
Declaration under Rule 4.17:


Published:

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

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

(54) Title: AN EMBRYO PREFERRED PROMOTER AND METHOD OF USING SAME

(57) Abstract: A regulatory region is shown, a nucleotide sequence of approximately 3kb which provides improved seed preferred, and particularly embryo preferred expression in plants. Methods of use are also shown in preferentially expressing a heterologous protein to the embryo tissue of a plant. The sequence is particularly useful in expression of heterologous proteins to the embryo of monocotyledonous plants, particularly cereals, and maize.
AN EMBRYO PREFERRED PROMOTER AND METHOD OF USING SAME

REFERENCE TO RELATED APPLICATIONS

This application claims priority to previously filed and co-pending application USSN 60/570,665, filed May 13, 2004, the contents of which are incorporated in their entirety.

BACKGROUND OF THE INVENTION

Promoters are vital molecular tools that have been applied widely in plant biotechnology to control the expression of introduced genes. There are many applications for promoters in driving gene expression in plant tissues. These include the synthesis of selectable and selectable markers to identify transgenic plants (Jefferson et al., 1987; Wohleben et al., 1988) and the over-expression of control point enzymes to modify metabolic flux through key pathways, so affecting the yields of important plant products (Nessler, 1994; Lessard et al., 2002). Other uses of plant promoters include the expression of genes conferring resistance to pests, thus conferring protection (Estruch et al., 1997), and the expression of non-native enzymes to facilitate the production of foreign metabolites in particular plant species (Poirier et al., 1995; Ye et al., 2000). A further application of plant promoters is to over-express controlling regulatory genes affecting aspects of plant physiology such as flowering time and so modify plant growth characteristics (Weigel and Nilsson, 1995).

Promoters are also used to repress the expression of specific genes by driving the synthesis of interfering RNA species (Waterhouse et al., 2001), thus affecting plant metabolic and developmental pathways (Yu and Kumar, 2003). Although high levels of expression may not be necessary for all of the above applications, there is clearly a need for promoters showing activity in plant tissues.

Apart from these and other applications of promoters to modify plant traits, promoters are also required for plants to act as production systems for heterologous proteins. Plants have been used to produce a wide range of recombinant proteins of potential economic and/or medicinal importance. These include research chemicals (Hood et al., 1997; Zhong et al., 1999), processing enzymes that are used, for example, in the pharmaceutical industry (Woodard et al., 2003), industrial enzymes
that are deployed in large-scale processing operations such as bleaching (Hood et al., 2003; Bailey et al., 2004), candidate vaccine antigens for animal or plant disease prevention (Mason et al., 1992; Haq et al., 1995; Carrillo et al., 1998; Streitfield et al., 2001), and therapeutic pharmaceuticals including antibodies (Daniell et al., 2001; Hood et al., 2002). The expressed proteins may either be purified from the plant tissues (Hood et al., 1997; Woodard et al., 2003) or, if as with vaccines the final application allows it, the recombinant plant material may be processed into a suitable form for use or even deployed directly (Streitfield et al., 2002; Lamphear et al., 2002). For these and other protein products to be produced in plant systems it is necessary that promoters drive a sufficiently high level of expression to ensure commercial viability.

Spatial and temporal control is also often important in driving gene expression in plants. For example selectable and scoreable markers must be expressed at a suitable time and in an appropriate tissue to allow for screening, and controlling enzymes and regulatory factors must be produced in metabolically active and physiologically responsive tissues, respectively. Similarly, genes conferring host protection must be expressed in the target tissues for the pathogen or pest, and plant produced protein products should be expressed in tissues suitable for protein accumulation and storage. Furthermore, since certain protein products may have detrimental affects on plant health and yield when expressed in metabolically active plant tissues that are essential for survival and growth, promoters may be favored that are active in the chosen plant storage tissues but show low or no activity in other, non-storage tissues.

Promoters that preferentially express relatively high levels of foreign proteins in tissues suitable for stable protein accumulation and storage are particularly useful for commercial protein production. The seed tissues of the cereals are especially well suited to the large-scale production of recombinant proteins. Thus, there is a requirement for promoters that show a seed tissue preferred expression pattern in plants and particularly cereals and drive relatively high levels of protein accumulation in these tissues.

Several promoters of plant and plant pathogen (bacterial and viral) origin have been used to direct transgene expression in plants. Prominent examples include the French bean beta-phaseolin promoter (Bustos et al., 1989), the mannopeine synthase promoter of Agrobacterium tumefaciens (Leung et al., 1991), and the 35S promoter of
cauliflower mosaic virus (Guilleuy et al., 1982). These and several other promoters in widespread use in plants were originally developed and utilized in dicot species. Promoter sequences from one species are predictably used in other species (see discussion below). The cereals comprise particularly important crops and there is therefore a pressing need for promoters that have high activity and/or tissue preference in monocots. Cereals, such as grasses, are cultivated for their grain. Since the nutritional value of cereals is in their seeds, and these tissues are also well suited for recombinant protein accumulation and storage, promoters that are active in cereal seed tissues are especially useful.

Two broad classes of promoters are typically deployed: constitutive and tissue preferred. Constitutive promoters, such as maize polyubiquitin-1 drive expression in the seed but also in other tissues (Christensen et al., 1992). A drawback with such constitutive promoters is that expression in tissues other than seed storage tissues may result in plant health being compromised, for example if a potentially toxic protein is expressed in metabolically active tissues required for germination or growth (Hood et al., 2003). Furthermore, constitutive expression may result in the expressed foreign protein being synthesized in pollen grains and thus being difficult to contain. By contrast, seed preferred promoters limit all or the bulk of transgene expression to seed tissues, so avoiding such concerns. Tissue preferred expression can include seed preferred expression. An example of one such promoter providing seed preferred expression is the phaseolin promoter. See, Bustos et al. “Regulation of β-glucuronidase Expression in Transgenic Tobacco Plants by an A/T-Rich cis-Acting Sequence Found Upstream of a French Bean β-Phaseolin Gene” The Plant Cell Vol. 1, 839-853 (1989).

The principle tissue types in maize seeds are the embryo, the endosperm including a surrounding aleurone cell layer and the maternally derived pericarp. Of these, the endosperm and to a lesser extent the embryo, comprise most of the volume of the seed. Thus, endosperm and embryo promoters are particularly important for modifying seed characteristics and contents. The proximal 1.1 kb of a maize 27 kD γ-zein promoter (Russell and Fromm, 1997) and the proximal 1.45 kb of a maize globulin-1 promoter (Belanger and Kriz, 1991; Genbank accession L22344) are prominent examples of seed preferred promoters that have been used to express transgenes in the seeds of monocots.
The endosperm is comprised almost entirely of nutritional reserves, primarily of complex carbohydrate and insoluble protein, but the embryo also contains considerable stores, mainly of oils and soluble proteins. Globulin-1 is one of the most abundant proteins in maize embryo tissue. It is largely limited to this tissue and becomes particularly concentrated in the scutellum late in embryo development. Given the high concentration of this protein observed in embryo tissues a maize globulin-1 promoter was identified as being a good candidate to direct high levels of transgene expression in the embryo. An approximately 1.45 kb extent of a maize globulin-1 promoter/leader has been cloned (Belanger and Kriz, 1991; Genbank accession L22344) and used to drive high levels of transgene expression preferentially in maize seeds (Hood et al., 2003; Woodard et al., 2003). However, still more active promoters are very desirable for some applications, such as the expression of cost sensitive foreign proteins in cereal seeds.

However, despite these examples, there is currently a very limited repertoire of promoters for preferentially expressing foreign proteins in the seed tissues of plants, and in particular, cereals. There is a need for further promoters that express transgenes at similar or higher levels to those currently deployed and with similar or improved tissue specificity. The best promoters would facilitate the expression of foreign proteins in seeds at higher levels than are currently achieved, while restricting expression specifically or predominantly to seed tissues. Also, a range of new promoters would allow the expression of multiple copies of a single transgene in seeds without the need to repeatedly use the same promoter. This should reduce silencing phenomena associated with promoter methylation (De Wilde et al., 2000), and thereby it should also serve to boost expression. Similarly, multiple distinct transgenes could be simultaneously expressed from different promoters in seed tissues, allowing more complex traits and foreign protein products to be reliably introduced into seeds.

All references cited herein are incorporated herein by reference.

SUMMARY OF THE INVENTION

A Zea mays regulatory region has been identified and has preferential expression to the embryo of a plant. It has been found to drive one of the most prevalent messages in developing maize embryos. This invention describes a sequence proximal to a gene with high homology to a maize abscisic acid-inducible gene with preferential transgene expression in plant embryo tissues. In an
embodiment, it is used to drive expression preferentially to embryos in monocotyledonous plants, particularly cereal plants, and most preferentially, in maize.

DESCRIPTION OF THE DRAWINGS

Figure 1A shows the nucleotide sequence (SEQ ID NO: 1) and Figure 1B shows the encoded amino acid sequence (SEQ ID NO: 2) of the β-glucuronidase gene used in experiments.

Figure 2 is a vector map of the reporter construct pPGN9087 (promoter of the maize gene fused to uidA).

Figure 3 shows the nucleotide sequence of the proximal approximately 3 kb of DNA upstream of the translation start codon of the maize gene (SEQ ID NO: 3). The predicted minimal extent of the untranslated leader sequence is given in bold type and the translation start codon is capitalized.

Figure 4 is a graph showing recombinant protein level data derived from single and bulk seed analysis of transgenic maize carrying the promoter -uidA reporter fusion. Data points included to determine mean values (1-7) are as described in the legend to Table 1. Standard deviations of the means are shown. The cut off value for detecting expression is 0.006% of total soluble protein, and non-transgenic control lines never approach this level.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Nucleotide sequences are described herein that regulate transcription with preferential expression to plant seed tissue, and preferential expression to plant embryo tissue in the seed. These novel nucleotide sequences are those natively associated with the nucleotide sequence coding for a Zea mays gene and comprise SEQ ID NO: 3.

A genomics approach is described to identify further sequences that can drive high levels of transgene expression in maize embryo tissues. The cloning of one such sequence identified by this approach is set forth, the proximal approximately 3 kb of a maize promoter identified upstream of an open reading frame of a maize gene. The gene has over 99% DNA sequence similarity to a cDNA of the maize abscisic acid-inducible gene Emb564 (Williams and Tsang, 1991), but the promoter has not been identified to date. Demonstrated here are transgenic plants generated using this sequence that can express the β-glucuronidase (uidA) reporter gene at similar levels to those achieved using the maize globulin-1 promoter (Belanger and Kriz, 1991;...
Genbank accession L22344), which has previously been deployed to express transgenes in maize seeds (Hood et al., 2003; Woodard et al., 2003). Furthermore, the 3 kb promoter sequence cloned here is highly embryo preferred in its expression pattern. Thus, this 3 kb maize promoter sequence is well suited to drive transgene expression in maize and other plant seeds.

The here cloned promoter is particularly useful for the expression of gene sequences in cereal plants and especially in maize plants. However, it can be used in any plant species, including, for example, a monocotyledonous plant such as wheat, rye, rice, oat, barley, turfgrass, sorghum, millet or sugarcane. Alternatively, the plant may be a dicotyledonous plant, for example, tobacco, tomato, potato, soybean, cotton, canola, sunflower or alfalfa. Maize promoters have been used repeatedly to drive expression of genes in non-maize plants, including tobacco (Yang and Russell, 1990; Geffers et al., 2000; Vilardell et al., 1991), cultured rice cells (Vilardell et al., 1991), wheat (Oldach et al., 2001; Brinch-Pedersen et al., 2003), rice (Cornejo et al., 1993; Takimoto et al., 1994), sunflower (Rousell et al., 1988) and protoplasts of carrot (Rousell et al., 1988).

The nucleotide sequences of the invention can be used to isolate corresponding sequences from other organisms, particularly other plants, more particularly other monocots, or to synthesize synthetic sequences. In this manner, methods such as PCR, hybridization, synthetic gene construction and the like can be used to identify or generate such sequences based on their sequence homology to the sequences set forth herein. Sequences isolated or constructed based on their sequence identity to the whole of or any portion of the maize embryo preferred promoter sequences set forth herein are encompassed by the present invention.

In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any plant of interest. Methods for designing PCR primers and PCR cloning are generally known in the art and are disclosed (Sambrook et al., 1989; Innis et al., 1990; Innis et al., 1995; Innis et al., 1999). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like.

In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences
present in a population of cloned genomic DNA fragments or cDNA fragments (i.e.,
genomic or cDNA libraries) from a chosen organism. The hybridization probes may
be genomic DNA fragments, cDNA fragments, RNA fragments, or other
oligonucleotides, and may be labeled with a detectable group such as $^{32}$P, or any other
detectable marker. Thus, for example, probes for hybridization can be made by
labeling synthetic oligonucleotides based on the DNA sequences of the invention.
Methods for preparation of probes for hybridization and for construction of cDNA
and genomic libraries are generally known in the art and are disclosed (Sambrook et
al., 1989).

For example, the embryo preferred promoter sequence disclosed herein, or one
or more portions thereof, may be used as a probe capable of specifically hybridizing
to corresponding sequences. To achieve specific hybridization under a variety of
conditions, such probes include sequences that are unique among the sequences to be
screened and are preferably at least about 10 nucleotides in length, and most
preferably at least about 20 nucleotides in length. Such sequences may alternatively
be used to amplify corresponding sequences from a chosen plant by PCR. This
technique may be used to isolate sequences from a desired plant or as a diagnostic
assay to determine the presence of sequences in a plant. Hybridization techniques
include hybridization screening of DNA libraries plated as either plaques or colonies
(Sambrook et al., 1989).

Hybridization of such sequences may be carried out under stringent
conditions. By "stringent conditions" or "stringent hybridization conditions" is
intended conditions under which a probe will hybridize to its target sequence to a
detectably greater degree than to other sequences (e.g., at least 2-fold over
background). Stringent conditions are sequence-dependent and will be different in
different circumstances. By controlling the stringency of the hybridization and/or
washing conditions, target sequences that are 100% complementary to the probe can
be identified (homologous probing). Alternatively, stringency conditions can be
adjusted to allow some mismatches in sequences so that lower degrees of similarity
are detected (heterologous probing). Generally, a probe is less than about 1000
nucleotides in length, preferably less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is
less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or
other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes
(e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37°C, and a wash in 1X to 2X SSC (20X SSC=3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C.

Specificity is also the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the $T_m$ can be approximated from the equation $T_m = 81.5°C+16.6 \log M + 0.41(%GC) - 0.61(%form.) - 500/L$, where $M$ is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, %form. is the percentage of formamide in the hybridization solution, and $L$ is the length of the hybrid in base pairs (Meinkoth and Wahl, 1984). The $T_m$ is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. $T_m$ is reduced by about 1°C for each 1% of mismatching; thus, $T_m$, hybridization, and/or wash conditions can be adjusted for sequences of the desired identity to hybridize. For example, if sequences with 90% identity are sought, the $T_m$ can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ($T_m$) for the specific sequence and its complement at a defined ionic strength and pH.

However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point ($T_m$); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point ($T_m$); low stringency conditions can utilize a hybridization and/or wash at 11 to 20°C lower than the thermal melting point ($T_m$). Using the equation, hybridization and wash compositions, and desired $T_m$, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a $T_m$ of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the
hybridization of nucleic acids is found in Ausubel et al. (1993) and Sambrook et al. (1989).

Thus, isolated sequences that have promoter activity and which hybridize under stringent conditions to the promoter sequences disclosed herein, or to fragments thereof, are encompassed by the present invention. Such sequences will be at least, 90%, 95% to 98% identical or more. The promoter regions of the invention may be used to isolate substantially identical sequences from any plant species, including but not limited to any plant species described herein.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity" and (d) "percentage of sequence identity"; (a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length promoter sequence, or the complete promoter sequence.

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

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

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations
include, but are not limited to: CLUSTAL in the PC/Gene program (available from IntelliGenetics, Mountain View, CA, USA); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (CGC), 575 Science Drive, Madison, WI, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins and Sharp (1988), Higgins and Sharp (1989), Corpet (1988), Huang et al. (1992) and Pearson (1994). The ALIGN program is based on the algorithm of Myers and Miller (1988). The BLAST programs of Altschul et al. (1990) are based on the algorithm of Karlin and Altschul (1990). To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. (1997). Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules, see Altschul et al. (1997). When utilizing BLAST, Gapped BLAST or PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used, see the World Wide Web site ncbi.nlm.nih.gov. Alignment may also be performed manually by inspection.

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

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window.

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

The invention is further to "functional variants" of the regulatory sequence disclosed. Functional variants include, for example, regulatory sequences of the invention having one or more nucleotide substitutions, deletions or insertions and wherein the variant retains promoter activity, particularly the ability to drive expression preferentially to the embryo of a plant. Functional variants can be created by any of a number of methods available to one skilled in the art, such as by site-directed mutagenesis, induced mutation, identified as allelic variants, cleaving through use of restriction enzymes, or the like. Activity can likewise be measured by any variety of techniques, including measurement of reporter activity as is described at US Patent 6,844,484, Northern blot analysis, or similar techniques. The '484 patent describes the identification of functional variants of different promoters.

The promoter of the invention may be combined with any number of other components to be introduced into the plant, including combined with a gene of interest to be expressed in the plant. The "gene of interest" refers to a nucleotide sequence that encodes for a desired polypeptide or protein but also may refer to nucleotide sequences that do not constitute an entire gene, and which do not necessarily encode a polypeptide or protein. For example, when used in a homologous recombination process, the promoter may be placed in a construct with a sequence that targets an area of the chromosome in the plant but may not encode a protein. If desired, the gene of interest can be optimized for plant translation by optimizing the codons used for plants and the sequence around the translational start site for plants. Sequences resulting in potential mRNA instability can also be avoided.

By "promoter" is meant a regulatory region of DNA capable of regulating the transcription of a sequence linked thereto. It usually comprises a TATA box capable of directing RNA polymerase II to initiate RNA synthesis at the appropriate transcription initiation site for a particular coding sequence. The promoter is the minimal sequence sufficient to direct transcription in a desired manner. The term "regulatory region" is also used to refer to the sequence capable of initiating transcription in a desired manner. By "seed preferred" is intended favored expression
in the seed of the plant, and "embryo preferred" indicates favored expression in the embryo of the seed of the plant.

The promoter of the invention may also be used in conjunction with another promoter. In one embodiment, the plant selection marker and the gene of interest can be both functionally linked to the same promoter. In another embodiment, the plant selection marker and the gene of interest can be functionally linked to different promoters. In yet third and fourth embodiments, the expression vector can contain two or more genes of interest that can be linked to the same promoter or different promoters. For example, the promoter described here can be used to drive the gene of interest and the selectable marker, or a different promoter used for one or the other. These other promoter elements can be those that are constitutive or sufficient to render promoter-dependent gene expression controllable as being cell-type specific, tissue-specific or time or developmental stage specific, or being inducible by external signals or agents. Such elements may be located in the 5' or 3' regions of the gene.

Although the additional promoter may be the endogenous promoter of a structural gene of interest, the promoter can also be a foreign regulatory sequence. Promoter elements employed to control expression of product proteins and the selection gene can be any plant-compatible promoters. These can be plant gene promoters, such as, for example, the ubiquitin promoter (European patent application no. 0 342 926); the promoter for the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO) (Coruzzi et al., 1984; Broglie et al., 1984); or promoters from the tumor-inducing plasmids from Agrobacterium tumefaciens, such as the nopaline synthase, octopine synthase and mannopine synthase promoters (Velten and Schell, 1985) that have plant activity; or viral promoters such as the cauliflower mosaic virus (CaMV) 19S and 35S promoters (Guilley et al., 1982; Odell et al., 1985), the figwort mosaic virus FLt promoter (Maiti et al., 1997) or the coat protein promoter of TMV (Grdzelishvili et al., 2000). Alternatively, plant promoters such as heat shock promoters for example soybean hsp 17.5-E (Gurley et al., 1986); or ethanol-inducible promoters (Caddick et al., 1998) may be used. See International Patent Application No. WO 91/19806 for a review of illustrative plant promoters suitably employed in the present invention.

A promoter can additionally comprise other recognition sequences generally positioned upstream or 5' to the TATA box, referred to as upstream promoter elements, which influence the transcription initiation rate. It is recognized that having
identified the nucleotide sequences for the promoter region disclosed herein, it is
within the state of the art to isolate and identify further regulatory elements in the 5’
region upstream from the particular promoter region identified herein. Thus the
promoter region disclosed herein is generally further defined by comprising upstream
regulatory elements such as those responsible for tissue and temporal expression of
the coding sequence, enhancers and the like. In the same manner, the promoter
elements which enable expression in the desired tissue such as the embryo can be
identified, isolated, and used with other core promoters to confirm embryo-preferred
expression. By core promoter is meant the sequence sometimes referred to as the
TATA box (or similar sequence) which is common to promoters in most genes
encoding proteins. Thus the upstream region of the promoter can optionally be used
in conjunction with its own or core promoters from other sources

In general, the methods available for construction of recombinant genes,
optionally comprising various modifications for improved expression, can differ in
detail. However, conventionally employed methods include PCR amplification, or
the designing and synthesis of overlapping, complementary synthetic
oligonucleotides, which are annealed and ligated together to yield a gene with
convenient restriction sites for cloning, or subcloning from another already cloned
source, or cloning from a library. The methods involved are standard methods for a
molecular biologist (Sambrook et al., 1989).

One skilled in the art readily appreciates that the promoter can be used with
any of a variety of nucleotide sequences comprising the gene of interest to be
expressed in plants. For example, the gene of interest may encode a protein that is
useful for industrial or pharmaceutical purposes or the like, or to impact the plant
itself, such as through expression of a protein that provides disease resistance, insect
resistance, herbicide resistance, or impacts agronomic traits as well as grain quality
traits. The sequences used with the promoter can be native or non-native sequences to
the plant. DNA sequences native to plants as well as non-native DNA sequences can
be transformed into plants and used to modulate levels of native or non-native
proteins.

The gene of interest can also be a nucleotide sequence used to target an area of
the plant genome through homologous recombination. The promoter may be placed
in a construct with such sequence, which sequence will not necessarily encode a
protein. The sequence recombines in the genome and the promoter may be placed at
the desired site targeted by the sequences to regulate the desired endogenous nucleotide sequence.

Further, the promoter can be used to drive mRNA that can be used for a silencing system, such as antisense, and in that instance, no protein is produced. Means of increasing or inhibiting a protein are well known to one skilled in the art and, by way of example, may include, transgenic expression, antisense suppression, co-suppression methods including but not limited to: RNA interference, gene activation or suppression using transcription factors and/or repressors, mutagenesis including transposon tagging, directed and site-specific mutagenesis, chromosome engineering and, homologous recombination. In the case of use with homologous recombination, no in vivo construct will be required.

Once the gene is engineered to contain desired features, such as the desired subcellular localization sequences, it may then be placed into an expression vector by standard methods. The selection of an appropriate expression vector will depend upon the method of introducing the expression vector into host cells. A typical expression vector contains prokaryotic DNA elements coding for a bacterial origin of replication and an antibiotic resistance gene to provide for the growth and selection of the expression vector in the bacterial host; a cloning site for insertion of an exogenous DNA sequence; eukaryotic DNA elements that control initiation of transcription of the exogenous gene (such as the promoter of the invention or another promoter); and DNA elements that control the processing of transcripts, such as transcription termination/polyadenylation sequences. It also can contain such sequences as are needed for the eventual integration of the vector into the plant chromosome.

Clearly, many variations in use of the promoter of the invention are available to one skilled in the art.

In one embodiment, the expression vector also contains a gene encoding a selectable or scoreable marker that is operably or functionally linked to a promoter that controls transcription initiation, which can be the promoter of the invention or another promoter. By “operably linked” it is understood that the gene of interest (in this case the gene encoding a selectable or scoreable marker) is oriented in connection to the gene such that the promoter initiates transcription of the gene in order to allow its expression of the resulting protein in plants. For a general description of plant expression vectors and reporter genes, see Gruber et al. (1993). In one embodiment, the selective gene is a glufosinate-resistance encoding DNA and in another
embodiment it can be phosphinotricin acetyl transferase (*pat*) or a maize optimized *pat* gene under the control of the CaMV 3S promoter. Such *pat* genes confer resistance to the herbicide bialaphos (Gordon-Kamm *et al.*, 1990).

The expression vector can optionally also contain a signal sequence located between the promoter and the gene of interest. A signal sequence is a nucleotide sequence, translated to give an amino acid sequence, which is used by a cell to direct the protein or polypeptide of interest to be placed in a particular place within or outside the eukaryotic cell. One example of a plant signal sequence is the barley α-amylase secretion signal (Rogers, 1985). Many signal sequences are known in the art. See, for example Becker *et al.* (1992), Fontes *et al.* (1991), Matsuoka and Nakamura (1991), Gould *et al.* (1989), Creissen *et al.* (1992), Kalderon *et al.* (1984) and Stiefel *et al.* (1990).

Leader sequences can be included to enhance translation. Instead of, or in addition to the untranslated leader sequence of the promoter described here, other leader sequences may be substituted or added. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein *et al.* (1989); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Gallie *et al.* (1995)); human immunoglobulin heavy-chain binding protein (BiP) (Macejak *et al.* (1991)); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling *et al.* (1987)); tobacco mosaic virus leader (TMV) (Gallie, (1989)); and maize chlorotic mottle virus leader (MCMV) (Lommel *et al.* (1991)). See also, Della-Cioppa *et al.* (1987). Other methods known to enhance translation can also be utilized, for example, introns, and the like. Obviously, many variations on the promoters, selectable markers, signal sequences, leader sequences, termination sequences, introns, enhancers and other components of the construct are available to one skilled in the art.

Methods for introducing expression vectors into plant tissue available to one skilled in the art are varied and will depend on the plant selected. Procedures for transforming a wide variety of plant species are well known and described throughout the literature. See, for example, Miki and McHugh (2004); Klein *et al.* (1992); and Weising *et al.* (1988). For example, the DNA construct may be introduced into the genomic DNA of the plant cell using techniques such as microprojectile-mediated delivery (Klein *et al.* 1992), electroporation (Fromm *et al.*, 1985), polyethylene glycol
(PEG) precipitation (Mathur and Koncz, 1998), direct gene transfer (WO 85/01856 and EP -A- 275 069), in vitro protoplast transformation (U.S. Patent No. 4,684,611) and microinjection of plant cell protoplasts or embryogenic callus (Crossway, 1985). Co-cultivation of plant tissue with Agrobacterium tumefaciens is another option, where the DNA constructs are placed into a binary vector system (Ishida et al., 1996). The virulence functions of the Agrobacterium tumefaciens host will direct the insertion of the construct into the plant cell DNA when the cell is infected by the bacteria. See, for example, Fraley et al. (1983).

Standard methods for transformation of canola are described by Moloney et al. (1989). Corn transformation is described by Fromm et al. (1990) and Gordon-Kamm et al. (1990). Agrobacterium is primarily used in dicots, but certain monocots such as maize can be transformed by Agrobacterium. See, for example, U.S. Patent No. 5,550,318. Rice transformation is described by Hiei et al. (1994) and Lee et al. (1991). Wheat can be transformed by techniques similar to those used for transforming corn or rice. Sorghum transformation is described by Casas et al. (1993) and barley transformation is described by Wan and Lemaux (1994). Soybean transformation is described in a number of publications, including U.S. Patent No. 5,015,580.

In one preferred method, the Agrobacterium transformation methods of Ishida et al. (1996) and also described in U.S. Patent 5,591,616, are generally followed, with modifications that the inventors have found improve the number of transformants obtained. The Ishida method uses the A188 variety of maize that produces Type I callus in culture. In one preferred embodiment the Hi II maize line is used which initiates Type II embryogenic callus in culture (Armstrong et al., 1991).

While Ishida recommends selection on phosphinothricin when using the bar or pat gene for selection, another preferred embodiment provides use of bialaphos instead. In general, as set forth in the 5,591,616 patent, and as outlined in more detail below, dedifferentiation is obtained by culturing an explant of the plant on a dedifferentiation-inducing medium for not less than seven days, and the tissue during or after dedifferentiation is contacted with Agrobacterium having the gene of interest. The cultured tissue can be callus, an adventitious embryo-like tissue or suspension cells, for example. In this preferred embodiment, the suspension of Agrobacterium has a cell population of $10^6$ to $10^{11}$ cells/ml and are contacted for three to ten minutes with the tissue, or continuously cultured with Agrobacterium for not less than seven
days. The *Agrobacterium* can contain plasmid pTOK162, with the gene of interest between border sequences of the T region of the plasmid, or the gene of interest may be present in another plasmid-containing *Agrobacterium*. The virulence region may originate from the virulence region of a Ti plasmid or Ri plasmid. The bacterial strain used in the Ishida protocol is LBA4404 with the 40 kb super binary plasmid containing three *vir* loci from the hypervirulent A281 strain. The plasmid has resistance to tetracycline. The cloning vector cointegrates with the super binary plasmid. Since the cloning vector has an *E. coli* specific replication origin, but not an *Agrobacterium* replication origin, it cannot survive in *Agrobacterium* without cointegrating with the super binary plasmid. Since the LBA4404 strain is not highly virulent, and has limited application without the super binary plasmid, the inventors have found in yet another embodiment that the EHA101 strain is preferred. It is a disarmed helper strain derived from the hypervirulent A281 strain. The cointegrated super binary/cloning vector from the LBA4404 parent is isolated and electroporated into EHA101, selecting for spectinomycin resistance. The plasmid is isolated to assure that the EHA101 contains the plasmid. EHA101 contains a disarmed pTi that carries resistance to kanamycin. See, Hood *et al.* (1986).

Further, the Ishida protocol as described provides for growing fresh culture of the *Agrobacterium* on plates, scraping the bacteria from the plates, and resuspending in the co-culture medium as stated in the 5,591,616 patent for incubation with the maize embryos. This medium includes 4.3 g MS salts, 0.5 mg nicotinic acid, 0.5 mg pyridoxine hydrochloride, 1.0 ml thiamine hydrochloride, casamino acids, 1.5 mg 2,4-D, 68.5 g sucrose and 36 g glucose per liter, all at a pH of 5.8. In a further preferred method, the bacteria are grown overnight in a 1 ml culture and then a fresh 10 ml culture is re-inoculated the next day when transformation is to occur. The bacteria grow into log phase, and are harvested at a density of no more than OD<sub>600</sub>=0.5, preferably between 0.2 and 0.5. The bacteria are then centrifuged to remove the media and resuspended in the co-culture medium. Since Hi II is used, medium preferred for Hi II is used. This medium is described in considerable detail by Armstrong and Green (1985). The resuspension medium is the same as that described above. All further Hi II media are as described in Armstrong and Green (1985). The result is redifferentiation of the plant cells and regeneration into a plant. Redifferentiation is sometimes referred to as dedifferentiation, but the former term more accurately describes the process where the cell begins with a form and identity, is placed on a
medium in which it loses that identity, and becomes "reprogrammed" to have a new identity. Thus the scutellum cells become embryogenic callus.

In accordance with the present invention, a transgenic plant is produced that contains an introduced promoter. It can be combined with any one of the components set forth above. In a preferred embodiment, the promoter is driving expression of a nucleotide sequence such that the sequence encodes a protein preferentially expressed in the seed of the plant. Preferably, the plant is a cereal plant, and most preferably, a maize plant.

In a further embodiment, plant breeding can be used to introduce the nucleotide sequences into other plants once transformation has occurred. This can be accomplished by any means known in the art for breeding plants such as, for example, cross pollination of the transgenic plants that are described above with other plants, and selection for plants from subsequent generations which express the amino acid sequence. The plant breeding methods used herein are well known to one skilled in the art. For a discussion of plant breeding techniques, see Poehlman and Sleper (1995). Many crop plants useful in this method are bred through techniques that take advantage of the plant's method of pollination. A plant is self-pollinating if pollen from one flower is transferred to the same or another flower of the same plant. A plant is cross-pollinating if the pollen comes from a flower on a different plant. For example, in *Brassica*, the plant is normally self-sterile and can only be cross-pollinated unless, through discovery of a mutant or through genetic intervention, self-compatibility is obtained. In self-pollinating species, such as rice, oats, wheat, barley, peas, beans, soybeans, tobacco and cotton, the male and female plants are anatomically juxtaposed. During natural pollination, the male reproductive organs of a given flower pollinate the female reproductive organs of the same flower. Maize plants (*Zea mays L.*) can be bred by both self-pollination and cross-pollination techniques. Maize has male flowers, located on the tassel, and female flowers, located on the ear, on the same plant. It can self or cross-pollinate.

Pollination can be by any means, including but not limited to hand, wind or insect pollination, or mechanical contact between the male fertile and male sterile plant. For production of hybrid seeds on a commercial scale in most plant species pollination by wind or by insects is preferred. Stricter control of the pollination process can be achieved by using a variety of methods to make one plant pool male sterile, and the other the male fertile pollen donor. This can be accomplished by hand
detassling, cytoplasmic male sterility, or control of male sterility through a variety of methods well known to the skilled breeder. Examples of more sophisticated male sterility systems include those described by Brar et al., U.S. Patent Nos. 4,654,465 and 4,727,219 and Albertsen et al., U.S. Patent Nos. 5,859,341 and 6,013,859.

Backcrossing methods may be used to introduce the gene into the plants. This technique has been used for decades to introduce traits into a plant. An example of a description of this and other plant breeding methodologies that are well known can be found in references such as Neal (1988). In a typical backcross protocol, the original variety of interest (recurrent parent) is crossed to a second variety (nonrecurrent parent) that carries the single gene of interest to be transferred. The resulting progeny from this cross are then crossed again to the recurrent parent and the process is repeated until a plant is obtained wherein essentially all of the desired morphological and physiological characteristics of the recurrent parent are recovered in the converted plant, in addition to the single transferred gene from the nonrecurrent parent.

EXAMPLES

The following is presented as illustrative of an embodiment of the invention and does not limit the scope of the invention as otherwise set forth.

Materials and Methods

Construction of cDNA libraries representative of maize embryo tissues

Maize plants were grown from seed in moist soil under standard greenhouse conditions. Four lines of maize were grown, representative Lancaster, Stiff Stalk, high protein and high oil lines. Elite inbreds are commonly derived from germplasm pools known as Stiff Stalk and Lancaster. Stiff Stalk inbreds have been known for decades and are reported by the USDA to have been widely available for decades. They are derived from the Iowa Stiff Stalk synthetic population (Sprague, 1946). For example see PI accession no. 550481 and discussions of Stiff Stalk germplasm at U.S. Patents Nos. 5,706,603; 6,252,148; 6,245,975; 6,344,599 and 5,134,074. See also, Neuhauzen (1989). Lancaster inbreds are derived from the open pollinated variety Lancaster Surecrop (Anderson, 1944). See for example, PI 280061. High oil or high protein plants are those in which the oil or protein content of the seed is higher than lower oil or protein producing plants such as hybrid #2 yellow dent corn.

Plants were self-pollinated and individual plants were sacrificed at 10, 11, 12, 19, 28, 37 and 46 days post-pollination. Embryos were immediately harvested from these plants, frozen in liquid nitrogen and stored at -80°C. Embryos harvested from
distinct lines and at different time points were kept separate, except that embryos of
the same line harvested at 10, 11 and 12 days post-pollination were pooled. For each
of the five resulting time points (10 to 12 days, 19 days, 28 days, 37 days and 46 days
post-pollination) equal amounts of embryo tissues harvested from each of the four
maize lines were pooled. Total RNA was isolated from the pooled embryo tissues
using a phenol-based method (Chatterjee et al., 1996), and poly-A message was then
prepared from this RNA using Poly(A) Quik mRNA isolation columns (Stratagene;
La Jolla, CA). These poly-A RNA samples were used to prepare five cDNA libraries,
each representative of all four maize lines and each corresponding to a different time
point of embryo development. The libraries were constructed in the Lambda ZAP II
vector (Stratagene; La Jolla, CA).

DNA sequence analysis of representative clones from maize embryo libraries

For each of the five libraries, phagemids were excised from the phage vector.
Approximately 100 clones were randomly selected to represent each library and the
nucleotide sequences of the cDNA inserts were determined using the chain
termination approach using attached dyes by the ‘DNA Sequencing and Synthesis
Facility’ of Iowa State University (Ames, IA). Nucleotide sequences of clones were
compared using the ‘Sequencher’ package (Gene Codes Corporation; Ann Arbor,
MI).

Analysis of clone representation in embryo libraries by plaque hybridization

Equal aliquots of each of the five embryo developmental time point cDNA
libraries were pooled, and the pooled phage infected onto the bacterial strain XL1-
Blue MRF’ (Stratagene; La Jolla, CA) to generate approximately 40,000 plaques upon
plating. Phage DNA was transferred onto charged nylon filters (Amersham;
Piscataway, NJ) and cross-linked to the filters by exposure to ultraviolet light.
Radio-nucleotide ($^{32}$P) was incorporated into cDNA sequence of the maize gene by
random prime labeling (Feinberg and Vogelstein, 1983) using the ‘High Prime’
reagent mix (Roche Diagnostics GmbH; Mannheim, Germany) and the filters were
incubated with this probe. Filters were washed under conditions of high stringency
(15 mM NaCl, 1.5 mM sodium citrate [$\text{Na}_3\text{C}_6\text{H}_5\text{O}_7\cdot2\text{H}_2\text{O}$], 0.1% sodium dodecyl
sulfate, 65°C) and exposed to BioMax MS film (Kodak; Rochester, NY), to reveal
clones homologous to the maize abscisic acid-inducible gene cDNA.
Analysis of genome organization by DNA hybridization

DNA was prepared from maize leaves using a hexadecyltrimethyl-ammonium bromide based method (Stacey and Issac, 1994). DNA (15 μg samples) was digested with the restriction endonucleases EcoRI or HindIII and DNA fragments were size separated on 0.7% agarose gels. Vector DNA was similarly digested and 60 pg was size separated on the gels. The DNA was transferred onto charged nylon filters (Amersham; Piscataway, NJ) and cross-linked to the filters by exposure to ultraviolet light. Radionucleotide (\(^{32}\)P) was incorporated into cDNA sequence of the maize gene by random prime labeling using the ‘High Prime’ reagent mix (Roche Diagnostics GmbH; Mannheim, Germany) and the filters were incubated with this probe. Filters were washed under conditions of high stringency (15 mM NaCl, 1.5 mM sodium citrate [Na\(_3\)C\(_6\)H\(_5\)O\(_7\)·2H\(_2\)O], 0.1% sodium dodecyl sulfate, 65°C) and exposed to BioMax MS film (Kodak; Rochester, NY).

Analysis of message levels by RNA hybridization

Total RNA was isolated from maize tissues using a phenol-based method (Chatterjee et al., 1996). RNA (20 μg samples) was size separated on agarose / formaldehyde gels, transferred onto charged nylon filters (Amersham; Piscataway, NJ) and cross-linked to the filters by exposure to ultraviolet light. Radionucleotide labeled DNA probes were prepared by random prime labeling using the ‘High Prime’ reagent mix (Roche Diagnostics GmbH; Mannheim, Germany) and the filters were incubated with the maize cDNA sequence or 18S rRNA gene sequence. Filters were washed under conditions of high stringency (15 mM NaCl, 1.5 mM Na\(_3\)C\(_6\)H\(_5\)O\(_7\)·2H\(_2\)O, 0.1% sodium dodecyl sulfate, 65°C) and exposed to BioMax MS film (Kodak; Rochester, NY). DNA probes were stripped from filters by washing with near-boiling 0.1% sodium dodecyl sulfate.

Cloning of and nucleotide sequence determination of the promoter of a maize abscisic acid-inducible gene

DNA sequences upstream of the maize open reading frame were isolated from a maize Missouri-13 line genomic library in the Lambda FIX II vector (Stratagene; La Jolla, CA). The phage library was infected onto the bacterial strain XL1-Blue MRA (Stratagene; La Jolla, CA) and plated to generate plaques. Phage DNA was transferred onto charged nylon filters (Amersham; Piscataway, NJ) and cross-linked to the filters by exposure to ultraviolet light. Radionucleotide (\(^{32}\)P) was incorporated into cDNA.
sequence of the maize gene by random prime labeling using the 'High Prime' reagent mix (Roche Diagnostics GmbH; Mannheim, Germany) and the filters were incubated with this probe. Filters were washed under conditions of high stringency (15 mM NaCl, 1.5 mM sodium citrate [Na₂C₆H₃O₇·2H₂O], 0.1% sodium dodecyl sulfate, 65°C) and exposed to BioMax MS film (Kodak; Rochester, NY) to reveal sequences homologous to the maize cDNA. Homologous clones were recovered and the phage inserts mapped by comparing restriction endonuclease digests of the clones following size fractionation via agarose gel electrophoresis. The nucleotide sequence of DNA identified as extending approximately 3 kb 5' of the open reading frame of the maize gene was determined by the 'DNA Sequencing Facility' of Iowa State University (Ames, IA).

Construction of promoter-reporter gene fusions and introduction into plants

The here cloned untranslated leader sequence of the maize gene, plus proximal promoter sequence, together corresponding to approximately 3 kb of sequence 5' to the open reading frame, was fused to the β-glucoronidase (uidA) reporter gene of Escherichia coli (Jefferson et al., 1987). Note that while any version of the uidA gene would be workable in the invention, in this particular instance a version with a six histidine residue fusion to the C-terminus was used (SEQ ID NO: 4). (See Figure 1A showing nucleotide sequence used (SEQ ID NO: 1) and Figure 1B showing the corresponding amino acid sequence (SEQ ID NO: 2).) This tag allows for easy isolation from plant tissues using a nickel column, should purification be desired. To ensure appropriate message termination, the potato proteinase inhibitor II (PinII) transcription terminator region was added 3' of the reporter gene (An et al., 1989). This fusion was included on a vector that also carried the phosphinothricin N-acetyltransferase gene (pat) of Streptomyces viridochromogenes to confer herbicide resistance to transgenic plants. This gene confers resistance to bialaphos (Gordon-Kamm et al., 1990). The expression of the pat marker was controlled by the cauliflower mosaic virus 35S promoter and terminator sequences (Guilley et al., 1982; Odell et al., 1985). In addition, the vector contained border sequences flanking the transcription units. These borders allowed the transformation of vector DNA enclosed within them into the target plant's genome. The vector is shown in Figure 2.

The procedure for stable transformation was modified from that of Ishida et al. (1996) as described supra. Immature zygotic embryos from kernels of a Hi-II / elite
line were transformed with *A. tumefaciens* strain EHA101 containing the relevant maize upstream sequence / reporter fusion to generate transgenic events. T0 plants were regenerated from tissue culture of each event, transferred to soil in a greenhouse and pollinated using pollen from an elite inbred line to produce T1 seeds.

5 **Analysis of uidA reporter gene expression in transiently transformed embryos**

Transiently transformed embryos were stained for 24 hours with 0.5 mg ml\(^{-1}\) 5-bromo-4-chloro-3-indolyl-\(\beta\)-D-glucuronic acid: cyclohexylammonium salt, or X-gluc, (Inalco; Milan, Italy) and were subsequently transferred to 70% ethanol. Blue staining indicated the presence of GUS activity.

10 **Quantification of uidA reporter gene expression in seed tissues**

Six dry seeds from each ear were individually pulverized and extracted with 1 ml of lysis buffer (50 mM sodium phosphate pH 7.0, 1 mM EDTA, 10 mM \(\beta\)-mercaptoethanol). Furthermore, fifty seed pools from each ear were homogenized in a blender and three approximately 100 mg aliquots were extracted with the above lysis buffer. Single and pooled seed samples were placed in extraction tubes held in a rack, with a ball bearing added to each tube, and were then homogenized in a high-speed shaker for 20 seconds. Samples were centrifuged, and the supernatants recovered and stored on ice prior to analysis. Assays were performed in triplicate to determine GUS activity resulting from expression of the *uidA* reporter gene (Jefferson et al., 1987).

15 Total soluble protein (1 \(\mu\)g) was incubated in 100 \(\mu\)l of lysis buffer and the reaction was initiated with 25 \(\mu\)l of 5 mM 4-methylumbelliferyl \(\beta\)-D-glucuronide (Sigma; St. Louis, MO). The reaction was incubated for up to 20 min at 37°C. At specific time points 25 \(\mu\)l volumes of the reaction mixture were transferred to PolySorp 96-well plates (Nalge Nunc International; Rochester, NY) that had 175 \(\mu\)l of stop buffer (0.2M Na\(_2\)CO\(_3\)) per well. Fluorescence was measured at an excitation wavelength of 360 nm and an emission wavelength of 460 nm on a Microplate Fluorometer (Molecular Devices; Sunnyvale, CA). GUS protein levels were then calculated by comparison to a standard curve of 1 to 100 \(\mu\)M 4-methylumbelliferone (Sigma; St. Louis, MO). Protein concentrations were determined in duplicate using a dye-binding assay (Bradford, 1976).

20 **Statistical analysis of uidA reporter gene expression in transgenic seeds**

Following the six individual seed analysis the mean value for all seed expressing above a background cut off level was determined for each plant and
separately for the construct. Next, from the mean values for each plant, mean
expression levels were determined among all plants derived from a particular
independent transformation event and also from all plants derived from the construct.
If all seed from a particular plant expressed below the background cut off level, then
that plant was scored as zero and was included as such in the analysis. Finally, from
the mean values for each transformation event, mean levels were determined among
all events derived from the construct. If all plants from a particular event had been
scored as zero, then that event was scored as zero and was included as such in the
analysis.

The single seed data was also analyzed focusing on the highest individual seed
for each plant. From the highest individual seed values for each plant, mean of high
seed expression levels were determined among all plants derived from a particular
independent transformation event and also from all plants derived from the construct.
If the highest expressing seed from a particular plant expressed below the background
cut off level for the assay, then that plant was scored as zero and was included as such
in the analysis. Finally, from the mean of high seed values for each transformation
event, mean levels were determined among all events derived from the construct. If
all plants from a particular event had been scored as zero, then that event was scored
as zero and was included as such in the analysis.

The data for the 50 seed pools was similarly analyzed to give mean expression
levels for the construct derived from either expression levels determined for each
plant's pooled seed or from mean expression levels for each event, which themselves
were derived from expression levels for each plant's pooled seed. Note that for the 50
seed pool analysis pooled seed was not assayed from plants that had given no positive
seed by the six individual seed analysis. Rather, pools for these plants were assigned
an expression value of zero. These artificial zeros, together with any negative
expression data obtained by assaying pools were included in the analysis. Also, note
that due to lack of available seed some plants were not analyzed at the bulk seed level
even though they had some positive individual seed, and these were excluded from
any statistics on bulk seed analysis.

The highest recorded expression level for an individual seed observed with
regenerated plants that carried the construct was also noted. This gives an indication
of expression potential using the promoter sequence.
Analysis of *uidA* reporter gene expression in transgenic plant tissues

*T*₁ seeds were sectioned using a scalpel and were incubated with Jefferson’s buffer containing 0.5 mgml⁻¹ X-gluc (Jefferson *et al.*, 1987) for up to 3 hours at 37°C until a clear blue stain was visible. In addition, *T*₁ seeds were allowed to germinate and the resulting *T*₁ seedlings were screened for the presence of *pat*, and hence for the linked *uidA* reporter gene, by treating an area of leaf tissue with a 1% glufosinate solution and scoring for resistance to the herbicide. Resistant *T*₁ plants (hemizygous for *uidA*) were self-pollinated. Representative tissue samples were collected from selected non-seed tissues and were incubated overnight at 37°C with Jefferson’s buffer containing 0.5 mgml⁻¹ X-gluc (Jefferson *et al.*, 1987). Blue staining indicated GUS activity. Furthermore, developing *T*₂ seeds were harvested at defined time points and were similarly treated to reveal GUS activity, with sufficient incubation times to reveal any clear staining.

Results

Identification of a maize gene as being highly expressed in the developing embryo

The approach taken to identify promoters capable of driving foreign gene expression in maize embryo tissues was to examine relative levels of expression of native maize embryo genes. This was achieved by analyzing clone representation in cDNA libraries prepared from embryo tissues. To enable clones to be identified from various stages of seed development, libraries were prepared from embryo tissues harvested at five time points post-pollination. The selected time points were between 10 and 12 days post-pollination, and at 19, 28, 37 and 46 days post-pollination, the last time point corresponding to fully mature seed. Furthermore, in order to identify clones that would be of value in different corn germplasms, each of the above five embryo pools was made up equally of embryos isolated from each of four lines of maize, comprising a Lancaster line, a Stiff Stalk line, a high protein line and a high oil line.

For each of the five embryo cDNA libraries the DNA sequence of approximately one hundred randomly selected clones was determined. The approximately five hundred cDNA sequences that were so generated were analyzed to reveal the gene expression profile of developing maize embryos and to identify the most highly represented sequences. These sequences were considered to correspond to the most abundant clones or families of clones in the libraries and therefore to the most highly expressed genes or families of genes. Using this approach, a maize gene

---

25
with high homology to an abscisic-acid inducible gene (Williams and Tsang, 1991) was identified as being one of the most highly expressed sequences, with a total of twelve hits out of 530 cloned sequences, with these twelve sequences being over 99% similar to each other. This indicates that approximately 2.3% of mRNA molecules present in developing maize embryo tissues encode the gene product. However, the representation of this message varies throughout embryo development. No such sequences were identified among approximately one hundred randomly selected clones from the 10 to 12-day post-pollination cDNA library or among approximately one hundred randomly selected clones from the 19-day post-pollination cDNA library. By contrast, seven, two and three such sequences were identified among similar numbers of clones selected from the 28, 37 and 46-day post-pollination cDNA libraries, respectively. Thus, expression of this gene appears to increase later during embryo development, peaking at about 28 days post-pollination.

**Confirmation of the gene as being highly expressed in the developing maize embryo**

The gene was then confirmed as being highly expressed with a greater level of confidence. A region of the gene was screened for hybridization against a random plating of approximately 40,000 plaques of an equally represented combination of the five embryo cDNA libraries. Thus, a representative pool of plaques corresponding to all five time points throughout embryo development and all four lines of maize was assessed. Sequence of a strongly expressed gene should identify a relatively high proportion of plaques, comparable to its representation in the cDNA libraries. Since tens of thousands of plaques were screened there is a greater confidence that the result is representative of all sequences, compared to results obtained using the more restricted DNA sequencing approach described above to initially identify highly expressed clones. This plaque hybridization approach identified approximately 0.6% of the cDNA clones as corresponding to the gene.

However, a concern with the plaque hybridization approach is that cross hybridization of the selected clone with related but non-identical sequences may result in an overestimation of a particular clone’s representation in the libraries. To determine whether this is a serious limitation in the case of this gene, the approximate copy number of the gene sequence plus closely related sequences in the maize genome was determined. DNA hybridization analysis using cDNA sequence of the gene as a probe and genomic DNA prepared from leaf tissue of a standard maize laboratory line as the template identified only four or five annealing DNA fragments,
depending on the restriction enzyme used to digest the genomic DNA template. This is consistent with one or at most a few sequences being present in the maize genome corresponding to or highly similar to the gene. Thus, the estimation of clone representation for the gene using plaque hybridization data should not be greatly distorted by gene copy number considerations, particularly since some sequences identified by the copy number determination approach may represent pseudogenes that produce no transcripts.

In the seed the message of the gene is located in developing embryo tissues

The tissue and line specificity of expression for the gene was then assessed at the messenger RNA level by conducting a hybridization analysis using the gene cDNA sequence as a probe and RNA prepared from various tissues as the templates. For non-seed material the tissues providing the RNA were pooled samples collected from the four maize lines originally used to make the cDNA libraries. Expression was assessed in leaf, stem, root, tassel, anther, pollen, husk, silk, immature ear and cob tissues. However, in the case of seed tissues expression was assessed in 28-day post-pollination embryos isolated separately from each of the four maize lines used to make the cDNA libraries and in 28-day post-pollination embryos and endosperm tissues isolated from a standard maize laboratory line.

RNA hybridizing to the gene cDNA sequence was detected in 28-day post-pollination embryo tissue of all four maize lines used to make the cDNA libraries and of the standard laboratory line. By contrast, this message was not detected in endosperm tissue of the standard laboratory line, indicating that within the seed the gene is much more highly expressed in the embryo than the endosperm. No gene message was detected in leaf, stem, root, tassel, anther, pollen, husk, silk, immature ear or cob tissues pooled from the four lines used to make the cDNA libraries, indicating the strong seed preference in expression of this gene.

Novel sequences are located within the approximately 3 kb of sequence 5’ and proximal to the open reading frame of the gene

Since expression of the gene was only detectable in embryo tissue, and it was identified as being highly expressed by the library sampling approach deployed here, an extensive genomic clone spanning approximately 8.7 kb of proximal promoter sequence of the gene, but also including the untranslated leader together with approximately 0.3 kb downstream of the translation start codon, was isolated. These sequences were cloned from a library of genomic sequences prepared from leaf tissue
of a standard maize laboratory line, using the gene cDNA sequence as a probe. Plaques were thus identified in the genomic library as carrying homologous sequences to the cDNA. Genomic DNA extending approximately 3 kb upstream of the translation start codon for the gene was sub-cloned and the nucleotide sequence determined (Figure 3, SEQ ID NO: 3). By comparison with the here isolated gene cDNAs, at least the 101 nucleotides of genomic sequence proximal to the translation start codon must correspond to untranslated leader sequence. Also, comparison of these genomic and cDNA sequences implies that sequence between 119 and 257 nucleotides downstream of the 5' end of the open reading frame correspond to an intron. Discounting this intron, in the region that they overlap, the genomic and consensus of cDNA sequences show over 99% similarity.

Comparison of the genomic sequence cloned here with the previously cloned and reported cDNA sequence of the maize abscisic acid-inducible gene (Williams and Tsang, 1991; Genbank accession X55388) extends the predicted minimal extent of the leader by a further five nucleotides to a total of 106 nucleotides proximal to the translation start codon. The predicted minimal extent of the leader is further extended by 133 nucleotides, to a total of 239 nucleotides, by comparing the genomic sequence cloned here with the previously reported cDNA sequence AY104117 (Genbank accession) that also has over 99% similarity to the here cloned cDNAs.

Beyond the predicted leader sequence, more distal to the open reading frame, the here cloned genomic sequence shows only one region of extensive similarity to sequences in the Genbank or EMBL databases. This region lies between approximately 1220 and 1960 nucleotides upstream of the predicted translation start codon and has 87% DNA sequence similarity with the maize transposable element ILS-1 (Genbank accession ZMU07956). This same region of the here cloned genomic sequence shows a similar level of identity to sequence of the ADP-glucose pyrophosphorylase large subunit locus reported for five maize cultivars (Genbank accessions AF544157, AF544158, AF544159, AF544160 and AF544161), indicating that this reported locus also contains ILS-1 transposon-like sequence.

Promoter sequences of the abscisic acid-inducible gene can drive transgene expression in transiently transformed embryos

To assess the activity and specificity of the promoter sequence of the gene a transcription unit was made in which 2996 base pairs of sequence immediately 5' and proximal to the translation start codon of the gene was fused to DNA encoding the
uidA reporter gene. The sequence modification was made to the leader to facilitate sub-cloning, such that the adenine (a) and thymine (t) bases at positions -2 and -1 respectively (at the very end of the leader, immediately before the translation start codon) were replaced with two cytosine (c) bases. The potato protease inhibitor II (PinII) terminator sequence was positioned downstream of the uidA coding sequence. This transcription unit was included in a plant transformation vector. The resulting construct is shown in Figure 2.

As a potential guide to promoter activity, the construct was transiently introduced into developing maize embryos and stained for GUS activity. The promoter sequence of the gene drove uidA expression in transiently transformed embryos, indicating the potential of this sequence to regulate transgene expression in maize.

Promoter sequences of the gene can drive transgene expression in stably transformed seed tissue

The promoter -reporter fusion was then stably introduced into the maize genome by Agrobacterium mediated transformation. Following the transformation of developing embryo tissues, uidA expression was assessed in non-differentiated callus tissue prior to plant regeneration. GUS activity was detected in callus tissue derived from transformation experiments using the promoter -reporter fusion.

Plants were then regenerated from transformation events obtained using the vector. A total of 105 plants were regenerated from thirteen independent transformation events obtained using the gene promoter -uidA fusion. Seed was harvested, the soluble protein was extracted, and for each plant the level of GUS was determined in each of six randomly selected seeds and also on a pool of 50 randomly selected seeds.

The GUS levels for the transgenic seed are summarized in Table 1 and shown graphically in Figure 4.

Table 1. Recombinant protein level data derived from single and bulk seed analysis of transgenic maize carrying the gene promoter -uidA reporter fusion.

<table>
<thead>
<tr>
<th>Data included in analysis (see notes)</th>
<th>Sample size</th>
<th>Mean GUS level (% TSP)</th>
<th>Standard deviation (% TSP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>304</td>
<td>0.024</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1: Single seed analysis, where the mean level of GUS for the construct is calculated from all positive seed.

2: Single seed analysis, where the mean level of GUS for the construct is calculated from the mean level of GUS for all plants, itself derived from positive seed data only.

3: Single seed analysis, where the mean level of GUS for the construct is calculated from the mean levels of GUS for each independent transformation event, themselves calculated from the means of GUS for all plants regenerated from each event, which are derived from positive seed data only.

4: Single seed analysis: where the mean level of GUS for the construct is calculated from the highest recorded level of GUS for a seed from each plant.

5: Single seed analysis: where the mean level of GUS for the construct is calculated from the mean levels of GUS for each independent transformation event, themselves calculated from the highest recorded level of GUS for a seed from each plant regenerated from that event.

6: Bulk seed analysis: where the mean level of GUS for the construct is calculated from the level of GUS of each plant.

7: Bulk seed analysis: where the mean level of GUS for the construct is calculated from the mean levels of GUS for each independent transformation event, themselves calculated from the level of GUS of each plant regenerated from that event.

The mean GUS levels achieved using the promoter sequence was calculated in several alternative ways. Alternative methods of analysis were based on all seeds that had detectable levels of GUS for each plant or only on the seed that had the highest level of GUS for each plant. Also, the mean GUS expression level obtained using the construct was based either on mean expression levels for each independent transformation event, or for each transgenic plant, or on data for each seed. Negative GUS expression data was included in the analysis. Furthermore, the calculations were
either based on individual seed data or on bulk seed data where protein was extracted from a pool of 50 seed. In any bulk sample approximately half the seed are anticipated to be nulls, so that GUS levels calculated from bulk seed analyses are expected to be less than those calculated from single seed analyses.

The promoter sequence of the gene can clearly drive reporter gene expression in maize seed. Depending on the method of analysis, mean GUS levels for single T<sub>1</sub> seed vary from 0.022% to 0.027% of total soluble protein, whereas regardless of the method of analysis, mean GUS levels for bulk T<sub>1</sub> seed are 0.008% of total soluble protein. Also, as a guide to the potential of the promoter sequence of the gene to facilitate protein production in plants, the highest level of GUS recorded in a single seed was noted. This highest recorded level of GUS was 0.074% of total soluble protein.

The promoter of the maize gene drives embryo-preferred expression

The tissue specificity of expression using the promoter of the gene was then assessed. Three of the highest expressing lines for the construct, each from a separate transformation event, were grown in the next generation from T<sub>1</sub> seeds and were assessed in a wide range of non-seed tissues. Representative tissue samples were collected from leaves at 21 days post-germination and at 12 days post-pollination. Stem, root and silk tissues were also collected at 12 days post-pollination, and husk and cob tissues at 19 days post-pollination. Also, pollen and anther tissues were collected at the time of pollen shed. All tissue samples were treated to reveal any evidence of GUS activity. The promoter of the gene showed no indication of driving GUS activity in any of the above tissues, with the exception of cob tissue, which showed faint localized (non-uniform) staining. Thus, apart from in cob tissue, the promoter sequence does not drive expression in non-seed tissues, and even in the cob expression is weak.

Expression of the uidA reporter gene was also assessed in T<sub>1</sub> seed tissues harvested directly from the T<sub>0</sub> transgenic plants. Fully mature dried down seeds were sliced in half and treated to reveal GUS activity. Strong blue staining was observed in the embryo, but no staining was observed in endosperm or aleurone tissues. Thus, within the seed, expression appears to be localized to the embryo.

The promoter of the maize gene drives expression in embryo tissues throughout development
The specificity of the promoter of the gene was also assessed in seed tissues throughout development. The same plants were utilized as those used to examine non-seed tissue expression, described above. Three of the highest expressing lines, each from a separate transformation event, were grown from T₁ seeds. Seed tissues were collected at 12, 19, 27 or 28 and 37 days post-pollination, the final point corresponding approximately to seed maturity. Seed was then treated to reveal GUS activity. Also, seed tissues were assessed following a dry down period of approximately three weeks. For the 27/28 and 37-day post-pollination material and for the dried down material, the seeds were sliced in half prior to the treatment in order to more clearly reveal the pattern of embryo, endosperm and aleurone/pericarp expression. However, for 12 and 19-day post-pollination material, tissue specificity was determined by dissecting out the embryo from the surrounding endosperm prior to the treatment of each tissue type.

The staining pattern indicating GUS activity in seed tissues throughout development is summarized in Table 2.

**Table 2.** Tissue specificity of the gene promoter -uidA reporter fusions in developing T₂ seeds.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>GUS staining</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-day embryo</td>
<td>Very localizedᵇ</td>
</tr>
<tr>
<td>12-day endosperm</td>
<td>Negative</td>
</tr>
<tr>
<td>19-day embryo</td>
<td>Localized</td>
</tr>
<tr>
<td>19-day endosperm</td>
<td>Very faintᵇ</td>
</tr>
<tr>
<td>19-day aleurone/pericarp</td>
<td>Localized</td>
</tr>
<tr>
<td>~28-day embryo</td>
<td>Localized</td>
</tr>
<tr>
<td>~28-day endosperm</td>
<td>Faint</td>
</tr>
<tr>
<td>~28-day aleurone/pericarp</td>
<td>Localized</td>
</tr>
<tr>
<td>37-day embryo</td>
<td>Localized</td>
</tr>
<tr>
<td>37-day endosperm</td>
<td>Very faint</td>
</tr>
<tr>
<td>37-day aleurone/pericarp</td>
<td>Localized</td>
</tr>
<tr>
<td>Dried down embryo</td>
<td>Localized</td>
</tr>
<tr>
<td>Dried down endosperm</td>
<td>Negative</td>
</tr>
<tr>
<td>Dried down aleurone/pericarp</td>
<td>Negative</td>
</tr>
</tbody>
</table>
The time points are relative to pollination and the final samples were assessed after approximately 3 weeks dry down.

Only one of the three lines examined showed staining.

GUS activity was evident in embryo tissues 12 days after pollination with only one of the three lines tested, and even then only with occasional embryos. With the positive line expression is localized to the apical tip of the axial surface of the embryo. At this stage of development no staining was evident in the endosperm. The 19-day post-pollination developing seeds are much larger than the 12-day seeds, and the degree of staining was much greater for the older embryos with all three lines staining. The staining within the embryo was also less clearly restricted, although it was still somewhat localized to the apical region and was much more evident on the axial surface. At this stage some staining was evident in the aleurone/pericarp tissue for all three lines and faint diffuse staining was observed in the endosperm for one of the three lines only.

By 27/28 days post-pollination developing seeds have further enlarged and the degree of staining in the embryo increased, being particularly evident in the scutellum. Also, some localized GUS activity is present in the aleurone/pericarp at 27/28 days post-pollination, and some diffuse faint staining was evident in the endosperm, although the degree of staining is much fainter in the endosperm than in the embryo. The staining pattern in 37-day post-pollination seeds is very similar to that at 27/28 days post-pollination, although aleurone/pericarp and endosperm staining are even weaker at 37 days post-pollination. Given that the seeds do not increase in size during the intervening period, the similar expression pattern may reflect a continued unchanging pattern of uidA expression, or a drop off in expression without substantial GUS protein degradation.

Following dry down of seed the pattern of staining in the embryo was very similar to that observed in 27/28-day and 37-day post-pollination embryos. Since the dried down tissue is presumably not metabolically active, this staining pattern is taken to reflect the late seed stage GUS activity pattern. No staining was evident in the endosperm or aleurone/pericarp of dried seeds.

Overall, in developing seed tissues GUS activity is strongly embryo preferred, with the scutellum being the site of strongest activity within the embryo. Expression is
only just beginning to be evident at 12 days post-pollination, but from then on GUS staining is clear right through to the dried down seed stage.

Discussion

Promoter sequences that are active in plant tissues are vital tools in implementing a range of strategies to engineer plant characteristics. However, overexpression of transgenes throughout the plant can have undesired effects and consequences. Tissue preferred and tissue specific promoters are important for restricting the expression of selected transgenes to particular parts of the plant, thereby eliminating deleterious effects that might arise from constitutive expression. Promoters well suited to expressing transgenes specifically in target tissues are most clearly identified as those that drive the expression of native genes in those tissues. In the case of cereals, including maize, seed tissues are of particular interest for crop improvement and for acting as a repository for protein accumulation. Thus, promoters that are active in seed tissues are of considerable value for crop development and for innovations pertaining to seeds.

The above experiments confirmed that message of the gene identified here is one of the most highly prevalent messages in maize embryos by sampling cDNA libraries representing embryo tissues from diverse lines at different stages of development. The gene sequence was estimated to comprise from 0.6% to 2.3% of the total clones, depending on the method of screening. Furthermore, expression appeared specific to embryo tissues. From a genomic clone extending upstream of the maize gene’s translation start codon, approximately 3 kb of promoter sequence was isolated and the nucleotide sequence determined. When fused to the \textit{uidA} reporter gene and transformed back into maize, this promoter sequence resulted in clear reporter gene activity.

The promoter/leader of the gene cloned here appears to have high tissue specificity, with a reporter gene expression being seed preferred apart from some minor activity in the cob, though even this possibly represents an overflow from seed activity. The promoter sequence can drive embryo expression as early as 12 days after pollination, although activity is greatly increased by 19 days after pollination and increases further by 27 to 28 days after pollination. Activity continues throughout embryo development, but with \textit{uidA} as the reporter gene the presence of GUS activity in late stage embryos may reflect upon previously synthesized protein rather than active transcription and translation. The expression profile observed using the \textit{uidA}
reporter is in line with the observed abundance of cDNAs in the developmental seed libraries, where the native full length promoter of the gene appears not to be as active in young developing embryo tissue as in maturing tissue, particularly 28 day post-pollination tissue. During the later stages of seed development promoter activity within the embryo is strongest in the scutellum. Expression is also observed in the aleurone/pericarp, and at a very low level in the endosperm from 19 to 37 days post-pollination, although expression is not observed in these tissues in dried down seeds.

The strong highly embryo preferred activity of the promoter of the maize gene makes it an excellent choice for seed preferred/specific expression in plants, preferably in maize, and other cereals. Using the here cloned promoter, high levels of transgene expression are achievable in seeds.

References


activity following transfer to protoplasts by electroporation. Plant Mol. Biol. 18, 675-689.


Maiti, I.B., Gowda, S., Kierman, J., Ghosh, S.K. and Shepherd, R.J. (1997) Promoter/leader deletion analysis and plant expression vectors with the figwort mosaic virus (FMV) full length transcript (FLT) promoter containing single or double enhancer domains. Transgenic Res. 6, 143-156.


What is claimed is:

1. An isolated regulatory element that drives transcription in an embryo-preferred manner, the regulatory element comprising a nucleotide sequence selected from the group consisting of:
   (a) sequences natively associated with and regulating expression of DNA coding for a maize abscisic acid-inducible gene;
   (b) the nucleotide sequence of SEQ ID NO: 3; and
   (c) A sequence comprising a functional variant of the nucleotide sequence set forth in SEQ ID NO: 3.

2. An isolated regulatory element comprising the nucleotide sequence of SEQ ID NO: 3.

3. An isolated regulatory element comprising a functional variant of the nucleotide sequence set forth in SEQ ID NO: 3 capable of regulating transcription in an embryo-preferred manner.

4. An expressional cassette comprising a regulatory element that drives transcription in an embryo-preferred manner, the regulatory element comprising a nucleotide sequence selected from the group consisting of:
   (a) sequences natively associated with and regulating expression of DNA coding for a maize abscisic acid-inducible gene;
   (b) the nucleotide sequence of SEQ ID NO: 3; and
   (c) A sequence comprising a functional variant of the nucleotide sequence set forth in SEQ ID NO: 3.

5. An expression cassette comprising a regulatory element and a first nucleotide sequence operably linked to the regulatory element, said regulatory element comprising the nucleotide sequence of SEQ ID NO: 3.

6. An expression cassette comprising a regulatory element and a first nucleotide sequence operably linked to the regulatory element, said regulatory element comprising a functional variant of the nucleotide sequence of SEQ ID NO: 3 that regulates expression of the first nucleotide sequence in an embryo-preferred manner.

7. A transformation vector comprising an expression cassette, the expression cassette comprising a regulatory element selected from the group consisting of:
   (a) sequences natively associated with and regulating expression of DNA coding for a maize abscisic acid-inducible gene;
(b) the nucleotide sequence of SEQ ID NO: 3; and

(c) A sequence comprising a functional variant of the nucleotide sequence set forth in SEQ ID NO: 3.

8. A plant comprising the expression cassette of claim 4.

9. A plant cell comprising the expression cassette of claim 4.

10. Embryo of the plant of claim 9.

11. A method for selectively expressing a nucleotide sequence in a plant embryo, the method comprising transforming a plant cell with a transformation vector comprising an expression cassette, the expression cassette comprising a regulatory element selected from the group consisting of:

(a) sequences natively associated with and regulating expression of DNA coding for a maize abscisic acid-inducible gene;

(b) the nucleotide sequence of SEQ ID NO: 3; and

(c) A sequence comprising a functional variant of the nucleotide sequence set forth in SEQ ID NO: 3.

12. An isolated regulatory element comprising a nucleotide sequence having 90% identity to SEQ ID NO: 3.

13. An isolated regulatory element comprising a nucleotide sequence which hybridizes to SEQ ID NO: 3 under highly stringent conditions of a wash of 50% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60°C.
Figure 1B

MRPVETPTREIKLGDGLWAFLDRENCGIDQRWSES
ALQESRAIAVPGSNDQFDADIRNYAGNVWYQRREVFI
PKGWAQRIVRLRFDAVTYYGKVWNNQEVMEHQGGY
TPFEADVTPYVIAGKSVRITVCVNNENLWQTIPGTVT
DENKGKQKSYFHDFFNYAGIHRSVMLYTPNTWVDDEIT
VVTVAQDCNHASVDCQVVVANGDVSVELRDAQDQV
ATGQGTSGLQLQVYVPHLCQPGEGYLYELCVTAKSQTEC
DIYPLRVGIRSVAVKQQFLINHHPFYFTGFFRGHEADDL
RGKGDNVLMTVDHALMDWIGANSYRTSHPYAEEMLD
WDADEGVIQVIDETAAVGFNLSLGGFEEAGNKPKELYS
EEAVNGETQQAAHLQAIKELIARDKNHPSVVMWSIANEP
DTRPVQVHGLISPLAEATRKLDPTRPCITCVNMFCDAHT
DTSIDLFDVCLNRYYGWVQSGDETAEKVLKELLA
WQEKLHQPIIITEYGVDTLAGLHSMYTDWMSEXYQCA
WLMYHRVFDRVSAVVGEQVWNFADFBTSQGILRVGG
NKGGIFTTRDRKPKSAAFLLLQKWRTGMPNFGKPKQQGKG
QHHHHHH
Figure 2

![Diagram of pPGN9087](image)

- T-DNA right border
- T-DNA left border
- ColE1 replicon
- streptothricin acetyltransferase
- streptomycin/spectinomycin nucleotidytransferase
- promoter/leader
- beta-glucuronidase
- potato proteinase inhibitor II terminator
- cauliflower mosaic virus 35S terminator
- maize optimized phosphinothricin N-acetyltransferase

**pPGN9087**

12852 bp
Figure 4

![Graph showing GUS level (µg TSP) by approach to determine the mean.](image-url)
INTERNATIONAL SEARCH REPORT

International application No.
PCT/US05/16232

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/11, 15/92, A01H 5/00
US Cl. : 536/241; 435/320.1, 419, 468, 800/287, 298

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/241; 435/320 1, 419, 468, 800/287, 298

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic database consulted during the international search (name of database and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Database GenEmbl on STIC, European Bioinformatics Institute, (Hinxton, Cambridge, UK), Accession No. AC149655, BIRKEN et al., Gene Sequence, 17 January 2004 (20.01.2004).</td>
<td>1, 3-4, 6-7, 13</td>
</tr>
<tr>
<td>X</td>
<td>Database EST on STIC, National Center for Biotechnology Information (Bethesda, MD, USA), Accession No. CG000752, WHITEMAN et al., Gene Sequence, 19 August 2003 (19.08.2003).</td>
<td>1, 3-4, 6-7, 12-13</td>
</tr>
<tr>
<td>X</td>
<td>LIU et al. Tissue-specific and ABA-regulated Maize Glb1 gene expression in transgenic tobacco. Plant Cell Reports. 1996, Vol. 16, pages 158-162, the entire article.</td>
<td>1, 3-4, 6-11</td>
</tr>
</tbody>
</table>

* Further documents are listed in the continuation of Box C. See patent family annex.

*I* Special categories of cited documents

"X" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another document or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search
08 September 2005 (08 09 2005)

Date of mailing of the international search report
04 NOV 2005

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (571) 273-3201

Authorized officer
Cynthia Collins
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (April 2005)
Continuation of B. FIELDS SEARCHED Item 3:
STN (agricola, biosis, biotechno, cabo, capplus,medline): inventor names, maize, zea, embryo, promoter. STIC sequence search for SEQ ID NO:3.