



(43) International Publication Date
29 November 2012 (29.11.2012)

(10) International Publication Number
WO 2012/159891 A1

- (51) **International Patent Classification:**
C07K 14/405 (2006.01) *C12N 15/82* (2006.01)
- (21) **International Application Number:**
PCT/EP2012/058643
- (22) **International Filing Date:**
10 May 2012 (10.05.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/488,290 20 May 2011 (20.05.2011) US
- (71) **Applicant** (for all designated States except US): **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).
- (72) **Inventors; and**
- (75) **Inventors/Applicants** (for US only): **DEBRECHT, Andrew, August** [US/US]; 3054 E. Cornwallis Road, Durham, NC 27709 (US). **BETTS, Scott** [US/US]; 3054 E. Cornwallis Road, Durham, NC 27709 (US). **ZHOU, Nan** [US/US]; 3054 E. Cornwallis Road, Durham, NC 27709 (US).
- (74) **Agent:** **RADKOV, Stoyan**; Schwarzwaldallee 215, CH-4058 Basel (CH).
- (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, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, 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))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** ENDOSPERM-SPECIFIC PLANT PROMOTERS AND USES THEREFOR

(57) **Abstract:** Promoters that include nucleic acid sequences with a sequence identity to any of SEQ ID NOs: 1-3 of greater than 98% over at least 100, 200, 500, 1000, or 1500 consecutive nucleotides are provided. Also provided are transcription terminators, nucleic acids that include the promoters and/or terminators, expression cassettes that include the disclosed nucleic acid molecules, transformed plants containing the disclosed nucleic acid molecules, cells containing the disclosed promoters operably linked to heterologous nucleotide sequences, methods for producing a transformed plants, and methods for identifying non-*Zea mays* nucleic acid molecules with transcriptional regulatory activity.



WO 2012/159891 A1

DESCRIPTION

ENDOSPERM-SPECIFIC PLANT PROMOTERS AND USES THEREFOR

TECHNICAL FIELD

5 The presently disclosed subject matter relates generally to new plant promoters and terminators, nucleic acids identified therewith, and compositions comprising the identified nucleic acids. More particularly, the presently disclosed subject matter relates to methods for discovering *Zea mays* promoters and to nucleic acid molecules comprising promoters and terminators discovered using
10 the disclosed methods. Also provided are expression cassettes comprising the disclosed nucleic acids; cells and plants containing the disclosed nucleic acids as well as methods for generating the same; and methods for identifying transcriptional regulatory elements.

BACKGROUND

15 An objective of crop trait functional genomics is to identify crop trait genes of interest, for example, genes capable of conferring useful agronomic traits in crop plants. Such agronomic traits include, but are not limited to, enhanced yield, whether in quantity or quality; enhanced nutrient acquisition and metabolic efficiency; enhanced or altered nutrient composition of plant tissues used for
20 food, feed, fiber, or processing; enhanced utility for agricultural or industrial processing; enhanced resistance to plant diseases; enhanced tolerance of adverse environmental conditions including, but not limited to, drought, excessive cold, excessive heat, or excessive soil salinity or extreme acidity or alkalinity; and alterations in plant architecture or development, including changes in
25 developmental timing. The deployment of such identified trait genes by either transgenic or non-transgenic approaches can materially improve crop plants for the benefit of agriculture.

 The identification of genes that are important for crop development is thus an ongoing effort in the agricultural community. Additional information can also
30 be derived by analyzing the genomes of important plants. For example, the identification of transcriptional regulatory elements that can direct the expression

of linked nucleotide sequences can also lead to the ability to manipulate the plant genome to express polypeptides of interest in desirable spatial and/or temporal manners.

Among the transcriptional regulatory elements that can be employed, promoters play important roles, as transcription initiation is often a rate-limiting step in the expression of polypeptides of interest, particularly relative to subsequent stages such as the translation of mRNAs. As such, the characteristics of the promoter can be an important consideration in the selection of how to best accomplish protein production through transgenesis.

Promoters are capable of regulating transcription initiation in several ways. For example, certain promoters can be induced by the presence of particular compounds and/or external stimuli, such that they direct expression of operably linked nucleotide sequences only in specific tissues and/or cell types and/or during specific stages of development. Other promoters are capable of constitutively expressing linked sequences. Thus, the transcription of a coding sequence of interest can be regulated by operably linking the coding sequence to whichever promoter can provide the desired regulatory characteristics. As such, different promoters can be employed in different ways to enhance the agronomic, pharmaceutical, and/or nutritional value of crops.

What are needed, then, are new methods and compositions for expressing heterologous nucleotide sequences in plant cells. To meet these needs, the presently disclosed subject matter provides in some embodiments promoter sequences for directing expression of heterologous nucleotide sequences in plant cells. The presently disclosed subject matter addresses these problems associated with the expression of nucleotide sequences in transgenic plants, as well as other problems.

SUMMARY

This Summary lists several embodiments of the presently disclosed subject matter, and in many cases lists variations and permutations of these embodiments. This Summary is merely exemplary of the numerous and varied

embodiments. Mention of one or more representative features of a given embodiment is likewise exemplary. Such an embodiment can typically exist with or without the feature(s) mentioned; likewise, those features can be applied to other embodiments of the presently disclosed subject matter, whether listed in
5 this Summary or not. To avoid excessive repetition, this Summary does not list or suggest all possible combinations of such features.

In some embodiments, the presently disclosed subject matter provides isolated nucleic acid molecules comprising a plant promoter. In some
10 embodiments, the isolated nucleic acid molecules comprise a nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 1-3. In some embodiments, the nucleotide sequence is 100% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 1-3.

In some embodiments, the isolated nucleic acid molecules of the presently
15 disclosed subject matter further comprise a first exon, an intron, and optionally a second exon or a fragment thereof, operably linked thereto. In some embodiments, the promoter and the first exon and intron, and optionally the second exon or the fragment thereof if present, are derived from the same locus of a non-transgenic plant. In some embodiments, (a) the first exon is the 5' most
20 exon of the locus; and/or (b) the intron is the 5' most intron; and/or (c) the second exon or the fragment thereof, if present, is the exon immediately downstream of the intron in the genome of the non-transgenic plant.

In some embodiments, the isolated nucleic acid molecules further
25 comprise a transcription terminator operably linked to the plant promoter. In some embodiments, the transcription terminator and the plant promoter are derived from the same locus of a non-transgenic plant.

In some embodiments, the isolated nucleic acid molecules further
comprise a heterologous nucleotide sequence operably linked to the plant promoter. In some embodiments, the isolated nucleic acid molecule comprises:

(a) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 1 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 4; or

5 (b) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 5; or

10 (c) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 6.

The presently disclosed subject matter also provides in some embodiments isolated nucleic acid molecules comprising the following elements
15 operably linked in 5' to 3' order: (i) a plant promoter derived from a genetic locus; (ii) a heterologous nucleotide sequence; and (iii) a transcription terminator, wherein the isolated nucleic acid molecule comprises a nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 1-3. In some embodiments, the intron, the transcription terminator, or
20 both are also derived from the genetic locus. In some embodiments, the isolated nucleic acid molecule comprises:

(a) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 1 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive
25 nucleotides to SEQ ID NO: 4; or

(b) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 2 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 5; or

(c) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 6.

5 In some embodiments, the at least one exon is the first exon of the genetic locus, the intron is the first exon of the genetic locus, or both. In some embodiments, the heterologous nucleotide sequence encodes an RNA molecule selected from the group consisting of an mRNA for a polypeptide of interest and an inhibitory RNA. In some embodiments,

10 The presently disclosed subject matter also provides in some embodiments isolated nucleic acid molecules comprising a plant transcription terminator. In some embodiments, the isolated nucleic acid molecule comprises a nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 4-6. In some embodiments, the isolated
15 nucleic acid molecule comprises a nucleotide sequence that is 100% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 4-6.

The presently disclosed subject matter also provides in some embodiments isolated nucleic acid molecules comprising plant promoters operably linked to heterologous nucleotide sequences. In some embodiments,
20 the isolated nucleic acid molecule comprises:

(a) a nucleotide sequence having promoter activity that is at least 95% identical to any of SEQ ID NOs: 1-3 over at least 300 consecutive nucleotides; or

(b) a nucleotide sequence having promoter activity that comprises a fragment of at least about 100 consecutive nucleotides of any of SEQ ID NOs: 1-
25 3; or

(c) a nucleotide sequence with a sequence identity of greater than 95% to one of SEQ ID NOs: 1-3 over the full length of the one of SEQ ID NOs: 1-3, wherein the nucleotide sequence has promoter activity; or

(d) the nucleotide sequence of any of SEQ ID NOs: 1-3.

In some embodiments, the heterologous nucleotide sequence comprises one or more cloning sites. In some embodiments, the heterologous nucleotide sequence comprises a polylinker. In some embodiments, the heterologous nucleotide sequence comprises a coding sequence. In some embodiments, the heterologous nucleotide sequence encodes a reporter. In some embodiments, the reporter is a β -glucuronidase polypeptide.

In some embodiments, the isolated nucleic acid molecule further comprises a transcription terminator operably linked to the plant promoter and the heterologous nucleotide sequence. In some embodiments, the transcription terminator is derived from the same locus from which the plant promoter is derived. In some embodiments, the plant promoter comprises:

- (a) SEQ ID NO: 1 and the transcription terminator comprises SEQ ID NO: 4; or
- (b) SEQ ID NO: 2 and the transcription terminator comprises SEQ ID NO: 5; or
- (c) SEQ ID NO: 3 and the transcription terminator comprises SEQ ID NO: 6.

In some embodiments, the transcription terminator is heterologous to the gene from which the isolated promoter is derived.

The presently disclosed subject matter also provides in some embodiments expression cassettes comprising the disclosed isolated nucleic acid molecules. In some embodiments, the isolated nucleic acid molecule further comprises a plant promoter operably linked to its endogenous first exon, its endogenous first intron, its endogenous second exon or a fragment thereof, its endogenous transcription terminator, or any combination thereof.

The presently disclosed subject matter also provides in some embodiments transformed plants containing the disclosed isolated nucleic acid molecules. In some embodiments, the transformed plants comprise heterologous nucleotide sequences, which in some embodiments can be oriented to express antisense RNA molecules. In some embodiments, the

heterologous nucleotide sequence is expressed constitutively with the exception of pollen, in which expression of the heterologous nucleotide sequence is low or is absent. In some embodiments, the heterologous nucleotide sequence is expressed constitutively with the exception of pollen. In some embodiments, the transformed plant is maize.

The presently disclosed subject matter also provides in some embodiments cells containing the isolated nucleic acid molecules disclosed herein operably linked to a heterologous nucleotide sequence. In some embodiments, the cell is selected from the group consisting of a bacterial cell, a mammalian cell, an insect cell, a plant cell, and a fungal cell. In some embodiments, the cell is *Agrobacterium tumefaciens*.

The presently disclosed subject matter also provides in some embodiments methods for producing transformed plants. In some embodiments, the methods comprise (a) providing a nucleic acid construct comprising a presently disclosed isolated nucleic acid molecule; and (b) transforming a plant with the nucleic acid construct.

The presently disclosed subject matter also provides in some embodiments methods for producing biomolecules in transgenic plants. In some embodiments, the methods comprising transforming a plant with a nucleic acid construct comprising an isolated nucleic acid molecule as disclosed herein, wherein the heterologous nucleotide sequence encodes the biomolecule and is expressed in the transgenic plant. In some embodiments, the nucleic acid construct comprises a promoter, an intron, and/or a transcription terminator, and in some embodiments one, two, or all three of these elements are derived from a single genetic locus. In some embodiments, the nucleic acid construct comprises:

(a) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 1 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 4; or

(b) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 2 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 5 or

- 5 (c) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 6.

10 In some embodiments, the nucleic acid construct further comprises a heterologous nucleotide sequence, which in some embodiments encodes an RNA molecule selected from the group consisting of an mRNA for a polypeptide of interest and an inhibitory RNA.

The presently disclosed subject matter also provides in some embodiments methods for identifying a non-*Zea mays* nucleic acid molecule with transcriptional regulatory activity. In some embodiments, the methods comprise

15 (a) providing a first nucleic acid molecule that comprises a nucleotide sequence that is at least 95% identical to at least 50 consecutive bases of any of SEQ ID NOs: 1-3; (b) hybridizing the first nucleic acid molecule to a plurality of nucleic acid molecules, wherein the plurality of nucleic acid molecules is isolated from a species other than *Zea mays*; (c) isolating a second nucleic acid molecule

20 present in the plurality of nucleic acid molecules that specifically hybridizes to the first nucleic acid molecule; and (d) assaying the second nucleic acid molecule for transcriptional regulatory activity, whereby a non-*Zea mays* nucleic acid molecule with transcriptional regulatory activity is identified. In some embodiments, the

25 plurality of nucleic acid molecules comprises a library of genomic DNA from a plant species other than *Zea mays*.

In some embodiments, the presently disclosed methods further comprise screening a genomic library from the species other than *Zea mays* and identifying a genomic DNA molecule present in the genomic library that is at least

30 1000 nucleotides in length and that comprises a nucleotide sequence that is at

least 95% identical to the nucleotide sequence of the second nucleic acid molecule over at least 500 nucleotides.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

5 SEQ ID NO: 1 is a nucleotide sequence from *Zea mays* chromosome 4 that corresponds to nucleotides 73605-75604 of GENBANK® Accession No. AC193512.3.

 SEQ ID NO: 2 is a nucleotide sequence from *Zea mays* chromosome 4 that corresponds to the reverse complement of nucleotides 6815-8814 of
10 GENBANK® Accession No. AC193444.3.

 SEQ ID NO: 3 is a nucleotide sequence from *Zea mays* chromosome 4 that corresponds to the reverse complement of nucleotides 16935-18934 of GENBANK® Accession No. AC229981.2.

 SEQ ID NO: 4 is a nucleotide sequence from *Zea mays* chromosome 4
15 that corresponds to nucleotides 76409-77408 of GENBANK® Accession No. AC193512.3.

 SEQ ID NO: 5 is a nucleotide sequence from *Zea mays* chromosome 8 that corresponds to the reverse complement of nucleotides 4111-5110 of GENBANK® Accession No. AC193444.3.

20 SEQ ID NO: 6 is a nucleotide sequence from *Zea mays* chromosome 4 that corresponds to the reverse complement of nucleotides 15143-16142 of GENBANK® Accession No. AC229981.2.

 SEQ ID NO: 7 is an amino acid sequence of a putative signal sequence encoded by the Az1 locus that corresponds to SEQ ID NOs: 1 and 4.

25 SEQ ID NO: 8 is an amino acid sequence of the γ -zein signal sequence.

DETAILED DESCRIPTION

 The presently disclosed subject matter will be now be described more fully hereinafter with reference to the accompanying Examples, in which
30 representative embodiments of the presently disclosed subject matter are shown.

The presently disclosed subject matter can, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the presently disclosed subject matter to those skilled in the art.

I. Definitions

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the presently disclosed subject matter pertains. For clarity of the present specification, certain definitions are presented herein below.

Following long-standing patent law convention, the terms "a", "an", and "the" refer to "one or more" when used in this application, including the claims. For example, the phrase "a cell" refers to one or more cells, including, for example, tissues and organs unless the context in which the term appears clearly excludes such an interpretation. Similarly, the phrase "at least one", when employed herein to refer to an entity, refers to, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more of that entity, including but not limited to whole number values between 1 and 100 and greater than 100 as would be understood by one of ordinary skill in the art with respect to the context in which the phrase appears.

Unless otherwise indicated, all numbers expressing quantities of ingredients, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about". The term "about", as used herein when referring to a measurable value such as an amount of mass, weight, time, volume, concentration or percentage is meant to encompass variations of in some embodiments $\pm 20\%$, in some embodiments $\pm 10\%$, in some embodiments $\pm 5\%$, in some embodiments $\pm 1\%$, in some embodiments $\pm 0.5\%$, and in some embodiments $\pm 0.1\%$ from the specified amount, as such variations are appropriate to perform the disclosed methods. Accordingly, unless indicated to the contrary, the numerical parameters set forth

in this specification and attached claims are approximations that can vary depending upon the desired properties sought to be obtained by the presently disclosed subject matter.

As used herein, the term “and/or” when used in the context of a list of
5 entities, refers to the entities being present singly or in combination. Thus, for example, the phrase “A, B, C, and/or D” includes A, B, C, and D individually, but also includes any and all combinations and subcombinations of A, B, C, and D.

The term “comprising”, which is synonymous with “including” “containing”, or “characterized by”, is inclusive or open-ended and does not exclude additional,
10 unrecited elements and/or method steps. “Comprising” is a term of art that means that the named elements and/or steps are present, but that other elements and/or steps can be added and still fall within the scope of the relevant subject matter.

As used herein, the phrase “consisting of” excludes any element, step, or
15 ingredient not specifically recited. It is noted that, when the phrase “consists of” appears in a clause of the body of a claim, rather than immediately following the preamble, it limits only the element set forth in that clause; other elements are not excluded from the claim as a whole.

As used herein, the phrase “consisting essentially of” limits the scope of
20 the related disclosure or claim to the specified materials and/or steps, plus those that do not materially affect the basic and novel characteristic(s) of the disclosed and/or claimed subject matter. For example, a nucleic acid molecule of the presently disclosed subject matter can “consist essentially of” a promoter, a reporter gene coding sequence, and a transcriptional terminator. It is noted,
25 however, that additional nucleotides that are not specifically recited in the corresponding SEQ ID NOs. can also be present, provided that the additional nucleotides do not materially alter the activity of any of the promoter, the reporter gene coding sequence, and the transcriptional terminator.

With respect to the terms “comprising”, “consisting of”, and “consisting
30 essentially of”, where one of these three terms is used herein, the presently

disclosed and claimed subject matter can include the use of either of the other two terms. For example, in some embodiments, the presently disclosed subject matter relates to nucleic acid molecules that comprise plant promoters. It would be understood by one of ordinary skill in the art after review of the instant disclosure that the presently disclosed subject matter thus encompasses nucleic acid molecules that consist essentially of the plant promoters of the presently disclosed subject matter, as well as nucleic acid molecules that consist of the plant promoters of the presently disclosed subject matter.

As used herein, the phrases “associated with”, “operably linked”, and “operatively linked” refer to two or more nucleotide sequences that are related physically or functionally. For example, a promoter or other regulatory DNA sequence is said to be “associated with” a DNA sequence that encodes a RNA or a polypeptide if the two sequences are operably linked, and/or are situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence. Similarly, a coding sequence can be “associated with” or “operably linked” to a promoter which drives the expression of the coding sequence in particular cells or cell types. Also similarly, a transcription terminator can be operably linked to a promoter and to a coding sequence when transcription from the promoter through the coding sequence is terminated by the presence of the terminator. In some embodiments, however, the phrase “operably linked” refers to a nucleotide sequences that are present in a single nucleic acid molecule (for example, an expression cassette or an expression vector). In such embodiments, the phrase “operably linked” is synonymous with “physically linked”.

As used herein, the phrases “coding sequence” and “open reading frame” (ORF) are used interchangeably and refer to a nucleotide sequence that is transcribed into RNA, such as but not limited to mRNA, rRNA, tRNA, snRNA, sense RNA, or antisense RNA. In some embodiments, the RNA is then translated *in vivo* or *in vitro* to produce a polypeptide. In some embodiments, an ORF is a coding sequence of a reporter gene (e.g., cellobiohydrolase I (CBHI); β -

glucuronidase (GUS)). Thus, in some embodiments an "ORF" refers to a nucleotide sequence between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides (*i.e.*, a "codon") in a coding sequence that
5 specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

As used herein, the term "complementary" refers to two nucleotide sequences that comprise antiparallel nucleotide sequences capable of pairing with one another upon formation of hydrogen bonds between the complementary
10 base residues in the antiparallel nucleotide sequences. As is known in the art, the nucleotide sequences of two complementary strands are the reverse complement of each other when each is viewed in the 5' to 3' direction.

As is also known in the art, two sequences that hybridize to each other under a given set of conditions do not necessarily have to be 100% fully
15 complementary. As used herein, the terms "fully complementary" and "100% complementary" refer to sequences for which the complementary regions are 100% in Watson-Crick base-pairing: *i.e.*, that no mismatches occur within the complementary regions. However, as is often the case with recombinant molecules (for example, cDNAs) that are cloned into cloning vectors, certain of
20 these molecules can have non-complementary overhangs on either the 5' or 3' ends that result from the cloning event. In such a situation, it is understood that the region of 100% or full complementarity excludes any sequences that are added to the recombinant molecule (typically at the ends) solely as a result of, or to facilitate, the cloning event. Such sequences are, for example, polylinker
25 sequences, linkers with restriction enzyme recognition sites, etc.

As used herein, the phrase "corresponds to", and grammatical variants thereof, refers to nucleic acid sequences that come from the same locus, although the exact sequences might not be identical due to naturally occurring or induced sequence differences in the different molecules. For example, SEQ ID
30 NO: 1 is defined as "corresponding to" the nucleotides 73605-75604 of

GENBANK® Accession No. AC193512.3. This usage refers to the fact that SEQ ID NO: 1 and nucleotides 73605-75604 of GENBANK® Accession No. AC193512.3 are versions of a genomic nucleotide sequence from a particular genetic locus on *Zea mays* chromosome 4. Comparing the sequences of SEQ ID NO: 1 to nucleotides 73605-75604 of GENBANK® Accession No. AC193512.3 shows that these sequences are greater than 99% identical over 2000 bases. It is noted that due to various sources of nucleotide differences (including, but not limited to normal naturally occurring genomic variations), sequences with less than 100% identical can also "correspond to" each other provided that they are derived from the same genetic locus and, in some embodiments, are derived from the same basic region of the genetic locus.

Additionally, gene products and molecules derived therefrom can correspond to a particular genetic locus, which means that they are subsequences of, transcription and/or translation products of, and/or intentionally or unintentionally modified versions of sequences or subsequences of the genetic locus. An exemplary molecule that "corresponds to" a genetic locus is an RNA transcribed therefrom or a cDNA reverse transcribed from such an RNA, a polypeptide encoded thereby, etc.

As used herein, the term "exon" refers to a sequence of DNA which carries the coding sequence for a protein or part of it. Exons are separated by intervening, non-coding sequences (introns). For purposes of the presently disclosed subject matter, the definition of the term "exon" includes modifications to the nucleotide sequence of an exon derived from a target gene, provided the modified exon does not significantly reduce the activity of its associated 5' regulatory sequence.

As used herein, the term "intron" refers to an intervening section of DNA which occurs almost exclusively within eukaryotic genes, but which is not translated to amino acid sequences in the gene product as a consequence of being removed from a primary RNA transcript (sometimes referred to as a "heterogenous nuclear RNA", or "hnRNA") through the process of splicing.

Splicing removes introns, thereby connecting exons into a final messenger RNA (mRNA) form than can be translated into a polypeptide. For purposes of the presently disclosed subject matter, the definition of the term "intron" includes modifications to the nucleotide sequence of an intron derived from a target gene, provided the modified intron does not significantly reduce the activity of its associated 5' regulatory sequence.

As used herein, the term "expression cassette" refers to a nucleic acid molecule capable of directing expression of a particular nucleotide sequence (e.g., a coding sequence) in an appropriate host cell. In some embodiments, an expression cassette comprises a promoter operably linked to a coding sequence of a reporter gene, which in some embodiments is further operably linked to a transcription terminator. In some embodiments, it can also comprise sequences required for proper translation of the nucleotide sequence. The coding region usually encodes a polypeptide of interest but can also encode a functional RNA of interest, for example antisense RNA or a non-translated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest can be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette can also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. Typically, however, at least one component of the expression cassette is heterologous with respect to the host; for example, a particular DNA sequence of the expression cassette does not occur naturally in the host cell and was introduced into the host cell or an ancestor of the host cell by a transformation event. The expression of the nucleotide sequence in the expression cassette can be under the control of a promoter (for example, a promoter of any of SEQ ID NOs: 1-3, or functional fragments thereof). In the case of a multicellular organism such as a plant, the promoter can also be specific to a particular tissue, organ, or stage of development (for example, a plant seed).

As used herein, the term "fragment" refers to a sequence that comprises a subset of another sequence. When used in the context of a nucleic acid or amino acid sequence, the terms "fragment" and "subsequence" are used interchangeably. A fragment of a nucleotide sequence can be any number of
5 nucleotides that is less than that found in another nucleotide sequence, and thus includes, but is not limited to, the sequences of an exon or intron, a promoter, an enhancer, an origin of replication, a 5' or 3' untranslated region, a coding region, and a polypeptide binding domain. It is understood that a fragment or subsequence can also comprise less than the entirety of a nucleotide sequence,
10 for example, a portion of an exon or intron, promoter, enhancer, etc. Similarly, a fragment or subsequence of an amino acid sequence can be any number of residues that is less than that found in a naturally occurring polypeptide, and thus includes, but is not limited to, domains, features, repeats, etc. Also similarly, it is understood that a fragment or subsequence of an amino acid sequence need not
15 comprise the entirety of the amino acid sequence of the domain, feature, repeat, etc. A fragment can also be a "functional fragment", in which the fragment retains a specific biological function of the nucleotide sequence or amino acid sequence of interest. For example, a functional fragment of a transcription factor can include, but is not limited to, a DNA binding domain, a transactivating domain, or
20 both. Similarly, a functional fragment of a receptor tyrosine kinase can include, but is not limited to, a ligand binding domain, a kinase domain, an ATP binding domain, and combinations thereof.

As used herein, the term "gene" is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include, but are not
25 limited to, coding sequences and/or the regulatory sequences required for their expression. Genes can also include non-expressed DNA segments that, for example, form recognition sequences for a polypeptide. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and can include
30 sequences designed to have desired parameters. The term "gene" thus includes,

but is not limited to naturally occurring nucleotide sequences, as well as homologs, orthologs, paralog, fragments and other equivalents, variants, and analogs thereof.

5 The terms "heterologous" and "recombinant", when used herein to refer to a nucleotide sequence (*e.g.* a DNA sequence) or a gene, refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling or other
10 recombinant techniques. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign to the host cell, or naturally occurring in the host cell but in a position or form within the host cell in which the element is not ordinarily found in nature. Similarly, when used in the context of a polypeptide or amino
15 acid sequence, a heterologous polypeptide or amino acid sequence is a polypeptide or amino acid sequence that originates from a source foreign to the particular host cell (*e.g.*, is generated from a heterologous coding sequence) or, if from the same source, is modified from its original form. Thus, heterologous DNA segments can be expressed to yield heterologous polypeptides.

20 A "homologous" nucleotide (or amino acid) sequence is a nucleotide (or amino acid) sequence naturally associated with a host cell into which it is introduced and that is present in the chromosomal or extrachromosomal position in which it is normally found in nature.

The phrase "hybridizing specifically to" refers to the binding, duplexing, or
25 hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. The phrase "bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the

stringency of the hybridization media to achieve the desired detection of the target nucleotide sequence.

As used herein, the terms “mutation” and “mutant” carry their traditional connotations and refer to a change, inherited, naturally occurring, or introduced, in a nucleic acid or polypeptide sequence, and are used in their senses as generally known to those of skill in the art.

As used herein, the term “isolated”, when used in the context of an isolated nucleic acid molecule or an isolated polypeptide, is a nucleic acid molecule or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated nucleic acid molecule or polypeptide can exist in a purified form or can exist in a non-native environment such as, for example, in a transgenic host cell, a transgenic plant, or any other transgenic organism.

As used herein, the terms “cell”, “cell line”, and “cell culture” are used interchangeably, and all such designations include progeny. Thus, the words “transformants” and “transformed cells” include the primary subject cell and cultures derived therefrom without regard for the number of transfers and/or rounds of cell division that the originally manipulated cell or cells might have experienced. It is also understood that all progeny might not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same function or biological activity as screened for in the originally transformed cell are encompassed by the terms. Where distinct designations are intended, it will be clear from the context.

As used herein, the term “native” refers to a gene that is naturally present in the genome of a plant cell. Similarly, when used in the context of a polypeptide, a “native polypeptide” is a polypeptide that is encoded by a native gene of a plant cell’s genome. The term “endogenous” also refers to a gene (or a polypeptide encoded thereby) that is naturally present in the genome of a plant cell

As used herein, the term "naturally occurring" refers to an object (e.g., a nucleotide sequence) that is found in nature as distinct from being artificially produced by man. For example, a nucleotide sequence that is present in an organism in its natural state, which has not been intentionally modified or isolated by man in the laboratory, is naturally occurring. As such, a nucleotide sequence is considered "non-naturally occurring" if it is encoded by and/or present within a recombinant molecule, even if the nucleotide sequence is identical to a nucleotide sequence found in nature.

As used herein, the term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleotide sequence also implicitly encompasses conservatively modified variants thereof (e.g. degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly disclosed. Specifically, degenerate codon substitutions can be achieved by generating sequences in which the third position of one or more (or all) selected codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, 1991; Ohtsuka *et al.*, 1985; Rossolini *et al.*, 1994). The terms "nucleic acid" or "nucleotide sequence" can also be used interchangeably with gene, cDNA, and mRNA encoded by a gene.

As used herein, the phrase "percent identical," in the context of two nucleic acid sequences, refers to two or more sequences or subsequences that have in some embodiments at least 60% (e.g., 60, 63, 65, 67, or 69%), in some embodiments at least 70% (e.g., 70, 73, 75, 77, or 79%), in some embodiments at least 80% (e.g., 80, 83, 85, 86, 87, 88, or 89%), in some embodiments at least 90% (e.g., 90, 91, 92, 93, 94, 95, 96, 97, or 98%), and in some embodiments at least 99% nucleotide identity, respectively, when compared and aligned for maximum correspondence, as measured using one of the following sequence

comparison algorithms or by visual inspection. The percent identity exists in some embodiments over a region of the sequences that is at least about 50 nucleotides in length, in some embodiments over a region of at least about 100 nucleotides in length, in some embodiments over a region of at least about 250 nucleotides in length, in some embodiments over a region of at least about 500 nucleotides in length, in some embodiments over a region of at least about 1000 nucleotides in length, and in some embodiments, the percent identity exists over at least about 1500 residues. In some embodiments, the percent identity exists over the entire length of one or both of the sequences (e.g., any of SEQ ID NOs: 1-6).

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

Optimal alignment of sequences for comparison can be conducted, for example, by the local homology algorithm disclosed in Smith & Waterman, 1981, by the homology alignment algorithm disclosed in Needleman & Wunsch, 1970, by the search for similarity method disclosed in Pearson & Lipman, 1988, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the GCG® WISCONSIN PACKAGE®, available from Accelrys, Inc., San Diego, California, United States of America), or by visual inspection, in general Ausubel *et al.*, 2002 and Ausubel *et al.*, 2003.

It is noted that for several known algorithms, it is not possible to set the parameters such that a specific length of a given sequence (for example, the entire length of any of SEQ ID NOs: 1-6) will be included in the comparison between sequences. In such a circumstance, it might be necessary to align the sequences manually in order to determine the percent identity. For example,

SEQ ID NO: 1 is 2000 nucleotides in length. In some embodiments, a sequence that is less than 1900 nucleotides cannot be at least 95% identical to SEQ ID NO: 1 because the length of the latter is not at least 95% of the length of SEQ ID NO: 1. Similarly, in some embodiments a sequence to be compared to SEQ ID
5 NO: 1 must be at least 2000 nucleotides in length in order to be 100% identical to SEQ ID NO: 1.

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, 1990. Software for performing BLAST analysis is
10 publicly available through the website of the National Center for Biotechnology Information (NCBI) of the United States National Institutes of Health (NIH). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the
15 same length in a database sequence. T is referred to as the neighborhood word score threshold. See *generally*, Altschul *et al.*, 1990. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative
20 scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity
25 X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an
30 expectation (E) of 10, a cutoff of 100, M = 5, N = -4, and a comparison of both

strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. See Henikoff & Henikoff, 1992.

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

As used herein, the term "promoter" refers to a nucleotide sequence,
usually upstream (5') to a coding sequence, which controls the expression of the
coding sequence by providing the recognition for RNA polymerase and other
factors required for proper transcription. "Promoter regulatory sequences" consist
20 of proximal and more distal upstream elements. Promoter regulatory sequences
influence the transcription, RNA processing and/or stability, and/or translation of
the associated coding sequence. Regulatory sequences include enhancers,
promoters, untranslated leader sequences, introns, and polyadenylation signal
sequences. They include natural and synthetic sequences as well as sequences
25 that may be a combination of synthetic and natural sequences. An "enhancer" is
a DNA sequence that can stimulate promoter activity and may be an innate
element of the promoter or a heterologous element inserted to enhance the level
or tissue specificity of a promoter. It is capable of operating in both orientations
(normal or flipped), and is capable of functioning even when moved either
30 upstream or downstream from the promoter. The meaning of the term "promoter"

includes "promoter regulatory sequences". Exemplary promoters of the presently disclosed subject matter include nucleotide sequences that are in some embodiments at least 90%, in some embodiments at least 91%, in some embodiments at least 92%, in some embodiments at least 93%, in some
5 embodiments at least 94%, in some embodiments at least 95%, in some embodiments at least 96%, in some embodiments at least 97%, in some embodiments at least 98%, in some embodiments at least 99% identical, or that are 100% identical to any of SEQ ID NOs: 1-3. In some embodiments, the percent identity is calculated over 100 nucleotides, 200 nucleotides, 300
10 nucleotides, 500 nucleotides, 750 nucleotides, 1000 nucleotides, 1500 nucleotides, or the full length of any of SEQ ID NOs: 1-3.

Certain subsequences and/or fragments of known promoters have been shown to retain promoter activity, and for any given promoter, the smallest such subsequence and/or fragment is referred to as a "minimal" promoter. As used
15 herein, the term "minimal" refers to the shortest fragment of a regulatory polynucleotide molecule that is still effective in gene regulation. As such, a "minimal promoter" is the shortest identified fragment of a longer sequence length that retains promoter activity when operably linked to a transcribable polynucleotide molecule. Systematic mutagenesis of a particular promoter
20 region, also known as "promoter bashing", and testing the resulting effect on gene expression has been used to identify functional blocks in upstream regions of genes. In order to identify cis-acting elements in promoters, a series of truncated promoter fragments can be fused to a reporter gene via cloning and evaluated in transgenic plants.

25 As used herein, the term "minimal promoter" refers to the smallest piece of a promoter, such as a TATA element, that can support any transcription. A minimal promoter typically has greatly reduced promoter activity in the absence of upstream or downstream activation. In the presence of a suitable transcription factor, a minimal promoter can function to permit transcription.

An example of such a minimal promoter is the CaMV 35S minimal promoter (described in U.S. Patent No. 5,097,025, herein incorporated by reference in its entirety). Plant transformation vectors typically comprise at least one gene regulatory element operably linked to a structural coding sequence.

- 5 Because small vectors are desirable for plant transformation, it would be useful to use a smaller fragment of a larger known gene regulatory molecule that still retains its gene regulatory activity. Such novel regulatory elements comprise a “minimal”, or “core” region that retains gene regulatory activity.

- As used herein, the term “purified”, when applied to a nucleic acid or polypeptide, denotes that the nucleic acid or polypeptide is essentially free of other cellular components with which it is associated in the natural state. It can be in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A polypeptide that is the predominant species present in a preparation is substantially purified. The term “purified” denotes that a nucleic acid or polypeptide gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or polypeptide is in some embodiments at least about 50% pure, in some embodiments at least about 85% pure, in some embodiments at least about 90% pure, in some embodiments at least about 91% pure, in some embodiments at least about 92% pure, in some embodiments at least about 93% pure, in some embodiments at least about 94% pure, in some embodiments at least about 95% pure, in some embodiments at least about 96% pure, in some embodiments at least about 97% pure, in some embodiments at least about 98% pure, and in some embodiments at least about 99% pure.
- 10
15
20
25

- As used herein, the term “transformation” refers to a process for introducing a heterologous nucleic acid molecule (e.g., a DNA molecule) into a cell (e.g., a bacterial cell, a yeast cell, a plant cell, a plant tissue, and/or a plant).
- 30 In some embodiments, transformed plant cells, plant tissue, and/or plants are

understood to encompass not only the product of a particular transformation process, but also transgenic progeny thereof.

As used herein, the terms “transformed”, “transgenic”, and “recombinant” refer to a host cell or organism such as a bacterium or a plant cell (e.g., a plant) into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be transiently or stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A “non-transformed,” “non-transgenic”, or “non-recombinant” host refers to a wild-type organism, e.g., a bacterium or plant, which does not contain the heterologous nucleic acid molecule.

II. Identification of Transcriptional Regulatory Elements

II.A. Generally

In some embodiments, the presently disclosed subject matter provides isolated nucleic acids that comprise transcription regulatory elements. In some embodiments, a transcriptional regulatory element is a promoter. In some embodiments, a transcriptional regulatory element is a terminator.

In some embodiments, a transcriptional regulatory element is a promoter. Promoters can be classified into several main categories based on the type of RNA polymerase that binds them. Typically, protein encoding genes are regulated by promoters that bind to RNA Polymerase II, as opposed to some ribosomal and tRNA gene promoters, to which RNA Polymerase III typically bind.

With reference now to promoters to which RNA Polymerase II binds, there are several different subtypes of promoters, which can be classified based on the cell type(s) and/or developmental stage(s) in and/or during which the promoters are active in regulating transcription of operably linked nucleotide sequences.

For example, a promoter that is active under most or all conditions and in most or all cell types and developmental stages of an organism is referred to as a

"constitutive" promoter. It is understood, however, that "constitutive" is not to be interpreted in an absolute sense in that a promoter that is active in most or all cell types can be inactive or less active in certain cell types (e.g., pollen) and still be considered a constitutive promoter.

5 In contrast, promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, seeds, fibers, xylem vessels, tracheids, or sclerenchyma, are typically referred to as "tissue preferred promoters". Similarly, a "cell type" specific promoter primarily regulates expression of operably linked sequences in certain cell types in one or more organs, for example, vascular
10 cells in roots or leaves. "Inducible" or "regulatable" promoters are promoters that are under environmental control. Examples of environmental conditions that can effect transcription by inducible promoters include anaerobic conditions or the presence of light. Another type of promoter is a developmentally regulated promoter, for example, a promoter that drives expression during pollen
15 development. Tissue preferred, cell type specific, developmentally regulated, and inducible promoters constitute the class of "non-constitutive" promoters.

 The presently disclosed subject matter thus in some embodiments provides polynucleotide constructs comprising regulatory elements that can modulate expression of an operably linked transcribable polynucleotide molecule
20 and a transgenic plant stably transformed with the polynucleotide construct.

 From the examples given, the presently disclosed subject matter thus provides in some embodiments chimeric regulatory elements that are useful for modulating the expression of an operably linked transcribable polynucleotide molecule. In particular, the presently disclosed subject matter provides in some
25 embodiments methods for assembling polynucleotide constructs comprising the isolated regulatory elements and isolated promoter fragments, and for creating a transgenic plant stably transformed with the polynucleotide construct.

 In some embodiments, a promoter of the presently disclosed subject matter comprises, consists essentially of, or consists of any of SEQ ID NOs: 1-3.

30 II.B. Identification of Promoter Elements and Minimal Promoters

According to an exemplary embodiment of the presently disclosed subject matter, fragments and/or subsequences of the promoter sequences disclosed herein are tested to identify fragments and/or subsequences that retain some or all of the transcriptional regulatory activities of the promoters on which they are based. Thus, while the promoter sequences disclosed herein (*e.g.*, SEQ ID NOs: 1-3) have transcriptional regulatory activity, fragments and/or subsequences of these promoters would also be expected to have some promoter activity.

By way of example and not limitation, by deleting a portion of any one of SEQ ID NOs: 1-3 and transforming the resultant molecule(s) into a plant or a cell using known methods in the art, subsequences of SEQ ID NOs: 1-3 can be tested to determine to what extent they retain transcriptional activity. Such “promoter bashing” is described in the literature, see for example U.S. Patent No. 5,097,025 (herein incorporated by reference in its entirety), and can lead to the development of core sequences necessary and sufficient for desirable regulatory activity and/or minimal sequences sufficient for the desirable activity of the transgene inserted into the genome of plants of interest. Because regulatory expression elements comprise many motifs that can impact gene expression, various resultant fragments can have different levels of gene regulatory activity, each potentially providing benefits to different transformed plant species. Molecules of the presently disclosed subject matter thus comprise in some embodiments fragments that can themselves have gene regulatory activity.

In some embodiments, one of SEQ ID NOs: 1-3 is selected for further analysis to determine functional fragments thereof. As SEQ ID NOs: 1-3 correspond to the promoter sequences, subsequences of these sequences can be produced and tested, either *in vitro* or *in vivo*, for transcriptional regulatory activity.

Additionally, since promoter elements typically are located closer to the transcription start site, in some embodiments the promoter sequences disclosed in SEQ ID NOs: 1-3 are truncated to various degrees beginning at their 5' ends. By way of example and not limitation, any of SEQ ID NOs: 1-3 can be modified

by deleting any number of nucleotides from their 5' ends, which deletions can encompass in various embodiments any number of nucleotides from 1 to about 1950 or more, although it is noted that deletion of all 2000 of the 5'-most nucleotides of any of SEQ ID NOs: 1-3 would be expected to destroy the transcriptional regulatory activity of the sequences.

Thus, the identification of fragments and subsequences of any of SEQ ID NOs: 1-3 can include deletion of the 5'-most 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 1950, or more nucleotides, include any number of nucleotides between 1 and 2000.

It is further noted that promoter elements are frequently located several hundred to several thousand nucleotides upstream (*i.e.*, 5' to) transcriptional start sites, and thus subsequences of any of SEQ ID NOs: 1-3 that include internal deletions of any size between 1 and 2000 nucleotides can also be produced and tested for transcriptional regulatory using standard molecular biology techniques.

III. Isolated Nucleic Acid Molecules and Expression Vectors Comprising Isolated Nucleic Acid Molecules

In some embodiments, the presently disclosed subject matter provides isolated nucleic acid molecules comprising transcription regulatory elements selected from the group consisting of promoters, exons, introns, and transcription terminators. In some embodiments, the isolated nucleic acid molecules can comprise heterologous nucleotide sequences, which can include any nucleotide sequences for which the presence in a cell or a plant is desirable. In some embodiments, a heterologous nucleotide sequence encodes a biomolecule of interest, such as but not limited to a polypeptide of interest or a regulatory RNA (*e.g.*, an antisense RNA, an miRNA, a siRNA, etc.).

In some embodiments, an isolated nucleic acid molecule of the presently disclosed subject matter comprises a promoter. Exemplary promoters of the presently disclosed subject matter include nucleotide sequences that are in some embodiments at least 90%, in some embodiments at least 91%, in some

embodiments at least 92%, in some embodiments at least 93%, in some
embodiments at least 94%, in some embodiments at least 95%, in some
embodiments at least 96%, in some embodiments at least 97%, in some
embodiments at least 98%, in some embodiments at least 99% identical, or that
5 are 100% identical to any of SEQ ID NOs: 1-3. In some embodiments, the
percent identity is calculated over 100 nucleotides, 200 nucleotides, 300
nucleotides, 500 nucleotides, 750 nucleotides, 1000 nucleotides, 1500
nucleotides, or the full length of any of SEQ ID NOs: 1-3.

In some embodiments, the isolated nucleic acid molecules of the presently
10 disclosed subject matter further comprise one or more exons and/or introns
operably linked to, and under transcriptional regulatory control of, a promoter of
the presently disclosed subject matter including, but not limited to the exemplary
promoters listed hereinabove. Any exon and/or intron can be introduced into the
isolated nucleic acid molecule using standard molecular biology techniques (see
15 e.g., Sambrook & Russell, 2001). In some embodiments, an intron is operably
linked to the promoter.

In some embodiments, the one or more exons and/or introns that are
operably linked to an exemplary promoter of the presently disclosed subject
matter (e.g., any of SEQ ID NOs: 1-3) correspond to (i.e., are derived from) the
20 same genetic locus as the promoter of any of SEQ ID NOs: 1-3. Nucleic acid
molecules with one or more exons and/or introns that are operably linked to an
exemplary promoter can be generated using the general strategies discussed in
U.S. Patent Application Publication No. 20070006344 of Nuccio *et al.*, the entire
disclosure of which is incorporated herein by reference.

25 In some embodiments, a transcription terminator is operably linked to a
promoter of any of SEQ ID NOs: 1-3 in the isolated nucleic acid molecules of the
presently disclosed subject matter. Any transcription terminator can be operably
linked to a promoter of the presently disclosed subject matter in an isolated
nucleic acid molecule such that the transcription terminator functions to terminate
30 transcription controlled by the promoter using standard molecular biology

techniques (see e.g., Sambrook & Russell, 2001). Exemplary transcription terminators that are known to function in plants include, but are not limited to, the CaMV 35S terminator, the tm1 terminator, the pea rbcS E9 terminator, and the nopaline and octopine synthase terminator (see e.g., Odell *et al.*, 1985; Joshi,
5 1987; Rosenberg *et al.*, 1987; Ballas *et al.*, 1989; Mogen *et al.*, 1990; Munroe & Jacobson, 1990; Guerineau *et al.*, 1991; Proudfoot, 1991; Sanfaçon *et al.*, 1991).

In some embodiments, a transcription terminator is from the same genetic locus as the promoter to which it is operably linked. Thus, in some embodiments an isolated nucleic acid molecule of the presently disclosed subject matter
10 comprises:

- (a) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 1 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 4; or
- 15 (b) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 2 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 5; or
- (c) a first nucleotide sequence that is at least 95% identical over at
20 least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 6.

In some embodiments, a promoter of the presently disclosed subject matter is operably linked to a first exon, an intron, and optionally a second exon
25 or a fragment thereof, and further optionally a transcription terminator. Thus, the presently disclosed subject matter provides in some embodiments isolated nucleic acid molecules that comprise any of SEQ ID NOs: 1-3; optionally that also comprise any of SEQ ID NOs: 4-6.

In some embodiments, the isolated nucleic acid molecules of the presently
30 disclosed subject matter are incorporated into and/or form the basic structures of

expression cassettes designed to express heterologous sequences (*e.g.*, coding sequences) in plants. Thus, coding sequences intended for expression in transgenic plants can be first assembled into expression cassettes operably linked to one or more of the transcriptional regulatory elements disclosed herein.

- 5 The expression cassettes can also comprise any further sequences required or selected for the expression of the transgene. Such sequences include, but are not limited to, promoters, transcription terminators, extraneous sequences to enhance expression such as introns, vital sequences, and sequences intended for the targeting of the gene product to specific organelles and cell
10 compartments. These expression cassettes can then be easily transferred to the plant transformation vectors disclosed below. The following is a description of various components of typical expression cassettes.

III.A. Promoters

- The selection of the promoter used in expression cassettes can determine
15 the spatial and temporal expression pattern of the transgene in the transgenic plant. Selected promoters can express transgenes in specific cell types (for example, leaf epidermal cells, mesophyll cells, root cortex cells, and/or endosperm cells) or in specific tissues or organs (for example, roots, leaves, flowers, and/or seeds) and the selection can reflect the desired location for
20 accumulation of the gene product. Alternatively, the selected promoter can drive expression of the gene under various inducing conditions. Promoters vary in their strengths; *i.e.*, their abilities to promote transcription. Depending upon the host cell system utilized, any one of a number of suitable promoters can be used, including the gene's native promoter.

- 25 SEQ ID NOs: 1-3 comprise sequences that can be employed as promoters for expressing heterologous nucleic acid sequences in plants as set forth herein.

III.B. Transcriptional Terminators

- A variety of transcriptional terminators are available for use in expression
30 cassettes. These are responsible for termination of transcription and correct

mRNA polyadenylation. Appropriate transcriptional terminators are those that are known to function in plants and include the CaMV 35S terminator, the *tmf* terminator, the nopaline synthase (NOS) terminator, and the pea *rbcS* E9 terminator. These can be used in both monocotyledons and dicotyledons.

- 5 In addition, a gene's native transcription terminator can be used. In some embodiments, the native transcription terminator is selected from among SEQ ID NOs: 4-6.

III.C. Sequences for the Enhancement or Regulation of Expression

- 10 Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes of the presently disclosed subject matter to increase their expression in transgenic plants.

- Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize *Adhl* gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, 1987).
15 In the same experimental system, the intron from the maize *bronze1* gene had a similar effect in enhancing expression. Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

- A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV; the "W-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (see *e.g.* Gallie *et al.*, 1987; Skuzeski *et al.*, 1990). Other leader sequences
25 known in the art include, but are not limited to, picornavirus leaders, for example,
30

encephalomyocarditis virus (EMCV) leader (5' noncoding region; see Elroy-Stein *et al.*, 1989); potyvirus leaders, for example, from Tobacco Etch Virus (TEV; see Allison *et al.*, 1986); Maize Dwarf Mosaic Virus (MDMV; see Kong & Steinbiss 1998); human immunoglobulin heavy-chain binding polypeptide (BiP) leader (Macejak & Sarnow, 1991); untranslated leader from the coat polypeptide mRNA of alfalfa mosaic virus (AMV; RNA 4; see Jobling & Gehrke, 1987); tobacco mosaic virus (TMV) leader (Gallie *et al.*, 1989); and Maize Chlorotic Mottle Virus (MCMV) leader (Lommel *et al.*, 1991). See also, Della-Cioppa *et al.*, 1987.

In addition to incorporating one or more of the aforementioned elements into the 5' regulatory region of a target expression cassette of the presently disclosed subject matter, other elements can also be incorporated. Such elements include, but are not limited to, a minimal promoter. By minimal promoter it is intended that the basal promoter elements are inactive or nearly so in the absence of upstream or downstream activation. Such a promoter has low background activity in plants when there is no transactivator present or when enhancer or response element binding sites are absent. One minimal promoter that is particularly useful for target genes in plants is the Bz1 minimal promoter, which is obtained from the *bronze1* gene of maize. The Bz1 core promoter is obtained from the "myc" mutant Bz1-luciferase construct pBz1LucR98 via cleavage at the *NheI* site located at positions -53 to -58 (Roth *et al.*, 1991). The derived Bz1 core promoter fragment thus extends from positions -53 to +227 and includes the Bz1 intron-1 in the 5' untranslated region. Also useful for the presently disclosed subject matter is a minimal promoter created by use of a synthetic TATA element. The TATA element allows recognition of the promoter by RNA polymerase factors and confers a basal level of gene expression in the absence of activation (see generally, Mukumoto *et al.*, 1993; Green, 2000).

III.D. Targeting of the Gene Product Within the Cell

Various mechanisms for targeting gene products are known to exist in plants and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products

to the chloroplast is controlled by a signal sequence found at the amino terminal end of various polypeptides that is cleaved during chloroplast import to yield the mature polypeptides (see e.g., Comai *et al.*, 1988). These signal sequences can be fused to heterologous gene products to affect the import of heterologous products into the chloroplast (Van den Broeck *et al.*, 1985). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO) polypeptide, the chlorophyll a/b binding (CAB) polypeptide, the 5-enol-pyruvyl shikimate-3-phosphate (EPSP) synthase enzyme, the GS2 polypeptide and many other polypeptides which are known to be chloroplast localized. See also, the section entitled "Expression With Chloroplast Targeting" in Example 37 of U.S. Patent No. 5,639,949, herein incorporated by reference.

Other gene products can be localized to other organelles such as the mitochondrion and the peroxisome (e.g. Unger *et al.*, 1989). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous gene products to these organelles. Examples of such sequences are the nuclear-encoded ATPases and specific aspartate amino transferase isoforms for mitochondria. Targeting cellular polypeptide bodies has been disclosed by Rogers *et al.*, 1985.

In addition, sequences have been characterized that control the targeting of gene products to other cell compartments. Amino terminal sequences are responsible for targeting to the endoplasmic reticulum (ER), the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, 1990). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al.*, 1990).

By the fusion of the appropriate targeting sequences disclosed above to transgene sequences of interest it is possible to direct the transgene product to any organelle or cell compartment. For chloroplast targeting, for example, the chloroplast signal sequence from the RUBISCO gene, the CAB gene, the EPSP synthase gene, or the GS2 gene is fused in frame to the amino terminal ATG of

the transgene. The signal sequence selected can include the known cleavage site, and the fusion constructed can take into account any amino acids after the cleavage site that are required for cleavage. In some cases this requirement can be fulfilled by the addition of a small number of amino acids between the
5 cleavage site and the transgene ATG or, alternatively, replacement of some amino acids within the transgene sequence. Fusions constructed for chloroplast import can be tested for efficacy of chloroplast uptake by *in vitro* translation of *in vitro* transcribed constructions followed by *in vitro* chloroplast uptake using techniques disclosed by Bartlett *et al.*, 1982 and Wasmann *et al.*, 1986. These
10 construction techniques are well known in the art and are equally applicable to mitochondria and peroxisomes.

The above-disclosed mechanisms for cellular targeting can be utilized not only in conjunction with their cognate promoters, but also in conjunction with heterologous promoters so as to effect a specific cell-targeting goal under the
15 transcriptional regulation of a promoter that has an expression pattern different from that of the promoter from which the targeting signal derives.

III.E. Construction of Plant Transformation Vectors

Numerous transformation vectors available for plant transformation are known to those of ordinary skill in the plant transformation art, and the genes
20 pertinent to the presently disclosed subject matter can be used in conjunction with any such vectors. The selection of vector will depend upon the selected transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers might be employed. Selection markers used routinely in transformation include the *nptII*
25 gene, which confers resistance to kanamycin and related antibiotics (Messing & Vieira, 1982; Bevan *et al.*, 1983); the *bar* gene, which confers resistance to the herbicide phosphinothricin (White *et al.*, 1990; Spencer *et al.*, 1990); the *hph* gene, which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, 1984); the *dhfr* gene, which confers resistance to methotrexate
30 (Bourouis & Jarry, 1983); the EPSP synthase gene, which confers resistance to

glyphosate (U.S. Patent Nos. 4,940,935 and 5,188,642); and the mannose-6-phosphate isomerase gene, which provides the ability to metabolize mannose (U.S. Patent Nos. 5,767,378 and 5,994,629).

III.E.1. Vectors Suitable for *Agrobacterium* Transformation

5 Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, 1984). By way of example only, the construction of two typical vectors suitable for *Agrobacterium* transformation is disclosed.

10 IV.E.1.a. pCIB200 and pCIB2001

The binary vectors pCIB200 and pCIB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and are constructed in the following manner. pTJS75kan is created by *Nar* I digestion of pTJS75 (Schmidhauser & Helinski, 1985) allowing excision of the tetracycline-resistance
 15 gene, followed by insertion of an *Acc* I fragment from pUC4K carrying an NPTII sequence (Messing & Vieira, 1982; Bevan *et al.*, 1983; McBride & Summerfelt, 1990). *Xho* I linkers are ligated to the *EcoRV* fragment of PCIB7 which contains the left and right T-DNA borders, a plant selectable *nos/nptII* chimeric gene and the pUC polylinker (Rothstein *et al.*, 1987), and the *Xho* I-digested fragment are
 20 cloned into *Sal* I-digested pTJS75kan to create pCIB200 (see also European Patent Application EP 0 332 104, example 19). pCIB200 contains the following unique polylinker restriction sites: *EcoR* I, *Sst* I, *Kpn* I, *Bgl* II, *Xba* I, and *Sal* I. pCIB2001 is a derivative of pCIB200 created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001
 25 are *EcoR* I, *Sst* I, *Kpn* I, *Bgl* II, *Xba* I, *Sal* I, *Mlu* I, *Bcl* I, *Avr* II, *Apa* I, *Hpa* I, and *Stu* I. pCIB2001, in addition to containing these unique restriction sites, also has plant and bacterial kanamycin selection, left and right T-DNA borders for *Agrobacterium*-mediated transformation, the RK2-derived *trfA* function for mobilization between *E. coli* and other hosts, and the *OriT* and *OriV* functions

also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

III.E.1.b. pCIB10 and Hygromycin Selection Derivatives Thereof

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

III.E.2. Vectors Suitable for non-*Agrobacterium* Transformation

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector, and
15 consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones disclosed above that contain T-DNA sequences. Transformation techniques that do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake (*e.g.* polyethylene glycol (PEG) and electroporation), and microinjection. The choice of vector
20 depends largely on the species being transformed. Below, the construction of typical vectors suitable for non-*Agrobacterium* transformation is disclosed.

III.E.2.a. pCIB3064

pCIB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide BASTA® (glufosinate
25 ammonium or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter operably linked to the *E. coli* β -glucuronidase (GUS) gene and the CaMV 35S transcriptional terminator and is disclosed in the PCT International Publication WO 93/07278. In some embodiments, the CaMV 35S promoter and/or the CaMV 35S transcriptional terminator are replaced with a promoter as

set forth herein (e.g., SEQ ID NOs: 1-3) and/or a transcription terminator as set forth herein (e.g., SEQ ID NOs: 4-6).

The 35S promoter of the pCIB3064 vector contains two ATG sequences 5' of the start site. These sites are mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites *Ssp* I and *Pvu* II. The new restriction sites are 96 and 37 bp away from the unique *Sal* I site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 is designated pCIB3025. The GUS gene is then excised from pCIB3025 by digestion with *Sal* I and *Sac* I, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 is obtained from the John Innes Centre, Norwich, England, and the 400 bp *Sma* I fragment containing the *bar* gene from *Streptomyces viridochromogenes* is excised and inserted into the *Hpa* I site of pCIB3060 (Thompson *et al.*, 1987). This generated pCIB3064, which comprises the *bar* gene under the control of the CaMV 35S promoter and terminator for herbicide selection, a gene for ampicillin resistance (for selection in *E. coli*) and a polylinker with the unique sites *Sph* I, *Pst* I, *Hind* III, and *Bam* H I. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals (e.g., the promoters of SEQ ID NOs: 1-3 and/or transcription terminators of SEQ ID NOs: 4-6).

20 III.E.2.b. pSOG19 and pSOG35

pSOG35 is a transformation vector that utilizes the *E. coli* dihydrofolate reductase (DHFR) gene as a selectable marker conferring resistance to methotrexate. PCR is used to amplify the 35S promoter (-800 bp), intron 6 from the maize *Adh1* gene (-550 bp), and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250-bp fragment encoding the *E. coli* dihydrofolate reductase type II gene is also amplified by PCR and these two PCR fragments are assembled with a *Sac* I-*Pst* I fragment from pB1221 (BD Biosciences Clontech, Palo Alto, California, United States of America) that comprises the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generates pSOG19 that contains the 35S promoter in fusion

with the intron 6 sequence, the GUS leader, the DHFR gene, and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus (MCMV) generates the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *Hind* III, *Sph* I, *Pst* I, and *EcoR* I sites available for the cloning of foreign substances. The 35S promoter and/or the nopaline synthase terminator of any of these vectors can be replaced with the promoters of SEQ ID NOs: 1-3 and/or transcription terminators of SEQ ID NOs: 4-6 using standard recombinant DNA techniques.

10 III.F. Transformation

Once a nucleotide sequence of the presently disclosed subject matter has been cloned into an expression system, it is transformed into a plant cell. The receptor and target expression cassettes of the presently disclosed subject matter can be introduced into the plant cell in a number of art-recognized ways.

15 Methods for regeneration of plants are also well known in the art. For example, Ti plasmid vectors have been utilized for the delivery of foreign DNA, as well as direct DNA uptake, liposomes, electroporation, microinjection, and microprojectiles. In addition, bacteria from the genus *Agrobacterium* can be utilized to transform plant cells. Below are descriptions of representative techniques for transforming both dicotyledonous and monocotyledonous plants, as well as a representative plastid transformation technique.

III.F.1. Transformation of Dicotyledons

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques that do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of heterologous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation-mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are disclosed in Paszkowski *et al.*, 1984; Potrykus *et al.*, 1985; Reich

et al., 1986; and Klein *et al.*, 1987. In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

Agrobacterium-mediated transformation is a useful technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. *Agrobacterium* transformation typically involves the transfer of the binary vector carrying the foreign DNA of interest (e.g. pCIB200, pCIB2001, or in some embodiments a derivative thereof with a promoter of one of SEQ ID NOs: 1-3 and/or transcription terminator of one of SEQ ID NOs: 4-6) to an appropriate *Agrobacterium* strain which can depend on the complement of *vir* genes carried by the host *Agrobacterium* strain either on a co-resident Ti plasmid or chromosomally (e.g. strain CIB542 for pCIB200 and pCIB2001; Uknes *et al.*, 1993). The transfer of the recombinant binary vector to *Agrobacterium* is accomplished by a triparental mating procedure using *E. coli* carrying the recombinant binary vector, a helper *E. coli* strain that carries a plasmid such as pRK2013 and which is able to mobilize the recombinant binary vector to the target *Agrobacterium* strain. Alternatively, the recombinant binary vector can be transferred to *Agrobacterium* by DNA transformation (Höfgen & Willmitzer, 1988).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

Another approach to transforming plant cells with a gene involves propelling inert or biologically active particles at plant tissues and cells. This technique is disclosed in U.S. Patent Nos. 4,945,050; 5,036,006; and 5,100,792; all to Sanford *et al.* Generally, this procedure involves propelling inert or biologically active particles at the cells under conditions effective to penetrate the outer surface of the cell and afford incorporation within the interior thereof. When inert particles are utilized, the vector can be introduced into the cell by coating

the particles with the vector containing the desired gene. Alternatively, the target cell can be surrounded by the vector so that the vector is carried into the cell by the wake of the particle. Biologically active particles (e.g., dried yeast cells, dried bacterium, or a bacteriophage, each containing DNA sought to be introduced) can also be propelled into plant cell tissue.

III.F.2. Transformation of Monocotyledons

Transformation of most monocotyledon species has now also become routine. Exemplary techniques include direct gene transfer into protoplasts using PEG or electroporation, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (*i.e.* co-transformation), and both these techniques are suitable for use with the presently disclosed subject matter. Co-transformation can have the advantage of avoiding complete vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the selectable marker in subsequent generations, should this be regarded as desirable. However, a disadvantage of the use of co-transformation is the less than 100% frequency with which separate DNA species are integrated into the genome (Schocher *et al.*, 1986).

European Patent Applications EP 0 292 435, EP 0 392 225, and PCT International Patent Application Publication No. WO 93/07278 describe techniques for the preparation of callus and protoplasts from an elite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration of maize plants from transformed protoplasts. Gordon-Kamm *et al.*, 1990 and Fromm *et al.*, 1990 have published techniques for transformation of A188-derived maize line using particle bombardment. Furthermore, PCT International Patent Application Publication No. WO 93/07278 and Koziel *et al.*, 1993 describe techniques for the transformation of elite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length excised from a maize ear 14-15 days after pollination and a PDS-

1000/He Biolistic particle delivery device (Bio-Rad Laboratories, Hercules, California, United States of America) for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated
5 transformation has been disclosed for *Japonica*-types and *Indica*-types (Zhang *et al.*, 1988; Shimamoto *et al.*, 1989; Datta *et al.*, 1990) of rice. Both types are also routinely transformable using particle bombardment (Christou *et al.*, 1991). Furthermore, PCT International Patent Application Publication No. WO 93/21335 describes techniques for the transformation of rice via electroporation. Casas *et al.*,
10 *et al.*, 1993 discloses the production of transgenic sorghum plants by microprojectile bombardment.

European Patent Application EP 0 332 581 describes techniques for the generation, transformation, and regeneration of *Pooideae* protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat
15 transformation has been disclosed in Vasil *et al.*, 1992 using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil *et al.*, 1993 and Weeks *et al.*, 1993 using particle bombardment of immature embryos and immature embryo-derived callus.

A representative technique for wheat transformation, however, involves
20 the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, embryos (0.75-1 mm in length) are plated onto MS medium with 3% sucrose (Murashige & Skoog, 1962) and 3 mg/l 2,4-dichlorophenoxyacetic acid (2,4-D) for induction of somatic embryos, which is
25 allowed to proceed in the dark. On the chosen day of bombardment, embryos are removed from the induction medium and placed onto the osmoticum (*i.e.* induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 hours and are then bombarded. Twenty embryos per target plate are typical, although not
30 critical. An appropriate gene-carrying plasmid (such as pCIB3064, pSG35, or a

derivative thereof containing a promoter of one of SEQ ID NOs: 1-3 and/or a transcription terminator of one of SEQ ID NOs: 4-6) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with biolistics device using a burst pressure of about 1000 pounds per square inch (psi) using a standard 80 mesh screen. After bombardment, the embryos are placed back into the dark to recover for about 24 hours (still on osmoticum). After 24 hours, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l BASTA® in the case of pCIB3064 and derivatives there and 2 mg/l methotrexate in the case of pSOG35 and derivatives thereof). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contain half-strength MS, 2% sucrose, and the same concentration of selection agent.

Transformation of monocotyledons using *Agrobacterium* has also been disclosed. See PCT International Patent Application Publication No. WO 94/00977 and U.S. Patent No. 5,591,616, both of which are incorporated herein by reference. See also Negrotto *et al.*, 2000, incorporated herein by reference. Zhao *et al.*, 2000 specifically discloses transformation of sorghum with *Agrobacterium*. See also U.S. Patent No. 6,369,298.

Rice (*Oryza sativa*) can be used for generating transgenic plants. Various rice cultivars can be used (Hiei *et al.*, 1994; Dong *et al.*, 1996; Hiei *et al.*, 1997). Also, the various media constituents disclosed below can be either varied in quantity or substituted. Embryogenic responses are initiated and/or cultures are established from mature embryos by culturing on MS-CIM medium (MS basal salts, 4.3 g/liter; B5 vitamins (200 x), 5 ml/liter; sucrose, 30 g/liter; proline, 500 mg/liter; glutamine, 500 mg/liter; casein hydrolysate, 300 mg/liter; 2,4-D (1 mg/ml), 2 ml/liter; pH adjusted to 5.8 with 1 N KOH; Phytigel, 3 g/liter). Either

mature embryos at the initial stages of culture response or established culture lines are inoculated and co-cultivated with the *Agrobacterium tumefaciens* strain LBA4404 (*Agrobacterium*) containing the desired vector construction. *Agrobacterium* is cultured from glycerol stocks on solid YPC medium (plus 100 mg/L spectinomycin and any other appropriate antibiotic) for about 2 days at 28°C. *Agrobacterium* is re-suspended in liquid MS-CIM medium. The *Agrobacterium* culture is diluted to an OD₆₀₀ of 0.2-0.3 and acetosyringone is added to a final concentration of 200 µM. Acetosyringone is added before mixing the solution with the rice cultures to induce *Agrobacterium* for DNA transfer to the plant cells. For inoculation, the plant cultures are immersed in the bacterial suspension. The liquid bacterial suspension is removed and the inoculated cultures are placed on co-cultivation medium and incubated at 22°C for two days. The cultures are then transferred to MS-CIM medium with ticarcillin (400 mg/liter) to inhibit the growth of *Agrobacterium*. For constructs utilizing the PMI selectable marker gene (Reed *et al.*, 2001), cultures are transferred to selection medium containing mannose as a carbohydrate source (MS with 2% mannose, 300 mg/liter ticarcillin) after 7 days, and cultured for 3-4 weeks in the dark. Resistant colonies are then transferred to regeneration induction medium (MS with no 2,4-D, 0.5 mg/liter IAA, 1 mg/liter zeatin, 200 mg/liter TIMENTIN®, 2% mannose, and 3% sorbitol) and grown in the dark for 14 days. Proliferating colonies are then transferred to another round of regeneration induction media and moved to the light growth room. Regenerated shoots are transferred to GA7 containers with GA7-1 medium (MS with no hormones and 2% sorbitol) for 2 weeks and then moved to the greenhouse when they are large enough and have adequate roots. Plants are transplanted to soil in the greenhouse (T₀ generation) grown to maturity and the T₁ seed is harvested.

III.F.3. Transformation of Plastids

Seeds of *Nicotiana tabacum* c.v. 'Xanthi nc' are germinated seven per plate in a 1" circular array on T agar medium and bombarded 12-14 days after sowing with 1 µm tungsten particles (M10; Bio-Rad Laboratories, Hercules,

California, United States of America) coated with DNA from plasmids pPH143 and pPH145 essentially as disclosed (Svab & Maliga, 1993). Bombarded seedlings are incubated on T medium for two days after which leaves are excised and placed abaxial side up in bright light (350-500 $\mu\text{mol photons/m}^2/\text{s}$) on plates of RMOP medium (Svab *et al.*, 1990) containing 500 $\mu\text{g/ml}$ spectinomycin dihydrochloride (Sigma-Aldrich Chemical Co., St. Louis, Missouri, United States of America). Resistant shoots appearing underneath the bleached leaves three to eight weeks after bombardment are subcloned onto the same selective medium, allowed to form callus, and secondary shoots isolated and subcloned. Complete segregation of transformed plastid genome copies (homoplasmy) in independent subclones is assessed by standard techniques of Southern blotting (Sambrook & Russell, 2001). *BamH I/EcoR I*-digested total cellular DNA (Mettler, 1987) is separated on 1% Tris-borate-EDTA (TBE) agarose gels, transferred to nylon membranes (Amersham Biosciences, Piscataway, New Jersey, United States of America) and probed with ^{32}P -labeled random primed DNA sequences corresponding to a 0.7 kb *BamH I/Hind III* DNA fragment from pC8 containing a portion of the *rps7/12* plastid targeting sequence. Homoplasmic shoots are rooted aseptically on spectinomycin-containing MS/IBA medium (McBride *et al.*, 1994) and transferred to the greenhouse.

IV. Plants, Breeding, and Seed Production

IV.A. Plants

The presently disclosed subject matter also provides plants comprising the disclosed nucleic acids. In some embodiments, the modification includes overexpression, underexpression, antisense modulation, sense suppression, inducible expression, inducible repression, or inducible modulation of a gene.

IV.B. Breeding

The plants obtained via transformation with a nucleotide sequence of the presently disclosed subject matter can be any of a wide variety of plant species, including monocots and dicots; however, the plants used in the method for the presently disclosed subject matter are selected in some embodiments from the

list of agronomically important target crops set forth hereinabove. The expression of a gene of the presently disclosed subject matter in combination with other characteristics important for production and quality can be incorporated into plant lines through breeding. Breeding approaches and techniques are known in the art. See e.g., Welsh, 1981; Wood, 1983; Mayo, 1987; Singh, 1986; Wricke & Weber, 1986.

The genetic properties engineered into the transgenic seeds and plants disclosed above are passed on by sexual reproduction or vegetative growth and can thus be maintained and propagated in progeny plants. Generally, the maintenance and propagation make use of known agricultural methods developed to fit specific purposes such as tilling, sowing, or harvesting. Specialized processes such as hydroponics or greenhouse technologies can also be applied. As the growing crop is vulnerable to attack and damage caused by insects or infections as well as to competition by weed plants, measures are undertaken to control weeds, plant diseases, insects, nematodes, and other adverse conditions to improve yield. These include mechanical measures such as tillage of the soil or removal of weeds and infected plants, as well as the application of agrochemicals such as herbicides, fungicides, gametocides, nematocides, growth regulants, ripening agents, and insecticides.

Depending on the desired properties, different breeding measures are taken. The relevant techniques are well known in the art and include, but are not limited to, hybridization, inbreeding, backcross breeding, multiline breeding, variety blend, interspecific hybridization, aneuploid techniques, etc. Hybridization techniques can also include the sterilization of plants to yield male or female sterile plants by mechanical, chemical, or biochemical means. Cross-pollination of a male sterile plant with pollen of a different line assures that the genome of the male sterile but female fertile plant will uniformly obtain properties of both parental lines. Thus, the transgenic seeds and plants according to the presently disclosed subject matter can be used for the breeding of improved plant lines that, for example, increase the effectiveness of conventional methods such as

herbicide or pesticide treatment or allow one to dispense with said methods due to their modified genetic properties. Alternatively new crops with improved stress tolerance can be obtained, which, due to their optimized genetic "equipment", yield harvested product of better quality than products that were not able to tolerate comparable adverse developmental conditions (for example, drought).

IV.C. Seed Production

Some embodiments of the presently disclosed subject matter also provide seed and isolated product from plants that comprise an expression cassette comprising a promoter sequence operably linked to a heterologous nucleic acid of interest and further optionally linked to a transcription terminator of interest, the promoter of interest being selected from the group consisting of:

- (a) a promoter sequence that hybridizes under highly stringent conditions of hybridization of 65°C in 6x SSC, followed by a final washing step of at least 15 minutes at 65°C in 0.1x SSC to a promoter sequence as set forth in any of SEQ ID NOs: 1-3, or a fragment, domain, or feature thereof that has transcriptional regulatory activity; and
- (b) a promoter sequence that is an ortholog of any of SEQ ID NOs: 1-3, or a fragment, domain, or feature thereof that has transcriptional regulatory activity.

In some embodiments, the presently disclosed subject matter provides seed and isolated product from plants that comprise an expression cassette comprising a promoter sequence operably linked to a heterologous nucleic acid of interest and also a transcription terminator of interest, the transcription terminator of interest being selected from the group consisting of SEQ ID NOs: 3-6.

In some embodiments the isolated product comprises an enzyme, a nutritional polypeptide, a structural polypeptide, an amino acid, a lipid, a fatty acid, a polysaccharide, a sugar, an alcohol, an alkaloid, a carotenoid, a propanoid, a steroid, a pigment, a vitamin, or a plant hormone.

In seed production, germination quality, and uniformity of seeds are important product characteristics. As it is difficult to keep a crop free from other crop and weed seeds, to control seedborne diseases, and to produce seed with good germination, fairly extensive and well-defined seed production practices have been developed by seed producers who are experienced in the art of growing, conditioning, and marketing of pure seed. Thus, it is common practice for the farmer to buy certified seed meeting specific quality standards instead of using seed harvested from his own crop. Propagation material to be used as seeds is customarily treated with a protectant coating comprising herbicides, insecticides, fungicides, bactericides, nematocides, molluscicides, or mixtures thereof. Customarily used protectant coatings comprise compounds such as captan, carboxin, thiram (tetramethylthiuram disulfide; TMTD®; available from R. T. Vanderbilt Company, Inc., Norwalk, Connecticut, United States of America), methalaxyl (APRON XL®; available from Syngenta Corp., Wilmington, Delaware, United States of America), and pirimiphos-methyl (ACTELLIC®; available from Agrilience, LLC, St. Paul, Minnesota, United States of America). If desired, these compounds are formulated together with further carriers, surfactants, and/or application-promoting adjuvants customarily employed in the art of formulation to provide protection against damage caused by bacterial, fungal, or animal pests. The protectant coatings can be applied by impregnating propagation material with a liquid formulation or by coating with a combined wet or dry formulation. Other methods of application are also possible such as treatment directed at the buds or the fruit.

V. Additional Compositions and Methods

The presently disclosed subject matter also provides in some embodiments cells containing the isolated promoters disclosed herein. In some embodiments, the isolated promoters disclosed herein are operably linked to one or more heterologous nucleotide sequences and/or a transcriptional terminator as disclosed herein. In some embodiments, the cells are selected from the group

consisting of bacterial cells, mammalian cells, insect cells, plant cells, and fungal cells. In some embodiments, the cells are *Agrobacterium tumefaciens* cells.

The presently disclosed subject matter also provides in some embodiments methods for producing transformed plants. In some embodiments, the presently disclosed methods comprise providing a nucleic acid construct comprising an isolated promoter as disclosed herein (e.g., comprising a promoter sequence of any of SEQ ID NOs: 1-3 or a fragment thereof that has transcriptional regulatory activity, and optionally further comprising a transcription terminator as set forth in any of SEQ ID NOs: 4-6) and transforming a plant with the nucleic acid construct.

The presently disclosed subject matter also provides in some embodiments methods for identifying transcriptional regulatory elements with transcriptional regulatory activities. In some embodiments, the methods comprise (a) providing an expression construct comprising a nucleic acid molecule in which a putative transcriptional regulatory element, a coding sequence encoding a reporter, and optionally a transcription terminator are operably linked; (b) transforming a cell with the expression construct; and (c) assaying for expression of the reporter, wherein expression of the reporter is indicative of the putative transcriptional regulatory element having transcriptional regulatory activity.

The presently disclosed subject matter also provides in some embodiments methods for identifying non-*Zea mays* nucleic acid molecules with transcriptional regulatory activity, which in some embodiments are promoters from orthologous loci to those from which SEQ ID NOs: 1-3 have been isolated. In some embodiments, the methods comprise (a) providing a first nucleic acid molecule that comprises a nucleotide sequence that is at least 95% identical to at least 50 consecutive bases of any of SEQ ID NOs: 1-3; (b) hybridizing the first nucleic acid molecule to a plurality of nucleic acid molecules, wherein the plurality of nucleic acid molecules is isolated from a species other than *Zea mays*; (c) isolating a second nucleic acid molecule present in the plurality of nucleic acid molecules that specifically hybridizes to the first nucleic acid

molecule; and (d) assaying the second nucleic acid molecule for transcriptional regulatory activity, whereby a non-*Zea mays* nucleic acid molecule with transcriptional regulatory activity is identified.

In some embodiments, the plurality of nucleic acid molecules comprises a library of genomic DNA from a plant species other than *Zea mays*. Methods for preparing and screening genomic libraries are known to those of skill in the art and are described, for example, in Sambrook & Russell, 2001. In some embodiments, the methods further comprise screening a genomic library from the species other than *Zea mays* and identifying a genomic DNA molecule present in the genomic library that is at least 1000 nucleotides in length and that comprises a nucleotide sequence that is at least 95% identical to the nucleotide sequence of the second nucleic acid molecule over at least 500 nucleotides.

EXAMPLES

The following Examples have been included to illustrate representative and exemplary modes of the presently disclosed subject matter. In light of the present disclosure and the general level of skill in the art, those of skill will appreciate that the following Examples are intended to be exemplary only and that numerous changes, modifications, and alterations can be employed without departing from the spirit and scope of the presently disclosed subject matter.

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are disclosed in Sambrook & Russell, 2001; Silhavy *et al.*, 1984; Ausubel *et al.*, 2002; Ausubel *et al.*, 2003; Reiter *et al.*, 1992; and Schultz *et al.*, 1998.

Materials and Methods Employed in the EXAMPLES

General Construct Design. For each cassette, the 5'UTR and a variable number of upstream bases were selected to comprise a promoter of 2000 bp. The 3'UTR and a variable number of downstream bases were selected to

comprise a terminator of 1000 bp. Restriction enzyme recognition sites were engineered into the vector for cloning purposes.

prAz1. The prAz1 promoter nucleotide sequence present in this construct was modified to contain one point mutation to eliminate an internal restriction enzyme site. Specifically, the modification includes a G to A change at base pair 900 to eliminate a *BamHI* restriction site present therein. The resulting sequence is depicted as SEQ ID NO: 1.

tAz1. The tAz1 terminator nucleotide sequence present in this construct was modified to contain two mutations to disrupt (i.e., inactivate) outward facing ORFs. Specifically, the modifications include an A to T change at base pair 970 and an A to T change at base pair 993. The resulting sequence is depicted as SEQ ID NO: 4.

prAz2. The prAz2 promoter nucleotide sequence present in this construct was modified to contain one point mutation to disrupt outward facing ORF. Specifically, the modification includes a T to A change at base pair 66. The resulting sequence is depicted as SEQ ID NO: 2.

tAz2. The tAz1 terminator nucleotide sequence present in this construct was modified to contain four mutations to remove restriction sites and outward facing ORFs. Specifically, the modification includes a G to A change at base pair 255 to eliminate a *BamHI* restriction site present therein and a G to A change at base pair 658 to eliminate a *SanDI* restriction site present therein. Furthermore, the modifications include an A to T change at base pair 914 and an A to T change at base pair 983 to remove outward facing ORFs. The resulting sequence is depicted as SEQ ID NO: 5.

prAz3. The nucleotide sequence present in this construct was modified to contain one point mutation to eliminate an internal restriction enzyme site. Specifically, the modification includes a G to A change at base pair 129 to eliminate a *BamHI* restriction site present therein present therein. The resulting sequence is depicted as SEQ ID NO: 3.

tAz3. The tAz1 terminator nucleotide sequence present in this construct was modified to contain one mutation to remove outward facing ORF. Specifically, the modification includes an A to T change at base pair 969. The resulting sequence is depicted as SEQ ID NO: 6.

5 Az1 Native Signal Sequence Identification. The Az1 locus was determined to correspond to Accession No. Q946V4 in the Uniprot database. Uniprot Accession No. Q946V4 does not identify the signal sequence for the corresponding protein. However, several closely related database entries (Accession Nos. P04698, P04699, P04700, and PO6679, each with >85%
10 identity to Q946V4) suggested that the signal sequence could potentially be the first 21 amino acids of Q946V4. Consequently, the first 21 amino acids from the Az1 gene product, MASKTLLSLLALLALFVSATNA (SEQ ID NO: 7), were tested for the ability to act as a native signal sequence.

15 EXAMPLE 1

Identification of α -Zein Promoters

Microarray data generated using an Affymetrix platform was analyzed, and three α -zein genes that showed strong seed expression in endosperm, embryo, and pericarp of developing kernels were identified. Three probe sets representing genes Az1, Az2, and Az3, were found to have average expression signals at least 100-fold higher in seed samples than in other samples. All of these zein genes were highly expressed in seed, whereas their expression in other organ and tissue types was only detected at background levels.

A sequence search of cDNA and genomic DNA databases and
25 comparison with published sequences resulted in identification of a unique gene
locus matching each of the three α -zein gene candidates on B73 sequenced
BACs. The promoter region, 5'UTR, protein coding sequence, and 3'UTR were
annotated for each of Az1, Az2, and Az3.

Based on the annotated sequences, five constructs were designed using
30 putative gene regulatory sequences from these loci. These constructs are

described in Table 1. All vectors used cCel15112ER-03 as a reporter gene except Construct A due to the incorporation of a novel signal sequence into this Construct. Both cCel15112ER-03 and cCel15112ER-05 encode a cellobiohydrolase I (CBHI) enzyme. Both versions of the CBHI enzyme are maize codon optimized, the sequences set forth in the Sequence Listing of U.S. Patent Application Publication No. 20100319089 of Azhakanandam *et al.*, the entire disclosure, including the Sequence Listing, of which is incorporated herein by reference. Three vectors (Constructs B, D, and E) employed the native promoter with native 5'UTR and native termination sequences with native 3'UTR from the corresponding source α -zein gene as well as the γ -zein signal sequence as part of the CDS. The γ -zein signal sequence is MRVLLVALALLALAASATS (SEQ ID NO: 8) The native signal sequence from the Az1 locus replaced the γ -zein signal sequence in Construct A. The transcriptional terminator t35s, which is a standard CaMV 35S transcription terminator, replaced the native tAz1 terminator in Construct C.

Table 1
Promoters, Signal Sequences, and Termination Sequences

Construct	Promoter*	Signal Sequence	CDS	3'UTR**
A	prAz1	xAz1ss	cCel15112ER-05	tAz1
B	prAz1	xGzss	cCel15112ER-03	tAz1
C	prAz1	xGzss	cCel15112ER-03	t35s
D	prAz2	xGzss	cCel15112ER-03	tAz2
E	prAz3	xGzss	cCel15112ER-03	tAz3

* includes 5'UTR sequence

** includes transcriptional termination sequence

20

EXAMPLE 2

Production of Transgenic Plants

Transgenic plants were generated in the AX707 background by *Agrobacterium*-mediated transformation and phosphomannose isomerase (PMI) selection using standard techniques (see *e.g.*, Negrotto *et al.*, 2000; U.S. Patent

No. 7,119,255). Twenty pooled T1 seed from each single-copy, selfed event were ground to flour using a KLECO tissue pulverizer (Kleco, Visalia, California, United States of America). Duplicate flour samples assayed for CBHI levels in extracts were measured by ELISA. CBHI in these construct produced both a full-length polypeptide, which retained the C-terminal cellulose binding domain, and a truncated polypeptide that included the catalytic domain only. The ELISA employed a catalytic domain only standard for quantification. Thus, relative expression levels could be compared to evaluate the performance of events and constructs. The results from the ELISA using the truncated (catalytic domain only) CBHI as a standard are shown in Table 2. Samples were also run with a full-length standard, which produced similar relative expression levels.

Table 2Relative levels of CBHI catalytic domain in flour extracts

Construct	Sample	Average Relative Expression (ngCBHI / mg total soluble protein)	Standard Deviation
A	1	1.973	1.827
A	2*	1.000	0.370
A	3	0.650	0.095
A	4	0.632	0.191
A	5	0.524	0.106
A	6	0.481	0.047
A	7	0.459	0.170
A	8	0.364	0.360
A	9	0.363	0.121
A	10	0.083	0.004
B	1	1.359	0.721
B	2	1.035	0.317
B	3	1.029	0.142
B	4	0.936	0.596
B	5	0.865	0.176
B	6	0.784	0.125
B	7	0.697	0.037
B	8	0.628	0.154
B	9*	0.575	0.003
B	10	0.398	0.090
C	1	0.190	0.053
C	2	0.105	0.028

C	3	0.077	0.024
C	4	0.053	0.013
C	5*	0.049	0.024
C	6	0.041	0.013
C	7	0.032	0.002
C	8	0.017	0.009
C	9	0.010	0.003
D	1	3.054	0.157
D	2	1.701	0.670
D	3	1.493	0.447
D	4	1.457	0.802
D	5	1.241	0.454
D	6*	0.819	0.273
E	1	2.853	0.962
E	2	2.309	0.368
E	3	2.289	0.253
E	4	1.985	0.210
E	5*	1.200	0.601

Table 2 is shows the relative levels of CBHI catalytic domain in flour extracts of plants transformed with Constructs A-E (see Table 1). Units are relative expression levels quantified via ELISA (ng CBHI/mg total soluble protein). Events noted with asterisks were retested for the analysis in Table 4. Each data point represents the average of protein levels in the duplicate extractions of flour subsamples from 20 pooled seeds. Events were sorted in order of decreasing protein levels for each Construct.

Though the number of events screened was low, general trends were observed. The three constructs that combined the native promoter, 5'UTR, 3'UTR, and terminator sequence from the same gene (Constructs B, D, and E)

produced more events with higher expression levels. A particularly striking result was the poor performance of Construct C, in which the Az1 terminator in Construct B was replaced with the CaMV 35S terminator. This result indicated that using a native terminator from the same gene as the promoter enhanced transgene expression. Finally, use of the Az1 signal sequence in Construct A did not appear to enhance performance of this expression cassette as measured by reporter gene product accumulation relative to Construct B, which used the γ -zein signal sequence.

EXAMPLE 3

Comparison to a Known Strong Endosperm Promoter

How expression from these promoters compared to a high expressing event with a well-characterized and strong endosperm promoter was also tested. The reporter gene used here (cCel15112ER-03) had been previously expressed in transgenic corn under the control of the strong rice glutelin-1 promoter (osGT1, prGTL; Russell et al. 1997). The α -zein constructs described here were designed based on vector 15943. Vector 15943 is a binary vector containing the following components: a maize Ubiquitin promoter operably linked to an *E. coli* PMI gene operably linked to a synthetic NOS terminator and a rice glutelin promoter operably linked to a synthetic CBHI gene operably linked to a maize PEPC intron operably linked to a CMV 35S terminator. Two events (one medium expresser and one high expresser based on previous accumulated activity data) from this benchmark control 15943 were selected for screening. Two T₁ reference plants from each of these two events were selected for seed analysis. T₂ seed from these four T₁ reference plants were assayed along with T₁ seed from one randomly selected event from each α -zein construct above (Table 1 events identified with asterisks).

Seed from the five new constructs containing α -zein promoters and the medium-expressing benchmark controls containing the GTL promoter were from selfed, single copy parents, so the T₁ or T₂ seed, respectively, was expected to be segregating for these events. The high-expressing controls were from a

homozygous selfed plant which contained a 2-copy insert making the diploid embryo all 4-copy (and triploid endosperm 6 copy; see Table 3 for details).

Twenty seed samples for each line were pooled and ground to flour. Flour samples were analyzed in duplicate by ELISA. The results from the ELISA using the truncated (catalytic domain only) CBHI as a standard are shown in Table 3. Samples were also run with a full-length standard, which produced similar relative expression levels.

Table 3
15943 (prGTL) Control Events*

T ₂ Benchmark Seed ID	T0 Insertion	T1 Copy	T2 Seed	Embryo Copy	Endosperm Copy
1.1	1/0	1/0	Segregating	0, 1, 2	0, 1, 2, 3
1.2	1/0	1/0	Segregating	0, 1, 2	0, 1, 2, 3
1.3	2/0	2/2	Fixed	4	6
1.4	2/0	2/2	Fixed	4	6

* Numbers indicate copies of the desired expression cassette in each plant and tissue for each of the four seed IDs listed. 1/0 indicates presence of a single copy in hemizygous plant. 2/0 indicates a two copy insertion in hemizygous plant. 2/2 indicates a two copy event in a homozygous plant (and therefore fixed in T₂ seed).

Table 4
Relative Levels of CBHI in Flour Extracts¹

Construct (A-E) / Seed (1.1-1.4)	Components	Relative Expression - Catalytic Domain Standard (ng CBHI/mg TSP)	
		Result	σ
A	prAz1, ssAz1, tAz1	<u>4.4</u>	0.04
B	prAz1, ssGz, tAz1	<u>3.2</u>	0.08
C	prAz1, ssGz, t35S	0.4	0.13

D	prAz2, ssGz, tAz2	<u>7.3</u>	1.14
E	prAz3, ssGz, tAz3	<u>9.2</u>	0.17
1.1	prGTL, ssGz, t35S (MID)	1.7	0.66
1.2	prGTL, ssGz, t35S (MID)	1.0	0.05
1.3	prAz1, ssGz, t35S (HIGH)	2.4	0.35
1.4	prAz1, ssGz, t35S (HIGH)	2.8	0.29

¹ Results from ELISA using the truncated CBHI (catalytic domain) standard. The α -zein derived expression cassettes are indicated in the first column. The control expression cassettes with the rice glutelin-1 promoter (prGTL) and the CaMV 35S terminator (t35S) are shown below the double line.

- 5 *Note: A high expressing event was randomly selected for Construct A and a low expressing event was randomly selected for Construct B. Because of this, it appears as though the Az1 signal sequence (ssAz1) outperformed the γ -zein signal sequence (ssGz). This was not the case when looking across all events generated (see Table 2).
- 10 In a side-by-side comparison of the alpha-zein derived expression cassettes described herein and the well-characterized rice glutelin-1 derived reference cassettes, the alpha-zein derived sequences consistently outperformed the reference cassettes (see underlined protein level values in Table 4).

15 Discussion of the EXAMPLES

Disclosed herein are three new α -zein promoters, including the native terminators, which produce very high expression in transgenic seed. These cassettes outperformed best events with the rice glutelin-1 promoter by a factor of 2-4. This was a surprising finding, given that the T₁ seed of the new promoters was, on average, of a lower relative gene copy (by insert complexity and

20 zygosity) than the high expressing benchmark with prGTL.

Incorporation of the native terminator with prAz1 in place of the standard CaMV 35S terminator resulted in a fifteen-fold increase in expression.

Finally, incorporation of the native signal sequence from Az1 (ssAz1) contiguous with its promoter did not appear to result in enhanced expression compared to using the preferred and extensively utilized γ -zein signal sequence (ssGz).

5 As such, several promoters have been identified that are capable of regulating high level gene expression in plant endosperm and/or embryos. For example, the rice glutelin-1 (prGTL), maize γ -zein (prGZein), and maize globulin (prGlob) promoters have been used to direct high endosperm (prGTL & prGZein) and/or embryo (prGlob) expression in maize transgenics. However, additional
10 seed-specific promoters could also provide potential benefits for modulating gene expression in seeds. Disclosed herein are promoters that have been identified for three highly expressed α -zein gene products.

Disclosed herein is an effective approach for identifying promoter candidates that regulate specific expression patterns that can be used for trait
15 research and transgenic product development. Incorporation of the full genome sequence data, physical genome location, and extensive expression profiling data allowed for greater quality control and precision when selecting promoter leads for validation through either stable transformation or transient expression.

Also disclosed herein are expression cassettes designed with a “native
20 gene” approach that is a modification of that disclosed in U.S. Patent Application Publication No. 20070006344 of Nuccio *et al.*, the entire disclosure of which is incorporated herein by reference. Because zeins lack introns, the promoter design was simplified to include about 2 kilobases (kb) of 5' non-transcribed sequence that included the upstream untranslated region (5'UTR). The native
25 terminator for each promoter contained about 1 kb of downstream sequence including the 3'-untranslated region (3'UTR). For one α -zein candidate (Az1), expression of the native terminator was also compared to a construct that included the standard 35S viral terminator. All of the expression cassettes were designed with the standard γ -zein signal sequence. For the same α -zein

candidate above, the efficacy of utilizing a native signal sequence contiguous with its promoter and 5'UTR was also tested.

REFERENCES

- All references listed below, as well as all references cited in the instant disclosure, including but not limited to all patents, patent applications and publications thereof, scientific journal articles, and database entries (e.g., GENBANK® database entries and all annotations available therein) are incorporated herein by reference in their entireties to the extent that they supplement, explain, provide a background for, or teach methodology, techniques, and/or compositions employed herein.
- Allison RF, Johnston RE & Dougherty WG (1986) The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology* 154:9-20.
- Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic Local Alignment Search Tool. *J Mol Biol* 215:403-410.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA & Struhl K (2002) Short Protocols in Molecular Biology, Fifth ed. Wiley, New York, New York, United States of America.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA & Struhl K (2003) Current Protocols in Molecular Biology, John Wiley & Sons, Inc., New York, New York, United States of America.
- Ballas N, Broido S, Soreq H, & Loyter A (1989) Efficient functioning of plant promoters and poly(A) sites in *Xenopus* oocytes. *Nucleic Acids Res.* 17:7891-903.
- Bartlett SG, Grossman AR & Chua N-H (1982) in Methods in Chloroplast Molecular Biology, (Edelman M, Hallick RB & Chua N-H, eds.) Elsevier Biomedical Press, New York, New York, United States of America, pp. 1081-1091.

- Batzer MA, Carlton JE & Deininger PL (1991) Enhanced evolutionary PCR using oligonucleotides with inosine at the 3'-terminus. *Nucleic Acid Res.* 19:5081.
- 5 Bevan M (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucl. Acids Res* 12:8711-21.
- Bevan M, Flavell R B & Chilton M D (1983) A chimeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature* 304:184-187.
- Bloehinger K & Diggelmann H (1984) Hygromycin B phosphotransferase as a selectable marker for DNA transfer experiments with higher eukaryotic cells. *Mol Cell Biol* 4:2929-2931.
- 10 Bourouis M & Jarry B (1983) Vectors containing a prokaryotic dihydrofolate reductase gene transform *Drosophila* cells to methotrexate-resistance. *EMBO J* 2:1099-1104.
- Callis J, Fromm M & Walbot V (1987) Introns increase gene expression in cultured maize cells. *Genes Dev.* 1:1183-1200.
- 15 Christou P, Ford T & Kofron M (1991) Production of transgenic rice (*Oryza sativa* L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. *Bio/Technology* 9: 957-962.
- 20 Comai L, Larson-Kelly N, Kiser J, Mau CJ, Pokalsky AR, Shewmaker CK, McBride K, Jones A & Stalker DM (1988) Chloroplast transport of a ribulose biphosphate carboxylase small subunit-5-enolpyruvyl 3-phosphoshikimate synthase chimeric protein requires part of the mature small subunit in addition to the transit peptide. *J Biol Chem* 263:15104-15109.
- 25 Datta SK, Peterhans A, Datta K & Potrykus I (1990) Genetically engineered fertile Indica-rice recovered from protoplasts. *Bio/Technology* 8:736-740.
- Della-Cioppa G, Kishore GM, Beachy RN & Fraley RT (1987) Protein trafficking in plant cells. *Plant Physiol* 84:965-968.

- Dong, JJ, Teng WM, Buchholz WG & Hall TC (1996). *Agrobacterium*-mediated transformation of javanica rice. *Mol Breeding* 2:267-276.
- Elroy-Stein O, Fuerst TR & Moss B (1989) Cap-Independent Translation of mRNA Conferred by Encephalomyocarditis Virus 5' Sequence Improves the Performance of the Vaccinia Virus/Bacteriophage T7 Hybrid Expression System. *Proc Natl Acad Sci USA* 86:6126-6130.
- European Patent Applications EP 0 332 104; EP 0 292 435; EP 0 332 581; EP 0 392 225.
- Fromm ME, Morrish F, Armstrong C, Williams R, Thomas J & Klein TM. (1990) Inheritance and expression of chimeric genes in the progeny of transgenic maize plants. *Biotechnology (NY)* 8:833-839.
- Gallie DR, Kado CI, Hershey JWB, Wilson MA & Walbot V (1989) in Molecular Biology of RNA (Cech TR, ed.), Alan R. Liss, Inc., New York, New York, United States of America, pp. 237-256.
- Gallie DR, Sleat DE, Watts JW, Turner PC & Wilson TM (1987) A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression *in vivo*. *Nucl Acids Res* 15:8693-8711.
- GENBANK® Accession Nos AC1912673; AC1912673; AC1984173; AC1989014; AC1989713; AC2021643; AC2121944; AC2144283; AC2160673.
- Gordon-Kamm WJ, Spencer TM, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams WR Jr., Willetts NG, Rice TB, Mackey CJ, Krueger RW, Kausch AP & Lemaux PG (1990) Transformation of Maize Cells and Regeneration of Fertile Transgenic Plants. *Plant Cell* 2:603-618.
- Green MR (2000) TBP-associated factors (TAFIIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci* 25:59-63.
- Gritz L & Davies J (1983) Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* 25:179-188.

- Guerineau F, Brooks L, & Mullineaux P (1991) Effect of deletions in the cauliflower mosaic virus polyadenylation sequence on the choice of the polyadenylation sites in tobacco protoplasts. *Mol Gen Genet.* 226:141-144.
- 5 Henikoff S & Henikoff JG (1992) Amino Acid Substitution Matrices from Protein Blocks. *Proc Natl Acad Sci U S A* 89:10915-10919.
- Hiei Y, Komari T & Kubo T (1997) Transformation of rice mediated by *Agrobacterium tumefaciens*. *Plant Mol Biol* 35:205-18.
- Hiei Y, Ohta S, Komari T & Kumashiro Y (1994) Efficient transformation of rice
10 (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J* 6(2):271-282.
- Höfgen R & Willmitzer L (1988) Storage of competent cells for *Agrobacterium* transformation. *Nucl Acids Res* 16:9877.
- Holding DR & Larkins BA (2006) The Development and Importance of Zein
15 Protein Bodies in Maize Endosperm. *Maydica* 51:243-254.
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W and P Zimmermann (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics* 2008, Article ID 420747.
- 20 Jobling SA & Gehrke L (1987) Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. *Nature* 325:622-625.
- Joshi CP (1987) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucleic Acids Res.* 15:9627-9640.
- 25 Karlin S & Altschul SF (1993) Applications and Statistics for Multiple High-Scoring Segments in Molecular Sequences. *Proc Natl Acad Sci U S A* 90:5873-5877.
- Klein TM, Wolf ED, Wu R & Sanford JC (1987) High velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327:70-73.

- Koehler SM & Ho TM (1990) Hormonal regulation, processing, and secretion of cysteine proteinases in barley aleurone layers. *Plant Cell* 2:769-783.
- Kong P & Steinbiss HH (1998) Complete nucleotide sequence and analysis of the putative polyprotein of maize dwarf mosaic virus genomic RNA
5 (Bulgarian isolate). *Arch Virol* 143:1791-1799.
- Kozziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, Crossland L, Dawson J, Desai N, Hill M, Kadwell S, Launis L, Lewis K, Maddox D, McPherson K, Meghji MR, Merlin E, Rhodes R, Warren GW, Wright M & Evola SV (1993) Field performance of elite transgenic maize plants
10 expressing an insecticidal protein derived from *Bacillus thuringiensis*. *Bio/Technology* 11:194-200.
- Lommel SA, Kendall TL, Xiong Z & Nutter RC (1991) Identification of the maize chlorotic mottle virus capsid protein cistron and characterization of its subgenomic messenger RNA. *Virology* 81:382-385.
- 15 Macejak DG & Sarnow P (1991) Internal initiation of translation mediated by the 5' leader of a cellular mRNA. *Nature* 353:90-94.
- Mayo O (1987) The Theory of Plant Breeding, Second Edition, Clarendon Press, New York, New York, United States of America.
- McBride KE & Summerfelt KR (1990) Improved binary vectors for *Agrobacterium*-
20 mediated plant transformation. *Plant Mol Biol* 14: 269-276.
- Messing J & Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19:259-268.
- Mettler IJ (1987) A simple and rapid method for miniprep of DNA from tissue-cultured plant cells *Plant Mol Biol Reporter* 5:346-349.
- 25 Mogen BD, MacDonald MH, Graybosch R, & Hunt AG (1990) Upstream sequences other than AAUAAA are required for efficient messenger RNA 3'-end formation in plants. *Plant Cell*. 2:1261-1272.
- Mukumoto F, Hirose S, Imaseki H & Yamazaki K (1993) DNA sequence requirement of a TATA element-binding protein from Arabidopsis for
30 transcription *in vitro*. *Plant Mol Biol* 23: 995-1003.

- Munroe D & Jacobson A (1990) Tales of poly(A): a review. *Gene*. 91:151-158.
- Murashige T & Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiologia Plantarum* 15: 473-497 (1962)
- Needleman SB & Wunsch CD (1970) A General Method Applicable to the Search
5 for Similarities in the Amino Acid Sequence of Two proteins. *J Mol Biol*
48:443-453.
- Negrotto D, Jolley M, Beer S, Wenck AR & Hansen G (2000) The use of phosphomannose-isomerase as a selection marker to recover transgenic maize plant (*Zea mays* L.) via *Agrobacterium* transformation. *Plant Cell Reports* 19:798-803.
10
- Norris SR, Meyer SE & Callis J (1993) The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. *Plant Mol Biol* 21:895-906.
- Odell JT, Nagy F, & Chua NH (1985) Identification of DNA sequences required
15 for activity of the cauliflower mosaic virus 35S promoter. *Nature*. 313:810-812.
- Ohtsuka E, Matsuki S, Ikehara M, Takahashi Y & Matsubara K (1985) An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J. Biol. Chem.*
20 260:2605-2608.
- Paszkowski J, Shillito RD, Saul M, Mandak V, Hohn T, Hohn B & Potrykus I (1984) Direct gene transfer to plants. *EMBO J* 3:2717-2722.
- PCT International Patent Application Publication Nos. WO 93/07278; WO 93/21335; WO 94/00977.
- 25 Pearson WR & Lipman DJ (1988) Improved Tools for Biological Sequence Comparison. *Proc Natl Acad Sci U S A* 85:2444-2448.
- Potrykus I, Paszkowski J, Saul MW, Petruska J & Shillito RD (1985) Molecular and general genetics of a hybrid foreign gene introduced into tobacco by direct gene transfer. *Mol Gen Genet* 199:169-177.
- 30 Proudfoot N (1991) Poly(A) signals. *Cell*. 64:671-674.

- Reich TJ, Iyer VN & Miki BL (1986) Efficient transformation of alfalfa protoplasts by the intranuclear microinjection of Ti-plasmids. *Bio/Technology* 4:1001-1004.
- 5 Reiter RS, Young RM & Scolnik PA (1992) Genetic Linkage of the Arabidopsis Genome: Methods for Mapping with Recombinant Inbreds and Random Amplified Polymorphic DNAs (RAPDs), in Methods in Arabidopsis Research, World Scientific Press, River Edge, New Jersey, United States of America.
- 10 Rogers JC, Dean D & Heck GR (1985) Aleurain: a barley thiol protease closely related to mammalian cathepsin H. *Proc. Natl. Acad. Sci. USA* 82:6512-6516.
- Rosenberg AH, Lade BN, Chui DS, Lin SW, Dunn JJ, & Studier FW (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*. 56:125-135.
- 15 Rossolini GM, Cresti S, Ingianni A, Cattani P, Riccio ML & Satta G (1994) Use of deoxyinosine-containing primers vs degenerate primers for polymerase chain reaction based on ambiguous sequence information. *Mol Cell Probes* 8:91-98.
- 20 Roth BA, Goff SA, Klein TM & Fromm ME (1991) C1- and R-dependent expression of the maize Bz1 gene requires sequences with homology to mammalian myb and myc binding sites. *Plant Cell* 3:317-325.
- Rothstein SJ, Lahners KN, Lotstein RJ, Carozzi NB, Jayne SM & Rice DA (1987) Promoter cassettes, antibiotic-resistance genes, and vectors for plant transformation. *Gene* 53:153-161.
- 25 Russell DA & Fromm M (1997) Tissue-specific expression in transgenic maize of four endosperm promoters from maize and rice. *Transgenic Research* 6:157-168.
- 30 Sambrook J & Russell DW (2001) Molecular Cloning : A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, United States of America.

- Sanfaçon H, Brodmann P, Hohn T (1991) A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev.* 5:141-149.
- Schultz B et al. (1998) T-DNA tagging in *Arabidopsis thaliana*: Cloning by gene disruption, in Plant Molecular Biology Manual (2nd edition, Gelvin SB, Schilperoort RA & Verma DPS, eds.) Kluwer Academic Publishers, New York, New York, United States of America.
- Shimamoto K, Terada R, Izawa T & Fujimoto H (1989) Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* 338:274-276.
- Shinshi H, Neuhas JM, Ryals J & Meins F Jr. (1990) Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. *Plant Mol Biol* 14:357-368.
- Silhavy TJ, M.L. Berman, and L.W. Enquist (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America.
- Singh DP (1986) Breeding for Resistance to Diseases and Insect Pests, Springer-Verlag, New York, New York, United States of America.
- Skuzeski JM, Nichols LM & Gesteland RF (1990) Analysis of leaky viral translation termination codons in vivo by transient expression of improved beta-glucuronidase vectors. *Plant Mol Biol* 15:65-79.
- Smith TF & Waterman M (1981) Comparison of Biosequences. *Adv Appl Math* 2:482-489.
- Spencer TM, Gordon-Kamm WJ, Daines RJ, Start W & Lemaux P (1990). *Theor Appl Genet* 79:625-631.
- Stormo, GD (2000) Gene-Finding Approaches for Eukaryotes. *Genome research* 10:394-397.
- Svab Z, Hajdukiewicz P & Maliga P (1990) Stable transformation of plastids in higher plants. *Proc Natl Acad Sci USA* 87:8526–8530.
- Svab Z & Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc Natl Acad Sci USA* 90:913–917.

- Thompson CJ, Movva NR, Tizard R, Cramer R, Davies JE, Lauwereys M & Botterman J (1987) Characterization of the herbicide-resistance gene bar from *Streptomyces hygroscopicus*. *EMBO J* 6:2519-2523.
- U.S. Patent Application Publication No. 2007/0006344.
- 5 U.S. Patent Nos. 4,940,935; 4,945,050; 5,036,006; 5,100,792; 5,188,642; 5,591,616; 5,639,949; 5,767,378; 5,994,629; 6,369,298.
- Uknes S, Mauch-Mani B, Moyer M, Potter S, Williams S, Dincher S, Chandler D, Slusarenko A, Ward E & Ryals J (1992) Acquired resistance in *Arabidopsis*. *Plant Cell* 4:645-656.
- 10 Unger EA, Hand JM, Cashmore AR & Vasconcelos AC (1989) Isolation of a cDNA encoding mitochondrial citrate synthase from *Arabidopsis thaliana*. *Plant Mol Biol* 13:411-418.
- Van den Broeck G, Timko MP, Kausch AP, Cashmore AR, Van Montagu M & Herrera-Estrella L (1985) Targeting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase. *Nature* 313:358-363.
- 15 Vasil V, Castillo AM, Fromm ME & Vasil IK (1992) Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. *Bio/Technology* 10:667-674.
- 20 Vasil V, Srivastava V, Castillo AM, Fromm MR & Vasil IK (1993) Rapid production of transgenic plants by direct bombardment of cultured immature embryos. *Bio/Technology* 11:1553-1558.
- Wasmann CC, Reiss B, Bartlett SG & Bohnert HJ (1986) The importance of the transit peptide and the transported protein for protein import into chloroplasts. *Mol Gen Genet* 205:446-453.
- 25 Weeks JT, Anderson OD & Blechl AE (1993) Rapid Production of Multiple Independent Lines of Fertile Transgenic Wheat (*Triticum aestivum*). *Plant Physiol* 102:1077-1084.
- Welsh JR (1981) Fundamentals of Plant Genetics and Breeding, John Wiley & Sons, New York, New York, United States of America.
- 30

- White J, Chang SY & Bibb MJ (1990) A cassette containing the bar gene of *Streptomyces hygrosopicus*: a selectable marker for plant transformation. *Nucl Acids Res* 18:1062.
- 5 Woo Y-M, Hu DW-N, Larkins BA, and Jung R (2001) Genomics Analysis of Genes Expressed in Maize Endosperm Identifies Novel Seed Proteins and Clarifies Patterns of Zein Gene Expression. *Plant Cell* 13:2297-2317.
- Wood DR (ed.) (1983) Crop Breeding, American Society of Agronomy, Madison, Wisconsin, United States of America.
- 10 Wricke G & Weber WE (1986) Quantitative Genetics and Selection Plant Breeding, Walter de Gruyter and Co., Berlin, Germany.
- Zhang HM, Yang H, Rech EL, Golds TJ, Davis AS , Mulligan BJ, Cocking EC & Davey MR (1988) Transgenic rice plants produced by electroporation-mediated plasmid uptake into protoplasts. *Plant Cell Reports* 7: 379-384.
- 15 Zhao ZY, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J & Pierce D (2000) Agrobacterium-mediated sorghum transformation. *Plant Mol Biol* 44:789-98.

It will be understood that various details of the presently disclosed subject matter can be changed without departing from the scope of the presently disclosed subject matter. Furthermore, the foregoing description is for the purpose of illustration only, and not for the purpose of limitation.

CLAIMS

What is claimed is:

1. An nucleic acid molecule comprising a plant promoter, wherein the nucleic acid molecule comprises a nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 1-3.
5
2. The nucleic acid molecule of claim 1, wherein the nucleotide sequence is 100% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 1-3.
10
3. The nucleic acid molecule of claim 1, further comprising a first exon, an intron, and optionally a second exon or a fragment thereof, operably linked thereto.
15
4. The nucleic acid molecule of claim 3, wherein the promoter and the first exon and intron, and optionally the second exon or the fragment thereof if present, are derived from the same locus of a plant.
- 20 5. The nucleic acid molecule of claim 4, wherein:
 - (a) the first exon is the 5' most exon of the locus; and/or
 - (b) the intron is the 5' most intron; and/or
 - (c) the second exon or the fragment thereof, if present, is the exon immediately downstream of the intron in the genome of the plant.
25
6. The nucleic acid molecule of claim 1, further comprising a transcription terminator operably linked to the plant promoter.
7. The nucleic acid molecule of claim 6, wherein the transcription terminator and the plant promoter are derived from the same locus of a plant.
30

8. The nucleic acid molecule of claim 1, further comprising a heterologous nucleotide sequence operably linked to the plant promoter.
- 5 9. The nucleic acid molecule of claim 8, wherein the nucleic acid molecule comprises:
- 10 (a) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 1 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 4; or
- (b) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 5; or
- 15 (c) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 6.
- 20 10. An nucleic acid molecule comprising operably linked in 5' to 3' order:
- (i) a plant promoter derived from a genetic locus;
- (ii) a heterologous nucleotide sequence; and
- (iii) a transcription terminator,
- wherein the nucleic acid molecule comprises a nucleotide
- 25 sequence that is at least 95% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 1-3.
11. The nucleic acid molecule of claim 10, wherein the intron, the transcription terminator, or both are also derived from the genetic locus.
- 30

12. The nucleic acid molecule of claim 10, wherein the nucleic acid molecule comprises:
- 5 (a) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 1 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 4; or
- 10 (b) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 2 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 5; or
- 15 (c) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 6.
13. The nucleic acid molecule of claim 10, wherein the at least one exon is the first exon of the genetic locus, the intron is the first exon of the genetic locus, or both.
- 20 14. The nucleic acid molecule of claim 10, wherein the heterologous nucleotide sequence encodes an RNA molecule selected from the group consisting of an mRNA for a polypeptide of interest and an inhibitory RNA.
- 25 15. An nucleic acid molecule comprising a plant transcription terminator, wherein the nucleic acid molecule comprises a nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 4-6.

16. The nucleic acid molecule of claim 15, wherein the nucleic acid molecule comprises a nucleotide sequence that is 100% identical over at least 300 consecutive nucleotides to any of SEQ ID NOs: 4-6.
- 5 17. An nucleic acid molecule comprising a plant promoter operably linked to a heterologous nucleotide sequence, wherein the nucleic acid molecule comprises:
- 10 (a) a nucleotide sequence having promoter activity that is at least 95% identical to any of SEQ ID NOs: 1-3 over at least 300 consecutive nucleotides; or
- (b) a nucleotide sequence having promoter activity that comprises a fragment of at least about 100 consecutive nucleotides of any of SEQ ID NOs: 1-3; or
- 15 (c) a nucleotide sequence with a sequence identity of greater than 95% to one of SEQ ID NOs: 1-3 over the full length of the one of SEQ ID NOs: 1-3, wherein the nucleotide sequence has promoter activity; or
- (d) the nucleotide sequence of any of SEQ ID NOs: 1-3.
- 20 18. The nucleic acid molecule of claim 17, wherein the heterologous nucleotide sequence comprises one or more cloning sites.
19. The nucleic acid molecule of claim 18, wherein the heterologous nucleotide sequence comprises a polylinker.
- 25 20. The nucleic acid molecule of claim 17, wherein the heterologous nucleotide sequence encodes a reporter.
21. The nucleic acid molecule of claim 20, wherein the reporter is a β -glucuronidase polypeptide.
- 30

22. The nucleic acid molecule of claim 17, further comprising a transcription terminator operably linked to the plant promoter and the heterologous nucleotide sequence.
- 5
23. The nucleic acid molecule of claim 22, wherein the transcription terminator is derived from the same locus from which the plant promoter is derived.
24. The nucleic acid molecule of claim 23, wherein the plant promoter comprises:
- 10
- (a) SEQ ID NO: 1 and the transcription terminator comprises SEQ ID NO: 4; or
- (b) SEQ ID NO: 2 and the transcription terminator comprises SEQ ID NO: 5; or
- 15
- (c) SEQ ID NO: 3 and the transcription terminator comprises SEQ ID NO: 6.
25. The nucleic acid molecule of claim 17, wherein the transcription terminator is heterologous to the gene from which the promoter is derived.
- 20
26. An expression cassette comprising the nucleic acid molecule of claim 17.
27. The expression cassette of claim 26, wherein the nucleic acid molecule further comprises the plant promoter operably linked to its endogenous first exon, its endogenous first intron, its endogenous second exon or a fragment thereof, its endogenous transcription terminator, or any combination thereof.
- 25
28. A transformed plant containing the nucleic acid molecule of claim 17.
- 30

29. The transformed plant of claim 28, wherein the heterologous nucleotide sequence is oriented to express an antisense RNA molecule.
- 5 30. The transformed plant of claim 28, wherein the heterologous nucleotide sequence is expressed constitutively with the exception of pollen, in which expression of the heterologous nucleotide sequence is low or is absent.
- 10 31. The transformed plant of claim 28, wherein the heterologous nucleotide sequence is expressed constitutively with the exception of pollen.
32. The transformed plant of claim 28, wherein said transformed plant is maize.
- 15 33. A cell containing the nucleic acid molecule of claim 1 operably linked to a heterologous nucleotide sequence.
34. The cell of claim 33, wherein the cell is selected from the group consisting of a bacterial cell, a mammalian cell, an insect cell, a plant cell, and a fungal cell.
- 20 35. The cell of claim 34, wherein said cell is *Agrobacterium tumefaciens*.
36. A method for producing a transformed plant comprising:
25 (a) providing a nucleic acid construct comprising the nucleic acid molecule of claim 1; and
(b) transforming a plant with the nucleic acid construct.
37. A method for producing a biomolecule in a transgenic plant, comprising transforming a plant with a nucleic acid construct comprising the nucleic

acid molecule of claim 10, wherein the heterologous nucleotide sequence encodes the biomolecule and is expressed in the transgenic plant.

38. The method of claim 37, wherein the intron, the transcription terminator, or both are also derived from a single genetic locus.
39. The method of claim 38, wherein the nucleic acid construct comprises:
- (a) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 1 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 4; or
 - (b) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 2 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 5 or
 - (c) a first nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 3 operably linked to a second nucleotide sequence that is at least 95% identical over at least 300 consecutive nucleotides to SEQ ID NO: 6.
40. The method of claim 37, wherein the heterologous nucleotide sequence encodes an RNA molecule selected from the group consisting of an mRNA for a polypeptide of interest and an inhibitory RNA.
41. A method for identifying a non-*Zea mays* nucleic acid molecule with transcriptional regulatory activity, the method comprising:
- (a) providing a first nucleic acid molecule that comprises a nucleotide sequence that is at least 95% identical to at least 50 consecutive bases of any of SEQ ID NOs: 1-3;

- 5 (b) hybridizing the first nucleic acid molecule to a plurality of nucleic acid molecules, wherein the plurality of nucleic acid molecules is from a species other than *Zea mays*;
- (c) isolating a second nucleic acid molecule present in the plurality of nucleic acid molecules that specifically hybridizes to the first nucleic acid molecule; and
- (d) assaying the second nucleic acid molecule for transcriptional regulatory activity,
- 10 whereby a non-*Zea mays* nucleic acid molecule with transcriptional regulatory activity is identified.

42. The method of claim 41, wherein the plurality of nucleic acid molecules comprises a library of genomic DNA from a plant species other than *Zea mays*.

15 43. The method of claim 41, further comprising screening a genomic library from the species other than *Zea mays* and identifying a genomic DNA molecule present in the genomic library that is at least 1000 nucleotides in length and that comprises a nucleotide sequence that is at least 95%

20 identical to the nucleotide sequence of the second nucleic acid molecule over at least 500 nucleotides.

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/058643

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/405 C12N15/82
ADD.

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)
C07K C12N

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

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

EPO-Internal, Sequence Search, WPI Data, BIOSIS, CHEM ABS Data, EMBASE, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] 3 September 1992 (1992-09-03), "Z.mays ZSF4C4 gene for zein", XP002680724, retrieved from EBI accession no. EM PL:X55724 Database accession no. X55724 sequence -& LIU C-N ET AL: "MOLECULAR CHARACTERIZATION OF TWO TYPES OF 22 KILODALTON ALPHA-ZEINGENES IN A GENE CLUSTER IN MAIZE", MOLECULAR AND GENERAL GENETICS, SPRINGER VERLAG, BERLIN, DE, vol. 234, 1 January 1992 (1992-01-01), pages 244-253, XP002048189, ISSN: 0026-8925, DOI: 10.1007/BF00283845 ----- -/-	1-14, 17-43



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search

6 August 2012

Date of mailing of the international search report

24/10/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Marchesini, Patrizia

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2012/058643

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2004/121467 A1 (VOELKER TONI A [US] ET AL) 24 June 2004 (2004-06-24) example 8 -----	1-14, 17-43
Y	WO 2010/147825 A1 (PIONEER HI BRED INT [US]; ABBITT SHANE E [US]) 23 December 2010 (2010-12-23) claims 1-20 -----	1-14, 17-43

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2012/058643

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see additional sheet

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-43(partially)

Relates to an isolated nucleic sequence depicted by SEQ ID No. 1 and 4.

2-3. claims: 1-43(partially)

Relates to an isolated nucleic acid sequence depicted by SEQ ID No. 2 and 5 (inv. 2) or 3 and 6 (inv. 3).

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2012/058643

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2004121467	A1	24-06-2004	NONE

WO 2010147825	A1	23-12-2010	CA 2765034 A1 23-12-2010
			EP 2440663 A1 18-04-2012
			US 2010313301 A1 09-12-2010
			WO 2010147825 A1 23-12-2010
