	GAATATCATGTTACAATCTGTAGTAATATCATGTTACAATCTGTAGTTCA	708
R3	ATCTGTCAGATGCTAGAACTAGGTGGCTTTATTCTGTGTTCTTACATAGA	882
M	GTGTGTCATAC----ATCTTCATAGTTACGAGTTTAAGATGGATGGAAAT	734
	*** *	
R1	GTGCGT-CTCTCTATTTTGTAGGTCA---CTGTTGGTGTGATAGGTAC	644
R2	TCTATA-TAATCTATTGTGGTAATTTT---TTTTACTATCTGTGTGAAG	754
R3	TCTGTTGTCCTGTAGTTACTTATGTCAGTTTTGTTATTATCTGAAGATAT	932
M	ATCGATCTAGGATAGGTATACATGTTGATGTGGGTTTTACTGATGCATAT	784
	* * * * *	
R1	ACT--GATGT-----TATTGTGGTTTAGATCGTGTATCTAAC-ATAT	683
R2	ATT--ATTGC-----CACTAGTTCATTCTACTTATTTCTGAA-GTTC	793
R3	TTTTGGTTGT-----TGCTTGTTGATGTGGTGTGAGCTGTGA-GCAG	973
M	ACATGATGGCATATGCAGCATCTATTCATATGCTCTAACCTTGAGTACCT	834
	* * *	
R1	TGGAATAAT--TTGAT-TGACTGATTTCTGCTGTACTTG--CTTGGTATT	728
R2	AGGATACGT--GTGCTGTTACTACCTATCTGAATACATG--TGTGATGTG	839
R3	CGCTCTTATGATTAATGATGCTGTCCAATTGTAGTGTAG--TATGATGTG	1021
M	ATCTATTATAATAACAAGTATGTTTTATAATTATTTTGATCTTGATATA	884
	* * *	
R1	GTTATAAT-TT-CATGTTTCAT--AGTTGCTGACCATGCTTCGGTA-ATTG	773
R2	CCTGTTAC-T---ATCTTTTT--GAATACATGT-ATGTTCTGTTGGAATA	882
R3	ATTGATATGTT-CATCTATTTTGAGCTGACAGTACCGATATCGTAGGATC	1070
M	CTTGGATGATGGCATATGCAGC-AGCTATATGTGGATTTTTTTTAGCCCTG	933
	* * * *	
R1	TGTGTGCAG-----	782
R2	TGTTTGCTGTTTGATCCGTTGTTGTGTCCTTAATCTTGTGCTAGTTCTTA	932
R3	TGGTGCCAACTTATTCTCCAGCTGCTTTTTTTTACCTATGTTAATTCCAA	1120
M	CCTTCATACGCTATTTATTTGCTTGGTACTGTTTCTTTTGTGCGATGCTCA	983
R1	----- (SEQ ID NO:14)	
R2	CCCTATCTGTTTGGTGATTATTTCTTGCAG (SEQ ID NO:15)	962
R3	TCCTTTCTTGCCTCTTCCAG----- (SEQ ID NO:3)	1140
M	CCCTGT-TGTTTGGTGTACTTCTGCAG-- (SEQ ID NO:16)	1010

FIG. 3C
12/16

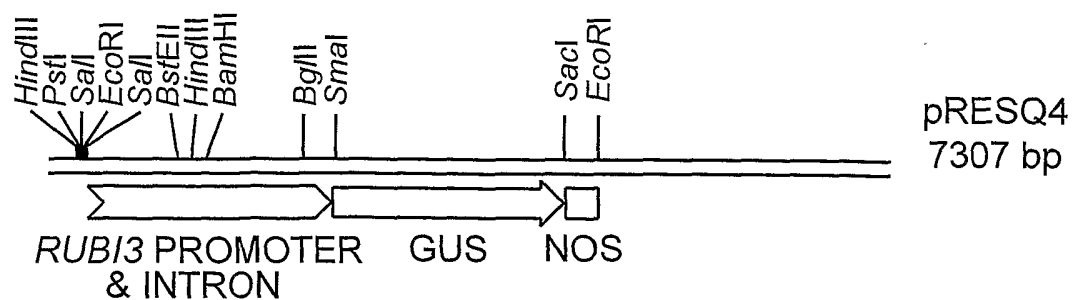


FIG. 4A

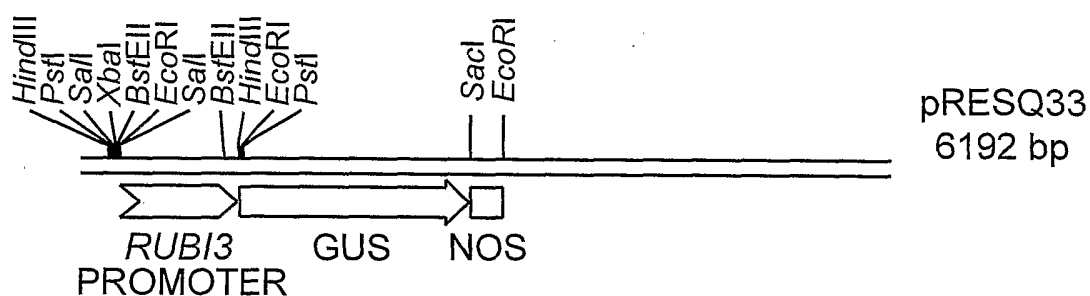


FIG. 4B

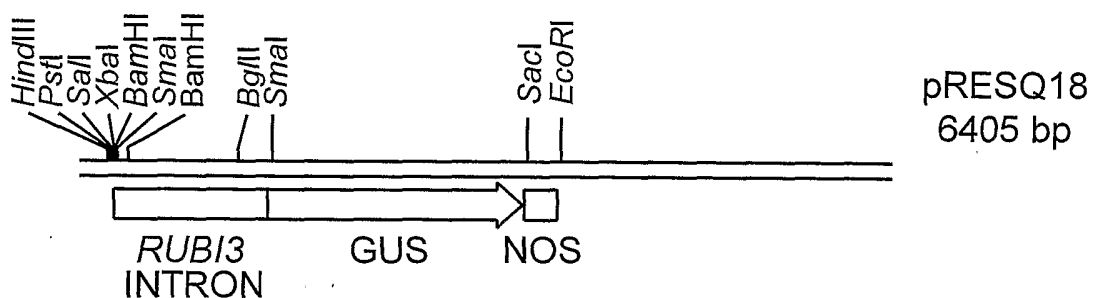


FIG. 4C

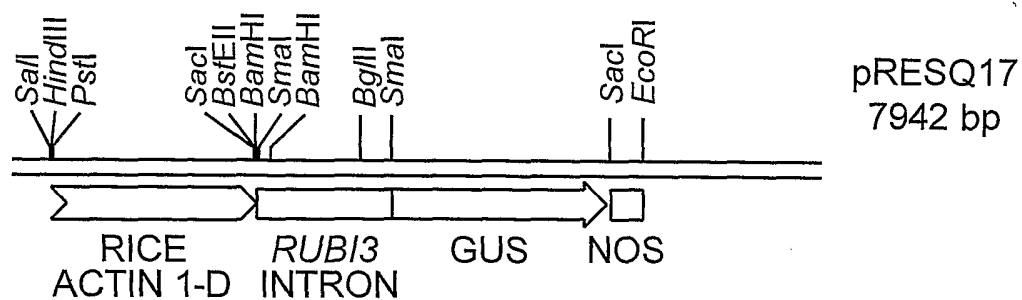


FIG. 4D

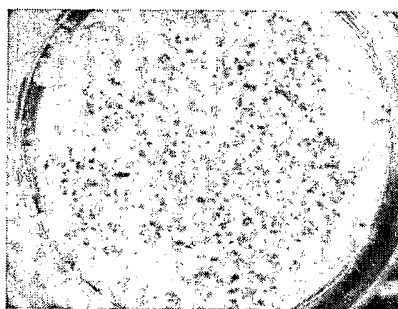
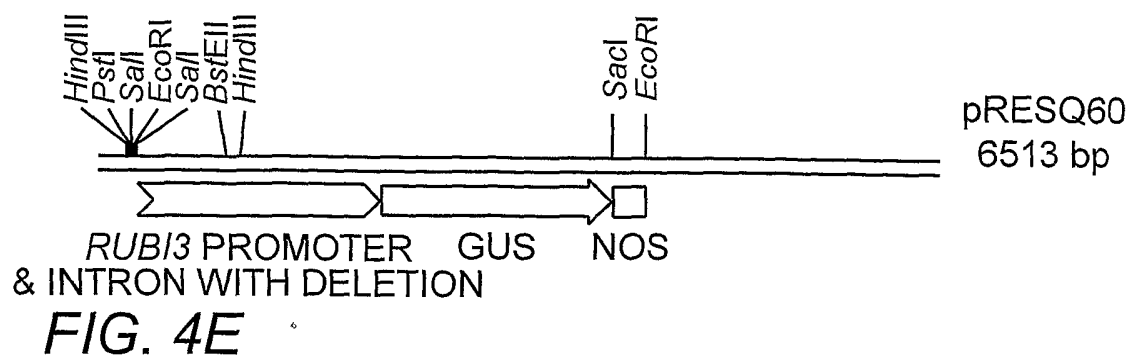


FIG. 5A

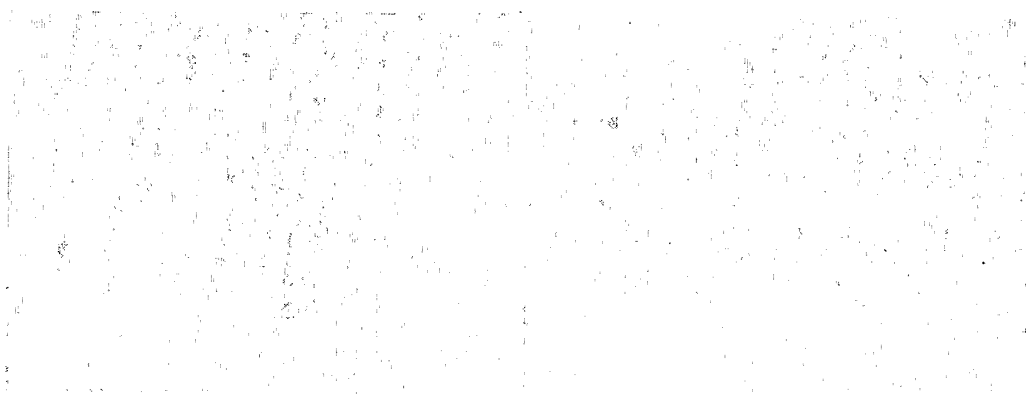


FIG. 5B

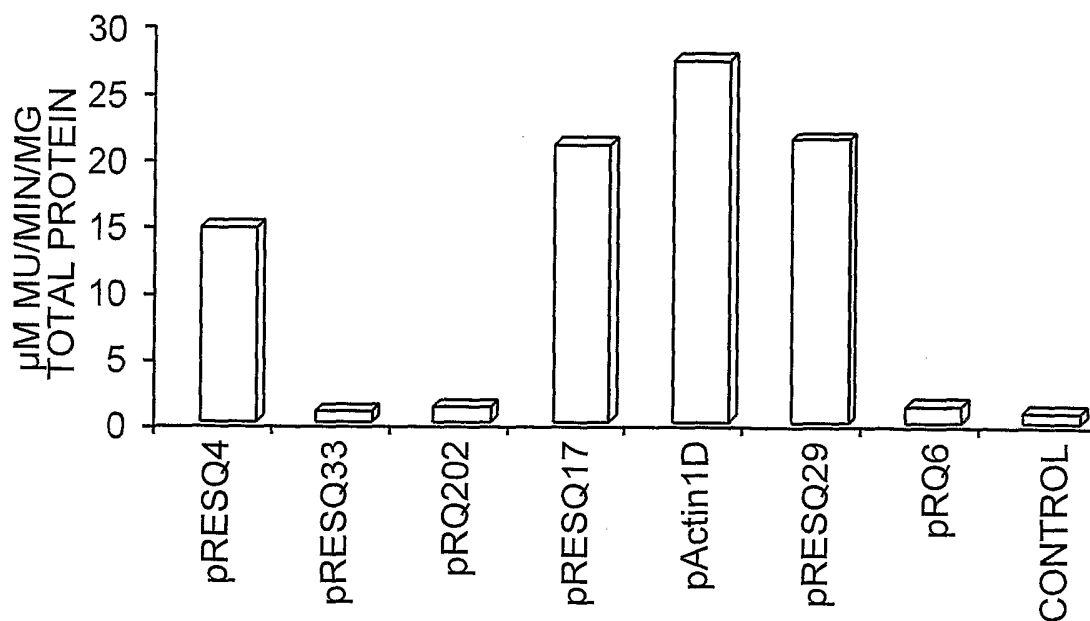


FIG. 6

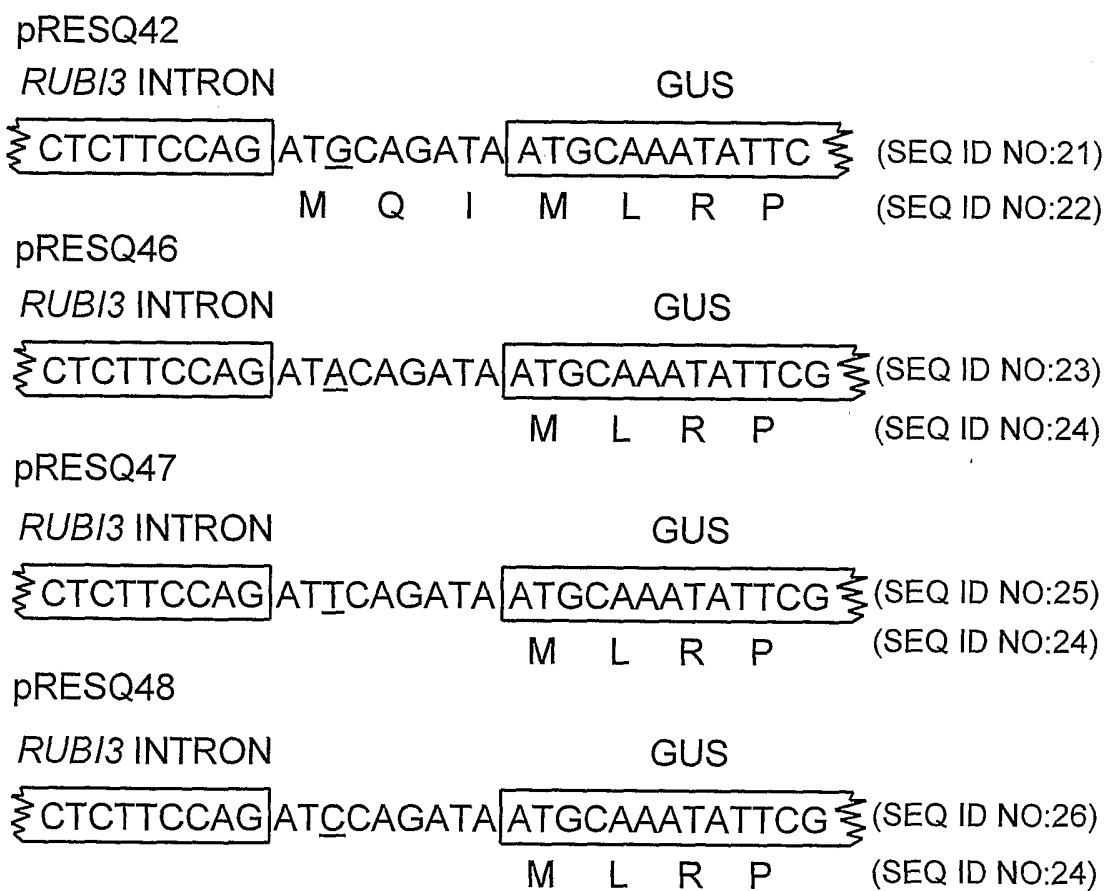


FIG. 7

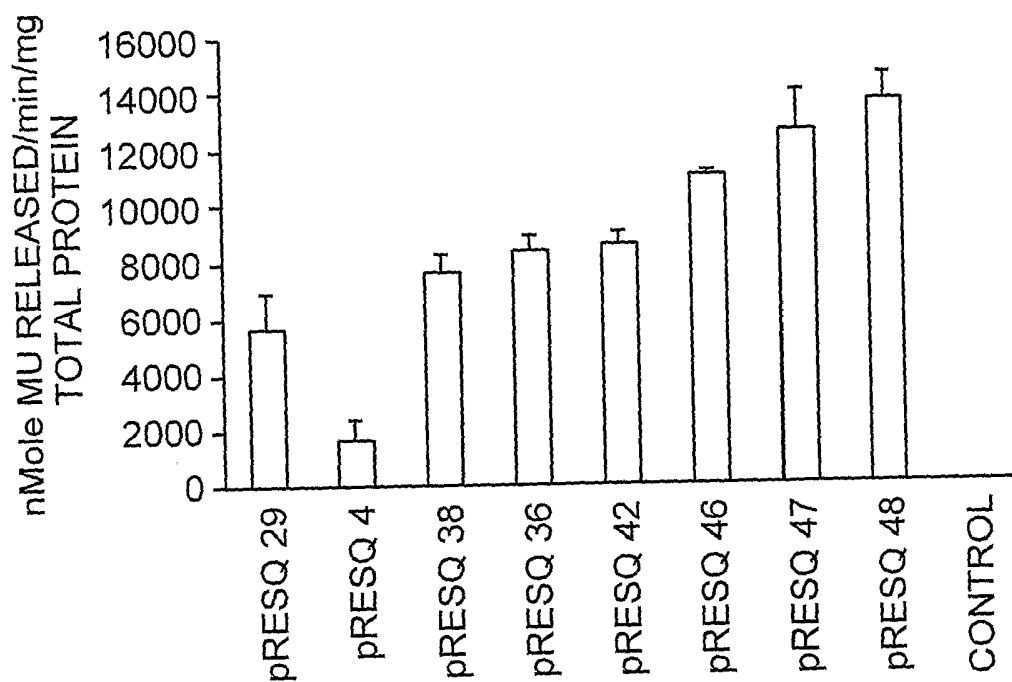


FIG. 8

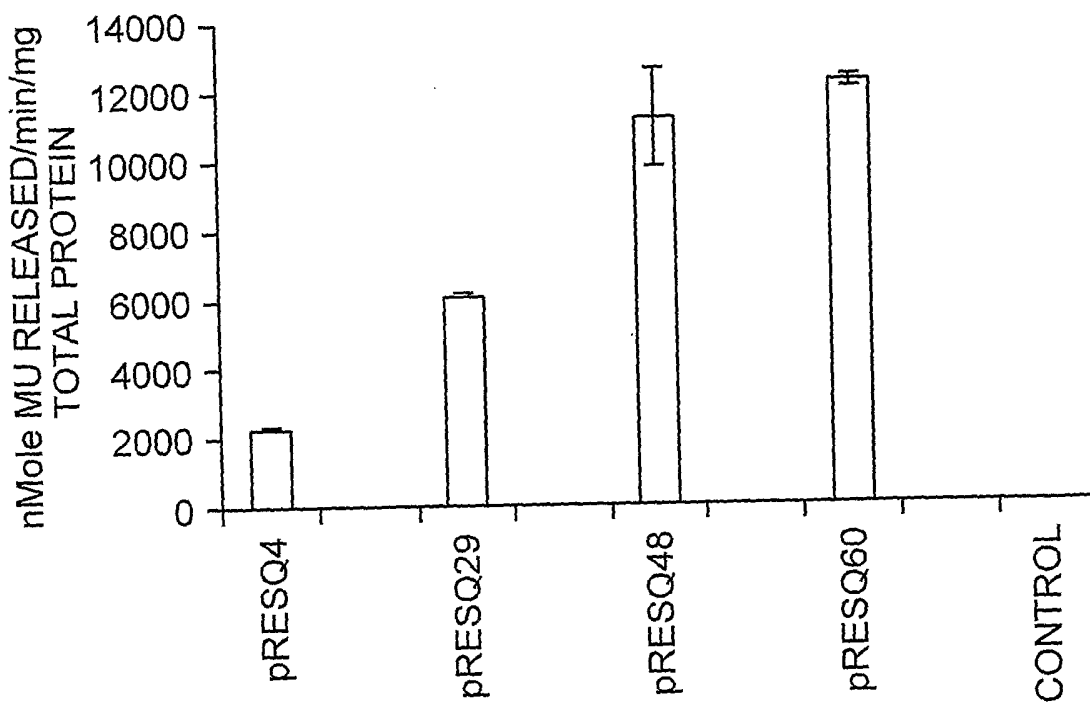


FIG. 9

SEQUENCE LISTING

<110> Elumalai, Sivamani
Qu, Rongda

<120> POLYUBIQUITIN rubi3 PROMOTER AND 5' REGULATORY SEQUENCE

<130> 5051.634

<150> US 60/471,007

<151> 2003-05-05

<160> 27

<170> PatentIn version 3.2

<210> 1

<211> 2015

<212> DNA

<213> Oryza sativa

<400> 1

```

ccacccaacc ccatatcgac agaggatgtg aagaacaggt aaatcacgca gaagaaccca      60
tctctgatag cagctatoga ttagaacaac gaatccatat tgggtccgtg ggaaatactt      120
actgcacagg aagggggggcg atctgaacgag gcccgcgcac cggcctcgac ccgaggccga      180
ggccgacgaa gcgcggcgga gtacggcgcc gcggcggcct ctgcccgtgc cctctgcgeg      240
tgggagggag aggccgcggt ggtggggggcg cgcgcgcgcg cgcgcgcagc tgggtgcggcg      300
gcgcgggggt cagccgcgga gccggcgggcg acggaggagc agggcggcgt ggacgcgaac      360
ttccgatcgg ttggtcagag tgcgcgagtt gggcttagcc aattaggtct caacaatcta      420
ttgggcccgt aaattcatgg gccctggttt gtctaggccc aatatcccgt tcatttcagc      480
ccacaaatat ttcccagag gattattaag gcccacacgc agcttatagc agatcaagta      540
cgatgtttcc tgatcgttgg atcggaacg tacggtcttg atcaggcatg ccgacttcgt      600
caaagagagg cggcatgacc tgacgcggag ttggttcggg gcaccgtctg gatggtcgta      660
ccgggaccgg acacgtgtcg cgcctccaac tacatggaca cgtgtggtgc tgccattggg      720
ccgtacgcgt ggcggtgacc gcaccggatg ctgcctcgca ccgccttgcc cagcctttat      780
atagagaggt tttctctcca ttaatcgcat agcgagtcga atcgaccgaa ggggaggggg      840
agcgaagcct tgcgttctct aatcgccctg tcaaggtaac taatcaatca cctcgtccta      900
atcctcgaat ctctcgtggt gcccgctctaa tctcgcgatt ttgatgctcg tgggtggaaag      960
cgtaggagga tcccgtgcga gttagtctca atctctcagg gtttcgtgcg attttagggg      1020
gatccacctc ttaatcgagt tacggtttcg tgcgatttta gggtaatcct cttaatctct      1080

```

cattgatttta gggtttcgtg agaatcgagg tagggatctg tgttatttat atcgatctaa 1140
tagatggatt ggttttgaga ttgttctgtc agatggggat tgtttcgata tattacccta 1200
atgatgtgtc agatggggat tgtttcgata tattacccta atgatgtgtc agatggggat 1260
tgtttcgata tattacccta atgatggata ataagagtag ttcacagtta tgttttgatc 1320
ctgccacata gtttgagttt tgtgatcaga tttagtttta cttatttgtg cttagttcgg 1380
atgggattgt tctgatattg ttccaataga tgaatagctc gttagggttaa aatctttagg 1440
ttgagttagg cgacacatag tttatttcoct ctggatttgg attggaattg tgttcttagt 1500
ttttttcccc tggatttggg ttggaattgt gtggagctgg gttagagaat tacatctgta 1560
tcgtgtacac ctacttgaac tgtagagctt gggttctaag gtcaatttaa tctgtattgt 1620
atctggctct ttgcctagtt gaactgtagt gctgatgttg tactgtgttt ttttaccctg 1680
tttatttgc tttactgtgc aaatcaaato tgtcagatgc tagaactagg tggctttatt 1740
ctgtgttctt acatagatct gttgtcctgt agttacttat gtcagttttg ttattatctg 1800
aagatatttt tggttgttgc ttgttgatgt ggtgtgagct gtgagcagcg ctcttatgat 1860
taatgatgct gtccaattgt agttagtat gatgtgattg atatgttcat ctattttgag 1920
ctgacagtac cgatatcgta ggatctggtg ccaacttatt ctccagctgc ttttttttac 1980
ctatgttaat tccaatcctt tcttgccctc tccag 2015

<210> 2
<211> 875
<212> DNA
<213> *Oryza sativa*

<400> 2
ccaccaacc ccatatcgac agaggatgtg aagaacaggt aaatcacgca gaagaacca 60
tctctgatag cagctatcga ttagaacaac gaatccatat tgggtccgtg ggaaatactt 120
actgcacagg aaggggggag atctgacgag gccccgccac cggcctcgac ccgaggccga 180
ggccgacgaa gcgccggcga gtacggcgcc gcggcgccct ctgcccgtag cctctgcgag 240
tgggagggag aggcgcggtt ggtgggggag cgcgcgcgag cgcgcgagc tgggtgcggcg 300
gcgcgggggg cagccgccga gccggcgagg acggaggagc agggcggcgt ggacgcgaac 360
ttccgatcgg ttgggtcagag tgcgcgagtt gggcttagcc aattaggctt caacaatcta 420
ttgggcccga aaattcatgg gccctgggtt gtctaggccc aatatcccgt tcatttcagc 480
ccacaaatat ttccccagag gattattaag gccacacgc agcttatagc agatcaagta 540

cgatgtttcc tgatcgttgg atcggaaacg tacgggtcttg atcaggcatg ccgaacttcgt 600
 caaagagagg cggcatgacc tgacgcggag ttggttccgg gcaccgtctg gatggtcgta 660
 ccgggaccgg acacgtgtcg cgctccaac tacatggaca cgtgtggtgc tgccattggg 720
 ccgtacgcgt ggcgtgacc gcaccggatg ctgcctcgca ccgccttgcc cagcctttat 780
 atagagaggt tttctctcca ttaatcgcat agcgagtoga atcgaccgaa ggggaggggg 840
 agcgaagcct tgcgttctct aatcgccctg tcaag 875

<210> 3
 <211> 1140
 <212> DNA
 <213> *Oryza sativa*

<400> 3
 gtaactaatc aatcacctcg tcctaatact cgaatctctc gtggtgcccg tctaatactcg 60
 cgattttgat gctcgtggtg gaaagcgtag gaggatcccg tgcgagttag tctcaatctc 120
 tcagggtttc gtgcgatttt agggatgatc acctcttaac cgagttacgg tttcgtgcga 180
 ttttagggta atcctcttaa tctctcattg atttaggggt tcgtgagaat cgaggtaggg 240
 atctgtgtta tttatatcga tctaatagat ggattgggtt tgagattgtt ctgtcagatg 300
 gggattgttt cgatatatta ccctaatagat gtgtcagatg gggattgttt cgatatatta 360
 ccctaatagat gtgtcagatg gggattgttt cgatatatta ccctaatagat ggataataag 420
 agtagttcac agttatgttt tgatcctgcc acatagtttg agttttgtga tcagatttag 480
 ttttacttat ttgtgcttag ttcgatggg attgttctga tattgttcca atagatgaat 540
 agctcgttag gttaaaatct ttaggttgag ttaggcgaca catagtttat ttcctctgga 600
 tttggattgg aattgtgttc ttagtttttt tcccctggat ttggattgga attgtgtgga 660
 gctgggttag agaattacat ctgtatctg tacacctact tgaactgtag agcttgggtt 720
 ctaagggtcaa tttaatctgt attgtatctg gctctttgcc tagttgaact gtagtgctga 780
 tgttgtactg tgttttttta ccggttttat ttgctttact cgtgcaaata aaatctgtca 840
 gatgctagaa ctaggtggct ttattctgtg ttcttacata gatctgttgt cctgtagtta 900
 cttatgtcag ttttgttatt atctgaagat atttttgggt gttgcttggt gatgtggtgt 960
 gagctgtgag cagcgtctct atgattaatg atgctgtcca attgtagtgt agtatgatgt 1020
 gattgatatg ttcactatatt ttgagctgac agtaccgata tcgtaggata tgggtgccaa 1080
 ttattctcca gctgcttttt ttacctatg ttaattccaa tcctttcttg cctcttcacg 1140

<210> 4
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 4
gtcgaccacc caaccccata tcgacagagg 30

<210> 5
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 5
cccgggctgg aagaggcaag aaaggattgg aattaac 37

<210> 6
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 6
cccgggtaac taatcaatca cctcgt 26

<210> 7
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic oligonucleotide primer

<400> 7
cttgacgagg cgattagaga acgc 24

<210> 8
<211> 3161
<212> DNA
<213> Oryza sativa

<400> 8
acacccaacc ccatatcgac agaggatgtg aagaacaggt aaatcacgca gaagaaccca 60

tctctgatag cagctatcga ttagaacaac gaatccatat tgggtccgtg ggaaatactt	120
actgcacagg aaggggggag atctgacgag gccccgccac cggcctcgac ccgaggccga	180
ggccgacgaa gcgcccgcga gtacggcgcc gggcgggcct ctgcccggtg cctctgcgcg	240
tgggagggag aggcgcgcgt ggtgggggag cgcgcgcgcg cgcgcgcgcg tgggagggag	300
gcgcgggggt cagccgccga gccggcggcg acggaggagc agggcggcgt ggacgcgaac	360
ttccgatcgg ttggtcagag tgcgcgagtt gggccttagcc aattaggtct caacaatcta	420
ttgggcccga aaattcatgg gccctgggtt gtctaggccc aatatccgt tcatctcagc	480
ccacaaatat ttcccagag gattattaag gccacacgc agcttatagc agatcaagta	540
cgatgtttcc tgatcgttgg atcgaaacg tacggtcttg atcaggcatg ccgacttcgt	600
caaagagagg cggcatgacc tgacgcggag ttggttcagg gcaccgtctg gatggtcgtg	660
ccgggaccgg acaagtgtcg cgcctccaac tacatggaca cgtgtggtgc tgccattggg	720
ccgtacgcgt ggcggtgacc gcaccggatg ctgcctcgca ccgccttgcc cagcctttat	780
atagagaggt tttctctcca ttaatogcat agcgagtcga atcgaccgaa ggggaggggg	840
agcgaagctt tgcgttctct aatcgccctg tcaaggtaac taatcaatca cctcgtccta	900
atcctcgaat ctctcgtggt gcccgctctaa tctcgcgatt ttgatgctcg tggggaag	960
cgtaggagga tcccgtcga gttagtctca atctctcagg gtttcgtgcg attttaggg	1020
gatccacctc ttaatcgagt tacggtttcg tgcgatttta gggtaatcct cttaatctct	1080
cattgattta gggtttcgtg agaatcgagg tagggatctg tgttatttat atcgatctaa	1140
tagatggatt ggttttgaga ttgttctgtc agatggggat tgtttcgata tattacccta	1200
atgatgtgtc agatggggat tgtttcgata tattacccta atgatgtgtc agatggggat	1260
tgtttcgata tattacccta atgatggata ataagagtag ttcacagtta tgttttgatc	1320
ctgccacata gtttgagttt tgtgatcaga tttagtttta cttatttbtg cttagtccg	1380
atgggattgt tctgatattg ttccaataga tgaatagctc gttagggtta aatctttagg	1440
ttgagttagg cgacacatag tttatttcct ctggatttgg attggaattg tgttcttagt	1500
ttttttcccc tggatttgga ttggaattgt gtggagctgg gttagagaat tacatctgta	1560
togtgtacac ctacttgaac tgtagagctt gggttctaa gtcaatttaa tctgtattgt	1620
atctggctct ttgcctagtt gaactgtagt gctgatgttg tactgtgtt ttttaccctg	1680
tttatttgct ttactcgtgc aaatcaaata tgtcagatgc tagaactagg tggctttatt	1740

```

ctgtgttctt acatagatct gttgtcctgt agttacttat gtcagttttg ttattatctg 1800
aagatatattt tggttgttgc ttgttgatgt ggtgtgagct gtgagcagcg ctcttatgat 1860
taatgatgct gtccaattgt agtgtagtat gatgtgattg atatgttcat ctattttgag 1920
ctgacagtac cgatatcgta ggatctgggtg ccaacttatt ctccagctgc ttttttttac 1980
ctatgttaat tccaatcctt tcttgccctc tccagatgca gatattcggtt aagaccctca 2040
ctggcaagac catcaccttg gaggttgagt tctccgatac gattgacaat gtgaaggcta 2100
agattcagga caaggagggc atccctccgg accagcaacg ccttatcttc gctggcaagc 2160
agcttgagga tgggcgtact ctgcgggatt ataacatcca gaaggagtcc accttgccacc 2220
ttgtcctccg ccttcgtgga ggcatgcaaa tattcgtgaa gaccctcacc ggcaagacca 2280
ttaccctgga ggtcgagtc tccgacacga tcgataatgt gaaggccaag atccaggaca 2340
aggaggggaat cccaccagac cagcagcgtc tcatctttgc tgggaagcag ctcgaggatg 2400
gccgcaccct tgcagactac aacatccaaa aggaatccac cctgcacctt gtctgcgcc 2460
tccggggcgg tatgcagatc tttgtgaaga cccttactgg caagacgac accttgaggg 2520
ttgagtcttc tgacacgac gacaatgtga aggccaagat ccaggacaag gagggatttc 2580
caccagacca gcagcggctc atcttcgctg gcaagcagct tgaggacggc cgcacccttg 2640
ctgactataa catccagaag gagtccacct tgcacttggc cctccgcctt cgtggaggta 2700
tgcagatctt cgtgaagacc ctccaccgca agaccatcac cctggagggtt gagtctctg 2760
acacgatcga caatgtgaag gccaagatcc aggacaagga ggggtattcca ccagaccagc 2820
agcgcctcgt cttcgctggc aagcagcttg aggatggctg cacccttgca gattacaaca 2880
tccagaagga gtccaccctg caccttgtcc tgcgccttcg tggaggatatg cagatcttcg 2940
tgaagaccct cactggcaag accatcactc tggagggttg gtcctctgac accatcgaca 3000
acgtgaaggc caagatccag gataaggagg gcatccctcc agaccagcag cgcctcatct 3060
tcgccggcaa gcagctcgag gatggccgca cccttgccga ctataacatc cagaaggagt 3120
ccaccctgca ccttgtcctg cgcctccgtg gtggtctctg a 3161

```

<210> 9

<211> 381

<212> PRT

<213> Oryza sativa

<400> 9

Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu

1	5	10	15
Val Glu Phe Ser Asp Thr Ile Asp Asn Val Lys Ala Lys Ile Gln Asp	20	25	30
Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys	35	40	45
Gln Leu Glu Asp Gly Arg Thr Leu Ala Asp Tyr Asn Ile Gln Lys Glu	50	55	60
Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe	65	70	75
Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Ser Ser	85	90	95
Asp Thr Ile Asp Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile	100	105	110
Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp	115	120	125
Gly Arg Thr Leu Ala Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His	130	135	140
Leu Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe Val Lys Thr Leu	145	150	155
Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Ser Ser Asp Thr Ile Asp	165	170	175
Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln	180	185	190
Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu	195	200	205
Ala Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg	210	215	220
Leu Arg Gly Gly Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr	225	230	235
			240

Ile Thr Leu Glu Val Glu Ser Ser Asp Thr Ile Asp Asn Val Lys Ala
245 250 255

Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Val
260 265 270

Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ala Asp Tyr Asn
275 280 285

Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly
290 295 300

Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
305 310 315 320

Val Glu Ser Ser Asp Thr Ile Asp Asn Val Lys Ala Lys Ile Gln Asp
325 330 335

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
340 345 350

Gln Leu Glu Asp Gly Arg Thr Leu Ala Asp Tyr Asn Ile Gln Lys Glu
355 360 365

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Leu
370 375 380

<210> 10
<211> 15
<212> DNA
<213> Drosophila sp.

<220>
<221> misc_feature
<222> (8)..(8)
<223> "n" represents A, T, C or G.

<400> 10
ctggaatntt ctaga

15

<210> 11
<211> 11
<212> DNA
<213> Artificial Sequence

<220>
 <223> Enhancer consensus sequence

<400> 11
 ggtgtgwww g 11

<210> 12
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic oligonucleotide primer

<400> 12
 cgaatcgacc gaaggggagg 20

<210> 13
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic oligonucleotide primer

<400> 13
 acacgatgat atgacagacg agc 23

<210> 14
 <211> 782
 <212> DNA
 <213> Oryza sativa

<400> 14
 gtatgcgatt ttctgacct ctccgttcct cgcgtttgat ttgatttccc ggcctgttcg 60
 tgattgtgag atgttggtg tagtctccgt ttgcgatct gtggtagatt tgaacagggt 120
 tagatggggg tgcggtgga tgctggatct gtgattatga gcgatgctga acgtggtcca 180
 agtattgatt ggttcggatc tagaagtaga agtagaactg tgctagggtt gtgatttggt 240
 ccgatctgtt caatcagtag gatttagtct ctgtttttct cgttgagcca agtagcagcg 300
 tcaggatatat ttgcttagg ttgtttttga ttcagtcctt ctagttgcat agatttact 360
 ctgttcattgt ttaatctaag ggctgcgtct tgttgattag tgattacata gcatagctgt 420
 caggatattt tacttgctta tgcctatctt atcaactgtt gcacctgtaa attctagcct 480
 atgttaatta acctgcctta tgtgctctcg ggatagtgt agtagttatt gaatcagttt 540
 gccgatggaa ttctagtagt tcatagacct gcagattatt tttgtgaact cgagcacggg 600

gcgctctctct attttggttag gtcactgttg gtgttgatag gtacactgat gttattgtgg 660
 tttagatcgt gtatctaaca tattggaata atttgattga ctgatttctg ctgtacttgc 720
 ttggtattgt tataatttca tgttcatagt tgctgaccat gcttcggtaa ttgtgtgtgc 780
 ag 782

<210> 15
 <211> 962
 <212> DNA
 <213> Oryza sativa

<400> 15
 gtacgctgct tctcctctcc tcgcttcggt tcgattcgat ttcggacggg tgagggtggt 60
 ttgttgctag atccgattgg tggtaggggt tgcgatgtg attatcgtga gatgtttagg 120
 ggttgtagat ctgatgggtg tgatttgggc acggttggtt cgatagggtg aatcgtgggt 180
 aggttttggg attggatggt ggttctgatg attgggggga atttttacgg ttagatgaat 240
 tgttggatga ttcgattggg gaaatcgggt tagatctggt ggggaattgt ggaactagtc 300
 atgcctgagt gattgggtgcg atttgtagcg tgttccatct agtaggcctt gttgcgagca 360
 tgttcagatc tactgttccg ctcttgattg agttattggt gccatgggtg ggtgcaaaca 420
 caggctgcaa tatgttatat ctgttttgtg tttgatgtag atctgtaggg tagttcttct 480
 tagacatggt tcaattatgt agcttgtcgt ttogatttga tgctcatatg ttcacagatt 540
 agataatgat gaactotttt aattaattgt caatggtaaa taggaagtct tatcgctata 600
 tctgtcataa tgatctcatg ttactatctg ccagtaattt atgctaagca ctatattaga 660
 atatcatggt acaatctgta gtaatatcat gttacaatct gtagttcatc tatataatct 720
 attgtggtaa tttcttttta ctatctgtgt gaagattatt gccactagtt cattctactt 780
 atttctgaag ttcaggatac gtgtgctggt actacctatc tgaatacatg tgtgatgtgc 840
 ctgttactat ctttttgaat acatgtatgt tctgttgga tatgtttgct gtttgatccg 900
 ttgttgtgtc cttaatcttg tgctagttct taccctatct gtttggtgat tatttcttgc 960
 ag 962

<210> 16
 <211> 1010
 <212> DNA
 <213> Zea mays

<400> 16

```

gtacgcccgt cgtcctcccc cccccccct ctctaccttc tctagatcgg cgttccggtc 60
catggttagg gcccggtagt tctacttctg ttcattgtttg tgttagatcc gtgttttgtgt 120
tagatccgtg ctgctagcgt tcgtacacgg atgcgacctg tacgtcagac acgtttctgat 180
tgctaacttg ccagtgtttc tctttgggga atcctgggat ggctctagcc gttccgcaga 240
cgggatcgat ttcattgattt tttttgtttc gttgcatagg gtttggtttg cctttttcct 300
ttatttcaat atatgccgtg cacttggttg tcgggtcatc ttttcatgct tttttttgtc 360
ttggttgtga tgatgtggtc tggttgggcg gtcgttctag atcggagtag aattctgttt 420
caaactacct ggtggattta ttaatttttg atctgtatgt gtgtgccata catattcata 480
gttacgaatt gaagatgatg gatggaaata tcgatctagg ataggtatac atgttgatgc 540
gggttttact gatgcatata cagagatgct ttttgttcgc ttggttgtga tgatgtggtg 600
tggttgggcg gtcgttcatt cgttctagat cggagtagaa tactgtttca aactacctgg 660
tgtatttatt aattttggaa ctgtatgtgt gtgtcataca tcttcatagt tacgagttta 720
agatggatgg aaatatcgat ctaggatagg tatacatggt gatgtgggtt ttactgatgc 780
atatacatga tggcatatgc agcatctatt catatgctct aaccttgagt acctatctat 840
tataataaac aagtatgttt tataattatt ttgatcttga tatacttgga tgatggcata 900
tgcagcagct atatgtggat ttttttagcc ctgccttcat acgctattta tttgcttggt 960
actgtttctt ttgtcgatgc tcacctgtt gtttggtgtt acttctgcag 1010

```

<210> 17
 <211> 2024
 <212> DNA
 <213> *Oryza sativa*

```

<400> 17
ccaccaacc ccatatcgac agaggatgtg aagaacaggt aaatcacgca gaagaaccca 60
tctctgatag cagctatoga ttagaacaac gaatccatat tgggtccgtg ggaaatactt 120
actgcacagg aaggggggcg atctgacgag gccccgccac cggcctcgac ccgaggccga 180
ggccgacgaa gcgcggcgga gtacggcgcc gcggcgccct ctgccctgac cctctgcgag 240
tgggagggag aggcgcgggt ggtgggggcg cgcgcgcgag cgcgcgcgag tgggtgcggcg 300
gcgcgggggt cagccgcgga gccggcgggc acggaggagc agggcggcgt ggacgcgaac 360
ttccgatcgg ttggtcagag tgcgcgagtt gggcttagcc aattaggtct caacaatcta 420
ttgggcccga aaattcatgg gccctggttt gtctaggccc aatatcccgt tcatttcagc 480

```

```

ccacaaatat ttccccagag gattattaag gccacacgc agcttatagc agatcaagta 540
cgatgtttcc tgategttgg atcggaacg tacggtcttg atcaggcatg ccgacttcgt 600
caaagagagg cggcatgacc tgacgcggag ttggttccgg gcaccgtctg gatggtcgta 660
ccgggaccgg acacgtgtcg cgcctccaac tacatggaca cgtgtggtgc tgccattggg 720
ccgtacgcgt ggcggtgacc gcaccggatg ctgcctcgca ccgccttgcc cacgctttat 780
atagagaggt tttctctcca ttaatcgcat agcgagtcga atcgaccgaa ggggaggggg 840
agcgaagctt tgcgttctct aatcgctcg tcaaggtaac taatcaatca cctcgtccta 900
atcctcgaat ctctcgtggt gcccgctcaa tctcgcgatt ttgatgctcg tggtggaag 960
cgtaggagga tcccgtgcga gttagtctca atctctcagg gtttcgtgcg attttaggg 1020
gatccacctc ttaatcgagt tacggtttcg tgcgatttta gggtaatcct cttaatctct 1080
cattgattta gggtttcgtg agaatcgagg tagggatctg tgttatttat atcgatctaa 1140
tagatggatt ggttttgaga ttgttctgtc agatggggat tgtttcgata tattacccta 1200
atgatgtgtc agatggggat tgtttcgata tattacccta atgatgtgtc agatggggat 1260
tgtttcgata tattacccta atgatggata ataagagtag ttcacagtta tgttttgatc 1320
ctgccacata gtttgagttt tgtgatcaga tttagtttta cttatttgtg cttagttcgg 1380
atgggattgt tctgatattg ttccaataga tgaatagctc gttagggtta aatctttagg 1440
ttgagttagg cgacacatag tttatttctc ctggatttgg attggaattg tgttcttagt 1500
ttttttcccc tggatttgga ttggaattgt gtggagctgg gttagagaat tacatctgta 1560
tcgtgtacac ctacttgaac tgtagagctt gggttctaag gtcaatttaa tctgtattgt 1620
atctggctct ttgcctagtt gaactgtagt gctgatgttg tactgtgttt ttttaccgt 1680
tttatttgc tttactcgtgc aaatcaaato tgtcagatgc tagaactagg tggctttatt 1740
ctgtgttctt acatagatct gttgtcctgt agttacttat gtcagttttg ttattatctg 1800
aagatatatt tggttgttgc ttgttgatgt ggtgtgagct gtgagcagcg ctcttatgat 1860
taatgatgct gtccaattgt agtgtagtat gatgtgattg atatgttcat ctattttgag 1920
ctgacagtac cgatatcgta ggatctggtg ccaacttatt ctccagctgc ttttttttac 1980
ctatgttaat tccaatcctt tcttgcctct tccagatgca gata 2024

```

<210> 18

<211> 2024

<212> DNA

<213> *Oryza sativa*

<400> 18

```

ccaccaacc ccataatcgac agaggatgtg aagaacaggt aaatcacgca gaagaaccca      60
tctctgatag cagctatcga ttagaacaac gaatccatat tgggtccgtg ggaaatactt      120
actgcacagg aaggggggag atctgacgag gccccgccac cggcctcgac ccgaggccga      180
ggccgacgaa gcgccggcga gtacggcgcc gcggcgccct ctgcccgtgc cctctgcgcg      240
tgaggaggag aggccgcggt ggtgggggag cgcgcgcgcg cgcgcgcgag tgggtgcggcg      300
gcgcgggggt cagccgccga gccggcgagg acggaggagc agggcgggcg ggaacgcgaac      360
ttccgatcgg ttggtcagag tgcgcgagtt gggccttagcc aattaggtct caacaatcta      420
ttgggccgta aaattcatgg gccctgggtt gtctaggccc aatatcccg tcatctcagc      480
ccacaaatat ttcccagag gattattaag gccacacgc agcttatagc agatcaagta      540
cgatgtttcc tgatcgttgg atcgaaaag tacggtcttg atcaggcatg ccgacttcgt      600
caaagagagg cggcatgacc tgacgcggag ttggttcagg gcaccgtctg gatggtcgta      660
ccgggaccgg acacgtgtcg cgcctccaac tacatggaca cgtgtggtgc tgccattggg      720
ccgtacgcgt ggcggtgacc gcaccggatg ctgcctcgca ccgccttgcc cacgctttat      780
atagagaggt tttctctcca ttaatcgcat agcgagtcga atcgaccgaa ggggaggggg      840
agcgaagctt tgcgttctct aatgcctcg tcaaggtaac taatcaatca cctcgtccta      900
atcctcgaat ctctcgtggg gcccgctcaa tctcgcgatt ttgatgctcg tgggtgaaag      960
cgtaggagga tcccgtcgca gttagtctca atctctcagg gtttcgtgcg attttagggg      1020
gatccacctc ttaatcgagt tacggtttcg tgcgatttta gggtaatcct cttaatctct      1080
cattgattha gggtttcgtg agaatcgagg tagggatctg tgttatttat atcgatctaa      1140
tagatggatt ggttttgaga ttgttctgtc agatggggat tgtttcgata tattacccta      1200
atgatgtgtc agatggggat tgtttcgata tattacccta atgatgtgtc agatggggat      1260
tgtttcgata tattacccta atgatggata ataagagtag ttcacagtta tgttttgatc      1320
ctgccacata gtttgagttt tgtgatcaga tttagtttta cttatttctg cttagtccgg      1380
atgggattgt tctgatattg ttccaataga tgaatagctc gttagggtta aatccttagg      1440
ttgagttagg cgacacatag tttatttcct ctggatttgg attggaattg tgttcttagt      1500
ttttttcccc tggatttgga ttggaattgt gtggagctgg gttagagaat tacatctgta      1560
tcgtgtacac ctacttgaac tgtagagctt gggttcctaag gtcaatttaa tctgtattgt      1620
atctggctct ttgcctagtt gaactgtagt gctgatgttg tactgtgttt ttttaccgt      1680

```

```

tttattttgct ttactcgtgc aaatcaaata tgtcagatgc tagaactagg tggcttttatt 740
ctgtgtttcct acatagatct gttgtcctgt agttacttat gtcagttttg ttattatctg 1800
aagatattttt tggttgttgc ttgttgatgt ggtgtgagct gtgagcagcg ctcttatgat 1860
taatgatgct gtccaattgt agtgtagtat gatgtgattg atatgttcat ctattttgag 1920
ctgacagtac cgatatcgta ggatctgggtg ccaacttatt ctccagctgc ttttttttac 1980
ctatgttaat tccaatcctt tcttgctctt tccagataca gata 2024

```

<210> 19

<211> 2024

<212> DNA

<213> *Oryza sativa*

<400> 19

```

ccaccaacc ccatatcgac agaggatgtg aagaacaggt aaatcacgca gaagaaccca 60
tctctgatag cagctatcga ttagaacaac gaatccatat tgggtccgtg ggaaatactt 120
actgcacagg aagggggggcg atctgacgag gccccgccac cggcctcgac ccgaggccga 180
ggccgacgaa gcgcgggcga gtacggcgcc gcggcgccct ctgcccgtgc cctctgcgcg 240
tgggagggag aggcgcgggt ggtggggggcg cgcgcgcgcg cgcgcgcgag tgggtgcggcg 300
gcgcgggggt cagccgcgga gccggcgggcg acggaggagc agggcgggcg ggacgcgaac 360
ttccgatcgg ttggtcagag tgcgcgagtt gggcttagcc aattaggtct caacaatcta 420
ttgggccgta aaattcatgg gccctggttt gtctaggccc aatatcccgt tcatctcagc 480
ccacaaatat ttcccagag gattattaag gcccacacgc agcttatagc agatcaagta 540
cgatgtttcc tgatcgtttg atcggaacg tacggctctg atcaggcatg ccgacttcgt 600
caaagagagg cggcatgacc tgacgcggag ttggttccgg gcaccgtctg gatggtcgta 660
ccgggaccgg acacgtgtcg cgcctccaac tacatggaca cgtgtggtgc tgccattggg 720
ccgtacgcgt ggcggtgacc gcaccggatg ctgcctcgca ccgccttgcc cagctttat 780
atagagaggt tttctctcca ttaatcgcat agcgagtcga atcgaccgaa ggggaggggg 840
agcgaagctt tgcgttctct aatcgctctg tcaaggtaac taatcaatca cctcgtccta 900
atcctogaat ctctcgtggg gcccgctctaa tctcgcgatt ttgatgctcg tgggtggaaag 960
cgtaggagga tcccgtcgca gttagtctca atctctcagg gtttcgtgcg attttagggg 1020
gatccacctc ttaatcgagt tacggtttctg tgcgatttta gggtaatcct cttaatctct 1080
cattgattta gggtttcgtg agaatcgagg tagggatctg tgttatttat atcgatctaa 1140

```

```

tagatgggatt ggttttgaga ttgttctgtc agatggggat tgtttcgata tattacccta 1200
atgatgtgtc agatggggat tgtttcgata tattacccta atgatgtgtc agatggggat 1260
tgtttcgata tattacccta atgatggata ataagagtag ttcacagtta tgttttgatc 1320
ctgccacata gtttgagttt tgtgatcaga tttagtttta cttatttgtg cttagttcgg 1380
atgggattgt tctgatattg ttccaataga tgaatagctc gttagggttaa aatctttagg 1440
ttgagttagg cgacacatag tttatttcct ctggatttgg attggaattg tgttcttagt 1500
ttttttcccc tggatttggg ttggaattgt gtggagctgg gttagagaat tacatctgta 1560
tcgtgtacac ctacttgaac tgtagagctt gggttctaaag gtcaatttaa tctgtattgt 1620
atctggctct ttgctagtt gaactgtagt gctgatgttg tactgtgttt ttttaccgt 1680
tttatttgc tttactgtgc aaatcaaata tgtcagatgc tagaactagg tggctttatt 1740
ctgtgttctt acatagatct gttgtcctgt agttacttat gtcagttttg ttattatctg 1800
aagatatatt tggttgttgc ttgttgatgt ggtgtgagct gtgagcagcg ctcttatgat 1860
taatgatgt gtccaattgt agttagtat gatgtgattg atatgttcat ctattttgag 1920
ctgacagtac cgatatcgta ggatctggtg ccaacttatt ctccagctgc ttttttttac 1980
ctatgttaat tccaatcctt tcttgccctt tccagattca gata 2024

```

<210> 20
 <211> 2024
 <212> DNA
 <213> *Oryza sativa*

```

<400> 20
ccaccaacc ccatatcgac agaggatgtg aagaacaggt aaatcacgca gaagaaccca 60
tctctgatag cagctatcga ttagaacaac gaatccatat tgggtccgtg ggaaatactt 120
actgcacagg aaggggggag atctgacgag gccccgcac cggcctcgac ccgaggcoga 180
ggccgacgaa gcgccggcga gtacggcgcc gcggcgccct ctgccgtgc cctctgcgcg 240
tgaggaggag aggcgcgggt ggtgggggag cgcgcgcgcg cgcgcgcagc tgggtgcggcg 300
gcgcgggggt cagccgccga gccggcgagg acggaggagc agggcggcgt ggacgcgaac 360
ttccgatcgg ttggtcagag tgcgcgagtt gggcttagcc aattaggtct caacaatcta 420
ttgggccgta aaattcatgg gccctggttt gtctaggccc aatatccgt tcatttcagc 480
ccacaaatat ttcccagag gattattaag gccacacgc agcttatagc agatcaagta 540
cgatgtttcc tgatcgttgg atcggaacg tacgggtctg atcaggcatg ccgaactcgt 600

```

```

caaagagagg cggcatgacc tgacgcggag ttgggtccgg gcaccgtctg gatggtcgta      660
ccgggaccgg acacgtgtcg cgcctccaac tacatggaca cgtgtggtgc tgccattggg      720
ccgtacgcgt ggcggtgacc gcaccggatg ctgcctcgca ccgccttgcc cacgccttat      780
atagagaggt tttctctcca ttaatcgcat agcgagtcga atcgaccgaa ggggaggggg      840
agcgaagctt tgcgttctct aatcgctctg tcaaggtaac taatcaatca cctcgtccta      900
atcctcgaat ctctcgtggt gcccgctctaa totcgogatt ttgatgctcg tgggtggaaag      960
cgtaggagga tcccgtgcga gttagtctca atctctcagg gtttcgtgog attttagggg     1020
gatccacctc ttaatcgagt tacgggttctg tgcgatttta gggtaatcct cttaatctct     1080
cattgattta gggtttcgtg agaatcgagg tagggatctg tgttatttat atcgatctaa     1140
tagatgggatt ggttttgaga ttgttctgtc agatggggat tgtttcgata tattacccta     1200
atgatgtgtc agatggggat tgtttcgata tattacccta atgatgtgtc agatggggat     1260
tgtttcgata tattacccta atgatggata ataagagtag ttcacagtta tgttttgatc     1320
ctgccacata gtttgagttt tgtgatcaga tttagtttta cttatttgtg cttagtctcg     1380
atgggattgt tctgatattg ttccaataga tgaatagctc gttagggttaa aatctttagg     1440
ttgagttagg cgacacatag tttatttctc ctggatttgg attggaattg tgttcttagt     1500
ttttttcccc tggatttggg ttggaattgt gtggagctgg gttagagaat tacatctgta     1560
tcgtgtacac ctacttgaac tgtagagctt gggttctaaag gtcaatttaa tctgtattgt     1620
atctggctct ttgcctagtt gaactgtagt gctgatgttg tactgtgttt ttttaccgt     1680
tttatttgc tttactcgtg aaatcaaata tgtcagatgc tagaactagg tggctttatt     1740
ctgtgttctt acatagatct gttgtcctgt agttacttat gtcagttttg ttattatctg     1800
aagatatttt tgggtgttgc ttgttgatgt ggtgtgagct gtgagcagcg ctcttatgat     1860
taatgatgct gtccaattgt agtgtagtat gatgtgattg atatgttcat ctattttgag     1920
ctgacagtac cgatatcgta ggatctgggt ccaacttatt ctccagctgc ttttttttac     1980
ctatgttaat tccaatcctt tottgctct tccagatcca gata                          .2024

```

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

<220>
 <223> Rubi3/GUS fusion

<400> 21
ctcttccaga tgcagataat gcaaattatc

30

<210> 22
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Rubi3/GUS fusion protein

<400> 22

Met Gln Ile Met Leu Arg Pro
1 5

<210> 23
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Rubi3/GUS fusion

<400> 23
ctcttccaga tacagataat gcaaattatc g

31

<210> 24
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Rubi3/GUS fusion protein

<400> 24

Met Leu Arg Pro
1

<210> 25
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Rubi3/GUS fusion

<400> 25
ctcttccaga ttcagataat gcaaattatc g

31

<210> 26
 <211> 31
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Rubi3/GUS fusion

<400> 26
 ctcttccaga tccagataat gcaaataattc g 31

<210> 27
 <211> 1241
 <212> DNA
 <213> Oryza sativa

<400> 27
 ccaccaacc ccatatcgac agaggatgtg aagaacaggt aaatcacgca gaagaaccca 60
 tctctgatag cagctatcga ttagaacaac gaatccatat tgggtccgtg ggaaatactt 120
 actgcacagg aagggggggcg atctgacgag gccccgccac cggcctcgac ccgaggccga 180
 ggccgacgaa gcgccggcgga gtacggcgcc gcggcggcct ctgcccgtgc cctctgcgcg 240
 tgggagggag aggcgcgggt ggtggggggcg cgcgcgcgcg cgcgcgcagc tgggtgcggcg 300
 gcgcgggggt cagccgccga gccggcggcg acggaggagc agggcggcgt ggacgcgaac 360
 ttccgatcgg ttggtcagag tgcgcgagtt gggcttagcc aattaggtct caacaatcta 420
 ttggggccgta aaattcatgg gccctggttt gtctaggccc aatatcccg tcatcttcagc 480
 ccacaaatat ttcccagag gattattaag gccacacgc agcttatagc agatcaagta 540
 cgatgtttcc tgatcgttgg atcgaaaacg tacggtcttg atcaggcatg ccgacttcgt 600
 caaagagagg cggcatgacc tgacgcggag ttggttccgg gcaccgtctg gatggtcgta 660
 ccgggaccgg acacgtgtcg cgcctccaac tacatggaca cgtgtggtgc tgccattggg 720
 ccgtacgcgt ggcggtgacc gcaccggatg ctgcctcgca ccgccttgcc cagcctttat 780
 atagagaggt tttctctcca ttaatcgcat agcgagtcga atcgaccgaa ggggaggggg 840
 agcgaagctt tgcgttctct aatcgctcgc tcaaggtaac taatcaatca cctcgtccta 900
 atcctcgaat ctctcgtggt gcccgctcaa tctcgcgatt ttgatgctcg tgggtggaaag 960
 cgtaggagga tcgatctgtt gtctgtagt tacttatgtc agttttgtta ttatctgaag 1020
 atattttttg ttgttgcttg ttgatgtggt gtgagctgtg agcagcgctc ttatgattaa 1080
 tgatgctgtc caattgtagt gtagtatgat gtgattgata tgttcata ttttgagctg 1140
 acagtaccga tatcgtagga tctggtgcc aattattctc cagctgcttt tttttaccta 1200

tgттааттс аатсттттс тсстстс агатссат а

1241