



(51) International Patent Classification:

C12N 15/82 (2006.01) C12N 15/63 (2006.01)
C12N 15/79 (2006.01)

(21) International Application Number:

PCT/US2017/026969

(22) International Filing Date:

11 April 2017 (11.04.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/330527 02 May 2016 (02.05.2016) US

(71) Applicant: **DOW AGROSCIENCES LLC** [US/US];
9330 Zionsville Road, Indianapolis, IN 46268 (US).

(72) Inventors: **GUPTA, Manju**; 9330 Zionsville Road, Indianapolis, IN 46268 (US). **KUMAR, Sandeep**; 9330 Zionsville Road, Indianapolis, IN 46268 (US). **CHEN, Wei**; 9330 Zionsville Road, Indianapolis, IN 46268 (US).

(74) Agent: **RUSSELL, Sean**; Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: PLANT PROMOTER AND 3'UTR FOR TRANSGENE EXPRESSION

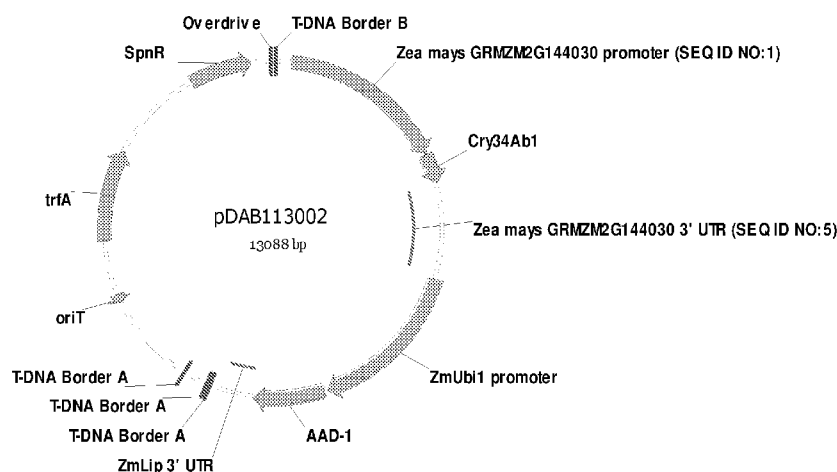


FIG. 1:

(57) Abstract: This disclosure concerns compositions and methods for promoting transcription of a nucleotide sequence in a plant or plant cell, employing a Zea mays GRMZM2G144030 promoter. Some embodiments relate to a Zea mays GRMZM2G144030 promoter that functions in plants to promote transcription of operably linked nucleotide sequences. Other embodiments relate to a Zea mays GRMZM2G144030 3'UTR that functions in plants to terminate transcription of operably linked nucleotide sequences.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

- *with international search report (Art. 21(3))*
- *with sequence listing part of description (Rule 5.2(a))*

PLANT PROMOTER AND 3'UTR FOR TRANSGENE EXPRESSION

[0001] This application claims a priority based on provisional application 62/330527 which was filed in the U.S. Patent and Trademark Office on May 2, 2016, the entire disclosure of which is hereby incorporated by reference.

INCORPORATION BY REFERENCE OF MATERIAL SUBMITTED ELECTRONICALLY

[0002] Incorporated by reference in its entirety is a computer-readable nucleotide/amino acid sequence listing submitted concurrently herewith and identified as follows: one 26.0 KB ACII (Text) file named "78826-US-PSP-20160126-Sequence-Listing-ST25.txt" created on March 4, 2016.

FIELD OF THE DISCLOSURE

[0003] The present disclosure generally relates to compositions and methods for promoting transcription of a nucleotide sequence in a plant or plant cell. Some embodiments relate to a novel *Zea mays* GRMZM2G144030 promoter and other *Zea mays* GRMZM2G144030 regulatory elements that function in plants to promote and/or terminate transcription of an operably linked nucleotide sequence. Particular embodiments relate to methods including a promoter (*e.g.*, to introduce a nucleic acid molecule into a cell) and cells, cell cultures, tissues, organisms, and parts of organisms comprising a promoter, as well as products produced therefrom. Other embodiments relate to methods including a 3'UTR (*e.g.*, to introduce a nucleic acid molecule into a cell) and cells, cell cultures, tissues, organisms, and parts of organisms comprising a promoter, as well as products produced therefrom.

BACKGROUND

[0004] Many plant species are capable of being transformed with transgenes to introduce agronomically desirable traits or characteristics. Plant species are developed and/or modified to have particular desirable traits. Generally, desirable traits include, for example, improving nutritional value quality, increasing yield, conferring pest or disease resistance, increasing drought and stress tolerance, improving horticultural qualities (*e.g.*, pigmentation and growth), imparting herbicide tolerance,

enabling the production of industrially useful compounds and/or materials from the plant, and/or enabling the production of pharmaceuticals.

[0005] Transgenic plant species comprising multiple transgenes stacked at a single genomic locus are produced via plant transformation technologies. Plant transformation technologies result in the introduction of a transgene into a plant cell, recovery of a fertile transgenic plant that contains the stably integrated copy of the transgene in the plant genome, and subsequent transgene expression via transcription and translation of the plant genome results in transgenic plants that possess desirable traits and phenotypes. However, mechanisms that allow the production of transgenic plant species to highly express multiple transgenes engineered as a trait stack are desirable.

[0006] Likewise, mechanisms that allow the expression of a transgene within particular tissues or organs of a plant are desirable. For example, increased resistance of a plant to infection by soil-borne pathogens might be accomplished by transforming the plant genome with a pathogen-resistance gene such that pathogen-resistance protein is robustly expressed within the roots of the plant. Alternatively, it may be desirable to express a transgene in plant tissues that are in a particular growth or developmental phase such as, for example, cell division or elongation. Furthermore, it may be desirable to express a transgene in leaf and stem tissues of a plant to provide tolerance against herbicides, or resistance against above ground insects and pests.

[0007] Therefore, a need exists for new gene regulatory elements that can drive the desired levels of expression of transgenes in specific plant tissues.

BRIEF SUMMARY

[0008] In embodiments of the subject disclosure, the subject disclosure relates to a nucleic acid vector comprising a promoter operably linked to a polylinker sequence, a non-GRMZM2G144030 gene; or a combination of the polylinker sequence and the non-GRMZM2G144030 gene, wherein said promoter comprises a polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:1. In aspects of this embodiment, the promoter is 2,082 bp in length. Further embodiments include a promoter that consists of a polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:1. In other aspects, the promoter

is operably linked to a selectable maker. In additional aspects, the promoter is operably linked to a transgene. Exemplary transgenes include; a selectable marker or a gene product conferring insecticidal resistance, herbicide tolerance, nitrogen use efficiency, small RNA expression, site specific nuclease, water use efficiency, nutritional quality or DNA binding protein. In further aspects, the nucleic acid vector comprises a 3' untranslated polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:5, wherein the 3' untranslated sequence is operably linked to said polylinker or said transgene. In other aspects, the nucleic acid vector comprises a 5' untranslated polynucleotide sequence, wherein the 5' untranslated sequence is operably linked to said polylinker or said transgene. In further aspects, the nucleic acid vector comprises an intron sequence. The promoter of the disclosure further drives transgene expression in leaf tissues.

[0009] In further embodiments of the subject disclosure, the disclosure relates to a non-*Zea mays* c.v. B73 plant comprising a polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:1 operably linked to a transgene. In an aspect of this embodiment, the plant is selected from the group consisting of wheat, rice, sorghum, oats, rye, bananas, sugar cane, soybean, cotton, Arabidopsis, tobacco, sunflower, and canola. In another aspect, the plant is a maize plant. In further aspects, the transgene is inserted into the genome of the plant. In other embodiments, the polynucleotide sequence having at least 90% sequence identity with SEQ ID NO:1 is a promoter. In another aspect, the plant comprises a 3' untranslated sequence comprising SEQ ID NO:5, wherein the 3' untranslated sequence is operably linked to said transgene. In further aspects, the promoter drives transgene expression leaf tissues. In an additional aspect of these embodiments, the promoter is 2,082 bp in length.

[0010] In other embodiments of the subject disclosure, a method for producing a transgenic plant cell is provided. The method including the steps of transforming a plant cell with a gene expression cassette comprising a *Zea mays* GRMZM2G144030 promoter operably linked to at least one polynucleotide sequence of interest; isolating the transformed plant cell comprising the gene expression cassette; and, producing a transgenic plant cell comprising the *Zea mays* GRMZM2G144030 promoter operably linked to at least one polynucleotide sequence of interest. In an aspect of this embodiment, the transformation of the plant cell is performed with a plant

transformation method. The plant transformation method being selected from any of the following transformation methods; *Agrobacterium*-mediated transformation method, a biolistics transformation method, a silicon carbide transformation method, a protoplast transformation method, and a liposome transformation method. In further aspects, the polynucleotide sequence of interest is preferentially expressed in leaf tissues. In additional aspects, the polynucleotide sequence of interest is stably integrated into the genome of the transgenic plant cell. In additional aspects the method further includes the steps of regenerating the transgenic plant cell into a transgenic plant, and obtaining the transgenic plant, wherein the transgenic plant comprises the gene expression cassette comprising the *Zea mays* GRMZM2G144030 promoter of claim 1 operably linked to at least one polynucleotide sequence of interest. In another aspect of the embodiment, the transgenic plant cell is a monocotyledonous transgenic plant cell or a dicotyledonous transgenic plant cell. Accordingly, the dicotyledonous transgenic plant cell can be an *Arabidopsis* plant cell, a tobacco plant cell, a soybean plant cell, a canola plant cell, and a cotton plant cell. Likewise, the monocotyledonous transgenic plant cell can be a maize plant cell, a rice plant cell, and a wheat plant cell. In other aspects, *Zea mays* GRMZM2G144030 promoter comprising the polynucleotide of SEQ ID NO:1. In an aspect the embodiments include a first polynucleotide sequence of interest operably linked to the 3' end of SEQ ID NO:1.

[0011] The subject disclosure further relates to a method for expressing a polynucleotide sequence of interest in a plant cell, the method comprising introducing into the plant cell a polynucleotide sequence of interest operably linked to a *Zea mays* GRMZM2G144030 promoter. In an aspect of this embodiment, the polynucleotide sequence of interest operably linked to the *Zea mays* GRMZM2G144030 promoter is introduced into the plant cell by a plant transformation method. In an additional aspect, the plant transformation method is selected from *Agrobacterium*-mediated transformation method, a biolistics transformation method, a silicon carbide transformation method, a protoplast transformation method, and a liposome transformation method. In further aspects, the polynucleotide sequence of interest is constitutively expressed throughout the plant cell. In yet other aspects, the polynucleotide sequence of interest is stably integrated into the genome of the plant cell. In an embodiment, the transgenic plant cell is a monocotyledonous plant

cell or a dicotyledonous plant cell. Accordingly, the dicotyledonous plant cell includes an *Arabidopsis* plant cell, a tobacco plant cell, a soybean plant cell, a canola plant cell, and a cotton plant cell. Likewise, the monocotyledonous plant cell includes a maize plant cell, a rice plant cell, and a wheat plant cell.

[0012] The subject disclosure further relates to a transgenic plant cell comprising a *Zea mays* GRMZM2G144030 promoter. In an aspect of this embodiment, the transgenic plant cell comprises a transgenic event. In other aspects, the transgenic event comprises an agronomic trait. Exemplary transgenic traits include an insecticidal resistance trait, herbicide tolerance trait, nitrogen use efficiency trait, water use efficiency trait, nutritional quality trait, DNA binding trait, selectable marker trait, small RNA trait, or any combination thereof. In an embodiment, the herbicide tolerant trait comprises an *aad-1* coding sequence. In other aspects, the transgenic plant cell produces a commodity product. For example, the commodity product can be protein concentrate, protein isolate, grain, meal, flour, oil, or fiber. In further aspects, the transgenic plant cell is selected from the group consisting of a dicotyledonous plant cell or a monocotyledonous plant cell. For example, the monocotyledonous plant cell may be a maize plant cell. In other aspects, the *Zea mays* GRMZM2G144030 promoter comprises a polynucleotide with at least 90% sequence identity to the polynucleotide of SEQ ID NO:1. In further aspects, the *Zea mays* GRMZM2G144030 promoter is 2,082 bp in length. In yet another aspect, the *Zea mays* GRMZM2G144030 promoter comprises a polynucleotide sequence with at least 90% sequence identity of SEQ ID NO:1. Further aspects include a first polynucleotide sequence of interest operably linked to the 3' end of SEQ ID NO:1. In an additional aspect, the agronomic trait is preferentially expressed in leaf tissues.

[0013] The subject disclosure further relates to an isolated polynucleotide comprising a nucleic acid sequence with at least 90% sequence identity to the polynucleotide of SEQ ID NO:1. In an aspect of this embodiment, the isolated polynucleotide has preferred expression in leaf tissues. In other aspects, the isolated polynucleotide has expression activity within a plant cell. In further aspects, the isolated polynucleotide comprises an open-reading frame polynucleotide

coding for a polypeptide; and a termination sequence. In an additional aspect, the polynucleotide of SEQ ID NO:1 is 2,082 bp in length.

[0014] The foregoing and other features will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

[0015] **Fig. 1:** This figure is a schematic of pDAB113002 which contains the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1 and the *Zea mays* GRMZM2G144030 3'UTR of SEQ ID NO:5 in a gene expression cassette that drives expression of the *cry3Ab1* transgene.

DETAILED DESCRIPTION

I. Overview of several embodiments

[0016] Development of transgenic plant products is becoming increasingly complex. Commercially viable transgenic plants now require the stacking of multiple transgenes into a single locus. Plant promoters used for basic research or biotechnological applications are generally unidirectional, directing only one gene that has been fused at its 3' end (downstream). Plant 3'UTRs used for basic research or biotechnological applications are generally unidirectional, terminating the expression of only one gene that has been fused at its 5' end (upstream). Accordingly, each transgene usually requires a promoter for expression and a 3'UTR for termination of expression, wherein multiple promoters and 3'UTRs are required to express multiple transgenes within one gene stack. With an increasing number of transgenes in gene stacks, the same promoter and 3'UTR is routinely used to obtain similar levels of expression patterns of different transgenes. Obtaining similar levels of transgene expression is necessary for the production of a single polygenic trait. Unfortunately, multi-gene constructs driven by the same promoter and 3'UTR are known to cause gene silencing resulting in less efficacious transgenic products in the field. The repeated promoter and 3'UTR elements may lead to homology-based gene silencing. In addition, repetitive sequences within a transgene may lead to gene intra locus homologous recombination resulting in polynucleotide

rearrangements. The silencing and rearrangement of transgenes will likely have an undesirable effect on the performance of a transgenic plant produced to express transgenes. Further, excess of transcription factor (TF)-binding sites due to promoter repetition can cause depletion of endogenous TFs leading to transcriptional inactivation. Given the need to introduce multiple genes into plants for metabolic engineering and trait stacking, a variety of promoters and 3'UTRs are required to develop transgenic crops that drive the expression of multiple genes.

[0017] A particular problem in promoter identification is the need to identify tissue-specific promoters, related to specific cell types, developmental stages and/or functions in the plant that are not expressed in other plant tissues. Tissue specific (*i.e.*, tissue preferred) or organ specific promoters drive gene expression in a certain tissue such as in the kernel, root, leaf, or tapetum of the plant. Tissue and developmental stage specific promoters can be initially identified from observing the expression of genes, which are expressed in particular tissues or at particular time periods during plant development. These tissue specific promoters are required for certain applications in the transgenic plant industry and are desirable as they permit specific expression of heterologous genes in a tissue and/or developmental stage selective manner, indicating expression of the heterologous gene differentially at various organs, tissues and/or times, but not in other tissue. For example, increased resistance of a plant to infection by soil-borne pathogens might be accomplished by transforming the plant genome with a pathogen-resistance gene such that pathogen-resistance protein is robustly expressed within the roots of the plant. Alternatively, it may be desirable to express a transgene in plant tissues that are in a particular growth or developmental phase such as, for example, cell division or elongation. Another application is the desirability of using tissue specific promoters to confine the expression of the transgenes encoding an agronomic trait in specific tissues types like developing parenchyma cells. As such, a particular problem in the identification of promoters is how to identify the promoters, and to relate the identified promoter to developmental properties of the cell for specific tissue expression.

[0018] Another problem regarding the identification of a promoter or 3'UTR is the requirement to clone all relevant cis-acting and trans-activating transcriptional control elements so that the cloned DNA fragment drives transcription in the wanted specific expression pattern.

Given that such control elements are located distally from the translation initiation or start site, the size of the polynucleotide that is selected to comprise the promoter is of importance for providing the level of expression and the expression patterns of the promoter polynucleotide sequence. Given that similar elements for termination of a 3'UTR are located distally from the translation termination or stop site, the size of the polynucleotide that is selected to comprise the 3'UTR is of importance for providing termination of the expression of a transgene encoded by a polynucleotide sequence. It is known that promoter and 3'UTR lengths include functional information, and different genes have been shown to have promoters longer or shorter than promoters of the other genes in the genome. Elucidating the transcription start site of a promoter and predicting the functional gene elements in the promoter region is challenging. Further adding to the challenge are the complexity, diversity and inherent degenerate nature of regulatory motifs and cis- and trans-regulatory elements (Blanchette, Mathieu, et al. "Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression." *Genome research* 16.5 (2006): 656-668). The cis- and trans-regulatory elements are located in the distal parts of the promoter which regulate the spatial and temporal expression of a gene to occur only at required sites and at specific times (Porto, Milena Silva, et al. "Plant promoters: an approach of structure and function." *Molecular biotechnology* 56.1 (2014): 38-49). Existing promoter analysis tools cannot reliably identify such cis regulatory elements in a genomic sequence, thus predicting too many false positives because these tools are generally focused only on the sequence content (Fickett JW, Hatzigeorgiou AG (1997) Eukaryotic promoter recognition. *Genome research* 7: 861–878). Accordingly, the identification of promoter regulatory elements requires that an appropriate sequence of a specific size is obtained that will result in driving expression of an operably linked transgene in a desirable manner. Furthermore, the identification of 3'UTR regulatory elements requires that an appropriate sequence of a specific size is obtained that will result in terminating the expression of an operably linked transgene in a desirable manner.

[0019] Provided are methods and compositions for overcoming such problems through the use of *Zea mays* GRMZM2G144030 promoter elements and other *Zea mays* GRMZM2G144030 regulatory elements to express transgenes in plant.

II. *Terms and Abbreviations*

[0020] Throughout the application, a number of terms are used. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

[0021] As used herein, the term “intron” refers to any nucleic acid sequence comprised in a gene (or expressed polynucleotide sequence of interest) that is transcribed but not translated. Introns include untranslated nucleic acid sequence within an expressed sequence of DNA, as well as the corresponding sequence in RNA molecules transcribed therefrom. A construct described herein can also contain sequences that enhance translation and/or mRNA stability such as introns. An example of one such intron is the first intron of gene II of the histone H3 variant of *Arabidopsis thaliana* or any other commonly known intron sequence. Introns can be used in combination with a promoter sequence to enhance translation and/or mRNA stability.

[0022] The term “isolated”, as used herein means having been removed from its natural environment, or removed from other compounds present when the compound is first formed. The term “isolated” embraces materials isolated from natural sources as well as materials (e.g., nucleic acids and proteins) recovered after preparation by recombinant expression in a host cell, or chemically-synthesized compounds such as nucleic acid molecules, proteins, and peptides.

[0023] The term “purified”, as used herein relates to the isolation of a molecule or compound in a form that is substantially free of contaminants normally associated with the molecule or compound in a native or natural environment, or substantially enriched in concentration relative to other compounds present when the compound is first formed, and means having been increased in purity as a result of being separated from other components of the original composition. The term “purified nucleic acid” is used herein to describe a nucleic acid sequence which has been separated, produced apart from, or purified away from other biological compounds including, but not limited to polypeptides, lipids and carbohydrates, while effecting a chemical or functional change in the component (e.g., a nucleic acid may be purified from a chromosome by removing protein contaminants and breaking chemical bonds connecting the nucleic acid to the

remaining DNA in the chromosome).

[0024] The term “synthetic”, as used herein refers to a polynucleotide (i.e., a DNA or RNA) molecule that was created via chemical synthesis as an *in vitro* process. For example, a synthetic DNA may be created during a reaction within an Eppendorf™ tube, such that the synthetic DNA is enzymatically produced from a native strand of DNA or RNA. Other laboratory methods may be utilized to synthesize a polynucleotide sequence. Oligonucleotides may be chemically synthesized on an oligo synthesizer via solid-phase synthesis using phosphoramidites. The synthesized oligonucleotides may be annealed to one another as a complex, thereby producing a “synthetic” polynucleotide. Other methods for chemically synthesizing a polynucleotide are known in the art, and can be readily implemented for use in the present disclosure.

[0025] The term "about" as used herein means greater or lesser than the value or range of values stated by 10 percent, but is not intended to designate any value or range of values to only this broader definition. Each value or range of values preceded by the term "about" is also intended to encompass the embodiment of the stated absolute value or range of values.

[0026] For the purposes of the present disclosure, a "gene," includes a DNA region encoding a gene product (see *infra*), as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

[0027] As used herein the terms "native" or "natural" define a condition found in nature. A "native DNA sequence" is a DNA sequence present in nature that was produced by natural means or traditional breeding techniques but not generated by genetic engineering (e.g., using molecular biology/transformation techniques).

[0028] As used herein a "transgene" is defined to be a nucleic acid sequence that encodes a gene product, including for example, but not limited to, an mRNA. In one embodiment the transgene is an exogenous nucleic acid, where the transgene sequence has been introduced into a

host cell by genetic engineering (or the progeny thereof) where the transgene is not normally found. In one example, a transgene encodes an industrially or pharmaceutically useful compound, or a gene encoding a desirable agricultural trait (e.g., an herbicide-resistance gene). In yet another example, a transgene is an antisense nucleic acid sequence, wherein expression of the antisense nucleic acid sequence inhibits expression of a target nucleic acid sequence. In one embodiment the transgene is an endogenous nucleic acid, wherein additional genomic copies of the endogenous nucleic acid are desired, or a nucleic acid that is in the antisense orientation with respect to the sequence of a target nucleic acid in a host organism.

[0029] As used herein the term "non-GRMZM2G144030 transgene" or "non-GRMZM2G144030 gene" is any transgene that has less than 80% sequence identity with the GRMZM2G144030 gene coding sequence (SEQ ID NO:4).

[0030] A "gene product" as defined herein is any product produced by the gene. For example the gene product can be the direct transcriptional product of a gene (e.g., mRNA, tRNA, rRNA, antisense RNA, interfering RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of a mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation. Gene expression can be influenced by external signals, for example, exposure of a cell, tissue, or organism to an agent that increases or decreases gene expression. Expression of a gene can also be regulated anywhere in the pathway from DNA to RNA to protein. Regulation of gene expression occurs, for example, through controls acting on transcription, translation, RNA transport and processing, degradation of intermediary molecules such as mRNA, or through activation, inactivation, compartmentalization, or degradation of specific protein molecules after they have been made, or by combinations thereof. Gene expression can be measured at the RNA level or the protein level by any method known in the art, including, without limitation, Northern blot, RT-PCR, Western blot, or in vitro, in situ, or in vivo protein activity assay(s).

[0031] As used herein the term "gene expression" relates to the process by which the coded

information of a nucleic acid transcriptional unit (including, *e.g.*, genomic DNA) is converted into an operational, non-operational, or structural part of a cell, often including the synthesis of a protein. Gene expression can be influenced by external signals; for example, exposure of a cell, tissue, or organism to an agent that increases or decreases gene expression. Expression of a gene can also be regulated anywhere in the pathway from DNA to RNA to protein. Regulation of gene expression occurs, for example, through controls acting on transcription, translation, RNA transport and processing, degradation of intermediary molecules such as mRNA, or through activation, inactivation, compartmentalization, or degradation of specific protein molecules after they have been made, or by combinations thereof. Gene expression can be measured at the RNA level or the protein level by any method known in the art, including, without limitation, Northern blot, RT-PCR, Western blot, or *in vitro*, *in situ*, or *in vivo* protein activity assay(s).

[0032] As used herein, “homology-based gene silencing” (HBGS) is a generic term that includes both transcriptional gene silencing and post-transcriptional gene silencing. Silencing of a target locus by an unlinked silencing locus can result from transcription inhibition (transcriptional gene silencing; TGS) or mRNA degradation (post-transcriptional gene silencing; PTGS), owing to the production of double-stranded RNA (dsRNA) corresponding to promoter or transcribed sequences, respectively. The involvement of distinct cellular components in each process suggests that dsRNA-induced TGS and PTGS likely result from the diversification of an ancient common mechanism. However, a strict comparison of TGS and PTGS has been difficult to achieve because it generally relies on the analysis of distinct silencing loci. In some instances, a single transgene locus can triggers both TGS and PTGS, owing to the production of dsRNA corresponding to promoter and transcribed sequences of different target genes. Mourrain *et al.* (2007) *Planta* 225:365-79. It is likely that siRNAs are the actual molecules that trigger TGS and PTGS on homologous sequences: the siRNAs would in this model trigger silencing and methylation of homologous sequences in *cis* and in *trans* through the spreading of methylation of transgene sequences into the endogenous promoter.

[0033] As used herein, the term “nucleic acid molecule” (or “nucleic acid” or “polynucleotide”) may refer to a polymeric form of nucleotides, which may include both sense and anti-sense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the

above. A nucleotide may refer to a ribonucleotide, deoxyribonucleotide, or a modified form of either type of nucleotide. A “nucleic acid molecule” as used herein is synonymous with “nucleic acid” and “polynucleotide”. A nucleic acid molecule is usually at least 10 bases in length, unless otherwise specified. The term may refer to a molecule of RNA or DNA of indeterminate length. The term includes single- and double-stranded forms of DNA. A nucleic acid molecule may include either or both naturally-occurring and modified nucleotides linked together by naturally occurring and/or non-naturally occurring nucleotide linkages.

[0034] Nucleic acid molecules may be modified chemically or biochemically, or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications (e.g., uncharged linkages: for example, methyl phosphonates, phosphotriesters, phosphoramidites, carbamates, etc.; charged linkages: for example, phosphorothioates, phosphorodithioates, etc.; pendent moieties: for example, peptides; intercalators: for example, acridine, psoralen, etc.; chelators; alkylators; and modified linkages: for example, alpha anomeric nucleic acids, etc.). The term “nucleic acid molecule” also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular, and padlocked conformations.

[0035] Transcription proceeds in a 5' to 3' manner along a DNA strand. This means that RNA is made by the sequential addition of ribonucleotide-5'-triphosphates to the 3' terminus of the growing chain (with a requisite elimination of the pyrophosphate). In either a linear or circular nucleic acid molecule, discrete elements (e.g., particular nucleotide sequences) may be referred to as being “upstream” or “5' ” relative to a further element if they are bonded or would be bonded to the same nucleic acid in the 5' direction from that element. Similarly, discrete elements may be “downstream” or “3' ” relative to a further element if they are or would be bonded to the same nucleic acid in the 3' direction from that element.

[0036] A base “position”, as used herein, refers to the location of a given base or nucleotide residue within a designated nucleic acid. The designated nucleic acid may be defined by alignment (see below) with a reference nucleic acid.

[0037] Hybridization relates to the binding of two polynucleotide strands via Hydrogen bonds. Oligonucleotides and their analogs hybridize by hydrogen bonding, which includes Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary bases. Generally, nucleic acid molecules consist of nitrogenous bases that are either pyrimidines (cytosine (C), uracil (U), and thymine (T)) or purines (adenine (A) and guanine (G)). These nitrogenous bases form hydrogen bonds between a pyrimidine and a purine, and the bonding of the pyrimidine to the purine is referred to as “base pairing.” More specifically, A will hydrogen bond to T or U, and G will bond to C. “Complementary” refers to the base pairing that occurs between two distinct nucleic acid sequences or two distinct regions of the same nucleic acid sequence.

[0038] “Specifically hybridizable” and “specifically complementary” are terms that indicate a sufficient degree of complementarity such that stable and specific binding occurs between the oligonucleotide and the DNA or RNA target. The oligonucleotide need not be 100% complementary to its target sequence to be specifically hybridizable. An oligonucleotide is specifically hybridizable when binding of the oligonucleotide to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA, and there is sufficient degree of complementarity to avoid non-specific binding of the oligonucleotide to non-target sequences under conditions where specific binding is desired, for example under physiological conditions in the case of in vivo assays or systems. Such binding is referred to as specific hybridization.

[0039] Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the chosen hybridization method and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (especially the Na⁺ and/or Mg²⁺ concentration) of the hybridization buffer will contribute to the stringency of hybridization, though wash times also influence stringency. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed in Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, chs. 9 and 11.

[0040] As used herein, “stringent conditions” encompass conditions under which hybridization will only occur if there is less than 50% mismatch between the hybridization molecule

and the DNA target. “Stringent conditions” include further particular levels of stringency. Thus, as used herein, “moderate stringency” conditions are those under which molecules with more than 50% sequence mismatch will not hybridize; conditions of “high stringency” are those under which sequences with more than 20% mismatch will not hybridize; and conditions of “very high stringency” are those under which sequences with more than 10% mismatch will not hybridize.

[0041] In particular embodiments, stringent conditions can include hybridization at 65°C, followed by washes at 65°C with 0.1x SSC/0.1% SDS for 40 minutes.

[0042] The following are representative, non-limiting hybridization conditions:

Very High Stringency: Hybridization in 5x SSC buffer at 65°C for 16 hours; wash twice in 2x SSC buffer at room temperature for 15 minutes each; and wash twice in 0.5x SSC buffer at 65°C for 20 minutes each.

High Stringency: Hybridization in 5x-6x SSC buffer at 65-70°C for 16-20 hours; wash twice in 2x SSC buffer at room temperature for 5-20 minutes each; and wash twice in 1x SSC buffer at 55-70°C for 30 minutes each.

Moderate Stringency: Hybridization in 6x SSC buffer at room temperature to 55°C for 16-20 hours; wash at least twice in 2x-3x SSC buffer at room temperature to 55°C for 20-30 minutes each.

[0043] In particular embodiments, specifically hybridizable nucleic acid molecules can remain bound under very high stringency hybridization conditions. In these and further embodiments, specifically hybridizable nucleic acid molecules can remain bound under high stringency hybridization conditions. In these and further embodiments, specifically hybridizable nucleic acid molecules can remain bound under moderate stringency hybridization conditions.

[0044] Oligonucleotide: An oligonucleotide is a short nucleic acid polymer. Oligonucleotides may be formed by cleavage of longer nucleic acid segments, or by polymerizing individual nucleotide precursors. Automated synthesizers allow the synthesis of oligonucleotides up to several hundred base pairs in length. Because oligonucleotides may bind to a complementary nucleotide sequence, they may be used as probes for detecting DNA or RNA. Oligonucleotides composed of DNA (oligodeoxyribonucleotides) may be used in PCR, a technique for the

amplification of small DNA sequences. In PCR, the oligonucleotide is typically referred to as a “primer”, which allows a DNA polymerase to extend the oligonucleotide and replicate the complementary strand.

[0045] As used herein, the term “sequence identity” or “identity”, as used herein in the context of two nucleic acid or polypeptide sequences, may refer to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window.

[0046] As used herein, the term “percentage of sequence identity” may refer to the value determined by comparing two optimally aligned sequences (*e.g.*, nucleic acid sequences, and amino acid sequences) over a comparison window, wherein the portion of the 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 nucleotide or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the comparison window, and multiplying the result by 100 to yield the percentage of sequence identity.

[0047] Methods for aligning sequences for comparison are well-known in the art. Various programs and alignment algorithms are described in, for example: Smith and Waterman (1981) *Adv. Appl. Math.* 2:482; Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443; Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:2444; Higgins and Sharp (1988) *Gene* 73:237-44; Higgins and Sharp (1989) *CABIOS* 5:151-3; Corpet *et al.* (1988) *Nucleic Acids Res.* 16:10881-90; Huang *et al.* (1992) *Comp. Appl. Biosci.* 8:155-65; Pearson *et al.* (1994) *Methods Mol. Biol.* 24:307-31; Tatiana *et al.* (1999) *FEMS Microbiol. Lett.* 174:247-50. A detailed consideration of sequence alignment methods and homology calculations can be found in, *e.g.*, Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10.

[0048] The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST™; Altschul *et al.* (1990)) is available from several sources, including the National Center for Biotechnology Information (Bethesda, MD), and on the internet, for use in connection with several sequence analysis programs. A description of how to determine sequence identity using this program is available on the internet under the “help” section for BLAST™. For

comparisons of nucleic acid sequences, the “Blast 2 sequences” function of the BLAST™ (Blastn) program may be employed using the default parameters. Nucleic acid sequences with even greater similarity to the reference sequences will show increasing percentage identity when assessed by this method.

[0049] As used herein the term “operably linked” relates to a first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked with a coding sequence when the promoter affects the transcription or expression of the coding sequence. When recombinantly produced, operably linked nucleic acid sequences are generally contiguous and, where necessary to join two protein-coding regions, in the same reading frame. However, elements need not be contiguous to be operably linked.

[0050] As used herein, the term “promoter” refers to a region of DNA that generally is located upstream (towards the 5’ region of a gene) of a gene and is needed to initiate and drive transcription of the gene. A promoter may permit proper activation or repression of a gene that it controls. A promoter may contain specific sequences that are recognized by transcription factors. These factors may bind to a promoter DNA sequence, which results in the recruitment of RNA polymerase, an enzyme that synthesizes RNA from the coding region of the gene. The promoter generally refers to all gene regulatory elements located upstream of the gene, including, upstream promoters, 5’ UTR, introns, and leader sequences.

[0051] As used herein, the term “upstream-promoter” refers to a contiguous polynucleotide sequence that is sufficient to direct initiation of transcription. As used herein, an upstream-promoter encompasses the site of initiation of transcription with several sequence motifs, which include TATA Box, initiator sequence, TFIIB recognition elements and other promoter motifs (Jennifer, E.F. *et al.*, (2002) *Genes & Dev.*, 16: 2583-2592). The upstream promoter provides the site of action to RNA polymerase II which is a multi-subunit enzyme with the basal or general transcription factors like, TFIIA, B, D, E, F and H. These factors assemble into a transcription pre initiation complex that catalyzes the synthesis of RNA from DNA template.

[0052] The activation of the upstream-promoter is done by the additional sequence of regulatory DNA sequence elements to which various proteins bind and subsequently interact with the transcription initiation complex to activate gene expression. These gene regulatory elements sequences interact with specific DNA-binding factors. These sequence motifs may sometimes be referred to as *cis*-elements. Such *cis*-elements, to which tissue-specific or development-specific transcription factors bind, individually or in combination, may determine the spatiotemporal expression pattern of a promoter at the transcriptional level. These *cis*-elements vary widely in the type of control they exert on operably linked genes. Some elements act to increase the transcription of operably-linked genes in response to environmental responses (*e.g.*, temperature, moisture, and wounding). Other *cis*-elements may respond to developmental cues (*e.g.*, germination, seed maturation, and flowering) or to spatial information (*e.g.*, tissue specificity). See, for example, Langridge *et al.*, (1989) Proc. Natl. Acad. Sci. USA 86:3219-23. These *cis*- elements are located at a varying distance from transcription start point, some *cis*- elements (called proximal elements) are adjacent to a minimal core promoter region while other elements can be positioned several kilobases upstream or downstream of the promoter (enhancers).

[0053] As used herein, the terms “5’ untranslated region” or “5’ UTR” or “5’UTR” is defined as the untranslated segment in the 5’ terminus of pre-mRNAs or mature mRNAs. For example, on mature mRNAs, a 5’ UTR typically harbors on its 5’ end a 7-methylguanosine cap and is involved in many processes such as splicing, polyadenylation, mRNA export towards the cytoplasm, identification of the 5’ end of the mRNA by the translational machinery, and protection of the mRNAs against degradation.

[0054] As used herein, the terms “transcription terminator” is defined as the transcribed segment in the 3’ terminus of pre-mRNAs or mature mRNAs. For example, longer stretches of DNA beyond “polyadenylation signal” site is transcribed as a pre-mRNA. This DNA sequence usually contains transcription termination signal for the proper processing of the pre-mRNA into mature mRNA.

[0055] As used herein, the term “3’ untranslated region” or “3’UTR” or “3’ UTR” is defined as the untranslated segment in a 3’ terminus of the pre-mRNAs or mature

mRNAs. For example, on mature mRNAs this region harbors the poly-(A) tail and is known to have many roles in mRNA stability, translation initiation, and mRNA export. In addition, the 3' UTR is considered to include the polyadenylation signal and transcription terminator.

[0056] As used herein, the term “polyadenylation signal” designates a nucleic acid sequence present in mRNA transcripts that allows for transcripts, when in the presence of a poly-(A) polymerase, to be polyadenylated on the polyadenylation site, for example, located 10 to 30 bases downstream of the poly-(A) signal. Many polyadenylation signals are known in the art and are useful for the present invention. An exemplary sequence includes AAUAAA and variants thereof, as described in Loke J., et al., (2005) *Plant Physiology* 138(3); 1457-1468.

[0057] A “DNA binding transgene” is a polynucleotide coding sequence that encodes a DNA binding protein. The DNA binding protein is subsequently able to bind to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), a RNA molecule (an RNA-binding protein), and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, etc.) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding, and protein-binding activity.

[0058] Examples of DNA binding proteins include; meganucleases, zinc fingers, CRISPRs, and TALE binding domains that can be “engineered” to bind to a predetermined nucleotide sequence. Typically, the engineered DNA binding proteins (*e.g.*, zinc fingers, CRISPRs, or TALEs) are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering DNA-binding proteins are design and selection. A designed DNA binding protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP, CRISPR, and/or TALE designs and binding data. See, for example, U.S. Patents 6,140,081; 6,453,242; and 6,534,261; see also WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496 and U.S. Publication Nos. 20110301073, 20110239315 and 20119145940.

[0059] A “zinc finger DNA binding protein” (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP. Zinc finger binding domains can be “engineered” to bind to a predetermined nucleotide sequence. Non-limiting examples of methods for engineering zinc finger proteins are design and selection. A designed zinc finger protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP designs and binding data. See, for example, U.S. Pat. Nos. 6,140,081; 6,453,242; 6,534,261 and 6,794,136; see also WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496.

[0060] In other examples, the DNA-binding domain of one or more of the nucleases comprises a naturally occurring or engineered (non-naturally occurring) TAL effector DNA binding domain. See, e.g., U.S. Patent Publication No. 20110301073, incorporated by reference in its entirety herein. The plant pathogenic bacteria of the genus *Xanthomonas* are known to cause many diseases in important crop plants. Pathogenicity of *Xanthomonas* depends on a conserved type III secretion (T3S) system which injects more than different effector proteins into the plant cell. Among these injected proteins are transcription activator-like (TALEN) effectors which mimic plant transcriptional activators and manipulate the plant transcriptome (see Kay *et al.*, (2007) *Science* 318:648-651). These proteins contain a DNA binding domain and a transcriptional activation domain. One of the most well characterized TAL-effectors is AvrBs3 from *Xanthomonas campestris* pv. *Vesicatoria* (see Bonas *et al.*, (1989) *Mol Gen Genet* 218: 127-136 and WO2010079430). TAL-effectors contain a centralized domain of tandem repeats, each repeat containing approximately 34 amino acids, which are key to the DNA binding specificity of these proteins. In addition, they contain a nuclear localization sequence and an acidic transcriptional activation domain (for a review see Schornack S, *et al.*, (2006) *J Plant Physiol* 163(3): 256-272). In addition, in the phytopathogenic bacteria *Ralstonia solanacearum* two genes, designated *brg11*

and *hpx17* have been found that are homologous to the AvrBs3 family of *Xanthomonas* in the *R. solanacearum* biovar strain GMI1000 and in the biovar 4 strain RS1000 (See Heuer *et al.*, (2007) *Appl and Enviro Micro* 73(13): 4379-4384). These genes are 98.9% identical in nucleotide sequence to each other but differ by a deletion of 1,575 bp in the repeat domain of *hpx17*. However, both gene products have less than 40% sequence identity with AvrBs3 family proteins of *Xanthomonas*. See, e.g., U.S. Patent Publication No. 20110301073, incorporated by reference in its entirety.

[0061] Specificity of these TAL effectors depends on the sequences found in the tandem repeats. The repeated sequence comprises approximately 102 bp and the repeats are typically 91-100% homologous with each other (Bonas *et al.*, *ibid*). Polymorphism of the repeats is usually located at positions 12 and 13 and there appears to be a one-to-one correspondence between the identity of the hypervariable diresidues at positions 12 and 13 with the identity of the contiguous nucleotides in the TAL-effector's target sequence (see Moscou and Bogdanove, (2009) *Science* 326:1501 and Boch *et al.*, (2009) *Science* 326:1509-1512). Experimentally, the natural code for DNA recognition of these TAL-effectors has been determined such that an HD sequence at positions 12 and 13 leads to a binding to cytosine (C), NG binds to T, NI to A, C, G or T, NN binds to A or G, and ING binds to T. These DNA binding repeats have been assembled into proteins with new combinations and numbers of repeats, to make artificial transcription factors that are able to interact with new sequences and activate the expression of a non-endogenous reporter gene in plant cells (Boch *et al.*, *ibid*). Engineered TAL proteins have been linked to a *FokI* cleavage half domain to yield a TAL effector domain nuclease fusion (TALEN) exhibiting activity in a yeast reporter assay (plasmid based target).

[0062] The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR Associated) nuclease system is a recently engineered nuclease system based on a bacterial system that can be used for genome engineering. It is based on part of the adaptive immune response of many bacteria and Archaea. When a virus or plasmid invades a bacterium, segments of the invader's DNA are converted into CRISPR RNAs (crRNA) by the 'immune' response. This crRNA then associates, through a region of partial complementarity, with another

type of RNA called tracrRNA to guide the Cas9 nuclease to a region homologous to the crRNA in the target DNA called a “protospacer.” Cas9 cleaves the DNA to generate blunt ends at the double-stranded break (DSB) at sites specified by a 20-nucleotide guide sequence contained within the crRNA transcript. Cas9 requires both the crRNA and the tracrRNA for site specific DNA recognition and cleavage. This system has now been engineered such that the crRNA and tracrRNA can be combined into one molecule (the “single guide RNA”), and the crRNA equivalent portion of the single guide RNA can be engineered to guide the Cas9 nuclease to target any desired sequence (see Jinek *et al.*, (2012) *Science* 337, pp. 816-821, Jinek *et al.*, (2013), *eLife* 2:e00471, and David Segal, (2013) *eLife* 2:e00563). Thus, the CRISPR/Cas system can be engineered to create a DSB at a desired target in a genome, and repair of the DSB can be influenced by the use of repair inhibitors to cause an increase in error prone repair.

[0063] In other examples, the DNA binding transgene is a site specific nuclease that comprises an engineered (non-naturally occurring) Meganuclease (also described as a homing endonuclease). The recognition sequences of homing endonucleases or meganucleases such as I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII and I-TevIII are known. *See also* U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort *et al.*, (1997) *Nucleic Acids Res.* 25:3379–30 3388; Dujon *et al.*, (1989) *Gene* 82:115–118; Perler *et al.*, (1994) *Nucleic Acids Res.* 22, 11127; Jasin (1996) *Trends Genet.* 12:224–228; Gimble *et al.*, (1996) *J. Mol. Biol.* 263:163–180; Argast *et al.*, (1998) *J. Mol. Biol.* 280:345–353 and the New England Biolabs catalogue. In addition, the DNA-binding specificity of homing endonucleases and meganucleases can be engineered to bind non-natural target sites. *See, for example*, Chevalier *et al.*, (2002) *Molec. Cell* 10:895-905; Epinat *et al.*, (2003) *Nucleic Acids Res.* 5 31:2952-2962; Ashworth *et al.*, (2006) *Nature* 441:656-659; Paques *et al.*, (2007) *Current Gene Therapy* 7:49-66; U.S. Patent Publication No. 20070117128. The DNA-binding domains of the homing endonucleases and meganucleases may be altered in the context of the nuclease as a whole (*i.e.*, such that the nuclease includes the cognate cleavage domain) or may be fused to a heterologous cleavage domain.

[0064] As used herein, the term “transformation” encompasses all techniques that a nucleic acid molecule can be introduced into such a cell. Examples include, but are not limited to: transfection with viral vectors; transformation with plasmid vectors; electroporation; lipofection; microinjection (Mueller *et al.*, (1978) Cell 15:579-85); *Agrobacterium*-mediated transfer; direct DNA uptake; WHISKERS™-mediated transformation; and microprojectile bombardment. These techniques may be used for both stable transformation and transient transformation of a plant cell. “Stable transformation” refers to the introduction of a nucleic acid fragment into a genome of a host organism resulting in genetically stable inheritance. Once stably transformed, the nucleic acid fragment is stably integrated in the genome of the host organism and any subsequent generation. Host organisms containing the transformed nucleic acid fragments are referred to as “transgenic” organisms. “Transient transformation” refers to the introduction of a nucleic acid fragment into the nucleus, or DNA-containing organelle, of a host organism resulting in gene expression without genetically stable inheritance.

[0065] An exogenous nucleic acid sequence. In one example, a transgene is a gene sequence (*e.g.*, an herbicide-resistance gene), a gene encoding an industrially or pharmaceutically useful compound, or a gene encoding a desirable agricultural trait. In yet another example, the transgene is an antisense nucleic acid sequence, wherein expression of the antisense nucleic acid sequence inhibits expression of a target nucleic acid sequence. A transgene may contain regulatory sequences operably linked to the transgene (*e.g.*, a promoter). In some embodiments, a polynucleotide sequence of interest is a transgene. However, in other embodiments, a polynucleotide sequence of interest is an endogenous nucleic acid sequence, wherein additional genomic copies of the endogenous nucleic acid sequence are desired, or a nucleic acid sequence that is in the antisense orientation with respect to the sequence of a target nucleic acid molecule in the host organism.

[0066] As used herein, the term a transgenic “event” is produced by transformation of plant cells with heterologous DNA, *i.e.*, a nucleic acid construct that includes a transgene of interest, regeneration of a population of plants resulting from the insertion of the transgene into the genome of the plant, and selection of a particular plant characterized by insertion into a particular genome location. The term “event” refers to the original transformant and progeny of the transformant that

include the heterologous DNA. The term “event” also refers to progeny produced by a sexual outcross between the transformant and another variety that includes the genomic/transgene DNA. Even after repeated back-crossing to a recurrent parent, the inserted transgene DNA and flanking genomic DNA (genomic/transgene DNA) from the transformed parent is present in the progeny of the cross at the same chromosomal location. The term “event” also refers to DNA from the original transformant and progeny thereof comprising the inserted DNA and flanking genomic sequence immediately adjacent to the inserted DNA that would be expected to be transferred to a progeny that receives inserted DNA including the transgene of interest as the result of a sexual cross of one parental line that includes the inserted DNA (e.g., the original transformant and progeny resulting from selfing) and a parental line that does not contain the inserted DNA.

[0067] As used herein, the terms “Polymerase Chain Reaction” or “PCR” define a procedure or technique in which minute amounts of nucleic acid, RNA and/or DNA, are amplified as described in U.S. Pat. No. 4,683,195 issued July 28, 1987. Generally, sequence information from the ends of the region of interest or beyond needs to be available, such that oligonucleotide primers can be designed; these primers will be identical or similar in sequence to opposite strands of the template to be amplified. The 5’ terminal nucleotides of the two primers may coincide with the ends of the amplified material. PCR can be used to amplify specific RNA sequences, specific DNA sequences from total genomic DNA, and cDNA transcribed from total cellular RNA, bacteriophage or plasmid sequences, etc. See generally Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, 51:263 (1987); Erlich, ed., *PCR Technology*, (Stockton Press, NY, 1989).

[0068] As used herein, the term “primer” refers to an oligonucleotide capable of acting as a point of initiation of synthesis along a complementary strand when conditions are suitable for synthesis of a primer extension product. The synthesizing conditions include the presence of four different deoxyribonucleotide triphosphates and at least one polymerization-inducing agent such as reverse transcriptase or DNA polymerase. These are present in a suitable buffer, which may include constituents which are co-factors or which affect conditions such as pH and the like at various suitable temperatures. A primer is preferably a single strand sequence, such that amplification efficiency is optimized, but double stranded sequences can be utilized.

[0069] As used herein, the term “probe” refers to an oligonucleotide that hybridizes to a target sequence. In the TaqMan® or TaqMan®-style assay procedure, the probe hybridizes to a portion of the target situated between the annealing site of the two primers. A probe includes about eight nucleotides, about ten nucleotides, about fifteen nucleotides, about twenty nucleotides, about thirty nucleotides, about forty nucleotides, or about fifty nucleotides. In some embodiments, a probe includes from about eight nucleotides to about fifteen nucleotides. A probe can further include a detectable label, e.g., a fluorophore (Texas-Red®, Fluorescein isothiocyanate, etc.). The detectable label can be covalently attached directly to the probe oligonucleotide, e.g., located at the probe’s 5’ end or at the probe’s 3’ end. A probe including a fluorophore may also further include a quencher, e.g., Black Hole Quencher™, Iowa Black™, etc.

[0070] As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence. Type -2 restriction enzymes recognize and cleave DNA at the same site, and include but are not limited to XbaI, BamHI, HindIII, EcoRI, XhoI, SalI, KpnI, AvaI, PstI and SmaI.

[0071] As used herein, the term "vector" is used interchangeably with the terms "construct", "cloning vector" and "expression vector" and means the vehicle by which a DNA or RNA sequence (e.g. a foreign gene) can be introduced into a host cell, so as to transform the host and promote expression (e.g. transcription and translation) of the introduced sequence. A "non-viral vector" is intended to mean any vector that does not comprise a virus or retrovirus. In some embodiments a "vector" is a sequence of DNA comprising at least one origin of DNA replication and at least one selectable marker gene. Examples include, but are not limited to, a plasmid, cosmid, bacteriophage, bacterial artificial chromosome (BAC), or virus that carries exogenous DNA into a cell. A vector can also include one or more genes, antisense molecules, and/or selectable marker genes and other genetic elements known in the art. A vector may transduce, transform, or infect a cell, thereby causing the cell to express the nucleic acid molecules and/or proteins encoded by the vector. The term "plasmid" defines a circular strand of nucleic acid capable of autosomal replication in either a prokaryotic or a eukaryotic host cell. The term includes nucleic acid which may be either DNA or RNA and may be single- or double-stranded. The plasmid of the definition may also include the sequences which

correspond to a bacterial origin of replication.

[0072] As used herein, the term "selectable marker gene" as used herein defines a gene or other expression cassette which encodes a protein which facilitates identification of cells into which the selectable marker gene is inserted. For example a "selectable marker gene" encompasses reporter genes as well as genes used in plant transformation to, for example, protect plant cells from a selective agent or provide resistance/tolerance to a selective agent. In one embodiment only those cells or plants that receive a functional selectable marker are capable of dividing or growing under conditions having a selective agent. Examples of selective agents can include, for example, antibiotics, including spectinomycin, neomycin, kanamycin, paromomycin, gentamicin, and hygromycin. These selectable markers include neomycin phosphotransferase (npt II), which expresses an enzyme conferring resistance to the antibiotic kanamycin, and genes for the related antibiotics neomycin, paromomycin, gentamicin, and G418, or the gene for hygromycin phosphotransferase (hpt), which expresses an enzyme conferring resistance to hygromycin. Other selectable marker genes can include genes encoding herbicide resistance including bar or pat (resistance against glufosinate ammonium or phosphinothricin), acetolactate synthase (ALS, resistance against inhibitors such as sulfonylureas (SUs), imidazolinones (IMIs), triazolopyrimidines (TPs), pyrimidinyl oxybenzoates (POBs), and sulfonylamino carbonyl triazolinones that prevent the first step in the synthesis of the branched-chain amino acids), glyphosate, 2,4-D, and metal resistance or sensitivity. Examples of "reporter genes" that can be used as a selectable marker gene include the visual observation of expressed reporter gene proteins such as proteins encoding β -glucuronidase (GUS), luciferase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), DsRed, β -galactosidase, chloramphenicol acetyltransferase (CAT), alkaline phosphatase, and the like. The phrase "marker-positive" refers to plants that have been transformed to include a selectable marker gene.

[0073] As used herein, the term "detectable marker" refers to a label capable of detection, such as, for example, a radioisotope, fluorescent compound, bioluminescent compound, a chemiluminescent compound, metal chelator, or enzyme. Examples of detectable markers include, but are not limited to, the following: fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), enzymatic labels (e.g., horseradish peroxidase, β -galactosidase, luciferase, alkaline phosphatase),

chemiluminescent, biotinyl groups, predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). In an embodiment, a detectable marker can be attached by spacer arms of various lengths to reduce potential steric hindrance.

[0074] As used herein, the terms "cassette", "expression cassette" and "gene expression cassette" refer to a segment of DNA that can be inserted into a nucleic acid or polynucleotide at specific restriction sites or by homologous recombination. As used herein the segment of DNA comprises a polynucleotide that encodes a polypeptide of interest, and the cassette and restriction sites are designed to ensure insertion of the cassette in the proper reading frame for transcription and translation. In an embodiment, an expression cassette can include a polynucleotide that encodes a polypeptide of interest and having elements in addition to the polynucleotide that facilitate transformation of a particular host cell. In an embodiment, a gene expression cassette may also include elements that allow for enhanced expression of a polynucleotide encoding a polypeptide of interest in a host cell. These elements may include, but are not limited to: a promoter, a minimal promoter, an enhancer, a response element, a terminator sequence, a polyadenylation sequence, and the like.

[0075] As used herein a "linker" or "spacer" is a bond, molecule or group of molecules that binds two separate entities to one another. Linkers and spacers may provide for optimal spacing of the two entities or may further supply a labile linkage that allows the two entities to be separated from each other. Labile linkages include photocleavable groups, acid-labile moieties, base-labile moieties and enzyme-cleavable groups. The terms "polylinker" or "multiple cloning site" as used herein defines a cluster of three or more Type -2 restriction enzyme sites located within 10 nucleotides of one another on a nucleic acid sequence. Constructs comprising a polylinker are utilized for the insertion and/or excision of nucleic acid sequences such as the coding region of a gene. In other instances the term "polylinker" as used herein refers to a stretch of nucleotides that are targeted for joining two sequences via any known seamless cloning method (i.e., Gibson Assembly®, NEBuilder HiFiDNA Assembly®, Golden Gate Assembly, BioBrick® Assembly, etc.). Constructs comprising a polylinker are utilized for the insertion and/or excision of nucleic acid

sequences such as the coding region of a gene.

[0076] As used herein, the term “control” refers to a sample used in an analytical procedure for comparison purposes. A control can be “positive” or “negative”. For example, where the purpose of an analytical procedure is to detect a differentially expressed transcript or polypeptide in cells or tissue, it is generally preferable to include a positive control, such as a sample from a known plant exhibiting the desired expression, and a negative control, such as a sample from a known plant lacking the desired expression.

[0077] As used herein, the term “plant” includes a whole plant and any descendant, cell, tissue, or part of a plant. A class of plant that can be used in the present invention is generally as broad as the class of higher and lower plants amenable to mutagenesis including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns and multicellular algae. Thus, “plant” includes dicot and monocot plants. The term “plant parts” include any part(s) of a plant, including, for example and without limitation: seed (including mature seed and immature seed); a plant cutting; a plant cell; a plant cell culture; a plant organ (e.g., pollen, embryos, flowers, fruits, shoots, leaves, roots, stems, and explants). A plant tissue or plant organ may be a seed, protoplast, callus, or any other group of plant cells that is organized into a structural or functional unit. A plant cell or tissue culture may be capable of regenerating a plant having the physiological and morphological characteristics of the plant from which the cell or tissue was obtained, and of regenerating a plant having substantially the same genotype as the plant. In contrast, some plant cells are not capable of being regenerated to produce plants. Regenerable cells in a plant cell or tissue culture may be embryos, protoplasts, meristematic cells, callus, pollen, leaves, anthers, roots, root tips, silk, flowers, kernels, ears, cobs, husks, or stalks.

[0078] Plant parts include harvestable parts and parts useful for propagation of progeny plants. Plant parts useful for propagation include, for example and without limitation: seed; fruit; a cutting; a seedling; a tuber; and a rootstock. A harvestable part of a plant may be any useful part of a plant, including, for example and without limitation: flower; pollen; seedling; tuber; leaf; stem; fruit; seed; and root.

[0079] A plant cell is the structural and physiological unit of the plant, comprising a

protoplast and a cell wall. A plant cell may be in the form of an isolated single cell, or an aggregate of cells (e.g., a friable callus and a cultured cell), and may be part of a higher organized unit (e.g., a plant tissue, plant organ, and plant). Thus, a plant cell may be a protoplast, a gamete producing cell, or a cell or collection of cells that can regenerate into a whole plant. As such, a seed, which comprises multiple plant cells and is capable of regenerating into a whole plant, is considered a “plant cell” in embodiments herein.

[0080] As used herein, the term “small RNA” refers to several classes of non-coding ribonucleic acid (ncRNA). The term *small RNA* describes the short chains of ncRNA produced in bacterial cells, animals, plants, and fungi. These short chains of ncRNA may be produced naturally within the cell or may be produced by the introduction of an exogenous sequence that expresses the short chain or ncRNA. The small RNA sequences do not directly code for a protein, and differ in function from other RNA in that small RNA sequences are only transcribed and not translated. The small RNA sequences are involved in other cellular functions, including gene expression and modification. Small RNA molecules are usually made up of about 20 to 30 nucleotides. The small RNA sequences may be derived from longer precursors. The precursors form structures that fold back on each other in self-complementary regions; they are then processed by the nuclease Dicer in animals or DCL1 in plants.

[0081] Many types of small RNA exist either naturally or produced artificially, including microRNAs (miRNAs), short interfering RNAs (siRNAs), antisense RNA, short hairpin RNA (shRNA), and small nucleolar RNAs (snoRNAs). Certain types of small RNA, such as microRNA and siRNA, are important in gene silencing and RNA interference (RNAi). Gene silencing is a process of genetic regulation in which a gene that would normally be expressed is “turned off” by an intracellular element, in this case, the small RNA. The protein that would normally be formed by this genetic information is not formed due to interference, and the information coded in the gene is blocked from expression.

[0082] As used herein, the term “small RNA” encompasses RNA molecules described in the literature as “tiny RNA” (Storz, (2002) *Science* 296:1260-3; Illangasekare *et al.*, (1999) *RNA* 5:1482-1489); prokaryotic “small RNA” (sRNA) (Wassarman *et al.*, (1999) *Trends Microbiol.*

7:37-45); eukaryotic “noncoding RNA (ncRNA)”;

“micro-RNA (miRNA)”;

“small non-mRNA (snmRNA)”;

“functional RNA (fRNA)”;

“transfer RNA (tRNA)”;

“catalytic RNA” [e.g., ribozymes, including self-acylating ribozymes (Illangaskare *et al.*, (1999) *RNA* 5:1482-1489); “small nucleolar RNAs (snoRNAs),” “tmRNA” (a.k.a. “10S RNA,” Muto *et al.*, (1998) *Trends Biochem Sci.* 23:25-29; and Gillet *et al.*, (2001) *Mol Microbiol.* 42:879-885); RNAi molecules including without limitation “small interfering RNA (siRNA),” “endoribonuclease-prepared siRNA (e-siRNA),” “short hairpin RNA (shRNA),” and “small temporally regulated RNA (stRNA),” “diced siRNA (d-siRNA),” and aptamers, oligonucleotides and other synthetic nucleic acids that comprise at least one uracil base.

[0083] Unless otherwise specifically explained, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which this disclosure belongs. Definitions of common terms in molecular biology can be found in, for example: Lewin, *Genes V*, Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Meyers (ed.), *Molecular Biology and Biotechnology: A Comprehensive Desk Reference*, VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

[0084] As used herein, the articles, “a,” “an,” and “the” include plural references unless the context clearly and unambiguously dictates otherwise.

III. *Zea mays* GRMZM2G144030 Promoter and Nucleic Acids Comprising the Same

[0085] Provided are methods and compositions for using a promoter and other regulatory elements from a *Zea mays* GRMZM2G144030 gene to express non transgenes in plants. In an embodiment, a promoter can be the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. In a further embodiment, a 3'UTR can be the *Zea mays* GRMZM2G144030 3'UTR of SEQ ID NO:5.

[0086] In an embodiment, a polynucleotide is provided comprising a promoter, wherein the promoter is at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, or 100% identical to SEQ ID NO:1. In an embodiment, a promoter is a *Zea mays* GRMZM2G144030 promoter comprising a polynucleotide of at least 80%, 85%, 90%, 91%, 92%,

93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, or 100% identity to the polynucleotide of SEQ ID NO:1. In an embodiment, an isolated polynucleotide is provided comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, or 100% identity to the polynucleotide of SEQ ID NO:1. In an embodiment, a nucleic acid vector is provided comprising a *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. In an embodiment, a polynucleotide is provided comprising a *Zea mays* GRMZM2G144030 promoter that is operably linked to a polylinker. In an embodiment, a gene expression cassette is provided comprising a *Zea mays* GRMZM2G144030 promoter that is operably linked to a non-GRMZM2G144030 transgene. In an embodiment, a nucleic acid vector is provided comprising a *Zea mays* GRMZM2G144030 promoter that is operably linked to a non-GRMZM2G144030 transgene. In one embodiment, the promoter consists of SEQ ID NO:1. In an illustrative embodiment, a nucleic acid vector comprises a *Zea mays* GRMZM2G144030 promoter that is operably linked to a transgene, wherein the transgene can be an insecticidal resistance transgene, an herbicide tolerance transgene, a nitrogen use efficiency transgene, a water use efficiency transgene, a nutritional quality transgene, a DNA binding transgene, a small RNA transgene, selectable marker transgene, or combinations thereof.

[0087] Transgene expression may also be regulated by the 3' untranslated gene region (i.e., 3' UTR) located downstream of the gene's coding sequence. Both a promoter and a 3' UTR can regulate transgene expression. While a promoter is necessary to drive transcription, a 3' UTR gene region can terminate transcription and initiate polyadenylation of a resulting mRNA transcript for translation and protein synthesis. A 3' UTR gene region aids stable expression of a transgene.

[0088] In an embodiment, a nucleic acid vector is provided comprising a *Zea mays* GRMZM2G144030 promoter as described herein and a 3' UTR. In an embodiment, the nucleic acid vector comprises a *Zea mays* GRMZM2G144030 3' UTR. In an embodiment, the *Zea mays* GRMZM2G144030 3' UTR is SEQ ID NO:5.

[0089] In an embodiment, a nucleic acid vector is provided comprising a *Zea mays* GRMZM2G144030 promoter as described herein and a 3' UTR, wherein the 3' UTR is at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, or 100% identical to the polynucleotide of SEQ ID NO:5. In an embodiment, a nucleic acid vector is provided comprising

a *Zea mays* GRMZM2G144030 promoter as described herein and the *Zea mays* GRMZM2G144030 3' UTR wherein the *Zea mays* GRMZM2G144030 promoter and the *Zea mays* GRMZM2G144030 3' UTR are both operably linked to opposite ends of a polylinker. In an embodiment, a gene expression cassette is provided comprising a *Zea mays* GRMZM2G144030 promoter as described herein and a *Zea mays* GRMZM2G144030 3' UTR, wherein the *Zea mays* GRMZM2G144030 promoter and the *Zea mays* GRMZM2G144030 3' UTR are both operably linked to opposite ends of a non-GRMZM2G144030 transgene. In one embodiment the 3' UTR, consists of SEQ ID NO:5. In one embodiment, a gene expression cassette is provided comprising a *Zea mays* GRMZM2G144030 promoter as described herein and a *Zea mays* GRMZM2G144030 3' UTR, wherein the *Zea mays* GRMZM2G144030 promoter comprises SEQ ID NO: 1 and the *Zea mays* GRMZM2G144030 3' UTR comprises SEQ ID NO: 5 wherein the promoter and 3' UTR are operably linked to opposite ends of a non-GRMZM2G144030 transgene. In an aspect of this embodiment the 3' UTR, consists of SEQ ID NO:5. In another aspect of this embodiment the promoter consists of SEQ ID NO:1. In an illustrative embodiment, a gene expression cassette comprises a *Zea mays* GRMZM2G144030 3' UTR that is operably linked to a transgene, wherein the transgene can be an insecticidal resistance transgene, an herbicide tolerance transgene, a nitrogen use efficiency transgene, a water use efficiency transgene, a nutritional quality transgene, a DNA binding transgene or protein, a small RNA transgene, a selectable marker transgene, or combinations thereof. In a further embodiment the transgene is operably linked to a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 3' UTR from the same GRMZM2G144030-like gene.

[0090] Transgene expression may also be regulated by an intron region located downstream of the promoter sequence. Both a promoter and an intron can regulate transgene expression. While a promoter is necessary to drive transcription, the presence of an intron can increase expression levels resulting in mRNA transcript for translation and protein synthesis. An intron gene region aids stable expression of a transgene. In a further embodiment an intron is operably linked to a *Zea mays* GRMZM2G144030 promoter.

[0091] Transgene expression may also be regulated by a 5' UTR region located downstream of the promoter sequence. Both a promoter and a 5' UTR can regulate transgene expression. While

a promoter is necessary to drive transcription, the presence of a 5' UTR can increase expression levels resulting in mRNA transcript for translation and protein synthesis. A 5' UTR gene region aids stable expression of a transgene. In a further embodiment an 5' UTR is operably linked to a *Zea mays* GRMZM2G144030 promoter.

[0092] A *Zea mays* GRMZM2G144030 promoter may also comprise one or more additional sequence elements. In some embodiments, a *Zea mays* GRMZM2G144030 promoter may comprise an exon (*e.g.*, a leader or signal peptide such as a chloroplast transit peptide or ER retention signal). For example and without limitation, a *Zea mays* GRMZM2G144030 promoter may encode an exon incorporated into the *Zea mays* GRMZM2G144030 promoter as a further embodiment.

[0093] In an embodiment, a nucleic acid vector comprises a gene expression cassette as disclosed herein. In an embodiment, a vector can be a plasmid, a cosmid, a bacterial artificial chromosome (BAC), a bacteriophage, a virus, or an excised polynucleotide fragment for use in direct transformation or gene targeting such as a donor DNA.

[0094] In accordance with one embodiment a nucleic acid vector is provided comprising a recombinant gene expression cassette wherein the recombinant gene expression cassette comprises a *Zea mays* GRMZM2G144030 promoter operably linked to a polylinker sequence, a non-GRMZM2G144030 transgene or combination thereof. In one embodiment the recombinant gene cassette comprises a *Zea mays* GRMZM2G144030 promoter operably linked to a non-GRMZM2G144030 transgene. In one embodiment the recombinant gene cassette comprises a *Zea mays* GRMZM2G144030 promoter as disclosed herein is operably linked to a polylinker sequence. The polylinker is operably linked to the *Zea mays* GRMZM2G144030 promoter in a manner such that insertion of a coding sequence into one of the restriction sites of the polylinker will operably link the coding sequence allowing for expression of the coding sequence when the vector is transformed or transfected into a host cell.

[0095] In accordance with one embodiment the *Zea mays* GRMZM2G144030 promoter comprises SEQ ID NO: 1 or a sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In accordance with one embodiment the promoter sequence has a total length of no more than 1.5, 2, 2.5, 3 or 4 kb. In accordance with one embodiment the *Zea mays*

GRMZM2G144030 promoter consists of SEQ ID NO: 1 or a 2,082 bp sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1.

[0096] In accordance with one embodiment a nucleic acid vector is provided comprising a gene cassette that consists of a *Zea mays* GRMZM2G144030 promoter, a non-GRMZM2G144030 transgene and a *Zea mays* GRMZM2G144030 3' UTR of SEQ ID NO:5. In an embodiment, the *Zea mays* GRMZM2G144030 3' UTR of SEQ ID NO:5 is operably linked to the 3' end of the non-GRMZM2G144030 transgene. In a further embodiment the 3' untranslated sequence comprises SEQ ID NO:5 or a sequence that has at least 80, 85, 90, 95, 99 or 100% sequence identity with SEQ ID NO:5. In accordance with one embodiment a nucleic acid vector is provided comprising a gene cassette that consists of SEQ ID NO: 1, or a 2,082 bp sequence that has at least 80, 85, 90, 95, or 99% sequence identity with SEQ ID NO: 1, a non-GRMZM2G144030 transgene and a *Zea mays* GRMZM2G144030 3' UTR, wherein SEQ ID NO: 1 is operably linked to the 5' end of the non-GRMZM2G144030 transgene and the 3' UTR of SEQ ID NO:5 is operably linked to the 3' end of the non-GRMZM2G144030 transgene. In a further embodiment the 3' untranslated sequence comprises SEQ ID NO:5 or a sequence that has at least 80, 85, 90, 95, 99 or 100% sequence identity with SEQ ID NO:5. In a further embodiment the *Zea mays* GRMZM2G144030 3' untranslated sequence consists of SEQ ID NO:5, or a 1,095 bp sequence that has at least 80, 85, 90, 95, or 99% sequence identity with SEQ ID NO:5.

[0097] In one embodiment a nucleic acid construct is provided comprising a promoter and a non-GRMZM2G144030 transgene and optionally one or more of the following elements:

- a) a 5' untranslated region;
- b) an intron; and
- c) a 3' untranslated region,

wherein,

the promoter consists of SEQ ID NO:1 or a sequence having at least 98% sequence identity with SEQ ID NO:1; and

the 3' untranslated region consists of SEQ ID NO:5 or a sequence having at least 98% sequence identity with SEQ ID NO:5; further wherein said promoter is operably linked to said

transgene and each optional element, when present, is also operably linked to both the promoter and the transgene. In a further embodiment a transgenic cell is provided comprising the nucleic acid construct disclosed immediately above. In one embodiment the transgenic cell is a plant cell, and in a further embodiment a plant is provided wherein the plant comprises said transgenic cells.

[0098] In one embodiment a nucleic acid construct is provided comprising a promoter and a non-GRMZM2G144030 transgene and optionally one or more of the following elements:

- a) a 5' untranslated region; and
- b) a 3' untranslated region,

wherein,

the promoter consists of SEQ ID NO:1 or a sequence having at least 98% sequence identity with SEQ ID NO:1; and

the 3' untranslated region consists of SEQ ID NO:5 or a sequence having at least 98% sequence identity with SEQ ID NO:5; further wherein said promoter is operably linked to said transgene and each optional element, when present, is also operably linked to both the promoter and the transgene. In a further embodiment a transgenic cell is provided comprising the nucleic acid construct disclosed immediately above. In one embodiment the transgenic cell is a plant cell, and in a further embodiment a plant is provided wherein the plant comprises said transgenic cells.

[0099] In one embodiment a nucleic acid construct is provided comprising a promoter and a polylinker and optionally one or more of the following elements:

- a) a 5' untranslated region;
- b) an intron; and
- c) a 3' untranslated region,

wherein,

the promoter consists of SEQ ID NO:1 or a sequence having at least 98% sequence identity with SEQ ID NO:1; and

the 3' untranslated region consists of SEQ ID NO:5 or a sequence having at least 98% sequence identity with SEQ ID NO:5; further wherein said promoter is operably linked to said polylinker and each optional element, when present, is also operably linked to both the promoter and the polylinker.

[00100] In accordance with one embodiment the nucleic acid vector further comprises a sequence encoding a selectable marker. In accordance with one embodiment the recombinant gene cassette is operably linked to an *Agrobacterium* T-DNA border. In accordance with one embodiment the recombinant gene cassette further comprises a first and second T-DNA border, wherein the first T-DNA border is operably linked to one end of the gene construct, and the second T-DNA border is operably linked to the other end of the gene construct. The first and second *Agrobacterium* T-DNA borders can be independently selected from T-DNA border sequences originating from bacterial strains selected from the group consisting of a nopaline synthesizing *Agrobacterium* T-DNA border, an octopine synthesizing *Agrobacterium* T-DNA border, a mannopine synthesizing *Agrobacterium* T-DNA border, a succinamopine synthesizing *Agrobacterium* T-DNA border, or any combination thereof. In one embodiment an *Agrobacterium* strain selected from the group consisting of a nopaline synthesizing strain, a mannopine synthesizing strain, a succinamopine synthesizing strain, or an octopine synthesizing strain is provided, wherein said strain comprises a plasmid wherein the plasmid comprises a transgene operably linked to a sequence selected from SEQ ID NO: 1 or a sequence having at least 80, 85, 90, 95, or 99% sequence identity with SEQ ID NO: 1.

[00101] Transgenes of interest that are suitable for use in the present disclosed constructs include, but are not limited to, coding sequences that confer (1) resistance to pests or disease, (2) tolerance to herbicides, (3) value added agronomic traits, such as; yield improvement, nitrogen use efficiency, water use efficiency, and nutritional quality, (4) binding of a protein to DNA in a site specific manner, (5) expression of small RNA, and (6) selectable markers. In accordance with one embodiment, the transgene encodes a selectable marker or a gene product conferring insecticidal resistance, herbicide tolerance, small RNA expression, nitrogen use efficiency, water use efficiency, or nutritional quality.

1. Insect Resistance

[00102] Various insect resistance coding sequences can be operably linked to the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. In an embodiment, a promoter can be the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1, or a sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In some embodiments, the sequences are operably

linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1 and a 5' UTR. The operably linked sequences can then be incorporated into a chosen vector to allow for identification and selection of transformed plants ("transformants"). Exemplary insect resistance coding sequences are known in the art. As embodiments of insect resistance coding sequences that can be operably linked to the regulatory elements of the subject disclosure, the following traits are provided. Coding sequences that provide exemplary Lepidopteran insect resistance include: *cryIA*; *cryIA.105*; *cryIAb*; *cryIAb*(truncated); *cryIAb-Ac* (fusion protein); *cryIAc* (marketed as Widestrike®); *cryIC*; *cryIF* (marketed as Widestrike®); *cryIFa2*; *cry2Ab2*; *cry2Ae*; *cry9C*; *mcryIF*; *pinII* (protease inhibitor protein); *vip3A(a)*; and *vip3Aa20*. Coding sequences that provide exemplary Coleopteran insect resistance include: *cry34Ab1* (marketed as Herculex®); *cry35Ab1* (marketed as Herculex®); *cry3A*; *cry3Bb1*; *dvsnf7*; and *mcry3A*. Coding sequences that provide exemplary multi-insect resistance include *ecry31.Ab*. The above list of insect resistance genes is not meant to be limiting. Any insect resistance genes are encompassed by the present disclosure.

2. Herbicide Tolerance

[00103] Various herbicide tolerance coding sequences can be operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1, or a sequence that has 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In an embodiment, a promoter can be the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. promoter comprising SEQ ID NO: 1, or a sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In some embodiments, the sequences are operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1 and a 5' UTR. The operably linked sequences can then be incorporated into a chosen vector to allow for identification and selection of transformed plants ("transformants"). Exemplary herbicide tolerance coding sequences are known in the art. As embodiments of herbicide tolerance coding sequences that can be operably linked to the regulatory elements of the subject disclosure, the following traits are provided. The glyphosate herbicide contains a mode of action by inhibiting the EPSPS enzyme (5-enolpyruvylshikimate-3-phosphate synthase). This enzyme is involved in the biosynthesis of aromatic amino acids that are essential for growth and development of plants. Various enzymatic mechanisms are known in the art that can be utilized to inhibit this

enzyme. The genes that encode such enzymes can be operably linked to the gene regulatory elements of the subject disclosure. In an embodiment, selectable marker genes include, but are not limited to genes encoding glyphosate resistance genes include: mutant EPSPS genes such as *2mEPSPS* genes, *cp4 EPSPS* genes, *mEPSPS* genes, *dgt-28* genes; *aroA* genes; and glyphosate degradation genes such as glyphosate acetyl transferase genes (*gat*) and glyphosate oxidase genes (*gox*). These traits are currently marketed as Gly-TolTM, Optimum® GAT®, Agrisure® GT and Roundup Ready®. Resistance genes for glufosinate and/or bialaphos compounds include *dsm-2*, *bar* and *pat* genes. The *bar* and *pat* traits are currently marketed as LibertyLink®. Also included are tolerance genes that provide resistance to 2,4-D such as *aad-1* genes (it should be noted that *aad-1* genes have further activity on arloxyphenoxypropionate herbicides) and *aad-12* genes (it should be noted that *aad-12* genes have further activity on pyridyloxyacetate synthetic auxins). These traits are marketed as Enlist® crop protection technology. Resistance genes for ALS inhibitors (sulfonylureas, imidazolinones, triazolopyrimidines, pyrimidinylthiobenzoates, and sulfonylamino-carbonyl-triazolinones) are known in the art. These resistance genes most commonly result from point mutations to the ALS encoding gene sequence. Other ALS inhibitor resistance genes include *hra* genes, the *csr1-2* genes, *Sr-HrA* genes, and *surB* genes. Some of the traits are marketed under the tradename Clearfield®. Herbicides that inhibit HPPD include the pyrazolones such as pyrazoxyfen, benzofenap, and topramezone; triketones such as mesotrione, sulcotrione, tembotrione, benzobicyclon; and diketonitriles such as isoxaflutole. These exemplary HPPD herbicides can be tolerated by known traits. Examples of HPPD inhibitors include *hppdPF_W336* genes (for resistance to isoxaflutole) and *avhppd-03* genes (for resistance to meotrione). An example of oxynil herbicide tolerant traits include the *bxn* gene, which has been showed to impart resistance to the herbicide/antibiotic bromoxynil. Resistance genes for dicamba include the dicamba monooxygenase gene (*dmo*) as disclosed in International PCT Publication No. WO 2008/105890. Resistance genes for PPO or PROTOX inhibitor type herbicides (e.g., acifluorfen, butafenacil, flupropazil, pentoxazone, carfentrazone, fluazolate, pyraflufen, aclonifen, azafenidin, flumioxazin, flumiclorac, bifenox, oxyfluorfen, lactofen, fomesafen, fluoroglycofen, and sulfentrazone) are known in the art. Exemplary genes conferring resistance to PPO include over expression of a wild-

type *Arabidopsis thaliana* PPO enzyme (Lermontova I and Grimm B, (2000) Overexpression of plastidic protoporphyrinogen IX oxidase leads to resistance to the diphenyl-ether herbicide acifluorfen. *Plant Physiol* **122**:75–83.), the *B. subtilis* PPO gene (Li, X. and Nicholl D. 2005. Development of PPO inhibitor-resistant cultures and crops. *Pest Manag. Sci.* 61:277-285 and Choi KW, Han O, Lee HJ, Yun YC, Moon YH, Kim MK, Kuk YI, Han SU and Guh JO, (1998) Generation of resistance to the diphenyl ether herbicide, oxyfluorfen, via expression of the *Bacillus subtilis* protoporphyrinogen oxidase gene in transgenic tobacco plants. *Biosci Biotechnol Biochem* **62**:558–560.) Resistance genes for pyridinoxy or phenoxy propionic acids and cyclohexones include the ACCase inhibitor-encoding genes (e.g., Acc1-S1, Acc1-S2 and Acc1-S3). Exemplary genes conferring resistance to cyclohexanediones and/or aryloxyphenoxypropanoic acid include haloxyfop, diclofop, fenoxypop, fluazifop, and quizalofop. Finally, herbicides can inhibit photosynthesis, including triazine or benzonitrile are provided tolerance by *psbA* genes (tolerance to triazine), *Is+* genes (tolerance to triazine), and *nitrilase* genes (tolerance to benzonitrile). The above list of herbicide tolerance genes is not meant to be limiting. Any herbicide tolerance genes are encompassed by the present disclosure.

3. Agronomic Traits

[00104] Various agronomic trait coding sequences can be operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1, or a sequence that has 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In an embodiment, a promoter can be the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. promoter comprising SEQ ID NO: 1, or a sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In some embodiments, the sequences are operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1 and a 5' UTR. The operably linked sequences can then be incorporated into a chosen vector to allow for identification and selection of transformed plants ("transformants"). Exemplary agronomic trait coding sequences are known in the art. As embodiments of agronomic trait coding sequences that can be operably linked to the regulatory elements of the subject disclosure, the following traits are provided. Delayed fruit softening as provided by the *pg* genes inhibit the production of polygalacturonase enzyme responsible for the breakdown of pectin molecules in the

cell wall, and thus causes delayed softening of the fruit. Further, delayed fruit ripening/senescence of *acc* genes act to suppress the normal expression of the native *acc synthase* gene, resulting in reduced ethylene production and delayed fruit ripening. Whereas, the *accd* genes metabolize the precursor of the fruit ripening hormone ethylene, resulting in delayed fruit ripening. Alternatively, the *sam-k* genes cause delayed ripening by reducing S-adenosylmethionine (SAM), a substrate for ethylene production. Drought stress tolerance phenotypes as provided by *cspB* genes maintain normal cellular functions under water stress conditions by preserving RNA stability and translation. Another example includes the *EcBetA* genes that catalyze the production of the osmoprotectant compound glycine betaine conferring tolerance to water stress. In addition, the *RmBetA* genes catalyze the production of the osmoprotectant compound glycine betaine conferring tolerance to water stress. Photosynthesis and yield enhancement is provided with the *bbx32* gene that expresses a protein that interacts with one or more endogenous transcription factors to regulate the plant's day/night physiological processes. Ethanol production can be increase by expression of the *amy797E* genes that encode a thermostable alpha-amylase enzyme that enhances bioethanol production by increasing the thermostability of amylase used in degrading starch. Finally, modified amino acid compositions can result by the expression of the *cordapA* genes that encode a dihydrodipicolinate synthase enzyme that increases the production of amino acid lysine. The above list of agronomic trait coding sequences is not meant to be limiting. Any agronomic trait coding sequence is encompassed by the present disclosure.

4. DNA Binding Proteins

[00105] Various DNA binding protein coding sequences can be operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1, or a sequence that has 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In an embodiment, a promoter can be the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. promoter comprising SEQ ID NO: 1, or a sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In some embodiments, the sequences are operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1 and a 5' UTR. The operably linked sequences can then be incorporated into a chosen vector to allow for identification and selectable of transformed plants ("transformants"). Exemplary DNA

binding protein coding sequences are known in the art. As embodiments of DNA binding protein coding sequences that can be operably linked to the regulatory elements of the subject disclosure, the following types of DNA binding proteins can include; Zinc Fingers, Talens, CRISPRs, and meganucleases. The above list of DNA binding protein coding sequences is not meant to be limiting. Any DNA binding protein coding sequences is encompassed by the present disclosure.

5. Small RNA

[00106] Various small RNAs can be operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1, or a sequence that has 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In an embodiment, a promoter can be the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. promoter comprising SEQ ID NO: 1, or a sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In some embodiments, the sequences are operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1 and a 5' UTR. The operably linked sequences can then be incorporated into a chosen vector to allow for identification and selection of transformed plants ("transformants"). Exemplary small RNA traits are known in the art. As embodiments of small RNA coding sequences that can be operably linked to the regulatory elements of the subject disclosure, the following traits are provided. For example, delayed fruit ripening/senescence of the *anti-efe* small RNA delays ripening by suppressing the production of ethylene via silencing of the ACO gene that encodes an ethylene-forming enzyme. The altered lignin production of *ccomt* small RNA reduces content of guaiacyl (G) lignin by inhibition of the endogenous S-adenosyl-L-methionine: trans-caffeoyl CoA 3-O-methyltransferase (CCOMT gene). Further, the Black Spot Bruise Tolerance in *Solanum verrucosum* can be reduced by the *Ppo5* small RNA which triggers the degradation of *Ppo5* transcripts to block black spot bruise development. Also included is the *dvsnf7* small RNA that inhibits Western Corn Rootworm with dsRNA containing a 240 bp fragment of the Western Corn Rootworm *Snf7* gene. Modified starch/carbohydrates can result from small RNA such as the *pPhL* small RNA (degrades PhL transcripts to limit the formation of reducing sugars through starch degradation) and *pR1* small RNA (degrades R1 transcripts to limit the formation of reducing sugars through starch degradation). Additional, benefits such as reduced acrylamide resulting from the *asn1* small RNA

that triggers degradation of Asn1 to impair asparagine formation and reduce polyacrylamide. Finally, the non-browning phenotype of *pgas ppo suppression* small RNA results in suppressing PPO to produce apples with a non-browning phenotype. The above list of small RNAs is not meant to be limiting. Any small RNA encoding sequences are encompassed by the present disclosure.

6. Selectable Markers

[00107] Various selectable markers also described as reporter genes can be operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1, or a sequence that has 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In an embodiment, a promoter can be the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1. promoter comprising SEQ ID NO: 1, or a sequence that has at least 80, 85, 90, 95 or 99% sequence identity with SEQ ID NO: 1. In some embodiments, the sequences are operably linked to the *Zea mays* GRMZM2G144030 promoter comprising SEQ ID NO: 1 and a 5' UTR. The operably linked sequences can then be incorporated into a chosen vector to allow for identification and selectable of transformed plants ("transformants"). Many methods are available to confirm expression of selectable markers in transformed plants, including for example DNA sequencing and PCR (polymerase chain reaction), Southern blotting, RNA blotting, immunological methods for detection of a protein expressed from the vector. But, usually the reporter genes are observed through visual observation of proteins that when expressed produce a colored product. Exemplary reporter genes are known in the art and encode β -glucuronidase (GUS), luciferase, green fluorescent protein (GFP), yellow fluorescent protein (YFP, Phi-YFP), red fluorescent protein (DsRFP, RFP, etc), β -galactosidase, and the like (See Sambrook, et al., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Press, N.Y., 2001, the content of which is incorporated herein by reference in its entirety).

[00108] Selectable marker genes are utilized for selection of transformed cells or tissues. Selectable marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO), spectinomycin/streptomycin resistance (AAD), and hygromycin phosphotransferase (HPT or HGR) as well as genes conferring resistance to herbicidal compounds. Herbicide resistance genes generally code for a modified target protein insensitive to the herbicide or for an enzyme that degrades or detoxifies the herbicide in the plant before it can act. For

example, resistance to glyphosate has been obtained by using genes coding for mutant target enzymes, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS). Genes and mutants for EPSPS are well known, and further described below. Resistance to glufosinate ammonium, bromoxynil, and 2,4-dichlorophenoxyacetate (2,4-D) have been obtained by using bacterial genes encoding PAT or DSM-2, a nitrilase, an AAD-1, or an AAD-12, each of which are examples of proteins that detoxify their respective herbicides.

[00109] In an embodiment, herbicides can inhibit the growing point or meristem, including imidazolinone or sulfonylurea, and genes for resistance/tolerance of acetohydroxyacid synthase (AHAS) and acetolactate synthase (ALS) for these herbicides are well known. Glyphosate resistance genes include mutant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPs) and *dgt-28* genes (via the introduction of recombinant nucleic acids and/or various forms of in vivo mutagenesis of native EPSPs genes), *aroA* genes and glyphosate acetyl transferase (GAT) genes, respectively). Resistance genes for other phosphono compounds include *bar* and *pat* genes from *Streptomyces* species, including *Streptomyces hygroscopicus* and *Streptomyces viridichromogenes*, and pyridinoxy or phenoxy propionic acids and cyclohexones (ACCase inhibitor-encoding genes). Exemplary genes conferring resistance to cyclohexanediones and/or aryloxyphenoxypropanoic acid (including haloxyfop, diclofop, fenoxypop, fluazifop, quizalofop) include genes of acetyl coenzyme A carboxylase (ACCase); *Acc1-S1*, *Acc1-S2* and *Acc1-S3*. In an embodiment, herbicides can inhibit photosynthesis, including triazine (*psbA* and *1s+* genes) or benzonitrile (nitrilase gene). Furthermore, such selectable markers can include positive selection markers such as phosphomannose isomerase (PMI) enzyme.

[00110] In an embodiment, selectable marker genes include, but are not limited to genes encoding: 2,4-D; neomycin phosphotransferase II; cyanamide hydratase; aspartate kinase; dihydrodipicolinate synthase; tryptophan decarboxylase; dihydrodipicolinate synthase and desensitized aspartate kinase; *bar* gene; tryptophan decarboxylase; neomycin phosphotransferase (NEO); hygromycin phosphotransferase (HPT or HYG); dihydrofolate reductase (DHFR); phosphinothricin acetyltransferase; 2,2-dichloropropionic acid dehalogenase; acetohydroxyacid synthase; 5-enolpyruvyl-shikimate-phosphate synthase (*aroA*); haloarylnitrilase; acetyl-coenzyme A

carboxylase; dihydropteroate synthase (sul I); and 32 kD photosystem II polypeptide (psbA). An embodiment also includes selectable marker genes encoding resistance to: chloramphenicol; methotrexate; hygromycin; spectinomycin; bromoxynil; glyphosate; and phosphinothricin. The above list of selectable marker genes is not meant to be limiting. Any reporter or selectable marker gene are encompassed by the present disclosure.

[00111] In some embodiments the coding sequences are synthesized for optimal expression in a plant. For example, in an embodiment, a coding sequence of a gene has been modified by codon optimization to enhance expression in plants. An insecticidal resistance transgene, an herbicide tolerance transgene, a nitrogen use efficiency transgene, a water use efficiency transgene, a nutritional quality transgene, a DNA binding transgene, or a selectable marker transgene can be optimized for expression in a particular plant species or alternatively can be modified for optimal expression in dicotyledonous or monocotyledonous plants. Plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in the particular plant species of interest. In an embodiment, a coding sequence, gene, or transgene is designed to be expressed in plants at a higher level resulting in higher transformation efficiency. Methods for plant optimization of genes are well known. Guidance regarding the optimization and production of synthetic DNA sequences can be found in, for example, WO2013016546, WO2011146524, WO1997013402, US Patent No. 6166302, and US Patent No. 5380831, herein incorporated by reference.

Transformation

[00112] Suitable methods for transformation of plants include any method by which DNA can be introduced into a cell, for example and without limitation: electroporation (see, e.g., U.S. Patent 5,384,253); micro-projectile bombardment (see, e.g., U.S. Patents 5,015,580, 5,550,318, 5,538,880, 6,160,208, 6,399,861, and 6,403,865); *Agrobacterium*-mediated transformation (see, e.g., U.S. Patents 5,635,055, 5,824,877, 5,591,616; 5,981,840, and 6,384,301); and protoplast transformation (see, e.g., U.S. Patent 5,508,184).

[00113] A DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as agitation with silicon carbide fibers (see, e.g., U.S. Patents 5,302,523

and 5,464,765), or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment (see, e.g., Klein et al. (1987) Nature 327:70-73). Alternatively, the DNA construct can be introduced into the plant cell via nanoparticle transformation (see, e.g., US Patent Publication No. 20090104700, which is incorporated herein by reference in its entirety).

[00114] In addition, gene transfer may be achieved using non-*Agrobacterium* bacteria or viruses such as *Rhizobium* sp. NGR234, *Sinorhizobium meliloti*, *Mesorhizobium loti*, potato virus X, cauliflower mosaic virus and cassava vein mosaic virus and/or tobacco mosaic virus, See, e.g., Chung et al. (2006) Trends Plant Sci. 11(1):1-4.

[00115] Through the application of transformation techniques, cells of virtually any plant species may be stably transformed, and these cells may be developed into transgenic plants by well-known techniques. For example, techniques that may be particularly useful in the context of cotton transformation are described in U.S. Patents 5,846,797, 5,159,135, 5,004,863, and 6,624,344; techniques for transforming Brassica plants in particular are described, for example, in U.S. Patent 5,750,871; techniques for transforming soy bean are described, for example, in U.S. Patent 6,384,301; and techniques for transforming maize are described, for example, in U.S. Patents 7,060,876 and 5,591,616, and International PCT Publication WO 95/06722.

[00116] After effecting delivery of an exogenous nucleic acid to a recipient cell, a transformed cell is generally identified for further culturing and plant regeneration. In order to improve the ability to identify transformants, one may desire to employ a selectable marker gene with the transformation vector used to generate the transformant. In an illustrative embodiment, a transformed cell population can be assayed by exposing the cells to a selective agent or agents, or the cells can be screened for the desired marker gene trait.

[00117] Cells that survive exposure to a selective agent, or cells that have been scored positive in a screening assay, may be cultured in media that supports regeneration of plants. In an embodiment, any suitable plant tissue culture media may be modified by including further substances, such as growth regulators. Tissue may be maintained on a basic media with growth regulators until sufficient tissue is available to begin plant regeneration efforts, or following repeated rounds of

manual selection, until the morphology of the tissue is suitable for regeneration (e.g., at least 2 weeks), then transferred to media conducive to shoot formation. Cultures are transferred periodically until sufficient shoot formation has occurred. Once shoots are formed, they are transferred to media conducive to root formation. Once sufficient roots are formed, plants can be transferred to soil for further growth and maturity.

Molecular Confirmation

[00118] A transformed plant cell, callus, tissue or plant may be identified and isolated by selecting or screening the engineered plant material for traits encoded by the marker genes present on the transforming DNA. For instance, selection can be performed by growing the engineered plant material on media containing an inhibitory amount of the antibiotic or herbicide to which the transforming gene construct confers resistance. Further, transformed plants and plant cells can also be identified by screening for the activities of any visible marker genes (e.g., the *β -glucuronidase*, *luciferase*, or *gfp* genes) that may be present on the recombinant nucleic acid constructs. Such selection and screening methodologies are well known to those skilled in the art. Molecular confirmation methods that can be used to identify transgenic plants are known to those with skill in the art. Several exemplary methods are further described below.

[00119] Molecular Beacons have been described for use in sequence detection. Briefly, a FRET oligonucleotide probe is designed that overlaps the flanking genomic and insert DNA junction. The unique structure of the FRET probe results in it containing a secondary structure that keeps the fluorescent and quenching moieties in close proximity. The FRET probe and PCR primers (one primer in the insert DNA sequence and one in the flanking genomic sequence) are cycled in the presence of a thermostable polymerase and dNTPs. Following successful PCR amplification, hybridization of the FRET probe(s) to the target sequence results in the removal of the probe secondary structure and spatial separation of the fluorescent and quenching moieties. A fluorescent signal indicates the presence of the flanking genomic/transgene insert sequence due to successful amplification and hybridization. Such a molecular beacon assay for detection of an amplification reaction is an embodiment of the subject disclosure.

[00120] Hydrolysis probe assay, otherwise known as TAQMAN[®] (Life Technologies,

Foster City, Calif.), is a method of detecting and quantifying the presence of a DNA sequence. Briefly, a FRET oligonucleotide probe is designed with one oligo within the transgene and one in the flanking genomic sequence for event-specific detection. The FRET probe and PCR primers (one primer in the insert DNA sequence and one in the flanking genomic sequence) are cycled in the presence of a thermostable polymerase and dNTPs. Hybridization of the FRET probe results in cleavage and release of the fluorescent moiety away from the quenching moiety on the FRET probe. A fluorescent signal indicates the presence of the flanking/transgene insert sequence due to successful amplification and hybridization. Such a hydrolysis probe assay for detection of an amplification reaction is an embodiment of the subject disclosure.

[00121] KASPar® assays are a method of detecting and quantifying the presence of a DNA sequence. Briefly, the genomic DNA sample comprising the integrated gene expression cassette polynucleotide is screened using a polymerase chain reaction (PCR) based assay known as a KASPar® assay system. The KASPar® assay used in the practice of the subject disclosure can utilize a KASPar® PCR assay mixture which contains multiple primers. The primers used in the PCR assay mixture can comprise at least one forward primers and at least one reverse primer. The forward primer contains a sequence corresponding to a specific region of the DNA polynucleotide, and the reverse primer contains a sequence corresponding to a specific region of the genomic sequence. In addition, the primers used in the PCR assay mixture can comprise at least one forward primers and at least one reverse primer. For example, the KASPar® PCR assay mixture can use two forward primers corresponding to two different alleles and one reverse primer. One of the forward primers contains a sequence corresponding to specific region of the endogenous genomic sequence. The second forward primer contains a sequence corresponding to a specific region of the DNA polynucleotide. The reverse primer contains a sequence corresponding to a specific region of the genomic sequence. Such a KASPar® assay for detection of an amplification reaction is an embodiment of the subject disclosure.

[00122] In some embodiments the fluorescent signal or fluorescent dye is selected from the group consisting of a HEX fluorescent dye, a FAM fluorescent dye, a JOE fluorescent dye, a TET fluorescent dye, a Cy 3 fluorescent dye, a Cy 3.5 fluorescent dye, a Cy 5 fluorescent dye, a

Cy 5.5 fluorescent dye, a Cy 7 fluorescent dye, and a ROX fluorescent dye.

[00123] In other embodiments the amplification reaction is run using suitable second fluorescent DNA dyes that are capable of staining cellular DNA at a concentration range detectable by flow cytometry, and have a fluorescent emission spectrum which is detectable by a real time thermocycler. It should be appreciated by those of ordinary skill in the art that other nucleic acid dyes are known and are continually being identified. Any suitable nucleic acid dye with appropriate excitation and emission spectra can be employed, such as YO-PRO-1®, SYTOX Green®, SYBR Green I®, SYTO11®, SYTO12®, SYTO13®, BOBO®, YOYO®, and TOTO®. In one embodiment, a second fluorescent DNA dye is SYTO13® used at less than 10 μM , less than 4 μM , or less than 2.7 μM .

[00124] In further embodiments, Next Generation Sequencing (NGS) can be used for detection. As described by Brautigma *et al.*, 2010, DNA sequence analysis can be used to determine the nucleotide sequence of the isolated and amplified fragment. The amplified fragments can be isolated and sub-cloned into a vector and sequenced using chain-terminator method (also referred to as Sanger sequencing) or Dye-terminator sequencing. In addition, the amplicon can be sequenced with Next Generation Sequencing. NGS technologies do not require the sub-cloning step, and multiple sequencing reads can be completed in a single reaction. Three NGS platforms are commercially available, the Genome Sequencer FLX™ from 454 Life Sciences / Roche, the Illumina Genome Analyser™ from Solexa and Applied Biosystems' SOLiD™ (acronym for: 'Sequencing by Oligo Ligation and Detection'). In addition, there are two single molecule sequencing methods that are currently being developed. These include the true Single Molecule Sequencing (tSMS) from Helicos Bioscience™ and the Single Molecule Real Time™ sequencing (SMRT) from Pacific Biosciences.

[00125] The Genome Sequencer FLX™ which is marketed by 454 Life Sciences/Roche is a long read NGS, which uses emulsion PCR and pyrosequencing to generate sequencing reads. DNA fragments of 300 – 800 bp or libraries containing fragments of 3 – 20 kb can be used. The reactions can produce over a million reads of about 250 to 400 bases per run for a total yield of 250 to 400 megabases. This technology produces the longest reads but the total sequence output

per run is low compared to other NGS technologies.

[00126] The Illumina Genome Analyser™ which is marketed by Solexa™ is a short read NGS which uses sequencing by synthesis approach with fluorescent dye-labeled reversible terminator nucleotides and is based on solid-phase bridge PCR. Construction of paired end sequencing libraries containing DNA fragments of up to 10 kb can be used. The reactions produce over 100 million short reads that are 35 – 76 bases in length. This data can produce from 3 – 6 gigabases per run.

[00127] The Sequencing by Oligo Ligation and Detection (SOLiD) system marketed by Applied Biosystems™ is a short read technology. This NGS technology uses fragmented double stranded DNA that are up to 10 kb in length. The system uses sequencing by ligation of dye-labelled oligonucleotide primers and emulsion PCR to generate one billion short reads that result in a total sequence output of up to 30 gigabases per run.

[00128] tSMS of Helicos Bioscience™ and SMRT of Pacific Biosciences™ apply a different approach which uses single DNA molecules for the sequence reactions. The tSMS Helicos™ system produces up to 800 million short reads that result in 21 gigabases per run. These reactions are completed using fluorescent dye-labelled virtual terminator nucleotides that is described as a ‘sequencing by synthesis’ approach.

[00129] The SMRT Next Generation Sequencing system marketed by Pacific Biosciences™ uses a real time sequencing by synthesis. This technology can produce reads of up to 1,000 bp in length as a result of not being limited by reversible terminators. Raw read throughput that is equivalent to one-fold coverage of a diploid human genome can be produced per day using this technology.

[00130] In another embodiment, the detection can be completed using blotting assays, including Western blots, Northern blots, and Southern blots. Such blotting assays are commonly used techniques in biological research for the identification and quantification of biological samples. These assays include first separating the sample components in gels by electrophoresis, followed by transfer of the electrophoretically separated components from the gels to transfer membranes that are made of materials such as nitrocellulose, polyvinylidene fluoride (PVDF), or

nylon. Analytes can also be directly spotted on these supports or directed to specific regions on the supports by applying vacuum, capillary action, or pressure, without prior separation. The transfer membranes are then commonly subjected to a post-transfer treatment to enhance the ability of the analytes to be distinguished from each other and detected, either visually or by automated readers.

[00131] In a further embodiment the detection can be completed using an ELISA assay, which uses a solid-phase enzyme immunoassay to detect the presence of a substance, usually an antigen, in a liquid sample or wet sample. Antigens from the sample are attached to a surface of a plate. Then, a further specific antibody is applied over the surface so it can bind to the antigen. This antibody is linked to an enzyme, and, in the final step, a substance containing the enzyme's substrate is added. The subsequent reaction produces a detectable signal, most commonly a color change in the substrate.

Transgenic Plants

[00132] In an embodiment, a plant, plant tissue, or plant cell comprises a *Zea mays* GRMZM2G144030 promoter. In one embodiment a plant, plant tissue, or plant cell comprises the *Zea mays* GRMZM2G144030 promoter of a sequence selected from SEQ ID NO:1 or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with a sequence selected from SEQ ID NO:1. In another embodiment a plant, plant tissue, or plant cell comprises the *Zea mays* GRMZM2G144030 3' UTR comprises a sequence selected from SEQ ID NO:5 or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with a sequence selected from SEQ ID NO:5. In another embodiment a plant, plant tissue, or plant cell comprises the *Zea mays* GRMZM2G144030 promoter from SEQ ID NO:1 operably linked to a 5' UTR. In an embodiment, a plant, plant tissue, or plant cell comprises a gene expression cassette comprising a sequence selected from SEQ ID NO:1, or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with a sequence selected from SEQ ID NO:1 that is operably linked to a non-GRMZM2G144030 transgene. In an illustrative embodiment, a plant, plant tissue, or plant cell comprises a gene expression cassette comprising a *Zea mays* GRMZM2G144030 promoter that is operably linked to a transgene, wherein the transgene can be an insecticidal resistance transgene, an herbicide tolerance

transgene, a nitrogen use efficiency transgene, a water use efficiency transgene, a nutritional quality transgene, a DNA binding transgene, a selectable marker transgene, or combinations thereof.

[00133] In accordance with one embodiment a plant, plant tissue, or plant cell is provided wherein the plant, plant tissue, or plant cell comprises a non-endogenous GRMZM2G144030 gene derived promoter sequence operably linked to a transgene, wherein the *Zea mays* GRMZM2G144030 promoter derived promoter sequence comprises a sequence of SEQ ID NO:1 or a sequence having at least 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO:1. In one embodiment a plant, plant tissue, or plant cell is provided wherein the plant, plant tissue, or plant cell comprises SEQ ID NO: 1, or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO: 1 operably linked to a non-GRMZM2G144030 transgene. In one embodiment the plant, plant tissue, or plant cell is a dicotyledonous or monocotyledonous plant or a cell or tissue derived from a dicotyledonous or monocotyledonous plant. In one embodiment the plant is selected from the group consisting of maize, wheat, rice, sorghum, oats, rye, bananas, sugar cane, soybean, cotton, sunflower, and canola. In one embodiment the plant is soybean. In accordance with one embodiment the plant, plant tissue, or plant cell comprises SEQ ID NO: 1 or a sequence having 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO:1 operably linked to a non-GRMZM2G144030 transgene. In one embodiment the plant, plant tissue, or plant cell comprises a promoter operably linked to a transgene wherein the promoter consists of SEQ ID NO: 1 or a sequence having 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO:1. In accordance with one embodiment the gene construct comprising *Zea mays* GRMZM2G144030 promoter sequence operably linked to a transgene is incorporated into the genome of the plant, plant tissue, or plant cell.

[00134] In one embodiment a non-*Zea mays* c.v. B73 plant, plant tissue, or plant cell is provided comprising SEQ ID NO: 1, or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO:1, operably linked to a transgene. In accordance with one embodiment the non-*Zea mays* c.v. B73 plant, plant tissue, or plant cell is a dicotyledonous or monocotyledonous plant or plant cell or tissue derived from a dicotyledonous or monocotyledonous plant. In one embodiment the plant is selected from the group consisting of maize, wheat, rice, sorghum, oats, rye, bananas, sugar cane, soybean, cotton, sunflower, and canola. In one embodiment

the plant is soybean. In accordance with one embodiment the promoter sequence operably linked to a transgene is incorporated into the genome of the plant, plant tissue, or plant cell.

[00135] In one embodiment a non-*Zea mays* c.v. B73 plant, plant tissue, or plant cell is provided that comprises SEQ ID NO: 1, or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO:1, operably linked to the 5' end of a transgene and a 3' untranslated sequence comprising SEQ ID NO:5 or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO:5, wherein the 3' untranslated sequence is operably linked to said transgene. In another embodiment a non-*Zea mays* c.v. B73 plant, plant tissue, or plant cell is provided that comprises SEQ ID NO: 1, or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with SEQ ID NO:1, operably linked to the 3' end of a 5' untranslated sequence, wherein the 5' untranslated sequence is operably linked to said transgene. In accordance with one embodiment the non-*Zea mays* c.v. B73 plant, plant tissue, or plant cell is a dicotyledonous or monocotyledonous plant or is a plant issue or cell derived from a dicotyledonous or monocotyledonous plant. In one embodiment the plant is selected from the group consisting of maize, wheat, rice, sorghum, oats, rye, bananas, sugar cane, soybean, cotton, sunflower, and canola. In one embodiment the plant is soybean. In accordance with one embodiment the promoter sequence operably linked to a transgene is incorporated into the genome of the plant, plant tissue, or plant cell.

[00136] In an embodiment, a plant, plant tissue, or plant cell according to the methods disclosed herein can be a monocotyledonous plant. The monocotyledonous plant, plant tissue, or plant cell can be, but not limited to corn, rice, wheat, sugarcane, barley, rye, sorghum, orchids, bamboo, banana, cattails, lilies, oat, onion, millet, switchgrass, turfgrass, and triticale.

[00137] In an embodiment, a plant, plant tissue, or plant cell according to the methods disclosed herein can be a dicotyledonous plant. The dicotyledonous plant, plant tissue, or plant cell can be, but not limited to alfalfa, rapeseed, canola, Indian mustard, Ethiopian mustard, soybean, sunflower, cotton, beans, broccoli, cabbage, cauliflower, celery, cucumber, eggplant, lettuce; melon, pea, pepper, peanut, potato, pumpkin, radish, spinach, sugarbeet, sunflower, tobacco, tomato, and watermelon.

[00138] One of skill in the art will recognize that after the exogenous sequence is stably

incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[00139] The present disclosure also encompasses seeds of the transgenic plants described above, wherein the seed has the transgene or gene construct containing the gene regulatory elements of the subject disclosure. The present disclosure further encompasses the progeny, clones, cell lines or cells of the transgenic plants described above wherein said progeny, clone, cell line or cell has the transgene or gene construct containing the gene regulatory elements of the subject disclosure.

[00140] The present disclosure also encompasses the cultivation of transgenic plants described above, wherein the transgenic plant has the transgene or gene construct containing the gene regulatory elements of the subject disclosure. Accordingly, such transgenic plants may be engineered to, *inter alia*, have one or more desired traits or transgenic events containing the gene regulatory elements of the subject disclosure, by being transformed with nucleic acid molecules according to the invention, and may be cropped or cultivated by any method known to those of skill in the art.

Method of Expressing a Transgene

[00141] In an embodiment, a method of expressing at least one transgene in a plant comprises growing a plant comprising a *Zea mays* GRMZM2G144030 promoter operably linked to at least one transgene or a polylinker sequence. In an embodiment, a method of expressing at least one transgene in a plant comprising growing a plant comprising *Zea mays* GRMZM2G144030 promoter and the *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene or a polylinker sequence. In an embodiment, a method of expressing at least one transgene in a plant comprises growing a plant comprising a *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene or a polylinker sequence. In one embodiment the *Zea mays* GRMZM2G144030 promoter consists of a sequence selected from SEQ ID NO:1 or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with a sequence selected from SEQ ID NO:1. In another embodiment the *Zea mays* GRMZM2G144030 3' UTR consists of a sequence selected from SEQ ID NO:5 or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with a sequence selected from SEQ ID NO:5. In an embodiment, a method of expressing at least one transgene in a

plant comprises growing a plant comprising a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant comprising growing a plant comprising a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprising culturing a plant tissue or plant cell comprising a *Zea mays* GRMZM2G144030 promoter operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprising culturing a plant tissue or plant cell comprising a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprising culturing a plant tissue or plant cell comprising a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprising culturing a plant tissue or plant cell comprising a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprising culturing a plant tissue or plant cell comprising a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene.

[00142] In an embodiment, a method of expressing at least one transgene in a plant comprises growing a plant comprising a gene expression cassette comprising a *Zea mays* GRMZM2G144030 promoter operably linked to at least one transgene. In one embodiment the *Zea mays* GRMZM2G144030 promoter consists of a sequence selected from SEQ ID NO:1 or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with a sequence selected from SEQ ID NO:1. In another embodiment the *Zea mays* GRMZM2G144030 3' UTR consists of a sequence selected from SEQ ID NO:5 or a sequence that has at least 80%, 85%, 90%, 95% or 99.5% sequence identity with a sequence selected from SEQ ID NO:5. In an embodiment, a method of expressing at least one transgene in a plant comprises growing a plant comprising a gene expression cassette

comprising a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant comprises growing a plant comprising a gene expression cassette comprising *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant comprises growing a plant comprising a gene expression cassette comprising *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant comprises growing a plant comprising a gene expression cassette comprising a *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprises culturing a plant tissue or plant cell comprising a gene expression cassette containing a *Zea mays* GRMZM2G144030 promoter operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprises culturing a plant tissue or plant cell comprising a gene expression cassette containing a *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprises culturing a plant tissue or plant cell comprising a gene expression cassette containing a *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprises culturing a plant tissue or plant cell comprising a gene expression cassette, a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 3' UTR operably linked to at least one transgene. In an embodiment, a method of expressing at least one transgene in a plant tissue or plant cell comprises culturing a plant tissue or plant cell comprising a gene expression cassette, a *Zea mays* GRMZM2G144030 promoter and a *Zea mays* GRMZM2G144030 5' UTR operably linked to at least one transgene.

[00143] The following examples are provided to illustrate certain particular features and/or embodiments. The examples should not be construed to limit the disclosure to the particular features or embodiments exemplified.

EXAMPLES

Example 1: Isolation of Novel Promoter and Other Regulatory Elements

[00144] Novel *Zea mays* GRMZM2G144030 gene regulatory elements were identified via analyzing publicly available transcriptome of maize seedlings. These regulatory elements were identified, isolated, and cloned to characterize the expression profile of the regulatory elements for use in transgenic plants. Transgenic maize lines stably transformed with a *cry3Ab1* gene isolated from *Bacillus thuringiensis* and an *aad-1* selectable marker gene derived from *Sphingobium herbicidovorans* were produced and the transgene expression levels and tissue specificity was assessed. As such novel *Zea mays* GRMZM2G144030 gene regulatory elements were identified and characterized. Disclosed are promoter and 3' UTR regulatory elements for use in gene expression constructs.

[00145] Three sources of data were considered to prioritize high expressing maize genes in shoots and roots of maize seedlings: 1) 35,000 maize gene sequences and their annotations present in the public maize database as of 2010 when this study was carried out; 2) gene expression data for total maize transcriptome for V4 shoots and roots (Wang, et al., [2009], *The Plant Cell*; 21: 1053–1069); and, 3) full-length cDNA sequences of 9,000 genes (Alexandrov, N. et al., [2009] *Plant Molecular Biology*; 69:179–194). In this study, the gene expression data was aligned to both 9,000 full-length cDNA sequences and 35,000 maize genes. Based on Fragments Per Kilobase of exon per Million fragments mapped (FPKM) values, a quantitative measure of gene expression, 500 best expressing genes were identified for maize shoot and root tissues each.

[00146] Since expression of transgenes require transgene expression throughout the life cycle of a maize plant, total mRNA was isolated from 3 different stages of leaf (V4, V12 and R3), 2 different stages of root development (V4 and V12), and one stage of pollen (R1) for transgene expression analysis. The *Zea mays* c.v. B73 maize genotype was used for all analysis. Of the 500 genes prioritized, about 150 of the best expressing genes were subjected to quantitative PCR for gene expression confirmation. The results identified a sub-set of these as the best expressing genes for leaf and root-preferred expression.

[00147] The promoter from the *Zea mays* GRMZM2G144030 gene regulatory elements

(SEQ ID NO:1) is a 2,082 bp polynucleotide sequence that was identified from the *Zea mays* c.v. B73 genomic DNA (gDNA) sequence. From the assessment of the contiguous chromosomal sequence that spanned millions of base pairs, a 2,082 bp polynucleotide sequence was identified and isolated, either by PCR amplification or DNA synthesis, for use in expression of heterologous coding sequences. This novel polynucleotide sequence was analyzed for use as a regulatory sequence to drive expression of a gene. As shown in the sequence (SEQ ID NO:2) below, the 2,082 bp *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1 is provided as base pairs 1 – 2,082. The native gene coding sequence of SEQ ID NO:4 is provided as base pairs 2,083 – 4,611 of SEQ ID NO:2 (the ATG start codon and the TAA termination codon are shown in capital letters). The 1,095 bp *Zea mays* GRMZM2G144030 3' UTR of SEQ ID NO:5 is provided as base pairs 4,612 – 5,706 of SEQ ID NO:2.

[00148] ttgtcacatcgtaattcgttaggtctacatcacgctaaggggttgatttggtgacaaggggatccaggaggattgga
 ggggattgaggggtccgatccctccggggtacccgggtcaccaaatcagctctaaggggtctgttcattttgttcaaatctacatggattgtgggg
 tatgaaacgaaggaggcctaagaacctttgacctgacattttccccactttgatctcacgtatatgcagtactgcaatgaggtcgcaaagaa
 gctgatcaatggagcagatcatagtttcgttgagcctgactcagactacagattttgacctgaagcaggtcattacttcgttgggtctagctagg
 actaaagattttgacgtgacaaaattttcactttaatccgatgagtgagatcatagagaagctgatcaatgggctagatcatcattcattgcgtc
 tagctgggagtaaagattttgatgtgacatttgcttgcgtttgatttggtcgggtgggtacttcgtacggctggctattgcttcagggttaaagtcgat
 tgtttagattgttacatttataatttataaaacaaaaatattgtagaattgatataaattagtagctattaaagctaaacgagcaaacaaagattttg
 atctgttgcgttgactttaatccatactgatttttactattttctataaacttcaatctgattagtgagacttcgtcttttgcctatcgatggctactt
 cgctgggtctgggtccaactagagggtgttttaggatgggtcaagggttcagttccagtatgaagctatatatttatacaatttaattagttttaagaa
 tatttttaatacataataaataaagtgacttatctaatgcattctaaaggcttacaatttcaactccacatttttagagcatctaatgacctattcta
 ctaacttcactaaattttctagagaaatagctatcctaaacagaatataaagattgcagcgggtatatgcctgtatattttttgtgtattgtattt
 atgtccagctctgcacataagatgtataaaaaactattagggctgatttgggtgaccggggatctcgaggggaagaaatcctcttgctattcaa
 aattgaattgcaagaggattttctccatggatcctctagagatccctaatcaccaaatcaaccctgatagtgagtccttgggtgtgtattgaacga
 gatatgctaataatatttaaacaagattataaaactagccttcgcctttctgacctttaaattgataactaaacttacaaccacatactacgatgctat
 aaggaaaacttcaggtgggggtgttgcaatccaagaatagtgcttctcacgacgaatctcaatctgatcggtatcggaataagagggtt
 cgaaaacgttccgcacgattgttttctatcgctgatcagcaaagggtgacgctgaaaagaaaaatgaaaagcttcgataagaaaaggaacg
 caagggtcgcaagcaagaaaaaaaaggaaggcgagccaaggaggaggagaaagatgcttttcagccgtcgatagcgtgattcaaacg

cgagaagatgccttcctctgcagccgtacacgcgtccgatgccccgacaaaaccgacgggtcagatacgcacaacatgcacctgaagg
cgccaaccgacatgcgggcccgtttccagcaagccccagagtcagtgaacacaccgtctctctcgggttggaagcgggtgggtgca
cctgctgtagcggtataaaaagcggaaaccctagcattcggcgcgagcttatcacttatccccatccgcacctctctccatctccctgct
cctccgctcctcttgatccagccaagccagggttaaattctgacctgctgcgcgtcagatttactccgaattcagttgatctgggttcagattgt
aatctctctgtttcctaaaatacggggaatccgaatcgaggctaatttcggatcttggtttgtttctttttttttgcagggatcgccATGtc
ggactcggaggagcaccacttcgaatcgaaggccgacgccggagcgtccaagacctaccgcagcaggctggcaccgtccgtaagaa
cggcttcacgtcatcaagaaccgccccgcaaggtatatgtgtgcgcagtttgggatcggctgtattccattgattagtgtgacctgtcgc
ttgtgctgtgaactgggtgatcgcagatgcagttgtctttggcatactgtgatagctgttaagtttgagggtgatccaatctatggagggtg
aatgtgtgggacttggtgatagataacctgggtggctagaaggagactttgctattgccagccggagtataattacaagtaggctttacagaa
taacagttcatctgatctgactttctggagaccagaaaaatacatttaaggcctgtttgattgctgcctaacttgccacactttgggtaactttt
cagcctatggtagttcttcaattcgagcgactaaccttagtcaaatgtggcacagttagtcacgaaccaacatgcccttagtcagagtagg
ttgcacctgccccatataagtagctgaaatgcaaatgggtgatcttttgattttccctttacaatgcgggaaaactagttggcatgttgagtta
attctttactcctctaaactgtcattgaacattcaagggtactcattattatctttgatgatggctgaatgtttgtaccccatgttctatggtgatatt
gtccaagtgatcagacaatttatcatttaagttcaacagagggttatccaatttccattattccaagataacgggaataagagatcacacagat
aatgtgttttaaggaaaataaaaatacttttatctttgtcagctctgtcatacaactaaactaatgttgcctgactcagtgatgttttcttgggagcta
taagctgttaaatgatgtttgaaatcaataaaaaaactaaaataacatatgccctacattttggacagagttttgatttgacccacaatgattgtaa
aacattgttcttattattgatttttaaatgagttgggttgtaattgggaagttatatgttaattttctgtttatcaataaaaccagtttgttgcataat
gaagtttcttttggcttgccaatccattgacaaagagaatgtgagtagtactgttttgcaattaccaaacagggtatctcttctttaattctaaacc
tcgaaaattttctatgtggtattgccataccgtattttatttttaggtgtcactgtattatcataaggtgatagtaagctcatttctgtttattaaccttc
tattttatgaacttctgatggatttgagtgatataatcagcaacagttcatgatgtgtccctaaaactatttggaaactttattgctcttaagtcgcagc
acatacgtcgtgtttattgagtttgggttggaatatggttgataaaaacaacttatgctacaggttggtggaggtttctacctcaaagactggta
agcatggccatgccaatgccactttgttgccatagacataattcaacgggaaaaagctcgaagatattgttcttcgtcacacaactgtgatgt
aagttaatttctgtgttaagctaagttcttttgatgggtcacttcatacgcactcacaagtttgacattttttaagattccgcatgtgaaccgtactg
agtaccagctgattgatataatcagaggatggatttgatgttctctaatcattaaaactacattcttcttgaaataaatgttaacttcagcttataatt
ttaatttaatttgggtcaggtgagccttcttacttcagatggcaacactaaggatgatcttagactcccaactgatgagactctgtggcccagggtc
tggttgaaaacattttcatcttaagcttgttggttttcttattgacttccaaataattacagatgggtggttatctttgttagcttattagtcatatgttcttc
ttggtagcttcttcttcttgatgtttgttaattgatctgggtcttataaggaaatgtggaaagtagggcctttagaactagacattccttatgtattta
gactgtctaccattctttttaaaaatgatatgacataacaacatgtcatataaaacatgctgaaactgcactaaaattggttgctttaaaactgaa

acgtatgcatgacctgcactttgtatgtgtctaggatacaatttatgttgattggaagtagtatgctgcataacttatgtctgttgagagttaattt
 atctacctcaagtggctgccccaaactaaataataacttctgcatcttgtatgcttggaattcagatcaaggaagggttgaaagcggcaaggat
 ctgtgtgactgtccagtctgctatgggggaggagcagatctgcgcgctgaaggatgttgccccaagTAActtccctgccttggaatac
 tgtatctcaaacctaaatcgaaaaagaagtgtatcaaggattgctacagagacatccatctggcttgagctggcttttgcctatggcaaacac
 aagtgcagggatcgctgggtgtgcaccgtgtctgcattatgtggttacattctggaccctgtattttctatgctttatgccctactgtttagtatt
 aattatcaataaatttgttgggacgggtgaattgcatgcattgcatgagggttcttgttcgttatttgtttctcttgcctatgctccagtattggtgat
 ggcagtttactcgtattttttgtacttgttcgaaaactggggcagttaccgcgcttgtgtccatccattgttctagtgttatccattcctct
 ttttccccctcaaaatataaaactataaagaagaaaatattatgacttacataaagaatcgtttagagaagcgggtaccaaacatgactcgacg
 gaaaattgctccaaaccaactattaactcgatgtttgtatttttcttctgccccaaacctgtcaaccgctttaggttaactttgcgggtatatgtt
 ctgaaagaaaactggtgtatagagctactccctaaattcaagttgtaagacattttgccatttgatataatgtaggtagaaaagatgtctttaa
 tctaaaacagaggaggatattatgtagctttttgtagataaaatgtcttatttaaacagaggagattttgagtagcttttgacgattgaatgctg
 atttggacggaggagctacttgtgtagctttttggaacgaagtacttgagtgtttttgaggattgatttctgtaataaaatgtacttgaccctt
 attataggtgtacatctggctgtgtcgggcttgggtggcatgacataatgcgtggatcgtgtctaggtcttaagcccacagattggtatggtatg
 acttatttagttaaggccattaaagcctgctaaacataaactgttgcgcagtgtgcttgtgtgtgggctttatcactgtatttgcatttcgt
 (SEQ ID NO:2)

Example 2: Vector Construction

[00149] The following vectors were built to incorporate the *Zea mays* GRMZM2G144030 promoter upstream of a transgene. The vector construct pDAB113002 contained a gene expression cassette, in which the *cry34ab1* transgene (reporter gene from *B. thuringiensis*) was driven by the *Zea mays* GRMZM2G144030 promoter of SEQ ID NO:1, and was flanked by the *Zea mays* GRMZM2G144030 3' UTR of SEQ ID NO:5. A diagram of this gene expression cassette is shown in **Fig. 1** and is provided as SEQ ID NO:3. The vector also contained a selectable marker gene expression cassette that contained the *aad-1* transgene (U.S. Pat. No. 7,838,733) driven by the *Zea mays* Ubiquitin 1 promoter (Christensen *et al.*, [1992] *Plant Molecular Biology* 18; 675-689) and was terminated by the *Zea mays* Lipase 3' -UTR (U.S. Pat. No. 7,179,902). A diagram of this gene expression cassette is shown in **Fig. 1** and is provided as SEQ ID NO:15.

[00150] This construct was built by synthesizing the newly designed *Zea mays*

GRMZM2G144030 promoter and *Zea mays* GRMZM2G144030 3' UTR sequences by an external provider (Geneart via Life Technologies, Carlsbad, CA), and cloning the promoter into a Gateway™ (Life Technologies) donor vector using the GeneArt® Seamless Cloning and Assembly Kit (Life technologies) and restriction enzymes. The resulting donor vector was integrated into a final binary destination vector using the Gateway™ cloning system (Life Technologies). Clones of pDAB113002 were obtained and confirmed via restriction enzyme digestions and sequencing. The resulting construct contained a promoter that could robustly drive expression of a transgenes which was operably linked to the 3' end of the promoter.

[00151] A control construct, pDAB101556, was assembled containing the *Phi-yfp* transgene (Shagin *et al.*, [2004] *Mol Biol Evol* 21; 841-50) driven by the *Zea mays* Ubiquitin-1 Promoter (Christensen *et al.*, [1992] *Plant Molecular Biology* 18; 675-689) and *Zea mays* Peroxidase5 3'UTR regulatory elements. This control construct contained the same *aad-1* expression cassette as present in pDAB113002. This control construct was transformed into plants using the same reagents and protocols as those for pDAB113002.

Example 3: Maize Transformation

Transformation of *Agrobacterium tumefaciens*:

[00152] The binary expression vector was transformed into *Agrobacterium tumefaciens* strain DA113192 (RecA deficient ternary strain) (Int'l. Pat. Pub. No. WO2012016222). Bacterial colonies were selected, and binary plasmid DNA was isolated and confirmed via restriction enzyme digestion.

Agrobacterium Culture Initiation:

[00153] *Agrobacterium* cultures were streaked from glycerol stocks onto AB minimal medium (Gelvin, S., 2006, *Agrobacterium* Virulence Gene Induction, in Wang, K., ed., *Agrobacterium Protocols Second Edition Vol. 1*, Humana Press, p. 79; made without sucrose and with 5 g/L glucose and 15 g/L Bacto™ Agar) and incubated at 20°C in the dark for 3 days. *Agrobacterium* cultures were then streaked onto a plate of YEP medium (Gelvin, S., 2006,

Agrobacterium Virulence Gene Induction, in Wang, K., ed., *Agrobacterium Protocols Second Edition Vol. 1*, Humana Press, p. 79) and incubated at 20 °C in the dark for 1 day.

[00154] On the day of an experiment, a mixture of Inoculation medium (2.2 g/L MS salts, 68.4 g/L sucrose, 36 g/L glucose, 115 mg/L L-proline, 2 mg/L glycine, 100 mg/L myo-Inositol, 0.05 mg/L nicotinic acid, 0.5 mg/L pyridoxine HCl, 0.5 mg/L thiamine HCl) and acetosyringone was prepared in a volume appropriate to the size of the experiment. A 1 M stock solution of acetosyringone in 100% dimethyl sulfoxide was added to the Inoculation medium to make a final acetosyringone concentration of 200 µM.

[00155] For each construct, 1-2 loops of *Agrobacterium* from the YEP plate were suspended in 15 ml of the inoculation medium/acetosyringone mixture inside a sterile, disposable, 50 ml centrifuge tube and the optical density of the solution at 600 nm (O.D.₆₀₀) was measured in a spectrophotometer. The suspension was then diluted down to 0.25-0.35 O.D.₆₀₀ using additional Inoculation medium/acetosyringone mixture. The tube of *Agrobacterium* suspension was then placed horizontally on a platform shaker set at about 75 rpm at room temperature for between 1 and 4 hours before use.

Maize Transformation:

[00156] Experimental constructs were transformed into maize via *Agrobacterium*-mediated transformation of immature embryos isolated from the inbred line, *Zea mays* c.v. B104. The method used is similar to those published by Ishida et al., (1996) *Nature Biotechnol*; 14:745–750 and Frame et al., (2006) *Plant Cell Rep*; 25: 1024–1034, but with several modifications and improvements to make the method amenable to high-throughput transformation. An example of a method used to produce a number of transgenic events in maize is given in U.S. Pat. App. Pub. No. US 2013/0157369 A1, beginning with the embryo infection and co-cultivation steps.

[00157] Putative T₀ transgenic plantlets were transplanted from Phytatrays™ (Sigma-Aldrich; St. Louis, MO) to small 3.5” plastic pots (T. O. Plastics ; Clearwater, MN) filled with growing media (Premix BX; Premier Tech Horticulture), covered with humidomes (Arco Plastics Ltd.), and then hardened-off in a growth room (28°C day/24°C night, 16-hour photoperiod, 50-

70% RH, 200 $\mu\text{Em}^{-2} \text{ sec}^{-1}$ light intensity). When plants reached the V3-V4 developmental stage (3-4 leaf collars visible), they were transplanted into Sunshine Custom Blend 160™ soil mixture and grown to flowering in the greenhouse (Light Exposure Type: Photo or Assimilation; High Light Limit: 1200 PAR; 16-hour day length; 27°C day/24°C night). The plants were analyzed for transgene copy number by qPCR assays using primers designed to detect relative copy numbers of the transgenes, and putative single copy events selected for advancement were transplanted into 5 gallon pots.

Example 4: Molecular Confirmation of Gene Copy Number and Protein Expression

Transgene presence and copy number estimation:

[00158] The maize plants were sampled at V2-3 leaf stage and screened for transgene presence and their copy number using *cry34Ab1* and *aad-1* quantitative PCR assays. Total DNA was extracted from the leaf samples, using MagAttract DNA extraction kit™ from Qiagen™ as per manufacturer's instructions.

[00159] DNA fragments were then amplified with TaqMan® primer/probe sets containing a FAM-labeled fluorescent probe for the *cry34Ab1* gene and a HEX-labeled fluorescent probe for the endogenous *invertase* reference gene. The following primers were used for the *cry34Ab1* and *invertase* gene amplifications.

Cry34Ab1 Primers/probes:

SEQ ID NO:6 (TQ.8v6.1.F): GCCATACCCTCCAGTTG

SEQ ID NO:7 (TQ.8v6.1.R): GCCGTTGATGGAGTAGTAGATGG

Probe: SEQ ID NO:8 (TQ.8v6.1.MGB.P): 5'-FAM/ CCGAATCCAACGGCTTCA / MGB/-3'

Invertase Primers:

SEQ ID NO:9 (InvertaseF): TGGCGGACGACGACTTGT

SEQ ID NO:10 (InvertaseR): AAAGTTTGGAGGCTGCCGT

SEQ ID NO:11 (InvertaseProbe): 5'-HEX/CGA GCA GAC CGC CGT GTA CTT /3BHQ_1/ -3'

[00160] PCR reactions were carried out in a final volume of 10 μ l containing 5 μ l of Roche LightCycler 480 Probes Master Mix™ (Roche Applied Sciences, Indianapolis, IN; Catalog 04887301001); 0.4 μ l each of TQ.8v6.1.F, TQ.8v6.1.R, InvertaseF, and InvertaseR primers from 10 μ M stocks to a final concentration of 400 nM; 0.4 μ l each of the probes, TQ.8v6.1.MGB.P and InvertaseProbe, from 5 μ M stocks to a final concentration of 200 nM, 0.1 μ l of 10% polyvinylpyrrolidone (PVP) to a final concentration of 0.1%; 2 μ l of 10 ng/ μ l genomic DNA and 0.5 μ l water. DNA was amplified in a Roche LightCycler 480 System™ under the following conditions: 1 cycle of 95°C for 10 min; 40 cycles of the following 3-steps: 95°C for 10 seconds; 58°C for 35 seconds and 72°C for 1 second, and a final cycle of 4°C for 10 seconds. *cry34Ab1* copy number was determined by comparing Target/Reference values for the unknown samples (output by the LightCycler 480) to Target/Reference values of *cry34Ab1* copy number controls.

[00161] *Aad-1* gene detection was carried out as described above for the *cry34Ab1* gene using the *invertase* endogenous reference gene. *Aad-1* primer sequences were as follows; PCR cycles remained the same:

SEQ ID NO:12 (AAD1 Forward Primer): TGTTTCGGTTCCTCTACCAA

SEQ ID NO:13 (AAD1 Reverse Primer): CAACATCCATCACCTTGACTGA

SEQ ID NO:14 (AAD1 Probe): 5'-FAM/CACAGAACCGTCGCTTCAGCAACA-MGB/BHQ3'

T₀ plant screening for transgene expression:

[00162] T₀ plants containing the *cry34Ab1* and *aad-1* transgenes and growing in the greenhouse were sampled at the V4-5 developmental stage for leaf ELISA assays. Four leaf punches were sampled. Protein extracts for ELISA assays were prepared by adding one 1/8" stainless steel bead (Hoover Precision Products, Cumming, GA, USA) to each 1.2 ml tube containing the leaf punches and 300 μ l extraction buffer (1X PBST [Fischer Scientific, St. Louis, MO] supplemented with 0.05% Tween 20 and 0.5% BSA). The samples were processed in a Genogrinder™ (SPEX

SamplePrep, Metuchen, NJ) at 1,500 rpm for 4 minutes. The samples were centrifuged at 4,000 rpm for 2 minutes in a Sorvall Legend XFR™ centrifuge. Following this step, an additional 300 µl extraction buffer was added to the samples and they were processed once more in the Genogrinder™ at 1,500 rpm for 2 minutes. The samples were centrifuged at 4,000 rpm for 7 minutes. The supernatant was collected and completed ELISA at different dilutions along with Cry34Ab1 and AAD-1 protein standards. Cry34Ab1 (Agdia, Inc.; Cat No 04500/4800), AAD-1 (Acadia BioScience, LLC; Cat No ABS-041) ELISA assays were performed as per the manufacturer's instructions and the ELISA results were expressed either as ng/cm² leaf surface area or as parts per million (or ng target protein per mg of total plant protein).

[00163] Another set of plants were sampled at V4-5 for the entire root mass. The samples were instantly frozen and lyophilized for a week and then ground. The ELISA was performed as described above for the leaf samples. Total root protein estimation were carried out with Bradford detection method as per the manufacturer's instructions (Thermo Scientific/Pierce, USA). Root ELISA results were expressed as parts per million (or ng target protein per mg of total plant protein).

T₁ plant screening for transgene detection and gene expression:

[00164] T₀ plants were reciprocally crossed with *Zea mays* c.v. B104 to obtain T₁ seed. Three to five T₁ lines (or events) of each of the regulatory element constructs were advanced for protein expression studies. About forty T₁ seed of each of the events were sown, the seedlings at the V2-3 stage of development were sprayed with Assure® II to kill null plants. All surviving plants were sampled for transgene copy number assays, which were carried out as described above.

[00165] For transgene gene expression analysis, plants were sampled at multiple growth and developmental stages as follows: leaf (V4, V12 and R3); root (V4); stem, pollen, silk (all at R1) and, kernel and cob (all at R3). All tissues were sampled in tubes embedded in dry ice; which were then transferred to -80°C immediately following the sampling completion. Frozen tissues were lyophilized prior to protein extraction for ELISA.

[00166] Protein extraction for leaf ELISA was carried out as described for T₀ samples as

described in the previous section. For instance, protein extraction for various tissue type ELISA was carried out by grinding the lyophilized tissue in 50 ml tubes in a paint shaker for 30 seconds in the presence of eight 0.25" ceramic beads (MP Biomedicals, USA). The step was repeated for certain tissues needing further grinding for another 30 seconds. Protein was then extracted in 2 ml polypropylene tubes containing enough garnet powder to cover the curved bottom portion of the tubes. The coarsely ground tissue was transferred to the 2 ml tubes to fill up to 0.3 ml mark. One 0.25" ceramic ball was then added to each tube and 0.6 ml of the extraction buffer (200 μ l of protease inhibitor cocktail [Research Products International Corp., Solon, OH, USA], 200 μ l of 500 mM EDTA, 15.5 mg DTT powder and PBST to 20 ml). All tubes were kept on ice for 10 minutes and then processed for 45 seconds in the Genogrinder™. Next, 40 μ l of 10% Tween-20 was added and another 300 μ l extraction buffer to the tubes and grinded samples for another 45 seconds. The tubes were centrifuged at 13,000 rpm for 7-14 minutes. The supernatant was carefully transferred to a new tube. For ELISA assays, the extract was diluted in the extraction buffer as needed. The ELISA results were expressed either as ng/cm² leaf surface area or as parts per million (or ng protein per mg of total plant protein).

Example 5: Expression of Genes Operably Linked to the *Zea mays* GRMZM2G144030 Promoter and 3'UTR Regulatory Elements

[00167] Maize plants were transformed with a gene expression construct that contained the *Zea mays* GRMZM2G144030 promoter and the *Zea mays* GRMZM2G144030 as described above. The ELISA analysis confirmed that the novel promoter drove expression of a transgene, and that the novel 3'UTR effectively terminated expression of the transgene. The quantitative measurements of Cry34Ab1 protein obtained from transgenic plants comprising novel promoter constructs are shown in Table 1. The data showed that Cry34Ab1 protein in the plants containing the novel *Zea mays* GRMZM2G144030 promoter and the *Zea mays* GRMZM2G144030 3' UTR (i.e., pDAB113002) is expressed preferentially in root, stem, cob, kernel and silk tissues as compared to marginal-low expression in tissues like leaf.

TABLE 1: *Zea mays* GRMZM2G144030 Promoter T0 Expression of Cry34Ab1 and AAD-1. The leaf ELISA was measured in ng of target protein per cm² of leaf surface area while the remaining tissue type ELISA was measured in ng of target protein per mg of total protein.

Construct No.	Plant Stage	Total Events Analyzed	Total Samples Analyzed	Cry34Ab1 Mean	Cry34Ab1 STD	AAD-1 Mean	AAD-1 STD
101556	Leaf V4	2	30	0	0	49	32
113002	Leaf V4	5	76	28	9	79	41
101556	Leaf V12	2	18	0	0	184	119
113002	Leaf V12	5	27	36	12	219	98
101556	R3 Leaf	2	6	0	0	349	198
113002	R3 Leaf	5	13	63	12	467	188
101556	Root V4	2	4	0	0	1437	584
113002	Root V4	2	6	1119	331	2835	2104
113002	Root V12	2	11	1063	325	ND	ND
113002	Pollen (R1)	3	6	Yes*	ND	ND	ND
113002	Stem (R1)	3	6	2192	692	ND	ND
113002	Kernel (20 DAP)	3	6	1491	599	ND	ND
113002	Silk (R1)	3	6	1019	679	ND	ND
113002	Cob (20 DAP)	3	6	1127	334	ND	ND

* Only qualitative assay was completed

ND = not determined

[00168] The Cry34Ab1 ELISA results indicated that the *Zea mays* GRMZM2G144030 promoter regulatory element (SEQ ID NO:1) and the *Zea mays* GRMZM2G144030 3'UTR (SEQ ID NO:5) drove root, stem, cob, kernel and silk preferred expression of Cry34Ab1 in T₁ events that were transformed with construct, pDAB113002. The events produced from the transformation also expressed AAD-1 protein in both leaf and root tissues, as expected. In summary, the *Zea mays* GRMZM2G144030 promoter was developed to show high levels of expression of a transgene in leaf tissues in a plant species.

Example 6: Crop Transformation of Genes Operably Linked to the *Zea mays* GRMZM2G144030

Promoter

[00169] Soybean may be transformed with genes operably linked to the *Zea mays* GRMZM2G144030 promoter by utilizing the same techniques previously described in Example #11 or Example #13 of patent application WO 2007/053482.

[00170] Cotton may be transformed with genes operably linked to the *Zea mays* GRMZM2G144030 promoter by utilizing the same techniques previously described in Examples #14 of U.S. Patent No. 7,838,733 or Example #12 of patent application WO 2007/053482 (Wright *et al.*).

[00171] Canola may be transformed with genes operably linked to the *Zea mays* GRMZM2G144030 promoter by utilizing the same techniques previously described in Example #26 of U.S. Patent No. 7,838,733 or Example #22 of patent application WO 2007/053482 (Wright *et al.*).

[00172] Wheat may be transformed with genes operably linked to the *Zea mays* GRMZM2G144030 promoter by utilizing the same techniques previously described in Example #23 of patent application WO 2013/116700A1 (Lira *et al.*).

[00173] Rice may be transformed with genes operably linked to the *Zea mays* GRMZM2G144030 promoter by utilizing the same techniques previously described in Example #19 of patent application WO 2013/116700A1 (Lira *et al.*).

Example 7: *Agrobacterium*-mediated Transformation of Genes Operably Linked to the *Zea mays* GRMZM2G144030 Promoter

[00174] In light of the subject disclosure, additional crops can be transformed according to embodiments of the subject disclosure using techniques that are known in the art. For *Agrobacterium*-mediated transformation of rye, see, e.g., Popelka JC, Xu J, Altpeter F., "Generation of rye with low transgene copy number after biolistic gene transfer and production of (*Secale cereale* L.) plants instantly marker-free transgenic rye," *Transgenic Res.* 2003 Oct;12(5):587-96.). For *Agrobacterium*-mediated transformation of sorghum, see, e.g., Zhao *et al.*, "*Agrobacterium*-mediated sorghum transformation," *Plant Mol Biol.* 2000 Dec;44(6):789-98. For *Agrobacterium*-mediated

transformation of barley, see, e.g., Tingay et al., “*Agrobacterium tumefaciens*-mediated barley transformation,” *The Plant Journal*, (1997) 11: 1369–1376. For *Agrobacterium*-mediated transformation of wheat, see, e.g., Cheng et al., “Genetic Transformation of Wheat Mediated by *Agrobacterium tumefaciens*,” *Plant Physiol.* 1997 Nov;115(3):971-980. For *Agrobacterium*-mediated transformation of rice, see, e.g., Hiei et al., “Transformation of rice mediated by *Agrobacterium tumefaciens*,” *Plant Mol. Biol.* 1997 Sep;35(1-2):205-18.

[00175] The latin names for these and other plants are given below. It should be clear that other (non-*Agrobacterium*) transformation techniques can be used to transform genes operably linked to the *Zea mays* GRMZM2G144030 promoter f, for example, into these and other plants. Examples include, but are not limited to; Maize (*Zea mays*), Wheat (*Triticum* spp.), Rice (*Oryza* spp. and *Zizania* spp.), Barley (*Hordeum* spp.), Cotton (*Abroma augusta* and *Gossypium* spp.), Soybean (*Glycine max*), Sugar and table beets (*Beta* spp.), Sugar cane (*Saccharum officinarum* and other spp.) Feather palm (*Arenga pinnata*), Tomato (*Lycopersicon esculentum* and other spp., *Physalis ixocarpa*, *Solanum incanum* and other spp., and *Cyphomandra betacea*), Potato (*Solanum tuberosum*), Sweet potato (*Ipomoea batatas*), Rye (*Secale* spp.), Peppers (*Capsicum annuum*, *chinense*, and *frutescens*), Lettuce (*Lactuca sativa*, *perennis*, and *pulchella*), Cabbage (*Brassica* spp.), Celery (*Apium graveolens*), Eggplant (*Solanum melongena*), Peanut (*Arachis hypogea*), Sorghum (*Sorghum* spp.), Alfalfa (*Medicago sativa*), Carrot (*Daucus carota*), Beans (*Phaseolus* spp. and other genera), Oats (*Avena sativa* and *strigosa*), Peas (*Pisum*, *Vigna*, and *Tetragonolobus* spp.), Sunflower (*Helianthus annuus*), Squash (*Cucurbita* spp.), Cucumber (*Cucumis sativa*), Tobacco (*Nicotiana* spp.), *Arabidopsis* (*Arabidopsis thaliana*), Turfgrass (*Lolium*, *Agrostis*, *Poa*, *Cynodon*, and other genera), Clover (*Trifolium*), Vetch (*Vicia*). Transformation of such plants, with genes operably linked to the *Zea mays* GRMZM2G144030 promoter, for example, is contemplated in embodiments of the subject disclosure.

[00176] Use of the *Zea mays* GRMZM2G144030 promoter to drive operably linked genes can be deployed in many deciduous and evergreen timber species. Such applications are also within the scope of embodiments of this disclosure. These species include, but are not limited to; alder (*Alnus* spp.), ash (*Fraxinus* spp.), aspen and poplar species (*Populus* spp.), beech (*Fagus* spp.), birch

(*Betula* spp.), cherry (*Prunus* spp.), eucalyptus (*Eucalyptus* spp.), hickory (*Carya* spp.), maple (*Acer* spp.), oak (*Quercus* spp.), and pine (*Pinus* spp.).

[00177] Use of the *Zea mays* GRMZM2G144030 promoter to drive operably linked genes can be deployed in ornamental and fruit-bearing species. Such applications are also within the scope of embodiments of this disclosure. Examples include, but are not limited to; rose (*Rosa* spp.), burning bush (*Euonymus* spp.), petunia (*Petunia* spp.), begonia (*Begonia* spp.), rhododendron (*Rhododendron* spp.), crabapple or apple (*Malus* spp.), pear (*Pyrus* spp.), peach (*Prunus* spp.), and marigolds (*Tagetes* spp.).

[00178] While a number of exemplary aspects and embodiments have been discussed above, those of skill in the art will recognize certain modifications, permutations, additions and sub-combinations thereof. It is therefore intended that the following appended claims and claims hereafter introduced are interpreted to include all such modifications, permutations, additions and sub-combinations as are within their true spirit and scope.

CLAIMS

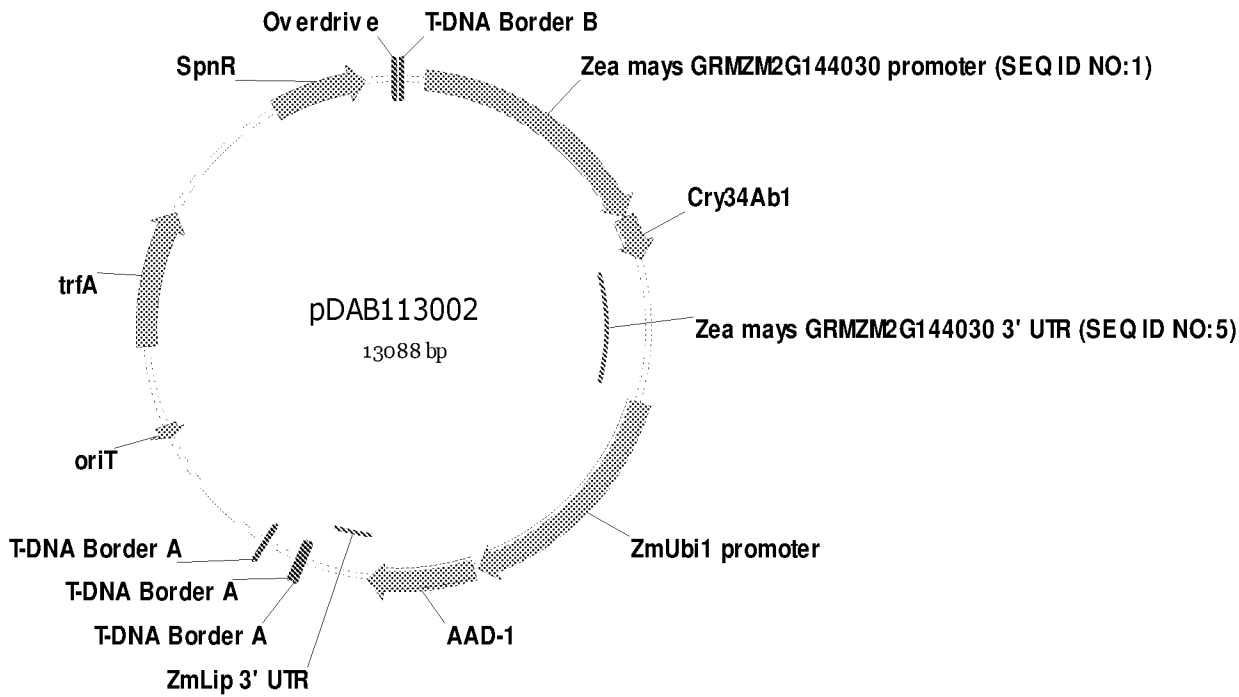
What is claimed is:

1. A nucleic acid vector comprising a promoter operably linked to:
 - a) a polylinker sequence;
 - b) a non-GRMZM2G144030 gene; or
 - c) a combination of a) and b), wherein said promoter comprises a polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:1.
2. The nucleic acid vector of claim 1, wherein said promoter is 2,082 bp in length.
3. The nucleic acid vector of claim 1, wherein said promoter consists of a polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:1.
4. The nucleic acid vector of any one of claims 1-3, further comprising a sequence encoding a selectable maker.
5. The nucleic acid vector of claim 1, wherein said promoter is operably linked to a transgene.
6. The nucleic acid vector of claim 5, wherein the transgene encodes a selectable marker or a gene product conferring insecticidal resistance, herbicide tolerance, nitrogen use efficiency, small RNA expression, site specific nuclease, water use efficiency, nutritional quality or DNA binding proteins.
7. The nucleic acid vector of any of claims 1-3, or 5, further comprising a 3' untranslated polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:5, wherein the 3' untranslated sequence is operably linked to said polylinker or said transgene.
8. The nucleic acid vector of any of claims 1-3, or 5, further comprising a 5' untranslated polynucleotide sequence, wherein the 5' untranslated sequence is operably linked to said polylinker or said transgene.

9. The nucleic acid vector of any of claims 1-3, or 5, further comprising an intron sequence.
10. The nucleic acid vector of claim 1, wherein said promoter drives transgene expression in leaf, root, stem, cob, kernel, pollen and silk tissues.
11. A non-*Zea mays* c.v. B73 plant comprising a polynucleotide sequence that has at least 90% sequence identity with SEQ ID NO:1 operably linked to a transgene.
12. The plant of claim 11, wherein said plant is selected from the group consisting of wheat, rice, sorghum, oats, rye, bananas, sugar cane, soybean, cotton, Arabidopsis, tobacco, sunflower, and canola.
13. The plant of claim 11, wherein said plant is maize.
14. The plant of any one of claims 11-13, wherein the transgene is inserted into the genome of said plant.
15. The plant of claim 11, wherein a promoter comprises a polynucleotide sequence having at least 90% sequence identity with SEQ ID NO:1 and said promoter is operably linked to a transgene.
16. The plant of claim 15, further comprising a 3' untranslated sequence comprising SEQ ID NO:5, wherein the 3' untranslated sequence is operably linked to said transgene.
17. The plant of claim 15, wherein said promoter drives transgene expression root, stem, cob, kernel and silk tissues.
18. The plant of claim 15, wherein said promoter is 2,082 bp in length.
19. A method for producing a transgenic plant cell, the method comprising the steps of:
 - a) transforming a plant cell with a gene expression cassette comprising a *Zea mays* GRMZM2G144030 promoter operably linked to at least one polynucleotide sequence of interest;
 - b) isolating the transformed plant cell comprising the gene expression cassette; and,

- c) producing a transgenic plant cell comprising the *Zea mays* GRMZM2G144030 promoter operably linked to at least one polynucleotide sequence of interest.
20. A method for expressing a polynucleotide sequence of interest in a plant cell, the method comprising introducing into the plant cell a polynucleotide sequence of interest operably linked to a *Zea mays* GRMZM2G144030 promoter.

FIG. 1:



International application No.

PCT/US17/26969

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 15/82, 15/79, 15/63 (2017.01)
CPC - C12N 15/82, 15/63, 15/8216, 15/8201, 15/8241

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)
See Search History document

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015/0067926 A1 (DOW AGROSCIENCES, LLC) March 5, 2015; paragraphs [0007], [0010], [0012], [0104], [0113], [0189], [0192], [0205]; claim 1	1-3, 4/1-3, 5-6, 7/1-3, 7/5, 8/1-3, 8/5, 9/1-3, 9/5, 10-13, 14/11-13, 15-20
A	US 2007/0130645 A1 (WU, et al.) June 7, 2007; paragraphs [0019]-[0022], [0027], [0220], [0232]; claim 1	1-3, 4/1-3, 5-6, 7/1-3, 7/5, 8/1-3, 8/5, 9/1-3, 9/5, 10-13, 14/11-13, 15-20
A	NELSON, et al. Methylation-Sensitive Linking Libraries Enhance Gene-Enriched Sequencing Of Complex Genomes And Map DNA Methylation Domains. BMC Genomics. 19 December 2008, Vol. 9, pages 1-16, doi:10.1186/1471-2164-9-621; page 4, second column, first paragraph	1-3, 4/1-3, 5-6, 7/1-3, 7/5, 8/1-3, 8/5, 9/1-3, 9/5, 10-13, 14/11-13, 15-18
A	WO 2015/031733 A1 (DOW AGROSCIENCES LLC) March 5, 2015; page 3, lines 4-11, 17-28; page 19, line 22; page 22, lines 18-19; page 29, lines 13-16; page 31, lines 5-13; page 39, lines 13-16; page 48, lines 5-34; page 52, lines 23-24; claim 1	1-3, 4/1-3, 5-6, 7/1-3, 7/5, 8/1-3, 8/5, 9/1-3, 9/5, 10-13, 14/11-13, 15-20
A	WO 2016/044090 A1 (E.I. DU PONT DE NEMOURS AND COMPANY) March 24, 2016; page 6, fourth paragraph-page 7, first paragraph.	19, 20

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

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"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
06 July 2017 (06.07.2017)

Date of mailing of the international search report

20 JUL 2017

Name and mailing address of the ISA/
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-8300

Authorized officer
Shane Thomas

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/26969

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments: