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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 October 2014 (02.10.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/160122 A1

(51) International Patent Classification:
C07K 14/415 (2006.01) *C12N 15/82* (2006.01)

(21) International Application Number:
PCT/US2014/025862

(22) International Filing Date:
13 March 2014 (13.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/782,509 14 March 2013 (14.03.2013) US

(71) Applicant: **PIONEER HI-BRED INTERNATIONAL, INC.** [US/US]; 7100 N.W. 62nd Avenue, Johnston, Iowa 50131-1014 (US).

(72) Inventors; and

(71) Applicants : **AYELE, Mulu** [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US). **FENG, Dongsheng** [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US). **HUNT, Joanne E.** [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US). **ROESLER, Keith R.** [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US). **SELINGER, David A.** [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US). **SIVASANKAR, Sobhana** [US/US]; c/o Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US).

(74) Agent: **VARLEY, Karen K.**; Pioneer Hi-Bred International, Inc., 7250 N.W. 62nd Avenue, Johnston, Iowa 50131-0552 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, 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, KM, ML, MR, NE, SN, TD, TG).

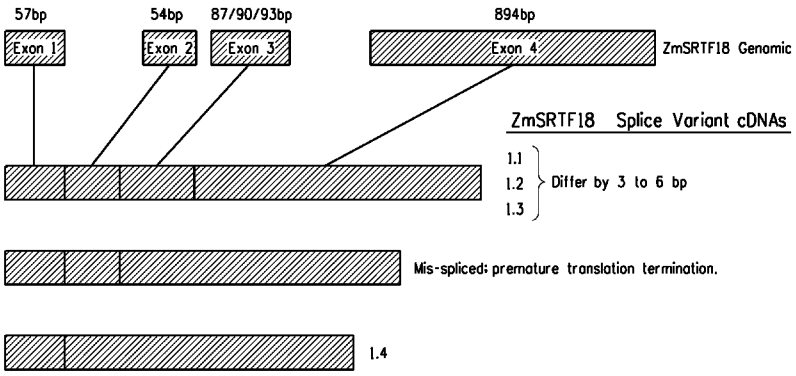
Published:

— *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**: MAIZE STRESS RELATED TRANSCRIPTION FACTOR 18 AND USES THEREOF



Splice Variants of ZmSRTF18

FIG. 2

(57) **Abstract**: A truncated stress-responsive transcription factor can be overexpressed without causing unfavorable pleiotropic effects. The truncation may result in deletion of a nuclear localization signal. The truncation may result in deletion of one or more regulatory motifs. The truncated transcription factor may be expressed under the control of a constitutive or tissue-preferred promoter. The transcription factor may be from maize.

WO 2014/160122 A1

MAIZE STRESS RELATED TRANSCRIPTION FACTOR 18 AND USES THEREOF

BACKGROUND

Plant stress tolerance reflects physiological, biochemical, and/or molecular responses to the environment to increase plant resiliency. Gene regulation in response to stress includes activation of genes encoding metabolic proteins, as well as activation of genes encoding signal proteins which regulate expression of downstream genes. These signal proteins may be transcription factors which interact with elements within promoter regions of target genes, thereby inducing or amplifying expression of those target genes to result in improved stress tolerance. The potential of such a functional cascade implicates transgenic manipulation of transcription factors as an important avenue for improvement of crop stress tolerance and ultimate yield performance.

SUMMARY

Among the embodiments of the disclosure are these:

1. A method of improving abiotic stress tolerance of a plant, the method comprising transforming said plant with a construct comprising a promoter operably linked to a polynucleotide encoding a truncated DREB transcription factor.
2. The method of embodiment 1, wherein the truncated DREB transcription factor lacks a functional N-terminal CBF domain.
3. The method of embodiment 1, wherein the truncation removes at least one nuclear localization signal present in the DREB transcription factor prior to truncation.
4. The method of embodiment 1, wherein the truncated DREB transcription factor lacks a functional N-terminal CBF domain or a functional nuclear localization signal which was present in the DREB transcription factor prior to truncation.
5. The method of embodiment 1, wherein the truncated DREB transcription factor lacks both a functional N-terminal CBF domain and at least one nuclear localization signal present in the DREB transcription factor prior to truncation.
6. The method of embodiment 1, wherein the truncated DREB transcription factor lacks both a function N-terminal CBF domain and all nuclear localization signals present in the DREB transcription factor prior to truncation.

7. The method of embodiment 1 wherein the polynucleotide encodes a polypeptide which is a truncation or variant of ZmSRTF18 (SEQ ID NO: 1) or ZmDREB2A (SEQ ID NO: 8).
8. The method of embodiment 7 wherein the sequence of the encoded polypeptide is SEQ ID NO: 2, 3, 4, 5, 6, 7 or 19.
9. The method of any of embodiments 1-8, wherein the promoter drives constitutive expression.
10. The method of any of embodiments 1-8, wherein the promoter drives tissue-preferred expression.
11. The method of any of embodiments 1-8, wherein the plant produces increased seed yield, relative to a control.
12. The method of any of embodiments 1-8, wherein the plant is maize, wheat, rice or sorghum.
13. The method of embodiment 11, wherein seed yield is increased under conditions of abiotic stress.
14. The method of embodiment 13, wherein abiotic stress includes high salt concentrations.
15. The method of embodiment 13, wherein abiotic stress includes chilling or freezing.
16. The method of embodiment 13, wherein abiotic stress includes reduced water availability at or about the time of anthesis or the time of grain fill.
17. The method of embodiment 13, wherein abiotic stress includes reduced nitrogen availability.
18. A method of improving abiotic stress tolerance of a plant, comprising transforming said plant with a construct comprising a promoter operably linked to a polynucleotide encoding a polypeptide comprising a truncated DREB transcription factor, wherein the truncation retains at least 12 but fewer than 65 amino acids N-terminal to the AP2 domain..
19. The method of embodiment 18 wherein the polynucleotide encodes a polypeptide which is a truncation or variant of ZmSRTF18 (SEQ ID NO: 1) or ZmDREB2A (SEQ ID NO: 8).
20. The method of embodiment 18 or 19, wherein the encoded polypeptide has the polypeptide sequence of SEQ ID NO: 2, 3, 4, 19, 20, 21, 24, 25, 26, or 28.
21. A method of reducing pleiotropy which results from ectopic expression of a DREB transcription factor, comprising expression of a truncated version of said transcription factor.

22. The method of embodiment 21 wherein the truncated version lacks at least one nuclear localization signal which is present in the non-truncated DREB transcription factor.

23. The method of embodiment 21 wherein the truncation deletes the first exon of the polypeptide.

24. The method of embodiment 21 wherein the truncation deletes the first 2 exons of the polypeptide.

25. A recombinant polynucleotide encoding a truncation or variant of a DREB transcription factor from maize, wherein overexpression of said polynucleotide in a maize plant increases grain yield and the maize plant does not pleiotropic effects.

26. A recombinant polynucleotide encoding a truncation or variant of a DREB transcription factor from maize, wherein overexpression of said polynucleotide in a maize plant increases grain yield under conditions of drought or reduced nitrogen availability.

27. A method of improving abiotic stress tolerance of a plant, the method comprising transforming said plant with a construct comprising a promoter operably linked to a polynucleotide encoding a DREB transcription factor lacking a functional N-terminal CBF domain or a functional nuclear localization signal.

28. The method of embodiment 13, wherein the plant is maize, wheat, rice, or sorghum.

29. The method of embodiment 12, wherein seed yield is increased under conditions of abiotic stress.

30. The method of embodiment 12, wherein abiotic stress includes high salt concentrations.

31. The method of embodiment 12, wherein abiotic stress includes chilling or freezing.

32. The method of embodiment 12, wherein abiotic stress includes reduced water availability at or about the time of anthesis or the time of grain fill.

33. The method of embodiment 12, wherein abiotic stress includes reduced nitrogen availability.

34. The method of embodiment 2, wherein the truncation results in loss of one or more nuclear localization signals.

35. A method of improving abiotic stress tolerance of a plant, comprising transforming said plant with a construct comprising a promoter operably linked to a polynucleotide encoding a a truncated DREB transcription factor, wherein the truncation retains at least 12 but fewer than 65 amino acids N-terminal to the AP2 domain.

36. The method of embodiment 35 wherein the polynucleotide encodes a polypeptide which is a truncation or variant of ZmSRTF18 (SEQ ID NO: 1) or ZmDREB2A (SEQ ID NO: 8).

37. The method of embodiment 35, wherein the encoded polypeptide does not comprise an N-terminal CBF domain.

38. A recombinant polynucleotide encoding a truncation or variant of a DREB transcription factor from maize, wherein overexpression of said polynucleotide in a maize plant increases grain yield and the maize plant does not exhibit pleiotropic effects.

39. A recombinant polynucleotide encoding a truncation or variant of a DREB transcription factor from maize, wherein overexpression of said polynucleotide in a maize plant increases grain yield under conditions of drought or reduced nitrogen availability.

40. The method of embodiment 1 wherein the polynucleotide is a homolog of ZmSRTF18.

41. The method of embodiment 40 wherein the polynucleotide is isolated from sorghum or pearl millet.

42. The method of embodiment 41 wherein the polynucleotide is at least 90% identical to SEQ ID NO: 24 or 25.

43. The method of embodiment 41 wherein the sequence of the polynucleotide is SEQ ID NO: 24 or 25.

44. The method of embodiment 40 wherein the polynucleotide lacks at least one nuclear localization signal present in the non-truncated homolog.

45. The method of embodiment 40 wherein the polynucleotide lacks a functional N-terminal CBF domain.

46. The method of claim 1 wherein the DREB transcription factor is a DREB2-type transcription factor.

47. The method of claim 1 wherein the DREB transcription factor lacks one or more of the DREB1/CBF signature sequences, prior to truncation.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is RT-PCR data for native ZmSRTF18 in maize inbred B73. Data indicate no strong induction by drought (panel A), slight induction by cold (panel B); and similar expression levels in leaf (L), midrib (MR), immature ear (IE), stem (S), and silk (SI) tissues, with slightly lower expression in kernel (K) tissues (panel C). Actin expression

data are provided as a control. Induction may be difficult to measure via RT-PCR in that the plant may respond to stress by producing a higher proportion of functional splice variants.

Figure 2 shows the structure of the native ZmSRTF18 gene and of splice variants 1.1, 1.2, 1.3, 1.4, and a prematurely-terminated, mis-spliced variant.

Figure 3 provides an alignment of ZmSRTF18 splice variant 1.1 (aka SPL VAR1; SEQ ID NO: 5), ZmSRTF18 splice variant 1.4 (aka SPL VAR4; SEQ ID NO: 6), ZmDREB2A (SEQ ID NO: 8; Qin et al. (2007) Plant Journal 50(1):54-69), and a radically truncated version of ZmSRTF18 known as ZmSRTF18-del (SEQ ID NO: 2).

Figure 4 provides evidence that treatment with ABA (abscisic acid) increases expression of ZmSRTF18.

Figure 5 (5A, 5B, 5C) provides fluorescence microscopy images demonstrating impact of nuclear localization signals on the subcellular location of proteins

Figure 6 (6A, 6B) provides results of gel-shift assays to evaluate the ability of ZmSRTF18 protein variants to bind DNA. The DNA used for these assays was a region of the maize RAB17 promoter that included the DRE core element. Figure 6A shows results of assay with no DTT added. Figure 6B shows results of assay in presence of 1 mM DTT.

Figure 7 provides western blot data showing detectable ZmSRTF18 protein in leaves of transgenic events, but not in leaves of null controls, resulting from overexpression with a constitutive promoter.

Figure 8 provides alignment of the N-terminal region of several truncations of ZmSRTF18 and two other DREB proteins.

BRIEF DESCRIPTION OF THE SEQUENCES

Table 1. Sequence Description

NAME / S	CHARACTERISTICS	SEQ ID NO:
ZmSRTF18 genomic	Genomic DNA sequence including 5' untranslated region, 3' untranslated region, and introns that can be alternatively spliced to give different amino acid sequences, including SEQ ID NO 5, 6 and 7.	1
ZmSRTF18-del	Radical truncation, shortest; lacks exons 1, 2, and 3, and	2

	part of 4. Lacks two putative NLSs. Retains 12 amino acids N-terminal to the AP2 domain.	
ZmSRTF18-del (ALT2)	Radical truncation but with 16-aa conserved region restored, plus a start methionine. Retains 28 amino acids N-terminal to the AP2 domain, plus a start methionine.	3
ZmSRTF18-del (ALT3)	Radical truncation but with 2 aa (including one cysteine) restored, plus a start methionine. Retains 14 amino acids N-terminal to the AP2 domain, plus a start methionine.	4
ZmSRTF18 (SPL VAR 1) ZmSRTF18 splice variant 1.1	Variants 1.1, 1.2, 1.3 differ by 3 to 6 base pairs (bp).	5
ZmSRTF18 (SPL VAR 4) ZmSRTF18 splice variant 1.4	Exons 1 and 4.	6
Prematurely-terminated, mis-spliced variant	Exons 1, 2, 4 Not functional. Protein does not include AP2 domain.	7
ZmDREB2A	An allele of the ZmSRTF18 gene induced by cold, dehydration, salt, heat.	8
SV40NLS-ZmSRTF18-del	A heterologous nuclear localization signal from a simian virus 40 protein is fused to ZmSRTF18-del.	9
AP2 domain	The highly conserved region of AP2/ERF transcription factors that is involved in binding DNA.	10
Pearl millet DREB2A	DQ227697	11
Barley DRF1.3	AY223807	12
Rice DREB2B		13
Wheat DREB1	DQ195068	14
Arabidopsis DREB2A		15
DNA for gel shift analysis	see Example 3	16
Rab17 promoter		17
ZmRNLS-ZmSRTF18-del	A heterologous nuclear localization signal from the maize R protein, fused to ZmSRTF18-del.	18
ZmSRTF18-del_E216Q	A modification of ZmSRTF18-del to remove an allergen match in an unintended open reading frame.	19
ZmSRTF18-del (ALT2)_E233Q	A modification of ZmSRTF18-del (ALT2) to remove an allergen match in an unintended open reading frame.	20
ZmSRTF18-del (ALT3) _E219Q	A modification of ZmSRTF18-del (ALT3) to remove an allergen match in an unintended open reading frame.	21
ZmRNLS-ZmSRTF18-del _E226Q	A modification of ZmRNLS-ZmSRTF18-del to remove an allergen match in an unintended open reading frame.	22
Sorghum DREB2	JF915841.1	23
Sorghum DREB2A-del	Truncated form.	24
Pearl millet DREB2A-del	Truncated form.	25
ZmSRTF18-del (ALT5)	A truncated form of ZmSRTF18 that retains 52 amino acids N-terminal to the AP2 domain, plus a start methionine.	26
ZmSRTF18-del (ALT6)	A truncated form of ZmSRTF18 that retains 65 amino acids N-terminal to the AP2 domain.	27
ZmSRTF18-del (ALT5)_E257Q	A modification of ZmSRTF18-del (ALT5) to remove an allergen match in an unintended open reading frame.	28

ZM-SRTF18 (SPL VAR2)	Variant 1.2	29
ZM-SRTF18 (SPL VAR3)	Variant 1.3	30

DETAILED DESCRIPTION

Constitutive overexpression of members of the CBF and DREB family of transcription factors for increasing stress tolerance is often associated with negative pleiotropic effects. For example, constitutive overexpression of Arabidopsis DREB1 or DREB2 in Arabidopsis resulted in slow growth (Liu et al, 1998, Plant Cell 10:1391-1406). Constitutive overexpression of tobacco CBF1 in tomato resulted in slow growth (Hsieh et al, 2002, Plant Physiol 129:1086-1094). Constitutive overexpression of rice OsDREB1A in Arabidopsis resulted in slow growth (Dubouzet et al, 2003, Plant J 33:751-763). Constitutive overexpression of wheat TaDREB1 in rice resulted in slow growth (Shen et al, 2003, Theor Appl Genet 106:923-930). Constitutive overexpression of Arabidopsis DREB1A in tobacco resulted in slow growth (Kasuga et al, 2004, Plant Cell Physiol 45:346-350). Constitutive overexpression of maize ZmDREB2A in Arabidopsis resulted in slow growth (Qin et al, 2007, Plant J 50:54-69). Constitutive overexpression of rice OsDREB2B in Arabidopsis resulted in slow growth (Matsukura et al, 2010, Mol Genet Genomics 283:185-196. Stress-inducible promoters have been used to overcome pleiotropy when transgenically expressing transcription factors.

Surprisingly, constitutive overexpression of a truncated maize DREB2 known as ZmSRTF18 has been found to improve maize grain yield in favorable environments, under conditions of reduced nitrogen availability, and under drought conditions, without significant negative pleiotropic effects. Further, this maize DREB2 displays a number of splice variants. Constitutive or targeted expression of selected splice variants or truncations may increase precision in regulating transcription-factor impacts, to provide improved plant performance, especially under conditions of abiotic stress.

Members of the AP2/ERF transcription factor family comprise a highly-conserved AP2 domain of about 58 amino acids (SEQ ID NO: 10). The AP2/ERF family is divided into four subfamilies: AP2; RAV; DREB; and ERF. For review, see Mizoi et al. (2012) Bioch.Biophys.Acta 1819:86-96.

DREB (Dehydration Responsive Element Binding) transcription factors bind to the C-repeat/DRE core element in the promoters of stress-responsive genes: TGGCCGAC or A/GCCGAC, respectively. See, for example, Srivastav et al. (2010) Plant Signaling and Behavior 5(7):775-784. The DREB subfamily is further divided into two subgroups:

DREB1 and DREB2. Typical features of a DREB1 (also known as CBF) protein are an N-terminal nuclear localization signal; an AP2 domain; and a C-terminal activation domain.

DREB1/CBF proteins , contain signature sequences PKKP/RAGRxKFxETRHP (abbreviated PKKPAGR) and DSAWR, which are located upstream and downstream, respectively, of the AP2/ERF DNA-binding domain. Canella *et al*.(2010) Biochim Biophys Acta.1799 (5-6):454-62. These signature sequences are highly conserved in DREB1/CBF proteins from diverse plant species. The two CBF signature sequences together may be referred to as the CBF domain. Unlike DREB1/CBF proteins, DREB2-type proteins do not contain the DREB1/CBF signature sequences. Qin, *et al.* (2007) *Plant J.* 50:54.

ZmSRTF18 has been isolated from maize; it is a member of the DREB2 subgroup and contains 4 exons. Figure 2 shows that predicted ZmSRTF18 splice variants 1.1, 1.2, and 1.3 are highly similar, containing all or part of each of the 4 exons, and encoding functional proteins. Splice variant 1.4 (also known as SPL VAR 4) contains only exons 1 and 4 but is functional.

Another maize splice variant comprising exons 1, 2, and 4 contains a premature translation termination codon; the encoded protein is nonfunctional. Splice variants 1.3, 1.4, and the prematurely-terminated transcript correspond generally to variants identified in barley and wheat.

Splice variants 1.3 and 1.4 have been confirmed by PCR cloning and sequencing. Confirmation of existence of splice variants 1.1 and 1.2 is technically difficult due to their similarity to each other and to splice variant 1.3, and due to the abundance of the misspliced variant.

A truncated variant, known as ZmSRTF18-del, is 71 amino acids shorter than the shortest naturally-occurring functional splice variant. ZmSRTF18-del lacks all amino acids encoded by exons 1, 2, and 3, and lacks 52 amino acids encoded by exon 4. This partial loss of exon 4 results in a loss of two putative nuclear localization signals (NLS). See Figure 3.

Manipulation of expression of these variants and/or of other ZmSRTF18 variants is useful to provide improved plant performance, including seed yield, especially under abiotic stress, such as drought or low-nitrogen conditions.

All references referred to are incorporated herein by reference.

Unless specifically defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Unless mentioned otherwise, the techniques employed or contemplated herein are standard methodologies well known to one of ordinary skill in the art. The materials, methods and examples are illustrative only and not limiting. The following is presented by way of illustration and is not intended to limit the scope of the disclosure.

Many modifications and other embodiments of the disclosures set forth herein will come to mind to one skilled in the art to which these disclosures pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the disclosures are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of botany, microbiology, tissue culture, molecular biology, chemistry, biochemistry and recombinant DNA technology, which are within the skill of the art.

Units, prefixes and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. The terms defined below are more fully defined by reference to the specification as a whole.

In describing the present disclosure, the following terms will be employed and are intended to be defined as indicated below.

By "microbe" is meant any microorganism (including both eukaryotic and prokaryotic microorganisms), such as fungi, yeast, bacteria, actinomycetes, algae and protozoa, as well as other unicellular structures.

By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the

polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-*Beta* Replicase systems, transcription-based amplification system (TAS) and strand displacement amplification (SDA). See, e.g., *Diagnostic Molecular Microbiology*: *Principles and Applications*, Persing, *et al.*, eds., American Society for Microbiology, Washington, DC (1993). The product of amplification is termed an amplicon.

The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refer to those nucleic acids that encode identical or conservatively modified variants of the amino acid sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively modified variation. Every nucleic acid sequence herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; one exception is *Micrococcus rubens*, for which GTG is the methionine codon (Ishizuka, *et al.*, (1993) *J. Gen. Microbiol.* 139:425-32)) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid, which encodes a polypeptide of the present disclosure, is implicit in each described polypeptide sequence and incorporated herein by reference.

As to amino acid sequences, one of skill will recognize that individual substitution, deletion or addition to a nucleic acid, peptide, polypeptide or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" when the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from 1 to 15 can be so altered. Thus, for example, 1, 2, 3, 4, 5, 7 or 10 alterations can be made. Conservatively modified variants typically provide similar biological activity as the unmodified polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80% or 90%, preferably 60-90% of the native protein for its native substrate.

Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);

2) Aspartic acid (D), Glutamic acid (E);

3) Asparagine (N), Glutamine (Q);

4) Arginine (R), Lysine (K);

5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V) and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

See also, Creighton, *Proteins*, W.H. Freeman and Co. (1984).

As used herein, "consisting essentially of" means the inclusion of additional sequences to an object polynucleotide or polypeptide where the additional sequences do not materially affect the basic function of the claimed polynucleotide or polypeptide sequences.

The term "construct" is used to refer generally to an artificial combination of polynucleotide sequences, i.e. a combination which does not occur in nature, normally comprising one or more regulatory elements and one or more coding sequences. The term may include reference to expression cassettes and/or vector sequences, as is appropriate for the context.

A "control" or "control plant" or "control plant cell" provides a reference point for measuring changes in phenotype of a subject plant or plant cell in which genetic alteration, such as transformation, has been effected as to a gene of interest. A subject plant or plant cell may be descended from a plant or cell so altered and will comprise the alteration.

A control plant or plant cell may comprise, for example: (a) a wild-type plant or cell, i.e., of the same genotype as the starting material for the genetic alteration which resulted in the subject plant or cell; (b) a plant or plant cell of the same genotype as the starting material but which has been transformed with a null construct (i.e., with a construct which has no known effect on the trait of interest, such as a construct comprising a marker gene); (c) a plant or plant cell which is a non-transformed segregant among progeny of a subject plant or plant cell; (d) a plant or plant cell genetically identical to the subject plant or plant cell but which is not exposed to conditions or stimuli that would induce expression

of the gene of interest; or (e) the subject plant or plant cell itself, under conditions in which the gene of interest is not expressed. A control plant may also be a plant transformed with an alternative construct.

By "encoding" or "encoded," with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the "universal" genetic code. However, variants of the universal code, such as is present in some plant, animal and fungal mitochondria, the bacterium *Mycoplasma capricolum* (Yamao, *et al*., (1985) *Proc. Natl. Acad. Sci. USA* 82:2306-9) or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed using these organisms.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present disclosure may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledonous plants or dicotyledonous plants as these preferences have been shown to differ (Murray, *et al*., (1989) *Nucleic Acids Res.* 17:477-98 and herein incorporated by reference). Thus, the maize preferred codon for a particular amino acid might be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants is listed in Table 4 of Murray, *et al*., *supra*.

As used herein, the term "endogenous", when used in reference to a gene, means a gene that is normally present in the genome of cells of a specified organism and is present in its normal state in the cells (i.e., present in the genome in the state in which it normally is present in nature).

The term "exogenous" is used herein to refer to any material that is introduced into a cell. The term "exogenous nucleic acid molecule" or "transgene" refers to any nucleic acid molecule that either is not normally present in a cell genome or is introduced into a cell. Such exogenous nucleic acid molecules generally are recombinant nucleic acid molecules, which are generated using recombinant DNA methods as disclosed herein or otherwise known in the art. In various embodiments, a transgenic non-human organism as disclosed herein, can contain, for example, a first transgene and a second transgene. Such first and second transgenes can be introduced into a cell, for example, a progenitor

cell of a transgenic organism, either as individual nucleic acid molecules or as a single unit (e.g., contained in different vectors or contained in a single vector, respectively). In either case, confirmation may be made that a cell from which the transgenic organism is to be derived contains both of the transgenes using routine and well-known methods such as expression of marker genes or nucleic acid hybridization or PCR analysis. Alternatively, or additionally, confirmation of the presence of transgenes may occur later, for example, after regeneration of a plant from a putatively transformed cell.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell which comprises a heterologous nucleic acid sequence of the disclosure, which contains a vector and supports the replication and/or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, plant, amphibian or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledonous plant cells, including but not limited to maize, sorghum, sunflower, soybean, wheat, alfalfa, rice, cotton, canola, barley, millet and tomato. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "hybridization complex" includes reference to a duplex nucleic acid structure formed by two single-stranded nucleic acid sequences selectively hybridized with each other.

The term "introduced" in the context of inserting a nucleic acid into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon or transiently expressed (e.g., transfected mRNA).

The terms "isolated" refers to material, such as a nucleic acid or a protein, which is substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The terms "non-naturally

occurring"; "mutated", "recombinant"; "recombinantly expressed"; "heterologous" or "heterologously expressed" are representative biological materials that are not present in its naturally occurring environment.

By "line" with reference to plants is meant a collection of genetically identical plants.

The term "NUE nucleic acid" means a nucleic acid comprising a polynucleotide ("NUE polynucleotide") encoding a full length or partial length polypeptide.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules, which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, (1987) *Guide To Molecular Cloning Techniques*, from the series *Methods in Enzymology,* vol. 152, Academic Press, Inc., San Diego, CA; Sambrook, *et al.*, (1989) *Molecular Cloning*: *A Laboratory Manual,* 2nd ed., vols. 1-3; and *Current Protocols in Molecular Biology*, Ausubel, *et al.*, eds, Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994 Supplement).

As used herein "operably linked" includes reference to a functional linkage between a first sequence, such as a promoter, and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, a cell present in or isolated from plant tissues including seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen and microspores. The class of plants which can be used in the methods of the disclosure is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants including species from the genera:

Cucurbita, Rosa, Vitis, Juglans, Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciahorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesis, Pelargonium, Panieum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Pisum, Phaseolus, Lolium, Oryza, Avena, Hordeum, Secale, Allium and *Triticum*. A particularly preferred plant is *Zea mays*.

As used herein, "yield" may include reference to bushels per acre of a grain crop at harvest, as adjusted for grain moisture (15% typically for maize, for example) and/or the volume of biomass generated (for forage crops such as alfalfa and plant root size for multiple crops). Grain moisture is measured in the grain at harvest. The adjusted test weight of grain is determined to be the weight in pounds per bushel, adjusted for grain moisture level at harvest. Biomass is measured as the weight of harvestable plant material generated.

As used herein, "polynucleotide" includes reference to a deoxyribopolynucleotide, ribopolynucleotide or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring nucleotides and/or allow translation into the same amino acid(s) as the naturally occurring nucleotide(s). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term may include reference to the specified sequence as well as the complementary sequence thereof.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses and bacteria which comprise genes expressed in plant cells such as *Agrobacterium* or *Rhizobium*. Examples are promoters that preferentially initiate transcription in certain tissues, such as leaves, roots,

seeds, fibres, xylem vessels, tracheids or sclerenchyma. Such promoters are referred to as "tissue preferred." A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" or "regulatable" promoter is a promoter which is under environmental control. Examples of environmental conditions that may 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 development. Tissue preferred, cell type specific, developmentally regulated and inducible promoters are members of the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter which is active in essentially all tissues of a plant, under most environmental conditions and states of development or cell differentiation.

The term "polypeptide" refers to one or more amino acid sequences. The term is also inclusive of fragments, variants, homologs, alleles or precursors (e.g., preproproteins or proproteins) thereof. A "NUE protein" comprises a polypeptide. Unless otherwise stated, the term "NUE nucleic acid" means a nucleic acid comprising a polynucleotide ("NUE polynucleotide") encoding a polypeptide.

As used herein "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention or may have reduced or eliminated expression of a native gene. The term "recombinant" as used herein does not encompass the alteration of the cell or vector by naturally occurring events (e.g., spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements, which permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed and a promoter.

The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 40% sequence identity, preferably 60-90% sequence identity and most preferably 100% sequence identity (i.e., complementary) with each other.

The terms "stringent conditions" or "stringent hybridization conditions" include reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which can be up to 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Optimally, the probe is approximately 500 nucleotides in length, but can vary greatly in length from less than 500 nucleotides to equal to the entire length of the target sequence.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide or Denhardt's. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 0.1X SSC at 60 to 65°C. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, (1984) *Anal. Biochem.*, 138:267-84: $T_m = 81.5°C + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\%\ form) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine

nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $\geq$90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3 or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9 or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15 or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T_m, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T_m of less than 45°C (aqueous solution) or 32°C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes,* part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," Elsevier, New York (1993); and *Current Protocols in Molecular Biology*, chapter 2, Ausubel, *et al*., eds, Greene Publishing and Wiley-Interscience, New York (1995). Unless otherwise stated, in the present application high stringency is defined as hybridization in 4X SSC, 5X Denhardt's (5 g Ficoll, 5 g polyvinypyrrolidone, 5 g bovine serum albumin in 500ml of water), 0.1 mg/ml boiled salmon sperm DNA, and 25 mM Na phosphate at 65°C and a wash in 0.1X SSC, 0.1% SDS at 65°C.

As used herein, "transgenic plant" includes reference to a plant which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses

or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition or spontaneous mutation.

As used herein, "vector" includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

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

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

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

Methods of alignment of nucleotide and amino acid sequences for comparison are well known in the art. The local homology algorithm (BESTFIT) of Smith and Waterman, (1981) *Adv. Appl. Math* 2:482, may conduct optimal alignment of sequences for comparison; by the homology alignment algorithm (GAP) of Needleman and Wunsch, (1970) *J. Mol. Biol.* 48:443-53; by the search for similarity method (Tfasta and Fasta) of Pearson and Lipman, (1988) *Proc. Natl. Acad. Sci. USA* 85:2444; by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California, GAP, BESTFIT, BLAST, FASTA and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG® programs (Accelrys, Inc., San Diego, CA).). The

CLUSTAL program is well described by Higgins and Sharp, (1988) *Gene* 73:237-44; Higgins and Sharp, (1989) *CABIOS* 5:151-3; Corpet, *et al.*, (1988) *Nucleic Acids Res.* 16:10881-90; Huang, *et al.*, (1992) *Computer Applications in the Biosciences* 8:155-65 and Pearson, *et al.*, (1994) *Meth. Mol. Biol.* 24:307-31. The preferred program to use for optimal global alignment of multiple sequences is PileUp (Feng and Doolittle, (1987) *J. Mol. Evol.*, 25:351-60 which is similar to the method described by Higgins and Sharp, (1989) *CABIOS* 5:151-53 and hereby incorporated by reference). The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, *Current Protocols in Molecular Biology*, Chapter 19, Ausubel *et al.*, eds., Greene Publishing and Wiley-Interscience, New York (1995).

GAP uses the algorithm of Needleman and Wunsch, *supra*, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package are 8 and 2, respectively. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 100. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold.

The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (see, Henikoff and Henikoff, (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters (Altschul, *et al.*, (1997) *Nucleic Acids Res.* 25:3389-402).

As those of ordinary skill in the art will understand, BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences, which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, (1993) *Comput. Chem.* 17:149-63) and XNU (Claverie and States, (1993) *Comput. Chem.* 17:191-201) low-complexity filters can be employed alone or in combination.

As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences which differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, (1988) *Computer Applic. Biol. Sci*. 4:11-17, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

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

The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has between 50-100% sequence identity, optionally at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of between 55-100%, such as at least 55%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, up to 100% identity.

The terms "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with between 55-100% sequence identity to a reference sequence, such as at least 55%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, up to 100% sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch, *supra*. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. In addition, a peptide can be substantially identical to a second peptide when they differ by a non-conservative change if the epitope that the antibody recognizes is substantially identical. Peptides which are "substantially similar" share sequences as noted above, except that residue positions, which are not identical, may differ by conservative amino acid changes.

Construction of Nucleic Acids

The isolated nucleic acids of the present disclosure can be made using (a) standard recombinant methods, (b) synthetic techniques or combinations thereof. In some embodiments, the polynucleotides of the present disclosure will be cloned, amplified or otherwise constructed from a fungus or bacteria.

UTRs and Codon Preference

In general, translational efficiency has been found to be regulated by specific sequence elements in the 5' non-coding or untranslated region (5' UTR) of the RNA. Positive sequence motifs include translational initiation consensus sequences (Kozak, (1987) *Nucleic Acids Res.*15:8125) and the 5<G> 7 methyl GpppG RNA cap structure (Drummond, *et al.*, (1985) *Nucleic Acids Res.* 13:7375). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing, *et al.*, (1987) *Cell* 48:691) and AUG sequences or short open reading frames preceded by an appropriate AUG in the 5' UTR (Kozak, *supra,* Rao, *et al.*, (1988) *Mol. and Cell. Biol*. 8:284). Accordingly, the present disclosure provides 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences.

Further, the polypeptide-encoding segments of the polynucleotides of the present disclosure can be modified to alter codon usage. Altered codon usage can be employed to alter translational efficiency and/or to optimize the coding sequence for expression in a desired host or to optimize the codon usage in a heterologous sequence for expression in maize. Codon usage in the coding regions of the polynucleotides of the present disclosure can be analyzed statistically using commercially available software packages such as "Codon Preference" available from the University of Wisconsin Genetics Computer Group. See, Devereaux, *et al.*, (1984) *Nucleic Acids Res.* 12:387-395) or MacVector 4.1 (Eastman Kodak Co., New Haven, CN). Thus, the present disclosure provides a codon usage frequency characteristic of the coding region of at least one of the polynucleotides of the present disclosure. The number of polynucleotides (3 nucleotides per amino acid) that can be used to determine a codon usage frequency can be any integer from 3 to the number of polynucleotides of the present disclosure as provided herein. Optionally, the polynucleotides will be full-length sequences. An exemplary number of sequences for statistical analysis can be at least 1, 5, 10, 20, 50 or 100.

Sequence Shuffling

The present disclosure provides methods for sequence shuffling using polynucleotides of the present disclosure, and compositions resulting therefrom. Sequence shuffling is described in PCT Publication Number 1996/19256. See also, Zhang, *et al.*, (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-9 and Zhao, *et al.*, (1998) *Nature Biotech* 16:258-61. Generally, sequence shuffling provides a means for generating libraries of polynucleotides having a desired characteristic, which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides, which comprise sequence regions, which have substantial sequence identity and can be homologously recombined in vitro or in vivo. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation or other expression property of a gene or transgene, a replicative element, a protein-binding element or the like, such as any feature which confers a selectable or detectable property. In some embodiments, the selected characteristic will be an altered K_m and/or K_{cat} over the wild-type protein as provided herein. In other embodiments, a protein or polynucleotide generated from sequence shuffling will have a ligand binding affinity greater than the non-shuffled wild-type polynucleotide. In yet other embodiments, a protein or polynucleotide generated from sequence shuffling will have an altered pH optimum as compared to the non-shuffled wild-type polynucleotide. The increase in such properties can be at least 110%, 120%, 130%, 140% or greater than 150% of the wild-type value.

Recombinant Expression Cassettes

The present disclosure further provides recombinant expression cassettes comprising a nucleic acid of the present disclosure. A nucleic acid sequence coding for the desired polynucleotide of the present disclosure, for example a cDNA or a genomic sequence encoding a polypeptide long enough to code for an active protein of the present disclosure, can be used to construct a recombinant expression cassette which can be introduced into the desired host cell. A recombinant expression cassette will typically comprise a polynucleotide of the present disclosure operably linked to transcriptional

initiation regulatory sequences which will direct the transcription of the polynucleotide in the intended host cell, such as tissues of a transformed plant.

For example, plant expression vectors may include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site and/or a polyadenylation signal.

Promoters, Terminators, Introns

A plant promoter fragment can be employed which will direct expression of a polynucleotide of the present disclosure in essentially all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (US Patent Number 5,683,439), the *Nos* promoter, the rubisco promoter, the GRP1-8 promoter, the 35S promoter from cauliflower mosaic virus (CaMV), as described in Odell, *et al*., (1985) *Nature* 313:810-2; rice actin (McElroy, *et al*., (1990) *Plant Cell* 163-171); ubiquitin (Christensen, *et al*., (1992) *Plant Mol. Biol.* 12:619-632 and Christensen, *et al*., (1992) *Plant Mol. Biol.* 18:675-89); pEMU (Last, *et al*., (1991) *Theor. Appl. Genet.* 81:581-8); MAS (Velten, *et al*., (1984) *EMBO J.* 3:2723-30) and maize H3 histone (Lepetit, *et al*., (1992) *Mol. Gen. Genet.* 231:276-85 and Atanassvoa, *et al*., (1992) *Plant Journal* 2(3):291-300); ALS promoter, as described in PCT Application Number WO 1996/30530 and other transcription initiation regions from various plant genes known to those of skill. For the present disclosure ubiquitin is the preferred promoter for expression in monocot plants.

Alternatively, the plant promoter can direct expression of a polynucleotide of the present disclosure in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters may be "inducible" promoters. Environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions or the presence of light. Examples of inducible promoters are the Adh1 promoter, which is inducible by hypoxia or cold stress, the Hsp70

promoter, which is inducible by heat stress and the PPDK promoter, which is inducible by light. Diurnal promoters that are active at different times during the circadian rhythm are also known (US Patent Application Publication Number 2011/0167517, incorporated herein by reference).

Examples of promoters under developmental control include promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds or flowers. The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may become fully or partially constitutive in certain locations.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from a variety of plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes or alternatively from another plant gene or less preferably from any other eukaryotic gene. Examples of such regulatory elements include, but are not limited to, 3' termination and/or polyadenylation regions such as those of the *Agrobacterium tumefaciens* nopaline synthase (nos) gene (Bevan, *et al.*, (1983) *Nucleic Acids Res.* 12:369-85); the potato proteinase inhibitor II (PINII) gene (Keil, *et al.*, (1986) *Nucleic Acids Res.* 14:5641-50 and An, *et al.*, (1989) *Plant Cell* 1:115-22) and the CaMV 19S gene (Mogen, *et al.*, (1990) *Plant Cell* 2:1261-72).

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg, (1988) *Mol. Cell Biol.* 8:4395-4405; Callis, *et al.*, (1987) *Genes Dev*. 1:1183-200). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of maize introns Adh1-S intron 1, 2 and 6, the Bronze-1 intron are known in the art. See generally, *The Maize Handbook*, Chapter 116, Freeling and Walbot, eds., Springer, New York (1994).

Signal Peptide Sequences

Plant signal sequences, including, but not limited to, signal-peptide encoding DNA/RNA sequences which target proteins to the extracellular matrix of the plant cell

(Dratewka-Kos, *et al.*, (1989) *J. Biol. Chem.* 264:4896-900), such as the *Nicotiana plumbaginifolia* extension gene (DeLoose, *et al.*, (1991) *Gene* 99:95-100); signal peptides which target proteins to the vacuole, such as the sweet potato sporamin gene (Matsuka, *et al.*, (1991) *Proc. Natl. Acad. Sci. USA* 88:834) and the barley lectin gene (Wilkins, *et al.*, (1990) *Plant Cell*, 2:301-13); signal peptides which cause proteins to be secreted, such as that of PRIb (Lind, *et al.*, (1992) *Plant Mol. Biol.* 18:47-53) or the barley alpha amylase (BAA) (Rahmatullah, *et al.*, (1989) *Plant Mol. Biol.* 12:119) or signal peptides which target proteins to the plastids such as that of rapeseed enoyl-Acp reductase (Verwaert, *et al.*, (1994) *Plant Mol. Biol.* 26:189-202) are useful in the disclosure.

Markers

The vector comprising the sequences from a polynucleotide of the present disclosure will typically comprise a marker gene, which confers a selectable phenotype on plant cells. The selectable marker gene may encode antibiotic resistance, with suitable genes including genes coding for resistance to the antibiotic spectinomycin (e.g., the aada gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance. Also useful are genes coding for resistance to herbicides which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Constructs described herein may comprise a polynucleotide of interest encoding a reporter or marker product. Examples of suitable reporter polynucleotides known in the art can be found in, for example, Jefferson *et al.* (1991) in Plant Molecular Biology Manual, ed. Gelvin *et al.* (Kluwer Academic Publishers), pp. 1-33; DeWet *et al.* Mol. Cell. Biol. 7:725-737 (1987); Goff *et al.* EMBO J. 9:2517-2522 (1990); Kain *et al.* BioTechniques 19:650-655 (1995); and Chiu *et al.* Current Biology 6:325-330 (1996). In certain embodiments, the polynucleotide of interest encodes a selectable reporter. These can include polynucleotides that confer antibiotic resistance or resistance to herbicides.

Examples of suitable selectable marker polynucleotides include, but are not limited to, genes encoding resistance to chloramphenicol, methotrexate, hygromycin, streptomycin, spectinomycin, bleomycin, sulfonamide, bromoxynil, glyphosate, and phosphinothricin.

In some embodiments, the expression cassettes disclosed herein comprise a polynucleotide of interest encoding scorable or screenable markers, where presence of the polynucleotide produces a measurable product. Examples include a β-glucuronidase, or uidA gene (GUS), which encodes an enzyme for which various chromogenic substrates are known (for example, U.S. Pat. Nos. 5,268,463 and 5,599,670); chloramphenicol acetyl transferase, and alkaline phosphatase. Other screenable markers include the anthocyanin/flavonoid polynucleotides including, for example, a R-locus polynucleotide, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues, the genes which control biosynthesis of flavonoid pigments, such as the maize C1 and C2 , the B gene, the p1 gene, and the bronze locus genes, among others. Further examples of suitable markers encoded by polynucleotides of interest include the cyan fluorescent protein (CYP) gene, the yellow fluorescent protein gene, a lux gene, which encodes a luciferase, the presence of which may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry, a green fluorescent protein (GFP), and DsRed2 (*Clontechniques*, 2001) where plant cells transformed with the marker gene are red in color, and thus visually selectable. Additional examples include a p-lactamase gene encoding an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin), a xylE gene encoding a catechol dioxygenase that can convert chromogenic catechols, an α-amylase gene, and a tyrosinase gene encoding an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone, which in turn condenses to form the easily detectable compound melanin.

The expression cassette can also comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). Additional selectable markers include phenotypic markers such as β-galactosidase and fluorescent proteins such as green fluorescent protein (GFP) (Su *et al.* (2004) *Biotechnol Bioeng 85:*610-9 and Fetter *et al.* (2004) *Plant Cell 16:*215-28), cyan florescent protein (CYP) (Bolte *et al.* (2004) *J. Cell Science 117*:943-54 and

Kato *et al.* (2002) *Plant Physiol 129*:913-42), and yellow florescent protein (PhiYFP™ from Evrogen, see, Bolte *et al.* (2004) *J. Cell Science 117*:943-54). For additional selectable markers, see generally, Yarranton (1992) *Curr. Opin. Biotech.* 3:506-511; Christopherson *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-6318; Yao *et al.* (1992) *Cell* 71:63-72; Reznikoff (1992) *Mol. Microbiol.* 6:2419-2422; Barkley *et al.* (1980) in *The Operon*, pp. 177-220; Hu *et al.* (1987) *Cell* 48:555-566; Brown *et al.* (1987) *Cell* 49:603-612; Figge *et al.* (1988) *Cell* 52:713-722; Deuschle *et al.* (1989) *Proc. Natl. Acad. Aci. USA* 86:5400-5404; Fuerst *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:2549-2553; Deuschle *et al.* (1990) *Science* 248:480-483; Gossen (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-1921; Labow *et al.* (1990) *Mol. Cell. Biol.* 10:3343-3356; Zambretti *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-3956; Baim *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-5076; Wyborski *et al.* (1991) *Nucleic Acids Res.* 19:4647-4653; Hillenand-Wissman (1989) *Topics Mol. Struc. Biol.* 10:143-162; Degenkolb *et al.* (1991) *Antimicrob. Agents Chemother.* 35:1591-1595; Kleinschnidt *et al.* (1988) *Biochemistry* 27:1094-1104; Bonin (1993) Ph.D. Thesis, University of Heidelberg; Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Oliva *et al.* (1992) *Antimicrob. Agents Chemother.* 36:913-919; Hlavka *et al.* (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.* (1988) *Nature* 334:721-724. Such disclosures are herein incorporated by reference. The above list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be used in the compositions and methods disclosed herein.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers, *et al.*, (1987) *Meth. Enzymol.* 153:253-77. These vectors are plant integrating vectors in that on transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl, *et al.*, (1987) *Gene* 61:1-11 and Berger, *et al.*, (1989) *Proc. Natl. Acad. Sci. USA*, 86:8402-6. Another useful vector herein is plasmid pBI101.2 that is available from CLONTECH Laboratories, Inc. (Palo Alto, CA).

Expression of Proteins in Host Cells

Using the nucleic acids of the present disclosure, one may express a protein of the present disclosure in a recombinantly engineered cell such as bacteria, yeast, insect, mammalian or preferably plant cells. The cells produce the protein in a non-natural condition (e.g., in quantity, composition, location and/or time), because they have been genetically altered through human intervention to do so.

It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of a nucleic acid encoding a protein of the present disclosure. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of isolated nucleic acids encoding a protein of the present disclosure will typically be achieved by operably linking, for example, the DNA or cDNA to a promoter, followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences and promoters useful for regulation of the expression of the DNA of the present disclosure. To obtain high level expression of a cloned gene, it is desirable to construct expression vectors which contain, at the minimum, a strong promoter, such as ubiquitin, to direct transcription, a ribosome binding site for translational initiation and a transcription/translation terminator. Constitutive promoters are classified as providing for a range of constitutive expression. Thus, some are weak constitutive promoters and others are strong constitutive promoters. Generally, by “weak promoter” is intended a promoter that drives expression of a coding sequence at a low level. By “low level” is intended at levels of about 1/10,000 transcripts to about 1/100,000 transcripts to about 1/500,000 transcripts. Conversely, a “strong promoter” drives expression of a coding sequence at a “high level,” or about 1/10 transcripts to about 1/100 transcripts to about 1/1,000 transcripts.

One of skill would recognize that modifications could be made to a protein of the present disclosure without diminishing its biological activity. Some modifications may be made to facilitate the cloning, expression or incorporation of the targeting molecule into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site or additional amino acids (e.g., poly His) placed on either terminus to create conveniently located restriction sites or termination codons or purification sequences.

Expression in Prokaryotes

Prokaryotic cells may be used as hosts for expression. Prokaryotes most frequently are represented by various strains of *E. coli*; however, other microbial strains may also be used. Commonly used prokaryotic control sequences which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta lactamase (penicillinase) and lactose (lac) promoter systems (Chang, *et al.*, (1977) *Nature* 198:1056), the tryptophan (trp) promoter system (Goeddel, *et al.*, (1980) *Nucleic Acids Res.* 8:4057) and the lambda derived P L promoter and N-gene ribosome binding site (Shimatake, *et al.*, (1981) *Nature* 292:128). The inclusion of selection markers in DNA vectors transfected in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline or chloramphenicol.

The vector is selected to allow introduction of the gene of interest into the appropriate host cell. Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA. Expression systems for expressing a protein of the present disclosure are available using *Bacillus sp.* and *Salmonella* (Palva, *et al.*, (1983) *Gene* 22:229-35; Mosbach, *et al.*, (1983) *Nature* 302:543-5). The pGEX-4T-1 plasmid vector from Pharmacia is the preferred *E. coli* expression vector for the present disclosure.

Expression in Eukaryotes

A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, the present disclosure can be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as expression systems for production of the proteins of the instant disclosure.

Synthesis of heterologous proteins in yeast is well known. Sherman, *et al.*, (1982) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory is a well recognized work describing the various methods available to produce the protein in yeast. Two widely utilized yeasts for production of eukaryotic proteins are *Saccharomyces cerevisiae* and *Pichia pastoris*. Vectors, strains and protocols for expression in *Saccharomyces* and *Pichia* are known in the art and available from commercial suppliers (e.g., Invitrogen). Suitable vectors usually have expression control sequences, such as promoters, including

3-phosphoglycerate kinase or alcohol oxidase and an origin of replication, termination sequences and the like as desired.

A protein of the present disclosure, once expressed, can be isolated from yeast by lysing the cells and applying standard protein isolation techniques to the lysates or the pellets. The monitoring of the purification process can be accomplished by using Western blot techniques or radioimmunoassay of other standard immunoassay techniques.

The sequences encoding proteins of the present disclosure can also be ligated to various expression vectors for use in transfecting cell cultures of, for instance, mammalian, insect or plant origin. Mammalian cell systems often will be in the form of monolayers of cells although mammalian cell suspensions may also be used. A number of suitable host cell lines capable of expressing intact proteins have been developed in the art, and include the HEK293, BHK21 and CHO cell lines. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter (e.g., the CMV promoter, a HSV *tk* promoter or *pgk* (phosphoglycerate kinase) promoter), an enhancer (Queen, *et al.*, (1986) *Immunol. Rev.* 89:49) and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites (e.g., an SV40 large T Ag poly A addition site) and transcriptional terminator sequences. Other animal cells useful for production of proteins of the present disclosure are available, for instance, from the American Type Culture Collection Catalogue of Cell Lines and Hybridomas (7^{th} ed., 1992).

Appropriate vectors for expressing proteins of the present disclosure in insect cells are usually derived from the SF9 baculovirus. Suitable insect cell lines include mosquito larvae, silkworm, armyworm, moth and *Drosophila* cell lines such as a Schneider cell line (see, e.g., Schneider, (1987) *J. Embryol. Exp. Morphol.* 27:353-65).

As with yeast, when higher animal or plant host cells are employed, polyadenlyation or transcription terminator sequences are typically incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague, *et al.*, (1983) *J. Virol.* 45:773-81). Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors (Saveria-Campo, "Bovine Papilloma Virus DNA a Eukaryotic Cloning Vector," in *DNA Cloning*: *A Practical Approach,* vol. II, Glover, ed., IRL Press, Arlington, VA, pp. 213-38 (1985)).

In addition, the gene of interest placed in the appropriate plant expression vector can be used to transform plant cells. The polypeptide can then be isolated from plant callus or the transformed cells can be used to regenerate transgenic plants. Such transgenic plants can be harvested, and the appropriate tissues (seed or leaves, for example) can be subjected to large scale protein extraction and purification techniques.

Plant Transformation Methods

Numerous methods for introducing heterologous genes into plants are known and can be used to insert a polynucleotide into a plant host, including biological and physical plant transformation protocols. See, e.g., Miki *et al.*, "Procedure for Introducing Foreign DNA into Plants," in *Methods in Plant Molecular Biology and Biotechnology*, Glick and Thompson, eds., CRC Press, Inc., Boca Raton, pp. 67-88 (1993). The methods chosen vary with the host plant and include chemical transfection methods such as calcium phosphate, microorganism-mediated gene transfer such as *Agrobacterium* (Horsch, *et al.*, (1985) *Science* 227:1229-31), electroporation, micro-injection and biolistic bombardment.

Expression cassettes and vectors and *in vitro* culture methods for plant cell or tissue transformation and regeneration of plants are known and available. See, e.g., Gruber, *et al.*, "Vectors for Plant Transformation," in *Methods in Plant Molecular Biology and Biotechnology*, *supra*, pp. 89-119.

The isolated polynucleotides or polypeptides may be introduced into the plant by one or more techniques typically used for direct delivery into cells. Such protocols may vary depending on the type of organism, cell, plant or plant cell, i.e., monocot or dicot, targeted for gene modification. Suitable methods of transforming plant cells include microinjection (Crossway, *et al.*, (1986) *Biotechniques* 4:320-334 and US Patent Number 6,300,543), electroporation (Riggs, *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, direct gene transfer (Paszkowski *et al.*, (1984) *EMBO J.* 3:2717-2722) and ballistic particle acceleration (see, for example, Sanford, *et al.*, US Patent Number 4,945,050; WO 1991/10725 and McCabe, *et al.*, (1988) *Biotechnology* 6:923-926). Also see, Tomes, *et al.*, "Direct DNA Transfer into Intact Plant Cells Via Microprojectile Bombardment". pp. 197-213 in *Plant Cell, Tissue and Organ Culture, Fundamental Methods*. eds. Gamborg and Phillips. Springer-Verlag Berlin Heidelberg New York, 1995; US Patent Number 5,736,369 (meristem); Weissinger, *et al.*, (1988) *Ann. Rev. Genet.* 22:421-477; Sanford, *et al.*, (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou, *et al.*, (1988) *Plant Physiol.* 87:671-674 (soybean); Datta, *et al.*, (1990) *Biotechnology* 8:736-740 (rice); Klein, *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein, *et al.*, (1988)

Biotechnology 6:559-563 (maize); WO 91/10725 (maize); Klein, *et al.*, (1988) *Plant Physiol.* 91:440-444 (maize); Fromm, *et al.*, (1990) *Biotechnology* 8:833-839 and Gordon-Kamm, *et al.*, (1990) *Plant Cell* 2:603-618 (maize); Hooydaas-Van Slogteren and Hooykaas, (1984) *Nature* (London) 311:763-764; Bytebierm, *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet, *et al.*, (1985) *In The Experimental Manipulation of Ovule Tissues*, ed. G.P. Chapman, *et al.*, pp. 197-209. Longman, NY (pollen); Kaeppler, *et al.*, (1990) *Plant Cell Reports* 9:415-418 and Kaeppler, *et al.*, (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); US Patent Number 5,693,512 (sonication); D'Halluin, *et al.*, (1992) *Plant Cell* 4:1495-1505 (electroporation); Li, *et al.*, (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford, (1995) *Annals of Botany* 75:407-413 (rice); Osjoda, *et al.*, (1996) *Nature Biotech.* 14:745-750; Agrobacterium mediated maize transformation (US Patent Number 5,981,840); silicon carbide whisker methods (Frame, *et al.*, (1994) *Plant J.* 6:941-948); laser methods (Guo, *et al.*, (1995) *Physiologia Plantarum* 93:19-24); sonication methods (Bao, *et al.*, (1997) *Ultrasound in Medicine & Biology* 23:953-959; Finer and Finer, (2000) *Lett Appl Microbiol.* 30:406-10; Amoah, *et al.*, (2001) *J Exp Bot* 52:1135-42); polyethylene glycol methods (Krens, *et al.*, (1982) *Nature* 296:72-77); protoplasts of monocot and dicot cells can be transformed using electroporation (Fromm, *et al.*, (1985) *Proc. Natl. Acad. Sci. USA* 82:5824-5828) and microinjection (Crossway, *et al.*, (1986) *Mol. Gen. Genet.* 202:179-185), all of which are herein incorporated by reference.

Agrobacterium-mediated Transformation

The most widely utilized method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium*. *A. tumefaciens* and *A. rhizogenes* are plant pathogenic soil bacteria which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of plants. See, e.g., Kado, (1991) *Crit. Rev. Plant Sci.* 10:1. Descriptions of the *Agrobacterium* vector systems and methods for *Agrobacterium*-mediated gene transfer are provided in Gruber, *et al.*, *supra*; Miki, *et al.*, *supra* and Moloney, *et al.*, (1989) *Plant Cell Reports* 8:238.

Similarly, the gene can be inserted into the T-DNA region of a Ti or Ri plasmid derived from *A. tumefaciens* or *A. rhizogenes*, respectively. Thus, expression cassettes can be constructed as above, using these plasmids. Many control sequences are known which when coupled to a heterologous coding sequence and transformed into a host organism show fidelity in gene expression with respect to tissue/organ specificity of the

original coding sequence. See, e.g., Benfey and Chua, (1989) *Science* 244:174-81. Particularly suitable control sequences for use in these plasmids are promoters for constitutive or tissue-preferred expression of the gene in the various target plants. Other useful control sequences include a promoter and terminator from the nopaline synthase gene (NOS). The NOS promoter and terminator are present in the plasmid pARC2, available from the American Type Culture Collection and designated ATCC 67238. If such a system is used, the virulence (*vir*) gene from either the Ti or Ri plasmid must also be present, either along with the T-DNA portion, or via a binary system where the *vir* gene is present on a separate vector. Such systems, vectors for use therein, and methods of transforming plant cells are described in US Patent Number 4,658,082; US Patent Application Serial Number 913,914, filed October 1, 1986, as referenced in US Patent Number 5,262,306, issued November 16, 1993 and Simpson, *et al*., (1986) *Plant Mol. Biol.* 6:403-15 (also referenced in the '306 patent), all incorporated by reference in their entirety.

Once constructed, these plasmids can be placed into *A. rhizogenes* or *A. tumefaciens* and these vectors used to transform cells of plant species which are ordinarily susceptible to *Fusarium* or *Alternaria* infection. Several other transgenic plants are also contemplated by the present disclosure including but not limited to soybean, corn, sorghum, alfalfa, rice, clover, cabbage, banana, coffee, celery, tobacco, cowpea, cotton, melon and pepper. The selection of either *A. tumefaciens* or *A. rhizogenes* will depend on the plant being transformed thereby. In general *A. tumefaciens* is the preferred organism for transformation. Most dicotyledonous plants, some gymnosperms and a few monocotyledonous plants (e.g., certain members of the *Liliales* and *Arales*) are susceptible to infection with *A. tumefaciens*. *A. rhizogenes* also has a wide host range, embracing most dicots and some gymnosperms, which includes members of the *Leguminosae*, *Compositae,* and *Chenopodiaceae*. Monocot plants can also be transformed. EP Patent Application Number 604 662 A1 discloses a method for transforming monocots using *Agrobacterium*. EP Patent Application Number 672 752 A1 discloses a method for transforming monocots with *Agrobacterium* using the scutellum of immature embryos. Ishida, *et al.*, discuss a method for transforming maize by exposing immature embryos to *A. tumefaciens* (*Nature Biotechnology* 14:745-50 (1996)).

Once transformed, these cells can be used to regenerate transgenic plants. For example, whole plants can be infected with these vectors by wounding the plant and then introducing the vector into the wound site. Any part of the plant can be wounded, including leaves, stems and roots. Alternatively, plant tissue in the form of an explant,

such as cotyledonary tissue or leaf disks, can be inoculated with these vectors, and cultured under conditions which promote plant regeneration. Examples of such methods for regenerating plant tissue are disclosed in Shahin, (1985) *Theor. Appl. Genet.* 69:235-40; US Patent Number 4,658,082; Simpson, *et al.*, *supra* and US Patent Application Serial Numbers 913,913 and 913,914, both filed October 1, 1986, as referenced in US Patent Number 5,262,306, issued November 16, 1993, the entire disclosures therein incorporated herein by reference.

Direct Gene Transfer

Despite the fact that the host range for *Agrobacterium*-mediated transformation is broad, some major cereal crop species and gymnosperms have generally been recalcitrant to this mode of gene transfer, even though some success has recently been achieved in rice (Hiei, *et al.*, (1994) *The Plant Journal* 6:271-82). Several methods of plant transformation, collectively referred to as direct gene transfer, have been developed as an alternative to *Agrobacterium*-mediated transformation.

A generally applicable method of plant transformation is microprojectile-mediated transformation, where DNA is carried on the surface of microprojectiles measuring about 1 to 4 μm. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds of 300 to 600 m/s which is sufficient to penetrate the plant cell walls and membranes (Sanford, *et al.*, (1987) *Part. Sci. Technol.* 5:27; Sanford, (1988) *Trends Biotech* 6:299; Sanford, (1990) *Physiol. Plant* 79:206 and Klein, *et al.*, (1992) *Biotechnology* 10:268).

Another method for physical delivery of DNA to plants is sonication of target cells as described in Zang, *et al.*, (1991) *BioTechnology* 9:996. Alternatively, liposome or spheroplast fusions have been used to introduce expression vectors into plants. See, e.g., Deshayes, *et al.*, (1985) *EMBO J.* 4:2731 and Christou, *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:3962. Direct uptake of DNA into protoplasts using $CaCl_2$ precipitation, polyvinyl alcohol, or poly-L-ornithine has also been reported. See, e.g., Hain, *et al.*, (1985) *Mol. Gen. Genet.* 199:161 and Draper, *et al.*, (1982) *Plant Cell Physiol.* 23:451.

Electroporation of protoplasts and whole cells and tissues has also been described. See, e.g., Donn, *et al.*, (1990) *Abstracts of the VIIth Int'l. Congress on Plant Cell and Tissue Culture IAPTC*, A2-38, p. 53; D'Halluin, *et al.*, (1992) *Plant Cell* 4:1495-505 and Spencer, *et al.*, (1994) *Plant Mol. Biol.* 24:51-61.

Reducing the Activity and/or Level of a Polypeptide

Methods are provided to reduce or eliminate the activity of a polypeptide of the disclosure by transforming a plant cell with an expression cassette that expresses a polynucleotide that inhibits the expression of the polypeptide. The polynucleotide may inhibit the expression of the polypeptide directly, by preventing transcription or translation of the messenger RNA, or indirectly, by encoding a polypeptide that inhibits the transcription or translation of a gene encoding polypeptide. Methods for inhibiting or eliminating the expression of a gene in a plant are well known in the art and any such method may be used in the present disclosure to inhibit the expression of polypeptide.

In accordance with the present disclosure, the expression of a polypeptide may be inhibited so that the protein level of the polypeptide is, for example, less than 70% of the protein level of the same polypeptide in a plant that has not been genetically modified or mutagenized to inhibit the expression of that polypeptide. In particular embodiments of the disclosure, the protein level of the polypeptide in a modified plant according to the disclosure is less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5% or less than 2% of the protein level of the same polypeptide in a plant that is not a mutant or that has not been genetically modified to inhibit the expression of that polypeptide. The expression level of the polypeptide may be measured directly, for example, by assaying for the level of polypeptide expressed in the plant cell or plant, or indirectly, for example, by measuring the nitrogen uptake activity of the polypeptide in the plant cell or plant or by measuring the phenotypic changes in the plant. Methods for performing such assays are described elsewhere herein.

In other embodiments of the disclosure, the activity of the polypeptide is reduced or eliminated by transforming a plant cell with an expression cassette comprising a polynucleotide encoding a polypeptide that inhibits the activity of a polypeptide. The activity of a polypeptide is inhibited according to the present disclosure if the activity of the polypeptide is, for example, less than 70% of the activity of the same polypeptide in a plant that has not been modified to inhibit the activity of that polypeptide. In particular embodiments of the disclosure, the activity of the polypeptide in a modified plant according to the disclosure is less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10% or less than 5% of the activity of the same polypeptide in a plant that that has not been modified to inhibit the expression of that polypeptide. The activity of a polypeptide is "eliminated" according to the disclosure when it is not detectable by the assay methods described elsewhere herein. Methods of determining the alteration of activity of a polypeptide are described elsewhere herein.

In other embodiments, the activity of a polypeptide may be reduced or eliminated by disrupting the gene encoding the polypeptide. The disclosure encompasses mutagenized plants that carry mutations in genes, where the mutations reduce expression of the gene or inhibit the activity of the encoded polypeptide.

Thus, many methods may be used to reduce or eliminate the activity of a polypeptide. In addition, more than one method may be used to reduce the activity of a single polypeptide.

1. Polynucleotide-Based Methods:

In some embodiments of the present disclosure, a plant is transformed with an expression cassette that is capable of expressing a polynucleotide that inhibits the expression of a polypeptide of the disclosure. The term "expression" as used herein refers to the biosynthesis of a gene product, including the transcription and/or translation of said gene product. For example, for the purposes of the present disclosure, an expression cassette capable of expressing a polynucleotide that inhibits the expression of at least one polypeptide is an expression cassette capable of producing an RNA molecule that inhibits the transcription and/or translation of at least one polypeptide of the disclosure. The "expression" or "production" of a protein or polypeptide from a DNA molecule refers to the transcription and translation of the coding sequence to produce the protein or polypeptide, while the "expression" or "production" of a protein or polypeptide from an RNA molecule refers to the translation of the RNA coding sequence to produce the protein or polypeptide.

Examples of polynucleotides that inhibit the expression of a polypeptide are given below.

i. Sense Suppression/Cosuppression

In some embodiments of the disclosure, inhibition of the expression of a polypeptide may be obtained by sense suppression or cosuppression. For cosuppression, an expression cassette is designed to express an RNA molecule corresponding to all or part of a messenger RNA encoding a polypeptide in the "sense" orientation. Over expression of the RNA molecule can result in reduced expression of the native gene. Accordingly, multiple plant lines transformed with the cosuppression expression cassette are screened to identify those that show the desired degree of inhibition of polypeptide expression.

The polynucleotide used for cosuppression may correspond to all or part of the sequence encoding the polypeptide, all or part of the 5' and/or 3' untranslated region of a polypeptide transcript or all or part of both the coding sequence and the untranslated regions of a transcript encoding a polypeptide. In some embodiments where the polynucleotide comprises all or part of the coding region for the polypeptide, the expression cassette is designed to eliminate the start codon of the polynucleotide so that no protein product will be translated.

Cosuppression may be used to inhibit the expression of plant genes to produce plants having undetectable protein levels for the proteins encoded by these genes. See, for example, Broin, *et al.*, (2002) *Plant Cell* 14:1417-1432. Cosuppression may also be used to inhibit the expression of multiple proteins in the same plant. See, for example, US Patent Number 5,942,657. Methods for using cosuppression to inhibit the expression of endogenous genes in plants are described in Flavell, *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:3490-3496; Jorgensen, *et al.*, (1996) *Plant Mol. Biol.* 31:957-973; Johansen and Carrington, (2001) *Plant Physiol.* 126:930-938; Broin, *et al.*, (2002) *Plant Cell* 14:1417-1432; Stoutjesdijk, *et al.*, (2002) *Plant Physiol.* 129:1723-1731; Yu, *et al.*, (2003) *Phytochemistry* 63:753-763 and US Patent Numbers 5,034,323, 5,283,184 and 5,942,657, each of which is herein incorporated by reference. The efficiency of cosuppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the sense sequence and 5' of the polyadenylation signal. See, US Patent Application Publication Number 2002/0048814, herein incorporated by reference. Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous gene, optimally greater than about 65% sequence identity, more optimally greater than about 85% sequence identity, most optimally greater than about 95% sequence identity. See US Patent Numbers 5,283,184 and 5,034,323, herein incorporated by reference.

ii. Antisense Suppression

In some embodiments of the disclosure, inhibition of the expression of the polypeptide may be obtained by antisense suppression. For antisense suppression, the expression cassette is designed to express an RNA molecule complementary to all or part of a messenger RNA encoding the polypeptide. Over expression of the antisense RNA molecule can result in reduced expression of the target gene. Accordingly, multiple plant lines transformed with the antisense suppression expression cassette are screened to identify those that show the desired degree of inhibition of polypeptide expression.

The polynucleotide for use in antisense suppression may correspond to all or part of the complement of the sequence encoding the polypeptide, all or part of the complement of the 5' and/or 3' untranslated region of the target transcript or all or part of the complement of both the coding sequence and the untranslated regions of a transcript encoding the polypeptide. In addition, the antisense polynucleotide may be fully complementary (i.e., 100% identical to the complement of the target sequence) or partially complementary (i.e., less than 100% identical to the complement of the target sequence) to the target sequence. Antisense suppression may be used to inhibit the expression of multiple proteins in the same plant. See, for example, US Patent Number 5,942,657. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, 300, 400, 450, 500, 550 or greater may be used. Methods for using antisense suppression to inhibit the expression of endogenous genes in plants are described, for example, in Liu, *et al.*, (2002) *Plant Physiol.* 129:1732-1743 and US Patent Numbers 5,759,829 and 5,942,657, each of which is herein incorporated by reference. Efficiency of antisense suppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the antisense sequence and 5' of the polyadenylation signal. See, US Patent Application Publication Number 2002/0048814, herein incorporated by reference.

iii. Double-Stranded RNA Interference

In some embodiments of the disclosure, inhibition of the expression of a polypeptide may be obtained by double-stranded RNA (dsRNA) interference. For dsRNA interference, a sense RNA molecule like that described above for cosuppression and an antisense RNA molecule that is fully or partially complementary to the sense RNA molecule are expressed in the same cell, resulting in inhibition of the expression of the corresponding endogenous messenger RNA.

Expression of the sense and antisense molecules can be accomplished by designing the expression cassette to comprise both a sense sequence and an antisense sequence. Alternatively, separate expression cassettes may be used for the sense and antisense sequences. Multiple plant lines transformed with the dsRNA interference expression cassette or expression cassettes are then screened to identify plant lines that show the desired degree of inhibition of polypeptide expression. Methods for using dsRNA interference to inhibit the expression of endogenous plant genes are described in Waterhouse, *et al.*, (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964, Liu, *et al.*, (2002)

Plant Physiol. 129:1732-1743 and WO 1999/49029, WO 1999/53050, WO 1999/61631 and WO 2000/49035, each of which is herein incorporated by reference.

iv. Hairpin RNA Interference and Intron-Containing Hairpin RNA Interference

In some embodiments of the disclosure, inhibition of the expression of a polypeptide may be obtained by hairpin RNA (hpRNA) interference or intron-containing hairpin RNA (ihpRNA) interference. These methods are highly efficient at inhibiting the expression of endogenous genes. See, Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38 and the references cited therein.

For hpRNA interference, the expression cassette is designed to express an RNA molecule that hybridizes with itself to form a hairpin structure that comprises a single-stranded loop region and a base-paired stem. The base-paired stem region comprises a sense sequence corresponding to all or part of the endogenous messenger RNA encoded by the gene whose expression is to be inhibited, and an antisense sequence that is fully or partially complementary to the sense sequence. Alternatively, the base-paired stem region may correspond to a portion of a promoter sequence controlling expression of the gene whose expression is to be inhibited. Thus, the base-paired stem region of the molecule generally determines the specificity of the RNA interference. hpRNA molecules are highly efficient at inhibiting the expression of endogenous genes and the RNA interference they induce is inherited by subsequent generations of plants. See, for example, Chuang and Meyerowitz, (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk, *et al.*, (2002) *Plant Physiol.* 129:1723-1731 and Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38. Methods for using hpRNA interference to inhibit or silence the expression of genes are described, for example, in Chuang and Meyerowitz, (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk, *et al.*, (2002) *Plant Physiol.* 129:1723-1731; Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38; Pandolfini *et al.*, *BMC Biotechnology* 3:7 and US Patent Application Publication Number 2003/0175965, each of which is herein incorporated by reference. A transient assay for the efficiency of hpRNA constructs to silence gene expression *in vivo* has been described by Panstruga, *et al.*, (2003) *Mol. Biol. Rep.* 30:135-140, herein incorporated by reference.

For ihpRNA, the interfering molecules have the same general structure as for hpRNA, but the RNA molecule additionally comprises an intron that is capable of being spliced in the cell in which the ihpRNA is expressed. The use of an intron minimizes the

size of the loop in the hairpin RNA molecule following splicing, and this increases the efficiency of interference. See, for example, Smith, *et al.*, (2000) *Nature* 407:319-320. In fact, Smith, *et al.*, show 100% suppression of endogenous gene expression using ihpRNA-mediated interference. Methods for using ihpRNA interference to inhibit the expression of endogenous plant genes are described, for example, in Smith, *et al.*, (2000) *Nature* 407:319-320; Wesley, *et al.*, (2001) *Plant J.* 27:581-590; Wang and Waterhouse, (2001) *Curr. Opin. Plant Biol.* 5:146-150; Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38; Helliwell and Waterhouse, (2003) *Methods* 30:289-295 and US Patent Application Publication Number 2003/0180945, each of which is herein incorporated by reference.

The expression cassette for hpRNA interference may also be designed such that the sense sequence and the antisense sequence do not correspond to an endogenous RNA. In this embodiment, the sense and antisense sequence flank a loop sequence that comprises a nucleotide sequence corresponding to all or part of the endogenous messenger RNA of the target gene. Thus, it is the loop region that determines the specificity of the RNA interference. See, for example, WO 2002/00904; Mette, *et al.*, (2000) *EMBO J* 19:5194-5201; Matzke, *et al.*, (2001) *Curr. Opin. Genet. Devel.* 11:221-227; Scheid, *et al.*, (2002) *Proc. Natl. Acad. Sci., USA* 99:13659-13662; Aufsaftz, *et al.*, (2002) *Proc. Nat'l. Acad. Sci*. 99(4):16499-16506; Sijen, *et al.*, *Curr. Biol.* (2001) 11:436-440), herein incorporated by reference.

v. Amplicon-Mediated Interference

Amplicon expression cassettes comprise a plant-virus-derived sequence that contains all or part of the target gene but generally not all of the genes of the native virus. The viral sequences present in the transcription product of the expression cassette allow the transcription product to direct its own replication. The transcripts produced by the amplicon may be either sense or antisense relative to the target sequence (i.e., the messenger RNA for the polypeptide). Methods of using amplicons to inhibit the expression of endogenous plant genes are described, for example, in Angell and Baulcombe, (1997) *EMBO J.* 16:3675-3684, Angell and Baulcombe, (1999) *Plant J.* 20:357-362 and US Patent Number 6,646,805, each of which is herein incorporated by reference.

vi. Ribozymes

In some embodiments, the polynucleotide expressed by the expression cassette of the disclosure is catalytic RNA or has ribozyme activity specific for the messenger RNA of the polypeptide. Thus, the polynucleotide causes the degradation of the endogenous messenger RNA, resulting in reduced expression of the polypeptide. This method is described, for example, in US Patent Number 4,987,071, herein incorporated by reference.

vii. Small Interfering RNA or Micro RNA

In some embodiments of the disclosure, inhibition of the expression of a polypeptide may be obtained by RNA interference by expression of a polynucleotide encoding a micro RNA (miRNA). miRNAs are regulatory agents consisting of about 22 ribonucleotides. miRNA are highly efficient at inhibiting the expression of endogenous genes. See, for example Javier, *et al.*, (2003) *Nature* 425:257-263, herein incorporated by reference.

For miRNA interference, the expression cassette is designed to express an RNA molecule that is modeled on an endogenous miRNA gene. For example, the miRNA gene encodes an RNA that forms a hairpin structure containing a 22-nucleotide sequence that is complementary to an endogenous gene target sequence. For suppression of NUE expression, the 22-nucleotide sequence is selected from a NUE transcript sequence and contains 22 nucleotides of said NUE sequence in sense orientation and 21 nucleotides of a corresponding antisense sequence that is complementary to the sense sequence. A fertility gene, whether endogenous or exogenous, may be an miRNA target. miRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and the RNA interference they induce is inherited by subsequent generations of plants.

2. Polypeptide-Based Inhibition of Gene Expression

In one embodiment, the polynucleotide encodes a zinc finger protein that binds to a gene encoding a polypeptide, resulting in reduced expression of the gene. In particular embodiments, the zinc finger protein binds to a regulatory region of a gene. In other embodiments, the zinc finger protein binds to a messenger RNA encoding a polypeptide and prevents its translation. Methods of selecting sites for targeting by zinc finger proteins have been described, for example, in US Patent Number 6,453,242, and

methods for using zinc finger proteins to inhibit the expression of genes in plants are described, for example, in US Patent Application Publication Number 2003/0037355, each of which is herein incorporated by reference.

3. Polypeptide-Based Inhibition of Protein Activity

In some embodiments of the disclosure, the polynucleotide encodes an antibody that binds to at least one polypeptide and reduces the activity of the polypeptide. In another embodiment, the binding of the antibody results in increased turnover of the antibody-polypeptide complex by cellular quality control mechanisms. The expression of antibodies in plant cells and the inhibition of molecular pathways by expression and binding of antibodies to proteins in plant cells are well known in the art. See, for example, Conrad and Sonnewald, (2003) *Nature Biotech.* 21:35-36, incorporated herein by reference.

4. Gene Disruption

In some embodiments of the present disclosure, the activity of a polypeptide is reduced or eliminated by disrupting the gene encoding the polypeptide. The gene encoding the polypeptide may be disrupted by any method known in the art. For example, in one embodiment, the gene is disrupted by transposon tagging. In another embodiment, the gene is disrupted by mutagenizing plants using random or targeted mutagenesis and selecting for plants that have reduced nitrogen utilization activity.

i. Transposon Tagging

In one embodiment of the disclosure, transposon tagging is used to reduce or eliminate the activity of one or more polypeptide. Transposon tagging comprises inserting a transposon within an endogenous gene to reduce or eliminate expression of the polypeptide.

In this embodiment, the expression of one or more polypeptides is reduced or eliminated by inserting a transposon within a regulatory region or coding region of the gene encoding the polypeptide. A transposon that is within an exon, intron, 5' or 3' untranslated sequence, a promoter or any other regulatory sequence of a gene may be used to reduce or eliminate the expression and/or activity of the encoded polypeptide.

Methods for the transposon tagging of specific genes in plants are well known in the art. See, for example, Maes, *et al.*, (1999) *Trends Plant Sci.* 4:90-96; Dharmapuri and Sonti, (1999) *FEMS Microbiol. Lett*. 179:53-59; Meissner, *et al.*, (2000) *Plant J.* 22:265-274; Phogat, *et al.*, (2000) *J. Biosci.* 25:57-63; Walbot, (2000) *Curr. Opin. Plant Biol.* 2:103-107; Gai, *et al.*, (2000) *Nucleic Acids Res.* 28:94-96; Fitzmaurice, *et al.*, (1999) *Genetics* 153:1919-1928). In addition, the TUSC process for selecting Mu insertions in selected genes has been described in Bensen, *et al.*, (1995) *Plant Cell* 7:75-84; Mena, *et al.*, (1996) *Science* 274:1537-1540 and US Patent Number 5,962,764, each of which is herein incorporated by reference.

ii. Mutant Plants with Reduced Activity

Additional methods for decreasing or eliminating the expression of endogenous genes in plants are known in the art and can be similarly applied to the instant disclosure. These methods include other forms of mutagenesis, such as ethyl methanesulfonate-induced mutagenesis, deletion mutagenesis and fast neutron deletion mutagenesis used in a reverse genetics sense (with PCR) to identify plant lines in which the endogenous gene has been deleted. For examples of these methods see, Ohshima, *et al.*, (1998) *Virology* 243:472-481; Okubara, *et al.*, (1994) *Genetics* 137:867-874 and Quesada, *et al.*, (2000) *Genetics* 154:421-436, each of which is herein incorporated by reference. In addition, a fast and automatable method for screening for chemically induced mutations, TILLING (Targeting Induced Local Lesions In Genomes), using denaturing HPLC or selective endonuclease digestion of selected PCR products is also applicable to the instant disclosure. See, McCallum, *et al.*, (2000) *Nat. Biotechnol.* 18:455-457, herein incorporated by reference.

Mutations that impact gene expression or that interfere with the function of the encoded protein are well known in the art. Insertional mutations in gene exons usually result in null-mutants. Mutations in conserved residues are particularly effective in inhibiting the activity of the encoded protein. Conserved residues of plant polypeptides suitable for mutagenesis with the goal to eliminate activity have been described. Such mutants can be isolated according to well-known procedures and mutations in different loci can be stacked by genetic crossing. See, for example, Gruis, *et al.*, (2002) *Plant Cell* 14:2863-2882.

In another embodiment of this disclosure, dominant mutants can be used to trigger RNA silencing due to gene inversion and recombination of a duplicated gene locus. See, for example, Kusaba, *et al.*, (2003) *Plant Cell* 15:1455-1467.

The disclosure encompasses additional methods for reducing or eliminating the activity of one or more polypeptide. Examples of other methods for altering or mutating a genomic nucleotide sequence in a plant are known in the art and include, but are not limited to, the use of RNA:DNA vectors, RNA:DNA mutational vectors, RNA:DNA repair vectors, mixed-duplex oligonucleotides, self-complementary RNA:DNA oligonucleotides and recombinogenic oligonucleobases. Such vectors and methods of use are known in the art. See, for example, US Patent Numbers 5,565,350; 5,731,181; 5,756,325; 5,760,012; 5,795,972 and 5,871,984, each of which are herein incorporated by reference. See also, WO 1998/49350, WO 1999/07865, WO 1999/25821 and Beetham, *et al.*, (1999) *Proc. Natl. Acad. Sci. USA* 96:8774-8778, each of which is herein incorporated by reference.

iii. Modulating nitrogen utilization activity

In specific methods, the level and/or activity of a NUE regulator in a plant is decreased by increasing the level or activity of the polypeptide in the plant. The increased expression of a negative regulatory molecule may decrease the level of expression of downstream one or more genes responsible for an improved NUE phenotype.

Methods for increasing the level and/or activity of polypeptides in a plant are discussed elsewhere herein. Briefly, such methods comprise providing a polypeptide of the disclosure to a plant and thereby increasing the level and/or activity of the polypeptide. In other embodiments, a NUE nucleotide sequence encoding a polypeptide can be provided by introducing into the plant a polynucleotide comprising a NUE nucleotide sequence of the disclosure, expressing the NUE sequence, increasing the activity of the polypeptide and thereby decreasing the number of tissue cells in the plant or plant part. In other embodiments, the NUE nucleotide construct introduced into the plant is stably incorporated into the genome of the plant.

In other methods, the growth of a plant tissue is increased by decreasing the level and/or activity of the polypeptide in the plant. Such methods are disclosed in detail elsewhere herein. In one such method, a NUE nucleotide sequence is introduced into the

plant and expression of said NUE nucleotide sequence decreases the activity of the polypeptide and thereby increasing the tissue growth in the plant or plant part. In other embodiments, the NUE nucleotide construct introduced into the plant is stably incorporated into the genome of the plant.

As discussed above, one of skill will recognize the appropriate promoter to use to modulate the level/activity of a NUE in the plant. Exemplary promoters for this embodiment have been disclosed elsewhere herein.

In other embodiments, such plants have stably incorporated into their genome a nucleic acid molecule comprising a NUE nucleotide sequence of the disclosure operably linked to a promoter that drives expression in the plant cell.

iv. Modulating Root Development

Methods for modulating root development in a plant are provided. By "modulating root development" is intended any alteration in the development of the plant root when compared to a control plant. Such alterations in root development include, but are not limited to, alterations in the growth rate of the primary root, the fresh root weight, the extent of lateral and adventitious root formation, the vasculature system, meristem development or radial expansion.

Methods for modulating root development in a plant are provided. The methods comprise modulating the level and/or activity of the polypeptide in the plant. In one method, a sequence of the disclosure is provided to the plant. In another method, the nucleotide sequence is provided by introducing into the plant a polynucleotide comprising a nucleotide sequence of the disclosure, expressing the sequence and thereby modifying root development. In still other methods, the nucleotide construct introduced into the plant is stably incorporated into the genome of the plant.

In other methods, root development is modulated by altering the level or activity of the polypeptide in the plant. A change in activity can result in at least one or more of the following alterations to root development, including, but not limited to, alterations in root biomass and length.

As used herein, "root growth" encompasses all aspects of growth of the different parts that make up the root system at different stages of its development in both monocotyledonous and dicotyledonous plants. It is to be understood that enhanced root growth can result from enhanced growth of one or more of its parts including the primary root, lateral roots, adventitious roots, etc.

Methods of measuring such developmental alterations in the root system are known in the art. See, for example, US Patent Application Publication Number 2003/0074698 and Werner, *et al.*, (2001) *PNAS* 18:10487-10492, both of which are herein incorporated by reference.

As discussed above, one of skill will recognize the appropriate promoter to use to modulate root development in the plant. Exemplary promoters for this embodiment include constitutive promoters and root-preferred promoters. Exemplary root-preferred promoters have been disclosed elsewhere herein.

Stimulating root growth and increasing root mass by decreasing the activity and/or level of the polypeptide also finds use in improving the standability of a plant. The term "resistance to lodging" or "standability" refers to the ability of a plant to fix itself to the soil. For plants with an erect or semi-erect growth habit, this term also refers to the ability to maintain an upright position under adverse environmental conditions. This trait relates to the size, depth and morphology of the root system. In addition, stimulating root growth and increasing root mass by altering the level and/or activity of the polypeptide finds use in promoting *in vitro* propagation of explants.

Furthermore, higher root biomass production has a direct effect on the yield and an indirect effect of production of compounds produced by root cells or transgenic root cells or cell cultures of said transgenic root cells. One example of an interesting compound produced in root cultures is shikonin, the yield of which can be advantageously enhanced by said methods.

Accordingly, the present disclosure further provides plants having modulated root development when compared to the root development of a control plant. In some embodiments, the plant of the disclosure has an increased level/activity of a polypeptide of the disclosure and has enhanced root growth and/or root biomass. In other embodiments, such plants have stably incorporated into their genome a nucleic acid molecule comprising a nucleotide sequence of the disclosure operably linked to a promoter that drives expression in the plant cell.

v. Modulating Shoot and Leaf Development

Methods are also provided for modulating shoot and leaf development in a plant. By "modulating shoot and/or leaf development" is intended any alteration in the development of the plant shoot and/or leaf. Such alterations in shoot and/or leaf development include, but are not limited to, alterations in shoot meristem development, in leaf number, leaf size, leaf and stem vasculature, internode length and leaf senescence.

As used herein, "leaf development" and "shoot development" encompasses all aspects of growth of the different parts that make up the leaf system and the shoot system, respectively, at different stages of their development, both in monocotyledonous and dicotyledonous plants. Methods for measuring such developmental alterations in the shoot and leaf system are known in the art. See, for example, Werner, *et al.*, (2001) *PNAS* 98:10487-10492 and US Patent Application Publication Number 2003/0074698, each of which is herein incorporated by reference.

The method for modulating shoot and/or leaf development in a plant comprises modulating the activity and/or level of a polypeptide of the disclosure. In one embodiment, a sequence of the disclosure is provided. In other embodiments, the nucleotide sequence can be provided by introducing into the plant a polynucleotide comprising a nucleotide sequence of the disclosure, expressing the sequence and thereby modifying shoot and/or leaf development. In other embodiments, the nucleotide construct introduced into the plant is stably incorporated into the genome of the plant.

In specific embodiments, shoot or leaf development is modulated by altering the level and/or activity of the polypeptide in the plant. A change in activity can result in at least one or more of the following alterations in shoot and/or leaf development, including, but not limited to, changes in leaf number, altered leaf surface, altered vasculature, internodes and plant growth and alterations in leaf senescence when compared to a control plant.

As discussed above, one of skill will recognize the appropriate promoter to use to modulate shoot and leaf development of the plant. Exemplary promoters for this embodiment include constitutive promoters, shoot-preferred promoters, shoot meristem-preferred promoters and leaf-preferred promoters. Exemplary promoters have been disclosed elsewhere herein.

Increasing activity and/or level of a polypeptide of the disclosure in a plant may result in altered internodes and growth. Thus, the methods of the disclosure find use in producing modified plants. In addition, as discussed above, activity in the plant modulates both root and shoot growth. Thus, the present disclosure further provides methods for altering the root/shoot ratio. Shoot or leaf development can further be modulated by altering the level and/or activity of the polypeptide in the plant.

Accordingly, the present disclosure further provides plants having modulated shoot and/or leaf development when compared to a control plant. In some embodiments, the plant of the disclosure has an increased level/activity of a polypeptide of the disclosure. In other embodiments, a plant of the disclosure has a decreased level/activity of a polypeptide of the disclosure.

vi. Modulating Reproductive Tissue Development

Methods for modulating reproductive tissue development are provided. In one embodiment, methods are provided to modulate floral development in a plant. By "modulating floral development" is intended any alteration in a structure of a plant's reproductive tissue as compared to a control plant in which the activity or level of the polypeptide has not been modulated. "Modulating floral development" further includes any alteration in the timing of the development of a plant's reproductive tissue (e.g., a delayed or an accelerated timing of floral development) when compared to a control plant in which the activity or level of the polypeptide has not been modulated. Changes in timing of reproductive development may result in altered synchronization of development of male and female reproductive tissues. Macroscopic alterations may include changes in size, shape, number or location of reproductive organs, the developmental time period that these structures form or the ability to maintain or proceed through the flowering process in times of environmental stress. Microscopic alterations may include changes to the types or shapes of cells that make up the reproductive organs.

The method for modulating floral development in a plant comprises modulating activity in a plant. In one method, a sequence of the disclosure is provided. A nucleotide sequence can be provided by introducing into the plant a polynucleotide comprising a nucleotide sequence of the disclosure, expressing the sequence and thereby modifying floral development. In other embodiments, the nucleotide construct introduced into the plant is stably incorporated into the genome of the plant.

In specific methods, floral development is modulated by increasing the level or activity of the polypeptide in the plant. A change in activity can result in at least one or more of the following alterations in floral development, including, but not limited to, altered flowering, changed number of flowers, modified male sterility and altered seed set, when compared to a control plant. Inducing delayed flowering or inhibiting flowering can be used to enhance yield in forage crops such as alfalfa. Methods for measuring such developmental alterations in floral development are known in the art. See, for example, Mouradov, *et al.*, (2002) *The Plant Cell* S111-S130, herein incorporated by reference.

As discussed above, one of skill will recognize the appropriate promoter to use to modulate floral development of the plant. Exemplary promoters for this embodiment include constitutive promoters, inducible promoters, shoot-preferred promoters and inflorescence-preferred promoters.

In other methods, floral development is modulated by altering the level and/or activity of a sequence of the disclosure. Such methods can comprise introducing a nucleotide sequence into the plant and changing the activity of the polypeptide. In other methods, the nucleotide construct introduced into the plant is stably incorporated into the genome of the plant. Altering expression of the sequence of the disclosure can modulate floral development during periods of stress. Such methods are described elsewhere herein. Accordingly, the present disclosure further provides plants having modulated floral development when compared to the floral development of a control plant. Compositions include plants having an altered level/activity of the polypeptide of the disclosure and having an altered floral development. Compositions also include plants having a modified level/activity of the polypeptide of the disclosure wherein the plant maintains or proceeds through the flowering process in times of stress.

Methods are also provided for the use of the sequences of the disclosure to increase seed size and/or weight. The method comprises increasing the activity of the sequences in a plant or plant part, such as the seed. An increase in seed size and/or weight comprises an increased size or weight of the seed and/or an increase in the size or weight of one or more seed part including, for example, the embryo, endosperm, seed coat, aleurone or cotyledon.

As discussed above, one of skill will recognize the appropriate promoter to use to increase seed size and/or seed weight. Exemplary promoters of this embodiment include constitutive promoters, inducible promoters, seed-preferred promoters, embryo-preferred promoters and endosperm-preferred promoters.

A method for altering seed size and/or seed weight in a plant may increasing activity in the plant. In one embodiment, the nucleotide sequence can be provided by introducing into the plant a polynucleotide comprising a nucleotide sequence of the disclosure, expressing the sequence and thereby impacting seed weight and/or size. In certain embodiments, the nucleotide construct introduced into the plant is stably incorporated into the genome of the plant.

It is further recognized that increasing seed size and/or weight can also be accompanied by an increase in the speed of growth of seedlings or an increase in early vigor. As used herein, the term "early vigor" refers to the ability of a plant to grow rapidly during early development, and relates to the successful establishment, after germination, of a well-developed root system and a well-developed photosynthetic apparatus. In addition, an increase in seed size and/or weight can also result in an increase in plant yield when compared to a control.

Accordingly, the present disclosure further provides plants having an increased seed weight and/or seed size when compared to a control plant. In other embodiments, plants having an increased vigor and plant yield are also provided. In some embodiments, the plant of the disclosure has a modified level/activity of the polypeptide of the disclosure and has an increased seed weight and/or seed size. In other embodiments, such plants have stably incorporated into their genome a nucleic acid molecule comprising a nucleotide sequence of the disclosure operably linked to a promoter that drives expression in the plant cell.

vii. Method of Use for polynucleotide, expression cassettes, and additional polynucleotides

The nucleotides, expression cassettes and methods disclosed herein are useful in regulating expression of any heterologous nucleotide sequence in a host plant in order to vary the phenotype of a plant. Various changes in phenotype are of interest including modifying the fatty acid composition in a plant, altering the amino acid content of a plant, altering a plant's pathogen defense mechanism and the like. These results can be achieved by providing expression of heterologous products or increased expression of endogenous products in plants. Alternatively, the results can be achieved by providing for a reduction of expression of one or more endogenous products, particularly enzymes or cofactors in the plant. These changes result in a change in phenotype of the transformed plant.

Genes of interest are reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest change, and as developing nations open up world markets, new crops and technologies will emerge also. In addition, as our understanding of agronomic traits and characteristics such as yield and heterosis increases, the choice of genes for transformation will change accordingly. General categories of genes of interest include, for example, those genes involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. More specific categories of transgenes, for example, include genes encoding important traits for agronomics, insect resistance, disease resistance, herbicide resistance, sterility, grain characteristics and commercial products. Genes of interest include, generally, those involved in oil, starch, carbohydrate or nutrient metabolism as well as those affecting kernel size, sucrose loading and the like.

In certain embodiments the nucleic acid sequences of the present disclosure can be used in combination ("stacked") with other polynucleotide sequences of interest in order to create plants with a desired phenotype. The combinations generated can include multiple copies of any one or more of the polynucleotides of interest. The promoter which is operably linked to a polynucleotide sequence of interest can be any promoter that is active in plant cells. In some embodiments it is particularly advantageous to use a promoter that is active (or can be activated) in reproductive tissues of a plant (e.g., stamens or ovaries). As such, the promoter can be, for example, a constitutively active promoter, an inducible promoter, a tissue-specific promoter or a developmental stage specific promoter. Also, the promoter of the a exogenous nucleic acid molecule can be the same as or different from the promoter of a second exogenous nucleic acid molecule.

The polynucleotides of the present disclosure may be stacked with any gene or combination of genes to produce plants with a variety of desired trait combinations, including but not limited to traits desirable for animal feed such as high oil genes (e.g., US Patent Number 6,232,529); balanced amino acids (e.g., hordothionins (US Patent Numbers 5,990,389; 5,885,801; 5,885,802 and 5,703,409); barley high lysine (Williamson, *et al.*, (1987) *Eur. J. Biochem*. 165:99-106 and WO 1998/20122) and high methionine proteins (Pedersen, *et al.*, (1986) *J. Biol. Chem*. 261:6279; Kirihara, *et al.*, (1988) *Gene* 71:359 and Musumura, *et al.*, (1989) *Plant Mol. Biol*. 12:123)); increased digestibility (e.g., modified storage proteins (US Patent Application Serial Number 10/053,410, filed November 7, 2001) and thioredoxins (US Patent Application Serial Number 10/005,429, filed December 3, 2001)), the disclosures of which are herein incorporated by reference. The polynucleotides of the present disclosure can also be stacked with traits desirable for insect, disease or herbicide resistance (e.g., *Bacillus thuringiensis* toxic proteins (US Patent Numbers 5,366,892; 5,747,450; 5,737,514; 5723,756; 5,593,881; Geiser, *et al.*, (1986) *Gene* 48:109); lectins (Van Damme, *et al.*, (1994) *Plant Mol. Biol*. 24:825); fumonisin detoxification genes (US Patent Number 5,792,931); avirulence and disease resistance genes (Jones, *et al.*, (1994) *Science* 266:789; Martin, *et al.*, (1993) *Science* 262:1432; Mindrinos, *et al.*, (1994) *Cell* 78:1089); acetolactate synthase (ALS) mutants that lead to herbicide resistance such as the S4 and/or Hra mutations; inhibitors of glutamine synthase such as phosphinothricin or basta (e.g., bar gene); and glyphosate resistance (EPSPS gene)) and traits desirable for processing or process products such as high oil (e.g., US Patent Number 6,232,529); modified oils (e.g., fatty acid desaturase genes (US Patent Number 5,952,544; WO 1994/11516)); modified starches (e.g., ADPG pyrophosphorylases (AGPase), starch synthases (SS), starch branching enzymes (SBE) and starch debranching enzymes (SDBE)) and polymers or bioplastics (e.g., US Patent

Number 5,602,321; beta-ketothiolase, polyhydroxybutyrate synthase, and acetoacetyl-CoA reductase (Schubert, *et al.*, (1988) *J. Bacteriol*. 170:5837-5847) facilitate expression of polyhydroxyalkanoates (PHAs)), the disclosures of which are herein incorporated by reference. One could also combine the polynucleotides of the present disclosure with polynucleotides affecting agronomic traits such as male sterility (e.g., see, US Patent Number 5.583,210), stalk strength, flowering time or transformation technology traits such as cell cycle regulation or gene targeting (e.g., WO 1999/61619; WO 2000/17364; WO 1999/25821), the disclosures of which are herein incorporated by reference.

Transgenic plants comprising or derived from plant cells or native plants of this disclosure can be further enhanced with stacked traits, e.g., a crop plant having an enhanced trait resulting from expression of DNA disclosed herein in combination with herbicide tolerance and/or pest resistance traits. For example, plants with an altered trait of interest can be stacked with other traits of agronomic interest, such as a trait providing herbicide resistance and/or insect resistance, such as using a gene from *Bacillus thuringensis* to provide resistance against one or more of lepidopteran, coleopteran, homopteran, hemiopteran and other insects. Known genes that confer tolerance to herbicides such as e.g., auxin, HPPD, glyphosate, dicamba, glufosinate, sulfonylurea, bromoxynil and norflurazon herbicides can be stacked either as a molecular stack or a breeding stack with plants expressing the traits disclosed herein. Polynucleotide molecules encoding proteins involved in herbicide tolerance include, but are not limited to, a polynucleotide molecule encoding 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) disclosed in US Patent Numbers 39,247; 6,566,587 and for imparting glyphosate tolerance; polynucleotide molecules encoding a glyphosate oxidoreductase (GOX) disclosed in US Patent Number 5,463,175 and a glyphosate-N-acetyl transferase (GAT) disclosed in US Patent Numbers 7,622,641; 7,462,481; 7,531,339; 7,527,955; 7,709,709; 7,714,188 and 7,666,643, also for providing glyphosate tolerance; dicamba monooxygenase disclosed in US Patent Number 7,022,896 and WO 2007/146706 A2 for providing dicamba tolerance; a polynucleotide molecule encoding AAD12 disclosed in US Patent Application Publication Number 2005/731044 or WO 2007/053482 A2 or encoding AAD1 disclosed in US Patent Application Publication Number 2011/0124503 A1 or US Patent Number 7,838,733 for providing tolerance to auxin herbicides (2,4-D); a polynucleotide molecule encoding hydroxyphenylpyruvate dioxygenase (HPPD) for providing tolerance to HPPD inhibitors (e.g., hydroxyphenylpyruvate dioxygenase) disclosed in e.g., US Patent Number 7,935,869; US Patent Application Publication Numbers 2009/0055976 A1 and 2011/0023180 A1; each publication is herein incorporated by reference in its entirety.

Other examples of herbicide-tolerance traits that could be combined with the traits disclosed herein include those conferred by polynucleotides encoding an exogenous phosphinothricin acetyltransferase, as described in US Patent Numbers 5,969,213; 5,489,520; 5,550,318; 5,874,265; 5,919,675; 5,561,236; 5,648,477; 5,646,024; 6,177,616 and 5,879,903. Plants containing an exogenous phosphinothricin acetyltransferase can exhibit improved tolerance to glufosinate herbicides, which inhibit the enzyme glutamine synthase. Other examples of herbicide-tolerance traits include those conferred by polynucleotides conferring altered protoporphyrinogen oxidase (protox) activity, as described in US Patent Numbers 6,288,306 B1; 6,282,837 B1 and 5,767,373 and international publication WO 2001/12825. Plants containing such polynucleotides can exhibit improved tolerance to any of a variety of herbicides which target the protox enzyme (also referred to as “protox inhibitors”)

In one embodiment, sequences of interest improve plant growth and/or crop yields. For example, sequences of interest include agronomically important genes that result in improved primary or lateral root systems. Such genes include, but are not limited to, nutrient/water transporters and growth inducers. Examples of such genes include, but are not limited to, maize plasma membrane H^+-ATPase (MHA2) (Frias, *et al.*, (1996) *Plant Cell 8*:1533-44); AKT1, a component of the potassium uptake apparatus in *Arabidopsis*, (Spalding, *et al.*, (1999) *J Gen Physiol* 113:909-18); RML genes which activate cell division cycle in the root apical cells (Cheng, *et al.*, (1995) *Plant Physiol* 108:881); maize glutamine synthetase genes (Sukanya, *et al.*, (1994) *Plant Mol Biol* 26:1935-46) and hemoglobin (Duff, *et al.*, (1997) *J. Biol. Chem* 27:16749-16752, Arredondo-Peter, *et al.*, (1997) *Plant Physiol.* 115:1259-1266; Arredondo-Peter, *et al.*, (1997) *Plant Physiol* 114:493-500 and references sited therein). The sequence of interest may also be useful in expressing antisense nucleotide sequences of genes that negatively affect root development.

Additional, agronomically important traits such as oil, starch and protein content can be genetically altered in addition to using traditional breeding methods. Modifications include increasing content of oleic acid, saturated and unsaturated oils, increasing levels of lysine and sulfur, providing essential amino acids and also modification of starch. Hordothionin protein modifications are described in US Patent Numbers 5,703,049, 5,885,801, 5,885,802 and 5,990,389, herein incorporated by reference. Another example is lysine and/or sulfur rich seed protein encoded by the soybean 2S albumin described in US Patent Number 5,850,016 and the chymotrypsin inhibitor from barley described in Williamson, *et al.*, (1987) *Eur. J. Biochem.* 165:99-106, the disclosures of which are

herein incorporated by reference. Derivatives of the coding sequences can be made by site-directed mutagenesis to increase the level of preselected amino acids in the encoded polypeptide. For example, the gene encoding the barley high lysine polypeptide (BHL) is derived from barley chymotrypsin inhibitor, US Patent Application Serial Number 08/740,682, filed November 1, 1996, and WO 1998/20133, the disclosures of which are herein incorporated by reference. Other proteins include methionine-rich plant proteins such as from sunflower seed (Lilley, *et al.*, (1989) *Proceedings of the World Congress on Vegetable Protein Utilization in Human Foods and Animal Feedstuffs*, ed. Applewhite (American Oil Chemists Society, Champaign, Illinois), pp. 497-502; herein incorporated by reference); corn (Pedersen, *et al.*, (1986) *J. Biol. Chem.* 261:6279; Kirihara, *et al.*, (1988) *Gene* 71:359, both of which are herein incorporated by reference) and rice (Musumura, *et al.*, (1989) *Plant Mol. Biol.* 12:123, herein incorporated by reference). Other agronomically important genes encode latex, Floury 2, growth factors, seed storage factors and transcription factors.

Insect resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer and the like. Such genes include, for example, *Bacillus thuringiensis* toxic protein genes (US Patent Numbers 5,366,892; 5,747,450; 5,736,514; 5,723,756; 5,593,881 and Geiser, *et al.*, (1986) *Gene* 48:109) and the like.

Genes encoding disease resistance traits include detoxification genes, such as against fumonosin (US Patent Number 5,792,931); avirulence (avr) and disease resistance (R) genes (Jones, *et al.*, (1994) *Science* 266:789; Martin, *et al.*, (1993) *Science* 262:1432 and Mindrinos, *et al.*, (1994) *Cell* 78:1089) and the like.

Herbicide resistance traits may include genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance, in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides that act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene) or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin and the ALS-gene mutants encode resistance to the herbicide chlorsulfuron.

Sterility genes can also be encoded in an expression cassette and provide an alternative to physical emasculation. Examples of genes used in such ways include male

tissue-preferred genes and genes with male sterility phenotypes such as QM, described in US Patent Number 5,583,210. Other genes include kinases and those encoding compounds toxic to either male or female gametophytic development.

The quality of grain is reflected in traits such as levels and types of oils, saturated and unsaturated, quality and quantity of essential amino acids, and levels of cellulose. In corn, modified hordothionin proteins are described in US Patent Numbers 5,703,049, 5,885,801, 5,885,802 and 5,990,389.

Commercial traits can also be encoded on a gene or genes that could increase, for example, starch for ethanol production or provide expression of proteins. Another important commercial use of transformed plants is the production of polymers and bioplastics such as described in US Patent Number 5,602,321. Genes such as β-Ketothiolase, PHBase (polyhydroxyburyrate synthase) and acetoacetyl-CoA reductase (see, Schubert, *et al.*, (1988) *J. Bacteriol.* 170:5837-5847) facilitate expression of polyhyroxyalkanoates (PHAs).

Exogenous products include plant enzymes and products as well as those from other sources including procaryotes and other eukaryotes. Such products include enzymes, cofactors, hormones and the like. The level of proteins, particularly modified proteins having improved amino acid distribution to improve the nutrient value of the plant, can be increased. This is achieved by the expression of such proteins having enhanced amino acid content.

Genome Editing and Induced Mutagenesis

In general, methods to modify or alter the host endogenous genomic DNA are available. This includes altering the host native DNA sequence or a pre-existing transgenic sequence including regulatory elements, coding and non-coding sequences. These methods are also useful in targeting nucleic acids to pre-engineered target recognition sequences in the genome. As an example, the genetically modified cell or plant described herein is generated using “custom" meganucleases produced to modify plant genomes (see, e.g., WO 2009/114321; Gao, *et al.*, (2010) *Plant Journal* 1:176-187). *Other site*-directed engineering is through the use of zinc finger domain recognition coupled with the restriction properties of restriction enzyme. See, e.g., Urnov, *et al.*, (2010) *Nat Rev Genet.* 11(9):636-46; Shukla, *et al.*, (2009) *Nature* 459(7245):437-41.

“TILLING” or “Targeting Induced Local Lesions IN Genomics” refers to a mutagenesis technology useful to generate and/or identify and to eventually isolate mutagenised variants of a particular nucleic acid with modulated expression and/or activity (McCallum, *et al.*, (2000), *Plant Physiology* 123:439-442; McCallum, *et al.*, (2000)

Nature Biotechnology 18:455-457 and Colbert, *et al.*, (2001) *Plant Physiology* 126:480-484). Methods for TILLING are well known in the art (US Patent Number 8,071,840).

Other mutagenic methods can also be employed to introduce mutations in a disclosed gene. Methods for introducing genetic mutations into plant genes and selecting plants with desired traits are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as X-rays or gamma rays can be used.

Embodiments of the disclosure reflect the determination that the genotype of an organism can be modified to contain dominant suppressor alleles or transgene constructs that suppress (i.e., reduce, but not ablate) the activity of a gene, wherein the phenotype of the organism is not substantially affected.

Hybrid seed production requires elimination or inactivation of pollen produced by the female parent. Incomplete removal or inactivation of the pollen provides the potential for selfing, raising the risk that inadvertently self-pollinated seed will unintentionally be harvested and packaged with hybrid seed. Once the seed is planted, the selfed plants can be identified and selected; the selfed plants are genetically equivalent to the female inbred line used to produce the hybrid. Typically, the selfed plants are identified and selected based on their decreased vigor relative to the hybrid plants. For example, female selfed plants of maize are identified by their less vigorous appearance for vegetative and/or reproductive characteristics, including shorter plant height, small ear size, ear and kernel shape, cob color or other characteristics. Selfed lines also can be identified using molecular marker analyses (see, e.g., Smith and Wych, (1995) *Seed Sci. Technol.* 14:1-8). Using such methods, the homozygosity of the self-pollinated line can be verified by analyzing allelic composition at various loci in the genome.

Because hybrid plants are important and valuable field crops, plant breeders are continually working to develop high-yielding hybrids that are agronomically sound based on stable inbred lines. The availability of such hybrids allows a maximum amount of crop to be produced with the inputs used, while minimizing susceptibility to pests and environmental stresses. To accomplish this goal, the plant breeder must develop superior inbred parental lines for producing hybrids by identifying and selecting genetically unique individuals that occur in a segregating population. The present disclosure contributes to this goal, for example by providing plants that, when crossed, generate male sterile progeny, which can be used as female parental plants for generating hybrid plants.

A large number of genes have been identified as being tassel preferred in their expression pattern using traditional methods and more recent high-throughput methods. The correlation of function of these genes with important biochemical or developmental processes that ultimately lead to functional pollen is arduous when approaches are limited to classical forward or reverse genetic mutational analysis. As disclosed herein, suppression approaches in maize provide an alternative rapid means to identify genes that are directly related to pollen development in maize.

Promoters useful for expressing a nucleic acid molecule of interest can be any of a range of naturally-occurring promoters known to be operative in plants or animals, as desired. Promoters that direct expression in cells of male or female reproductive organs of a plant are useful for generating a transgenic plant or breeding pair of plants of the disclosure. The promoters useful in the present disclosure can include constitutive promoters, which generally are active in most or all tissues of a plant; inducible promoters, which generally are inactive or exhibit a low basal level of expression and can be induced to a relatively high activity upon contact of cells with an appropriate inducing agent; tissue-specific (or tissue-preferred) promoters, which generally are expressed in only one or a few particular cell types (e.g., plant anther cells) and developmental- or stage-specific promoters, which are active only during a defined period during the growth or development of a plant. Often promoters can be modified, if necessary, to vary the expression level. Certain embodiments comprise promoters exogenous to the species being manipulated. For example, the Ms45 gene introduced into ms45ms45 maize germplasm may be driven by a promoter isolated from another plant species; a hairpin construct may then be designed to target the exogenous plant promoter, reducing the possibility of hairpin interaction with non-target, endogenous maize promoters.

Exemplary constitutive promoters include the 35S cauliflower mosaic virus (CaMV) promoter promoter (Odell, *et al.*, (1985) *Nature* 313:810-812), the maize ubiquitin promoter (Christensen, *et al.*, (1989) *Plant Mol. Biol.* 12:619-632 and Christensen, *et al.*, (1992) *Plant Mol. Biol.* 18:675-689); the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO 1999/43838 and US Patent Number 6,072,050; rice actin (McElroy, *et al.*, (1990) *Plant Cell* 2:163-171); pEMU (Last, *et al.*, (1991) *Theor. Appl. Genet.* 81:581-588); MAS (Velten, *et al.*, (1984) *EMBO J.* 3:2723-2730); ALS promoter (US Patent Number 5,659,026); rice actin promoter (US Patent Number 5,641,876; WO 2000/70067), maize histone promoter (Brignon, *et al.*, (1993) *Plant Mol Bio* 22(6):1007-1015; Rasco-Gaunt, *et al.*, (2003) *Plant Cell Rep.* 21(6):569-576) and the like. Other constitutive promoters include, for example, those described in US Patent Numbers 5,608,144 and 6,177,611 and PCT Publication Number WO 2003/102198.

Tissue-specific, tissue-preferred or stage-specific regulatory elements further include, for example, the *AGL8/FRUITFULL* regulatory element, which is activated upon floral induction (Hempel, *et al.*, (1997) *Development* 124:3845-3853); root-specific regulatory elements such as the regulatory elements from the RCP1 gene and the LRP1 gene (Tsugeki and Fedoroff, (1999) *Proc. Natl. Acad., USA* 96:12941-12946; Smith and Fedoroff, (1995) *Plant Cell* 7:735-745); flower-specific regulatory elements such as the regulatory elements from the *LEAFY* gene and the *APETALA1* gene (Blazquez, *et al.*, (1997) *Development* 124:3835-3844; Hempel, *et al.*, *supra*, 1997); seed-specific regulatory elements such as the regulatory element from the oleosin gene (Plant, *et al.*, (1994) *Plant Mol. Biol.* 25:193-205) and dehiscence zone specific regulatory element. Additional tissue-specific or stage-specific regulatory elements include the Zn13 promoter, which is a pollen-specific promoter (Hamilton, *et al.*, (1992) *Plant Mol. Biol.* 18:211-218); the *UNUSUAL FLORAL ORGANS* (*UFO*) promoter, which is active in apical shoot meristem; the promoter active in shoot meristems (Atanassova, *et al.*, (1992) *Plant J.* 2:291), the cdc2 promoter and cyc07 promoter (see, for example, Ito, *et al.*, (1994) *Plant Mol. Biol.* 24:863-878; Martinez, *et al.*, (1992) *Proc. Natl. Acad. Sci., USA* 89:7360); the meristematic-preferred meri-5 and H3 promoters (Medford, *et al.*, (1991) *Plant Cell* 3:359; Terada, *et al.*, (1993) *Plant J.* 3:241); meristematic and phloem-preferred promoters of Myb-related genes in barley (Wissenbach, *et al.*, (1993) *Plant J.* 4:411); *Arabidopsis* cyc3aAt and cyc1At (Shaul, *et al.*, (1996) *Proc. Natl. Acad. Sci.* 93:4868-4872); *C. roseus* cyclins CYS and CYM (Ito, *et al.*, (1997) *Plant J.* 11:983-992); and *Nicotiana* CyclinB1 (Trehin, *et al.*, (1997) *Plant Mol. Biol.* 35:667-672); the promoter of the *APETALA3* gene, which is active in floral meristems (Jack, *et al.*, (1994) *Cell* 76:703; Hempel, *et al.*, *supra*, 1997); a promoter of an agamous-like (AGL) family member, for example, AGL8, which is active in shoot meristem upon the transition to flowering (Hempel, *et al.*, *supra*, 1997); floral abscission zone promoters; L1-specific promoters; the ripening-enhanced tomato polygalacturonase promoter (Nicholass, *et al.*, (1995) *Plant Mol. Biol.* 28:423-435), the E8 promoter (Deikman, *et al.*, (1992) *Plant Physiol.* 100:2013-2017) and the fruit-specific 2A1 promoter, U2 and U5 snRNA promoters from maize, the Z4 promoter from a gene encoding the Z4 22 kD zein protein, the Z10 promoter from a gene encoding a 10 kD zein protein, a Z27 promoter from a gene encoding a 27 kD zein protein, the A20 promoter from the gene encoding a 19 kD zein protein, and the like. Additional tissue-specific promoters can be isolated using well known methods (see, e.g., US Patent Number 5,589,379). Shoot-preferred promoters include shoot meristem-preferred promoters such as promoters disclosed in Weigel, *et al.*, (1992) *Cell* 69:843-859 (Accession Number M91208); Accession Number AJ131822; Accession Number Z71981; Accession Number

AF049870 and shoot-preferred promoters disclosed in McAvoy, *et al.*, (2003) *Acta Hort. (ISHS)* 625:379-385. Inflorescence-preferred promoters include the promoter of chalcone synthase (Van der Meer, *et al.*, (1992) *Plant J.* 2(4):525-535), anther-specific LAT52 (Twell, *et al.*, (1989) *Mol. Gen. Genet.* 217:240-245), pollen-specific Bp4 (Albani, *et al.*, (1990) *Plant Mol Biol.* 15:605, maize pollen-specific gene Zm13 (Hamilton, *et al.*, (1992) *Plant Mol. Biol.* 18:211-218; Guerrero, *et al.*, (1993) *Mol. Gen. Genet.* 224:161-168), microspore-specific promoters such as the apg gene promoter (Twell, *et al.*, (1993) *Sex. Plant Reprod.* 6:217-224) and tapetum-specific promoters such as the TA29 gene promoter (Mariani, *et al.*, (1990) *Nature* 347:737; US Patent Number 6,372,967) and other stamen-specific promoters such as the MS45 gene promoter, 5126 gene promoter, BS7 gene promoter, PG47 gene promoter (US Patent Number 5,412,085; US Patent Number 5,545,546; *Plant J* 3(2):261-271 (1993)), SGB6 gene promoter (US Patent Number 5,470,359), G9 gene promoter (US Patent Number 5,8937,850; US Patent Number 5,589,610), SB200 gene promoter (WO 2002/26789), or the like. Tissue-preferred promoters of interest further include a sunflower pollen-expressed gene SF3 (Baltz, *et al.*, (1992) *The Plant Journal* 2:713-721), B. napus pollen specific genes (Arnoldo, *et al.*, (1992) *J. Cell. Biochem*, Abstract Number Y101204). Tissue-preferred promoters further include those reported by Yamamoto, *et al.*, (1997) *Plant J.* 12(2):255-265 (psaDb); Kawamata, *et al.*, (1997) *Plant Cell Physiol.* 38(7):792-803 (PsPAL1); Hansen, *et al.*, (1997) *Mol. Gen Genet.* 254(3):337-343 (ORF13); Russell, *et al.*, (1997) *Transgenic Res.* 6(2):157-168 (waxy or ZmGBS; 27kDa zein, ZmZ27; osAGP; osGT1); Rinehart, *et al.*, (1996) *Plant Physiol.* 112(3):1331-1341 (Fbl2A from cotton); Van Camp, *et al.*, (1996) *Plant Physiol.* 112(2):525-535 (*Nicotiana* SodA1 and SodA2); Canevascini, *et al.*, (1996) *Plant Physiol.* 112(2):513-524 (*Nicotiana* ltp1); Yamamoto, *et al.*, (1994) *Plant Cell Physiol.* 35(5):773-778 (*Pinus* cab-6 promoter); Lam, (1994) *Results Probl. Cell Differ.* 20:181-196; Orozco, *et al.*, (1993) *Plant Mol Biol.* 23(6):1129-1138 (spinach rubisco activase (Rca)); Matsuoka, *et al.*, (1993) *Proc Natl. Acad. Sci. USA* 90(20):9586-9590 (PPDK promoter) and Guevara-Garcia, *et al.*, (1993) *Plant J.* 4(3):495-505 (*Agrobacterium* pmas promoter). A tissue-preferred promoter that is active in cells of male or female reproductive organs can be particularly useful in certain aspects of the present disclosure.

"Seed-preferred" promoters include both "seed-developing" promoters (those promoters active during seed development such as promoters of seed storage proteins) as well as "seed-germinating" promoters (those promoters active during seed germination). See, Thompson, *et al.*, (1989) *BioEssays* 10:108. Such seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message), cZ19B1

(maize 19 kDa zein), mi1ps (myo-inositol-1-phosphate synthase); see, WO 2000/11177 and US Patent Number 6,225,529. Gamma-zein is an endosperm-specific promoter. Globulin-1 (Glob-1) is a representative embryo-specific promoter. For dicots, seed-specific promoters include, but are not limited to, bean β-phaseolin, napin, β-conglycinin, soybean lectin, cruciferin, and the like. For monocots, seed-specific promoters include, but are not limited to, maize 15 kDa zein, 22 kDa zein, 27 kDa zein, gamma-zein, waxy, shrunken 1, shrunken 2, globulin 1, etc. See also, WO 2000/12733 and US Patent Number 6,528,704, where seed-preferred promoters from *end1* and *end2* genes are disclosed. Additional embryo specific promoters are disclosed in Sato, *et al.*, (1996) *Proc. Natl. Acad. Sci.* 93:8117-8122 (rice homeobox, OSH1) and Postma-Haarsma, *et al.*, (1999) *Plant Mol. Biol.* 39:257-71 (rice KNOX genes). Additional endosperm specific promoters are disclosed in Albani, *et al.*, (1984) *EMBO* 3:1405-15; Albani, *et al.*, (1999) *Theor. Appl. Gen.* 98:1253-62; Albani, *et al.*, (1993) *Plant J.* 4:343-55; Mena, *et al.*, (1998) *The Plant Journal* 116:53-62 (barley DOF); Opsahl-Ferstad, *et al.*, (1997) *Plant J* 12:235-46 (maize Esr) and Wu, *et al.*, (1998) *Plant Cell Physiology* 39:885-889 (rice GluA-3, GluB-1, NRP33, RAG-1).

An inducible regulatory element is one that is capable of directly or indirectly activating transcription of one or more DNA sequences or genes in response to an inducer. The inducer can be a chemical agent such as a protein, metabolite, growth regulator, herbicide or phenolic compound or a physiological stress, such as that imposed directly by heat, cold, salt, or toxic elements or indirectly through the action of a pathogen or disease agent such as a virus or other biological or physical agent or environmental condition. A plant cell containing an inducible regulatory element may be exposed to an inducer by externally applying the inducer to the cell or plant such as by spraying, watering, heating or similar methods. An inducing agent useful for inducing expression from an inducible promoter is selected based on the particular inducible regulatory element. In response to exposure to an inducing agent, transcription from the inducible regulatory element generally is initiated *de novo* or is increased above a basal or constitutive level of expression. Typically the protein factor that binds specifically to an inducible regulatory element to activate transcription is present in an inactive form which is then directly or indirectly converted to the active form by the inducer. Any inducible promoter can be used in the instant disclosure (See, Ward, *et al.*, (1993) *Plant Mol. Biol.* 22:361-366).

Examples of inducible regulatory elements include a metallothionein regulatory element, a copper-inducible regulatory element or a tetracycline-inducible regulatory element, the transcription from which can be effected in response to divalent metal ions,

copper or tetracycline, respectively (Furst, *et al.*, (1988) *Cell* 55:705-717; Mett, *et al.*, (1993) *Proc. Natl. Acad. Sci., USA* 90:4567-4571; Gatz, *et al.*, (1992) *Plant J.* 2:397-404; Roder, *et al.*, (1994) *Mol. Gen. Genet.* 243:32-38). Inducible regulatory elements also include an ecdysone regulatory element or a glucocorticoid regulatory element, the transcription from which can be effected in response to ecdysone or other steroid (Christopherson, *et al.*, (1992) *Proc. Natl. Acad. Sci., USA* 89:6314-6318; Schena, *et al.*, (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425; US Patent Number 6,504,082); a cold responsive regulatory element or a heat shock regulatory element, the transcription of which can be effected in response to exposure to cold or heat, respectively (Takahashi, *et al.*, (1992) *Plant Physiol.* 99:383-390); the promoter of the alcohol dehydrogenase gene (Gerlach, *et al.*, (1982) *PNAS USA* 79:2981-2985; Walker, *et al.*, (1987) *PNAS* 84(19):6624-6628), inducible by anaerobic conditions; and the light-inducible promoter derived from the pea rbcS gene or pea psaDb gene (Yamamoto, *et al.*, (1997) *Plant J.* 12(2):255-265); a light-inducible regulatory element (Feinbaum, *et al.*, (1991) *Mol. Gen. Genet.* 226:449; Lam and Chua, (1990) *Science* 248:471; Matsuoka, *et al.*, (1993) *Proc. Natl. Acad. Sci. USA* 90(20):9586-9590; Orozco, *et al.*,. (1993) *Plant Mol. Bio.* 23(6):1129-1138), a plant hormone inducible regulatory element (Yamaguchi-Shinozaki, *et al.*, (1990) *Plant Mol. Biol.* 15:905; Kares, *et al.*, (1990) *Plant Mol. Biol.* 15:225), and the like. An inducible regulatory element also can be the promoter of the maize In2-1 or In2-2 gene, which responds to benzenesulfonamide herbicide safeners (Hershey, *et al.*, (1991) *Mol. Gen. Gene.* 227:229-237; Gatz, *et al.*, (1994) *Mol. Gen. Genet.* 243:32-38) and the Tet repressor of transposon Tn10 (Gatz, *et al.*, (1991) *Mol. Gen. Genet.* 227:229-237). Stress inducible promoters include salt/water stress-inducible promoters such as P5CS (Zang, *et al.*, (1997) *Plant Sciences* 129:81-89); cold-inducible promoters, such as, cor15a (Hajela, *et al.*, (1990) *Plant Physiol.* 93:1246-1252), cor15b (Wlihelm, *et al.*, (1993) *Plant Mol Biol* 23:1073-1077), wsc120 (Ouellet, *et al.*, (1998) *FEBS Lett.* 423:324-328), ci7 (Kirch, *et al.*, (1997) *Plant Mol Biol.* 33:897-909), ci21A (Schneider, *et al.*, (1997) *Plant Physiol.* 113:335-45); drought-inducible promoters, such as, Trg-31 (Chaudhary, *et al.*, (1996) *Plant Mol. Biol.* 30:1247-57), rd29 (Kasuga, *et al.*, (1999) *Nature Biotechnology* 18:287-291); osmotic inducible promoters, such as Rab17 (Vilardell, *et al.*, (1991) *Plant Mol. Biol.* 17:985-93) and osmotin (Raghothama, *et al.*, (1993) *Plant Mol Biol* 23:1117-28) and heat inducible promoters, such as heat shock proteins (Barros, *et al.*, (1992) *Plant Mol.* 19:665-75; Marrs, *et al.*, (1993) *Dev. Genet.* 14:27-41), smHSP (Waters, *et al.*, (1996) *J. Experimental Botany* 47:325-338) and the heat-shock inducible element from the parsley ubiquitin promoter (WO 03/102198). Other stress-inducible promoters include rip2 (US Patent Number 5,332,808 and US Patent Application Publication Number 2003/0217393)

and rd29a (Yamaguchi-Shinozaki, *et al.*, (1993) *Mol. Gen. Genetics* 236:331-340). Certain promoters are inducible by wounding, including the *Agrobacterium* pmas promoter (Guevara-Garcia, *et al.*, (1993) *Plant J.* 4(3):495-505) and the *Agrobacterium* ORF13 promoter (Hansen, *et al.*, (1997) *Mol. Gen. Genet*. 254(3):337-343).

In certain embodiments, a promoter is selected based, for example, on whether male fertility or female fertility is to be impacted Thus, where the male fertility is to be impacted, (e.g., a BS7 gene and an SB200 gene), the promoter may be, for example, an MS45 gene promoter (US Patent Number 6,037,523), a 5126 gene promoter (US Patent Number 5,837,851), a BS7 gene promoter (WO 2002/063021), an SB200 gene promoter (WO 2002/26789), a TA29 gene promoter (*Nature* 347:737 (1990)), a PG47 gene promoter (US Patent Number 5,412,085; US Patent Number 5,545,546; *Plant J* 3(2):261-271 (1993)) an SGB6 gene promoter (US Patent Number 5,470,359) a G9 gene promoter (US Patent Numbers 5,837,850 and 5,589,610) or the like. Where female fertility is to be impacted, the promoter can target female reproductive genes, for example an ovary specific promoter. In certain embodiments, any promoter can be used that directs expression in the tissue of interest, including, for example, a constitutively active promoter such as an ubiquitin promoter, which generally effects transcription in most or all plant cells.

Additional regulatory elements active in plant cells and useful in the methods or compositions of the disclosure include, for example, the spinach nitrite reductase gene regulatory element (Back, *et al.*, (1991) *Plant Mol. Biol.* 17:9); a gamma zein promoter, an oleosin ole16 promoter, a globulin I promoter, an actin I promoter, an actin cl promoter, a sucrose synthetase promoter, an INOPS promoter, an EXM5 promoter, a globulin2 promoter, a b-32, ADPG-pyrophosphorylase promoter, an Ltpl promoter, an Ltp2 promoter, an oleosin ole17 promoter, an oleosin ole18 promoter, an actin 2 promoter, a pollen-specific protein promoter, a pollen-specific pectate lyase gene promoter or PG47 gene promoter, an anther specific RTS2 gene promoter, SGB6 gene promoter, or G9 gene promoter, a tapetum specific RAB24 gene promoter, an anthranilate synthase alpha subunit promoter, an alpha zein promoter, an anthranilate synthase beta subunit promoter, a dihydrodipicolinate synthase promoter, a Thi I promoter, an alcohol dehydrogenase promoter, a cab binding protein promoter, an H3C4 promoter, a RUBISCO SS starch branching enzyme promoter, an actin3 promoter, an actin7 promoter, a regulatory protein GF14-12 promoter, a ribosomal protein L9 promoter, a cellulose biosynthetic enzyme promoter, an S-adenosyl-L-homocysteine hydrolase promoter, a superoxide dismutase promoter, a C-kinase receptor promoter, a phosphoglycerate mutase promoter, a root-specific RCc3 mRNA promoter, a glucose-6

phosphate isomerase promoter, a pyrophosphate-fructose 6-phosphate-l-phosphotransferase promoter, a beta-ketoacyl-ACP synthase promoter, a 33 kDa photosystem 11 promoter, an oxygen evolving protein promoter, a 69 kDa vacuolar ATPase subunit promoter, a glyceraldehyde-3-phosphate dehydrogenase promoter, an ABA- and ripening- inducible-like protein promoter, a phenylalanine ammonia lyase promoter, an adenosine triphosphatase S-adenosyl-L-homocysteine hydrolase promoter, a chalcone synthase promoter, a zein promoter, a globulin-1 promoter, an auxin-binding protein promoter, a UDP glucose flavonoid glycosyl-transferase gene promoter, an NTI promoter, an actin promoter and an opaque 2 promoter.

Plants suitable for purposes of the present disclosure can be monocots or dicots and include, but are not limited to, maize, wheat, barley, rye, sweet potato, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, pepper, celery, squash, pumpkin, hemp, zucchini, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tomato, sorghum, sugarcane, sugar beet, sunflower, rapeseed, clover, tobacco, carrot, cotton, alfalfa, rice, potato, eggplant, cucumber, *Arabidopsis thaliana* and woody plants such as coniferous and deciduous trees. Thus, a transgenic plant or genetically modified plant cell of the disclosure can be an angiosperm or gymnosperm.

Angiosperms are divided into two broad classes based on the number of cotyledons, which are seed leaves that generally store or absorb food; a monocotyledonous angiosperm has a single cotyledon and a dicotyledonous angiosperm has two cotyledons. Angiosperms produce a variety of useful products including materials such as lumber, rubber and paper; fibers such as cotton and linen; herbs and medicines such as quinine and vinblastine; ornamental flowers such as roses and where included within the scope of the present disclosure, orchids and foodstuffs such as grains, oils, fruits and vegetables. Angiosperms encompass a variety of flowering plants, including, for example, cereal plants, leguminous plants, oilseed plants, hardwood trees, fruit-bearing plants and ornamental flowers, which general classes are not necessarily exclusive. Cereal plants, which produce an edible grain, include, for example, corn, rice, wheat, barley, oat, rye, orchardgrass, guinea grass and sorghum. Leguminous plants include members of the pea family (*Fabaceae*) and produce a characteristic fruit known as a legume. Examples of leguminous plants include, for example, soybean, pea, chickpea, moth bean, broad bean, kidney bean, lima bean, lentil, cowpea, dry bean and peanut, as well as alfalfa, birdsfoot trefoil, clover and sainfoin. Oilseed plants, which have seeds that are useful as a source of oil, include soybean, sunflower, rapeseed (canola) and

cottonseed. Angiosperms also include hardwood trees, which are perennial woody plants that generally have a single stem (trunk). Examples of such trees include alder, ash, aspen, basswood (linden), beech, birch, cherry, cottonwood, elm, eucalyptus, hickory, locust, maple, oak, persimmon, poplar, sycamore, walnut, sequoia and willow. Trees are useful, for example, as a source of pulp, paper, structural material and fuel.

Angiosperms produce seeds enclosed within a mature, ripened ovary. An angiosperm fruit can be suitable for human or animal consumption or for collection of seeds to propagate the species. For example, hops are a member of the mulberry family that are prized for their flavoring in malt liquor. Fruit-bearing angiosperms also include grape, orange, lemon, grapefruit, avocado, date, peach, cherry, olive, plum, coconut, apple and pear trees and blackberry, blueberry, raspberry, strawberry, pineapple, tomato, cucumber and eggplant plants. An ornamental flower is an angiosperm cultivated for its decorative flower. Examples of commercially important ornamental flowers include rose, lily, tulip and chrysanthemum, snapdragon, camellia, carnation and petunia plants and can include orchids. It will be recognized that the present disclosure also can be practiced using gymnosperms, which do not produce seeds in a fruit.

Homozygosity is a genetic condition existing when identical alleles reside at corresponding loci on homologous chromosomes. Heterozygosity is a genetic condition existing when different alleles reside at corresponding loci on homologous chromosomes. Hemizygosity is a genetic condition existing when there is only one copy of a gene (or set of genes) with no allelic counterpart on the sister chromosome.

The plant breeding methods used herein are well known to one skilled in the art. For a discussion of plant breeding techniques, see, Poehlman, (1987) *Breeding Field Crops* AVI Publication Co., Westport Conn. Many of the plants which would be most preferred in this method are bred through techniques that take advantage of the plant's method of pollination.

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

recurrent parent are recovered in the converted plant, in addition to the single transferred gene from the nonrecurrent parent.

By transgene is meant any nucleic acid sequence which has been introduced into the genome of a cell by genetic engineering techniques. A transgene may be a native DNA sequence or a heterologous DNA sequence. The term native DNA sequence can refer to a nucleotide sequence which is naturally found in the cell but that may have been modified from its original form.

Using well-known techniques, additional promoter sequences may be isolated based on their sequence homology. In these techniques, all or part of a known promoter sequence is used as a probe which selectively hybridizes to other sequences present in a population of cloned genomic DNA fragments (i.e. genomic libraries) from a chosen organism. Methods that are readily available in the art for the hybridization of nucleic acid sequences may be used to obtain sequences which correspond to these promoter sequences in species including, but not limited to, maize (corn; *Zea mays*), canola (*Brassica napus, Brassica rapa ssp.*), alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor, Sorghum vulgare*), sunflower (*Helianthus annuus*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea spp.*), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus spp.*), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa spp.*), avocado *(Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), oats, barley, vegetables, ornamentals and conifers. Preferably, plants include maize, soybean, sunflower, safflower, canola, wheat, barley, rye, alfalfa and sorghum.

The entire promoter sequence or portions thereof can be used as a probe capable of specifically hybridizing to corresponding promoter sequences. To achieve specific hybridization under a variety of conditions, such probes include sequences that are unique and are preferably at least about 10 nucleotides in length and most preferably at least about 20 nucleotides in length. Such probes can be used to amplify corresponding promoter sequences from a chosen organism by the well-known process of polymerase chain reaction (PCR). This technique can be used to isolate additional promoter sequences from a desired organism or as a diagnostic assay to determine the presence of the promoter sequence in an organism. Examples include hybridization screening of plated DNA libraries (either plaques or colonies; see e.g., Innis, *et al.*, (1990*) PCR Protocols, A Guide to Methods and Applications*, eds., Academic Press).

In general, sequences that correspond to a promoter sequence of the present disclosure and hybridize to a promoter sequence disclosed herein will be at least 50% homologous, 55% homologous, 60% homologous, 65% homologous, 70% homologous, 75% homologous, 80% homologous, 85% homologous, 90% homologous, 95% homologous and even 98% homologous or more with the disclosed sequence.

Fragments of a particular promoter sequence disclosed herein may operate to promote the pollen-preferred expression of an operably-linked isolated nucleotide sequence. These fragments will comprise at least about 20 contiguous nucleotides, preferably at least about 50 contiguous nucleotides, more preferably at least about 75 contiguous nucleotides, even more preferably at least about 100 contiguous nucleotides of the particular promoter nucleotide sequences disclosed herein. The nucleotides of such fragments will usually comprise the TATA recognition sequence of the particular promoter sequence. Such fragments can be obtained by use of restriction enzymes to cleave the naturally-occurring promoter sequences disclosed herein; by synthesizing a nucleotide sequence from the naturally-occurring DNA sequence or through the use of PCR technology. See particularly, Mullis, *et al*., (1987) *Methods Enzymol*. 155:335-350 and Erlich, ed. (1989) *PCR Technology* (Stockton Press, New York). Again, variants of these fragments, such as those resulting from site-directed mutagenesis, are encompassed by the compositions of the present disclosure.

Biologically active variants of the promoter sequence are also encompassed by the compositions of the present disclosure. A regulatory “variant” is a modified form of a promoter wherein one or more bases have been modified, removed or added. For example, a routine way to remove part of a DNA sequence is to use an exonuclease in combination with DNA amplification to produce unidirectional nested deletions of double-stranded DNA clones. A commercial kit for this purpose is sold under the trade name Exo-Size™ (New England Biolabs, Beverly, Mass.). Briefly, this procedure entails incubating exonuclease III with DNA to progressively remove nucleotides in the 3' to 5' direction at 5' overhangs, blunt ends or nicks in the DNA template. However, exonuclease III is unable to remove nucleotides at 3', 4-base overhangs. Timed digests of a clone with this enzyme produce unidirectional nested deletions.

One example of a regulatory sequence variant is a promoter formed by causing one or more deletions in a larger promoter. Deletion of the 5' portion of a promoter up to the TATA box near the transcription start site may be accomplished without abolishing promoter activity, as described by Zhu, *et al*., (1995) *The Plant Cell* 7:1681-89. Such variants should retain promoter activity, particularly the ability to drive expression in

specific tissues. Biologically active variants include, for example, the native regulatory sequences of the disclosure having one or more nucleotide substitutions, deletions or insertions. Activity can be measured by Northern blot analysis, reporter activity measurements when using transcriptional fusions, and the like. See, for example, Sambrook, *et al*., (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), herein incorporated by reference.

The nucleotide sequences for the pollen-preferred promoters disclosed in the present disclosure, as well as variants and fragments thereof, are useful in the genetic manipulation of any plant when operably linked with an isolated nucleotide sequence whose expression is to be controlled to achieve a desired phenotypic response.

The nucleotide sequence operably linked to the regulatory elements disclosed herein can be an antisense sequence for a targeted gene. By "antisense DNA nucleotide sequence" is intended a sequence that is in inverse orientation to the 5'-to-3' normal orientation of that nucleotide sequence. When delivered into a plant cell, expression of the antisense DNA sequence prevents normal expression of the DNA nucleotide sequence for the targeted gene. The antisense nucleotide sequence encodes an RNA transcript that is complementary to and capable of hybridizing with the endogenous messenger RNA (mRNA) produced by transcription of the DNA nucleotide sequence for the targeted gene. In this case, production of the native protein encoded by the targeted gene is inhibited to achieve a desired phenotypic response. Thus the regulatory sequences claimed herein can be operably linked to antisense DNA sequences to reduce or inhibit expression of a native or exogenous protein in the plant.

Regulation of gene expression may be measured in terms of its effect on individual cells. Successful modulation of a trait may be accomplished with high stringency, for example impacting expression in all or nearly all cells of a particular cell type, or with lower stringency. Within a particular tissue, for example, modulation of expression in 98%, 95%, 90%, 80% or fewer cells may result in the desired phenotype.

By "flowering stress" is meant that water is withheld from plants such that drought stress occurs at or around the time of anthesis.

By "grain fill stress" is meant that water is withheld from plants such that drought stress occurs during the time when seeds are accumulating storage products (carbohydrates, protein and/or oil).

By "rain-fed conditions" is meant that water is neither deliberately withheld nor artificially supplemented.

By "well-watered conditions" is meant that water available to the plant is generally adequate for optimum growth.

Drought stress conditions for maize may be controlled to result in a targeted yield reduction. For example, a 20%, 30%, 40%, 50%, 60%, 70%, or greater reduction in yield of control plants can be accomplished by providing measured amounts of water during specific phases of plant development.

Methods for modulating drought tolerance in plants are also features of the invention. The ability to introduce different degrees of drought tolerance into plants offers flexibility in the use of the invention: for example, introduction of strong drought tolerance for improved grain-filling or for silage in areas with longer or drier growing seasons, versus the introduction of a moderate drought tolerance for silage in agricultural areas with shorter growing seasons.

In addition to increasing tolerance to drought stress in plants of the invention compared to a control plant, the invention also enables higher density planting of plants of the invention, leading to increased yield per acre of corn. Most of the increased yield per acre of corn over the last century has come from increasing tolerance to density, which is a stress to plants. Methods for modulating plant stress response, e.g., increasing tolerance for density, are also a feature of the invention.

Plants are grown in the field under normal and drought-stress conditions. Under normal conditions, plants are watered with an amount sufficient for optimum growth and yield. For drought-stressed plants, water may be limited for a period starting approximately one week before pollination and continuing through three weeks after pollination. During the period of limited water availability, drought-stressed plants may show visible signs of wilting and leaf rolling. The degree of stress may be calculated as % yield reduction relative to that obtained under well-watered conditions. Transpiration, stomatal conductance and CO_2 assimilation are determined with a portable TPS-1 Photosynthesis System (PP Systems, Amesbury, MA). Each leaf on a plant may be measured, e.g. at forty days after pollination. Values typically represent a mean of six determinations.

In addition to increasing tolerance to drought stress and improving density stress tolerance in plants of the invention compared to a control plant, the invention also may provide greater nitrogen utilization efficiency (NUE). Plants in which NUE is improved may be more productive than control plants under comparable conditions of ample nitrogen availability and/or may maintain productivity under significantly reduced nitrogen

availability. Improved NUE may be reflected in one or more attributes such as increased biomass, increased grain yield, increased harvest index, increased photosynthetic rates and increased tolerance to biotic or abiotic stress. In particular, improving NUE in maize would increase harvestable yield per unit of input nitrogen fertilizer, both in developing nations where access to nitrogen fertilizer is limited and in developed nations where the level of nitrogen use remains high.

EXAMPLE 1. Improved grain yield from overexpression of truncated ZmSRTF18.

Maize was transformed with a construct comprising a constitutive promoter driving a polynucleotide encoding a novel truncated version of ZmSRTF18 (SEQ ID NO: 2). This truncated version, ZmSRTF18-del, is 71 amino acids shorter than the shortest naturally-occurring functional splice variant known (1.4, SEQ ID NO: 6). ZmSRTF18-del lacks all amino acids encoded by exon 1, 2, and 3, and lacks 52 amino acids encoded by exon 4. This partial loss of exon 4 results in loss of two putative nuclear localization signals. See Figure 3.

Maize plants transformed with a construct comprising a constitutive promoter driving ZmSRTF18-del did not display significant phenotypic changes from the wild-type (WT) control plants. The transformed maize plants did, however, exhibit increased grain yield relative to the control.

Yield data were collected with 4-6 replicates per location. Yield analysis was by ASREML (VSN International Ltd), calculating BLUPs (Best Linear Unbiased Prediction) (Cullis, B. Ret al (1998) *Biometrics* 54: 1–18, Gilmour, A. R. et al (2009). ASReml User Guide 3.0, Gilmour, A.R., et al (1995) *Biometrics* 51: 1440-50). A mixed model framework was used to perform the single and multi location analysis.

In the single-location analysis, main effect of construct is considered as a random effect. The blocking factors such as replicates and incblock (incomplete block design) within replicates are considered as random. Corrections are made for spatial variation in the field. In the multi-location analysis, main effect of location, construct and location x construct interaction are considered as fixed effects. The main effect of event and its interaction with location are considered as random effects. The blocking factors such as replicates and incblock within replicates are considered as random.

Single-location and multi-location analyses were performed and blup (Best Linear Unbiased Prediction) was calculated for each event. Significance between the event and WT was calculated using a p-value of 0.1 in a two-tailed test.

In a first evaluation across three test environments, test-cross hybrid plants comprising any one of five out of six transformation events produced on average 3.4% to 6.0% greater grain yield than the control. These test environments included (1) managed stress environment for flowering drought stress, (2) managed stress environment for grain-filling drought stress, and (3) one of the targeted population of environments which did not have managed stress. Yield improvement over the control was statistically significant for all six events in the first testing environment. The multi-locational improvement in yield for five of the six events was statistically significant as well, primarily driven by results from the first testing environment.

In a second evaluation in five test environments with two tester inbreds, test-cross hybrid plants comprising any one of seven out of ten transformation events produced a statistically significant 2.0% to 4.0% greater grain yield than the control. A slight increase in grain moisture at harvest was observed.

In the second evaluation, tests of the same events in two different hybrid combinations were generally neutral to slightly negative for grain yield, with one event showing a statistically significant 4.1% increase.

Yield tests in the third evaluation of ten constitutive ZmSRTF18-del events showed a consistent yield increase over the bulk-null control in 2 of 3 low-nitrogen environments, with an average advantage for the construct of 3.4 bushels per acre across all three low-nitrogen environments (significant at $P \leq 0.01$).

While not being bound by any theory, it is proposed that constitutive expression of ZmSRTF18-del in maize increased grain yield without inducing severe phenotypic changes because the lack of nuclear localization signals might have reduced transport of the protein into the nucleus, and alternatively or additionally, because the truncated protein binds to fewer promoter elements, and/or binds more weakly. The combination of low protein abundance in the nucleus and weaker binding to promoters may result in fewer downstream genes being expressed, and/or in lower expression levels of downstream genes, compared with results achieved using constitutive expression of the non-truncated wild type ZmSRTF18. Thus the constitutive expression of ZmSRTF18-del may have an impact on downstream gene expression that is more favorable for plant performance.

Alternatively or additionally, the truncated version of the transcription factor may act in a dominant-negative fashion, inactivating a transcription complex by, for example, preventing the native, nontruncated protein from entering the complex, by causing poor

binding to interacting proteins in the complex, or because the truncation changes the conformation of another protein in the complex.

Alternatively or additionally, conserved cysteine residues may impact transcription-factor function. Variants of ZmSRTF18 have been created to evaluate this aspect. As shown in Table 1, ZmSRTF18-del (ALT3) restores to the truncated protein a conserved cysteine residue which may be required to form a disulfide bond needed for proper function. The gel shift assays of Fig 6, when done without dithiothreitol (DTT) in order to avoid reducing disulfide bonds, showed increased binding to DNA for the ZmSRTF18-del (ALT3) protein compared with the ZmSRTF18-del protein. ZmSRTF18-del (ALT2) restores a 16-amino-acid region conserved in homologs from pearl millet (Pg-DREB2A; SEQ ID NO: 11), barley (HvDRF1.3; SEQ ID NO: 12), rice (OsDREB2B; SEQ ID NO: 13), wheat (TaDREB1, SEQ ID NO: 14), and Arabidopsis (AtDREB2A, SEQ ID NO: 15). This 16-amino-acid region includes the conserved cysteine residue and a putative nuclear localization signal.

Alternatively or additionally, the truncated transcription factor may bind poorly to the DRE promoter element by itself in well watered conditions, but in stress conditions, interacting proteins may be induced that bind to the truncated protein and increase its ability to bind the DRE promoter element.

EXAMPLE 2. Determination of subcellular location of ZmSRTF18-del.

As noted in Example 1, truncation which deletes two putative nuclear localization signals may reduce movement of the ZmSRTF18-del protein to the nucleus, resulting in a lower abundance of the truncated protein there, relative to abundance of the nontruncated protein that would be obtained following constitutive overexpression. This may result in increased grain yield without inducing the deleterious phenotypic changes typically associated with constitutive overexpression of a DREB2 transcription factor. To further evaluate subcellular location, the sequence SV40NLS-ZmSRTF18-del was designed (SEQ ID NO: 9). It includes a heterologous nuclear localization signal at the N terminus of ZmSRTF18-del.

Constructs were prepared comprising DNA encoding ZmSRTF18-del or SV40NLS-ZmSRTF18-del proteins fused to the fluorescent protein TagRFP. A control construct comprising TagRFP only, without Zm-SRTF18-del, was also prepared. Short linkers were included in the fusion constructs, and a constitutive promoter was used. Maize embryo scutellum tissue was transformed with the constructs. For example, DNA encoding a red fluorescent protein tag (TagRFP) was linked to DNA encoding ZmSRTF18-del, and maize

embryo scutellum tissue was transformed with the construct. Separately, maize embryo scutellum tissue was transformed with a similar construct but which also comprised a nuclear localization signal (SV40NLS-ZmSRTF18-del). Transformation with the TagRFP sequence alone provided a control. Transient expression in the transformed scutellum tissue was monitored by fluorescent microscopy to determine subcellular location of the protein produced by each of the three constructs.

Results are shown in Figure 5. In the left panel, TagRFP is observed throughout the cell. Similarly, in the center panel, ZmSRTF18-del-TagRFP is located throughout the cell, including the nucleus. In contrast, SV40NLS-ZmSRTF18-del-TagRFP is located only in the nucleus (right panel) within the cell. More intense staining was observed in the nucleolus. This study indicates that the heterologous nuclear localization signal in the SV40NLS-ZmSRTF18-del-TagRFP construct efficiently targeted the fusion protein to the nucleus. Moreover, it indicates that not all of the ZmSRTF18-del-TagRFP protein is located in the nucleus, where it could impact downstream gene transcription.

GS3 X Gaspe maize plants were transformed with the SV40NLS-ZmSRTF18-del gene driven by a constitutive promoter. No consistent significant decreases in growth were observed, despite the fact that the SV40NLS effectively targeted ZmSRTF18-del to the nucleus. This result suggested that the lack of pleiotropy observed in maize plants transformed with ZmSRTF18-del driven by a constitutive promoter may not be entirely due to loss of nuclear targeting.

EXAMPLE 3. Impact of truncation on promoter-binding

As noted in Example 1, the truncated protein ZmSRTF18-del may bind to fewer promoter elements, and/or bind more weakly, thus resulting in induction of fewer genes, or in weaker expression of genes, relative to constitutive overexpression of the wild type ZmSRTF18. Gel-shift assays can be used to study the impact of the truncation on promoter-binding ability and thus indicate impact of the truncation on regulation of downstream genes.

In gel-shift assays, also known as electrophoretic mobility shift assays (EMSA), DNA with bound protein will traverse the gel more slowly than DNA without bound protein. Thus, gel shift assays indicate whether a protein binds to a specific DNA fragment. The gel shift assays of Figure 6 were performed using the Molecular Probes™ EMSA Kit (E33075), which provides background information, materials, and methods for gel shift experiments. The assays were done in the presence or in the absence of DTT (dithiothreitol), as it is not

certain which condition most accurately reflects the in vivo condition. DNA was stained with SYBR® Green according to the manufacturer's protocol.

ZmSRTF18 protein variants used for the gel shift assays were obtained by expressing in E.coli with the pET28 vector from Novagen®. Proteins were expressed with a 20 amino acid N-terminal tag that included 6 histidines, and protein purification was done by cobalt column affinity chromatography.

The DNA used for the gel shift assays of Figure 6 was a 34-nucleotide region (SEQ ID NO: 16), including the DRE element ACCGAC, obtained from the maize RAB17 promoter, as identified by Srivastav et al. (2010) Plant Signaling and Behavior 5(7):775-784). See SEQ ID NO: 17 for Rab17 promoter sequence.

The following ZmSRTF18 protein variants were used for the gel shift assays, in order from shortest to longest amino acid sequence:

ZmSRTF18-del was the most severely truncated protein used (see Fig. 3 and SEQ ID NO 2).

ZmSRTF18-del(ALT3) (SEQ ID NO: 4) adds three amino acids (MGC) to the N-terminus of the truncated protein. This addition restores a conserved Cysteine residue that may form a disulfide bond needed for proper function.

ZmSRTF18-del (ALT2) (SEQ ID NO: 3) restores a 16-amino-acid region highly conserved in homologs from pearl millet, barley, rice, wheat, and Arabidopsis, and also adds a start methionine. This region includes the conserved Cysteine referred to in the ZmSRTF18-del(ALT3) description, and also includes a putative nuclear localization signal (see Figure 3 for NLS location).

ZmSRTF18 (SPL VAR4) is the functional, naturally occurring splice variant that comprises all of exons 1 and 4 and has no further deletions (see Figure 3).

ZmSRTF18 (SPL VAR4) (C71S) replaces the conserved Cysteine at position 71 with Serine, and is otherwise identical to ZmSRTF18 (SPL VAR4).

Results of the gel shift assays are shown in Figure 6. In the absence of DTT, the severely truncated protein ZmSRTF18-del did not appear to bind DNA. However, ZmSRTF18-del (ALT3) showed increased binding to the DRE core element, relative to ZmSRTF18-del. Thus, addition of only 3 amino acids at the N-terminus of ZmSRTF18-del had a large impact on function. The other three proteins all bound DNA effectively and showed obvious gel shifts. In the presence of DTT, both the ZmSRTF18-del and the

ZmSRTF18-del (ALT3) proteins bound DNA weakly. The other proteins had strong binding to DNA, with obvious gel shifts observed. Thus in either the presence or the absence of DTT, the severely truncated protein ZmSRTF18-del bound DNA more weakly, compared with the non-truncated protein ZmSRTF18 (SPL VAR4).

To determine whether DNA binding was specific, and to examine binding of the less severely truncated proteins ZmSRTF18-del (ALT5) (SEQ ID NO: 26) and ZmSRTF18-del (ALT6) (SEQ ID NO: 27), further gel shift experiments were performed in the presence of DTT with the same 34 bp DNA fragment, or with the corresponding fragment that had the DRE element ACCGAC mutated to TTTTTT. Conditions were also optimized to reduce nonspecific binding by adding 300 mM KCl and sheared salmon DNA. Strong , specific binding was observed for ZmSRTF18 (SPL VAR4), ZmSRTF18 (SPL VAR4) (C71S), and ZmSRTF18-del (ALT6), but not for ZmSRTF18-del, ZmSRTF18-del (ALT2), ZmSRTF18-del (ALT3), or ZmSRTF18-del (ALT5).

EXAMPLE 4. Expression Profiling of ZmSRTF18

Constitutive overexpression of ZmDREB2A in Arabidopsis increased expression of numerous late-embryogenesis-abundant (LEA) protein- genes and heat-shock-protein genes, as well as other genes. In light of the reduced binding of ZmSRTF18-del to core promoter elements, fewer downstream genes may be induced, relative to induction by overexpression of the wild-type ZmSRTF18. To test this hypothesis, leaves of late-vegetative-stage plants transgenic for a constitutive promoter driving ZmSRTF18-del, sampled under well-watered and drought-stress conditions, have been profiled for gene expression using the Illumina ® platform (Illumina, Inc., San Diego, CA). Two clusters of genes were identified that had significantly different expression in transgenic leaves compared with null leaves in drought stressed conditions. One cluster comprised 47 genes that had increased expression in drought relative to well watered conditions for both transgenic and null leaves, but the increases in expression were greater for transgenic compared with null leaves. The other cluster comprised 56 genes that had decreased expression in drought relative to well watered conditions for both transgenic and null leaves, but the decreases in expression were greater for transgenic compared with null leaves. The 47 and 56 genes considered were those with p values ≤ 0.1.

Two specific examples of genes in the first cluster were two genes encoding delta-1-pyrroline-5-carboxylate synthetase or P5CS, a key enzyme controlling flux through the proline biosynthetic pathway. In leaves from drought stressed conditions, one P5CS gene had approximately double the expression in transgenic leaves compared with null leaves.

The other P5CS gene had approximately 50% more expression in transgenic compared with null leaves in drought stressed leaves. Proline content was determined in transgenic and null leaves. Increased proline content was observed in transgenic leaves, consistent with the increased expression of P5CS genes. Previous overexpression of DREB1 and DREB2 genes in Arabidopsis or tobacco increased proline or both proline and P5CS gene expression (Gilmour et al (2000) Plant Physiol 124:1854; Chen et al (2007) Biochemical and Biophysical Research Communications 353:299). The observation that overexpressing the truncated DREB2 gene ZmSRTF18-del increased proline and P5CS gene expression as did previous overexpression of nontruncated DREB proteins suggested that the truncated ZmSRTF18-del was acting as a functional transcription factor, and was less likely to be acting in a dominant negative manner to prevent function of the native nontruncated ZmSRTF18 gene.

EXAMPLE 5. Morphologic measurements of ZmSRTF18-del transgenic maize plants.

Maize testcross hybrid plants transgenic for a construct comprising a constitutive promoter driving ZmSRTF18-del were grown under standard field conditions. Ear leaf length, ear leaf width, ear leaf area, and number of nodes per plant, representing two independent transgenic events, were measured and compared to controls. No significant difference was found for the measured traits, relative to the controls. Furthermore, no significant differences in plant height, ear height, or time to flowering were observed between transgenic and control plants for multiple events. These results are in marked contrast to the severe stunting and slow growth phenotypes observed following constitutive overexpression of the ZmDREB2A gene in Arabidopsis (Qin et al, 2007, Plant J 50:54).

To assess effects of constitutive overexpression of nontruncated ZmSRTF18 in maize, GS3 X Gaspe maize was transformed with ZmSRTF18 (SPL VAR1), SEQ ID NO. 5, driven by a constitutive promoter. A severe delay in maize growth and development, as indicated by increased time to pollen shed, was observed. Truncation of the ZmSRTF18-del gene appeared to minimize such deleterious phenotypes associated with constitutive overexpression of nontruncated ZmSRTF18.

To determine whether an N-terminal addition of 16 amino acids plus a start methionine to ZmSRTF18-del would impact growth and development, transgenic maize inbred plants constitutively overexpressing ZmSRTF18-del (ALT2) (SEQ ID NO:3) were measured. No significant difference in plant height was observed between transgenic and null plants measured at the V5, V11, or V15 growth stage.

In summary, nontruncated ZmSRTF18 (SPL VAR1) severely delayed maize growth and development, while two different truncated proteins, ZmSRTF18-del and ZmSRTF18-del (ALT2), did not delay maize growth and development when overexpressed constitutively.

EXAMPLE 6. Increased ZmSRTF18 expression following ABA treatment.

To determine whether native ZmSRTF18 expression was influenced by hormone treatments, leaf discs of nontransgenic maize hybrid plants at late vegetative stage were floated on water, or on a solution containing ABA (abscisic acid), or ethephon (Figure 4). Relative abundance of the ZmSRTF18 transcript was determined by real time quantitative PCR using the TaqMan® and fluorescent probe method (Cao and Shockey (2012) J Agric Food Chem 60:12296). There appeared to be an initial wounding effect, because at 3 hours, expression was high even in the water control discs. However, at 6 and 24 hours, there was clearly more ZmSRTF18 transcript as a result of ABA treatment, compared with the water control. The probe region was in exon 4 of ZmSRTF18, and thus would detect transcript of all splice variants.

EXAMPLE 7. Detection of protein by western blots.

Rabbit polyclonal antibodies were used to probe a western blot of leaf protein extracts from transgenic events constitutively overexpressing the ZmSRTF18-del gene, or from the corresponding event nulls (Figure 7). The antibodies were prepared against a short peptide corresponding to the region from P303 to D320 of Figure 3. Lanes 2 through 5 had different concentrations of a standard protein obtained by expressing in E. coli the ZmSRTF18-del protein with a 20 amino acid N-terminal tag, including 6 histidines that were used for purification with a cobalt column. In four of the five events examined, the ZmSRTF18-del protein was detected in transgenic events (Lanes 6, 8, 10, and 12), but not in the corresponding event nulls (Lanes 7, 9, 11, and 13). This demonstrated that the truncated protein could accumulate, and that the truncation did not appear to destabilize the protein. One unrelated cross-reacting protein was also detected in both transgenic and nulls.

EXAMPLE 8. Improvement of Frost Tolerance by Overexpression of SRTF18.

Maize plants transformed with a construct comprising a constitutive promoter driving ZmSRTF18-del were tested for frost tolerance in a seedling assay. Seedling frost tolerance is predictive of frost tolerance at the whole plant level and through the reproductive stages of the plant, such as for example, during grain filling. In an embodiment, transgenic and control (null or wild-type) seeds are planted in 4” pot as a

matched pair in the greenhouse. Transformed lines from the construct of interest are randomized across 10 flats, with 15 pots in each flat. Completely randomized block design is used to block transgenic and null plants at pot and flat level.

Seedlings are grown to about V3 stage and then transferred to a growth chamber for cold acclimation at about 10°C for 5 hours with light and at about 4°C for 16 hours without light. After cold acclimation, the seedlings are subjected to a freezing treatment at -3°C for up to 5.5 hours. After freezing treatment and a 3-4 day recovery period at normal room temperature, the seedlings are scored for survival.

A binary logistic regression model that uses either "1" for survival or "0" for a dead plant provides logarithm of probability ratio of survived/dead. The null hypothesis is that transgenic plants have the same survival as the controls. If the transgenic plants have higher survival than controls at either the 0.05 or 0.1 level, then the null hypothesis is rejected. Results are provided in Table 2.

Table 2. S% = % of plants surviving the treatment. Rep# = number of matched pairs (transgenic and control plants) in the test. CK = control. TG+ = transgenic.

Event	2 Experiments				
	TG+ S%	CK S%	S % Diff	Rep#	P value
18.2.1	56.7	38.4	18	28	0.1775
18.2.11	57.2	45.8	11	36	0.3400
18.2.14	53.1	39.1	14	37	0.2373
18.2.19	61.2	37.4	24	39	0.0407
18.2.28	58.4	31.9	27	15	0.3853
18.2.39	65.9	51.3	15	15	0.4413
18.2.7	58.6	69.5	-11	37	0.3329
18.3.16	82	65.9	16	15	0.3696
18.3.17	51.5	51.5	0	36	1.0000
18.3.25	44.7	41.9	3	37	0.8111
Construct	59.5	47.3	12	295	0.0115

EXAMPLE 9. ZmSRTF18 constructs to be tested.

Additional constructs to optimize ZM-SRTF18 expression for yield increase are being evaluated. Constructs with a variety of promoters driving expression of ZmSRTF18-del, including stronger or weaker constitutive promoters compared with the constitutive promoter of Example 1, a drought inducible promoter, a strong green tissue mesophyll cell promoter, root preferred promoters, and an ear preferred promoter are evaluated.

- A drought inducible promoter driving expression of ZmSRTF18 (SPL VAR1).
- A drought inducible promoter driving expression of ZmSRTF18-del.
- A constitutive promoter driving expression of ZmSRTF18-del.
- A constitutive promoter driving expression of ZmSRTF18-del (ALT2).
- A constitutive promoter driving expression of ZmSRTF18-del (ALT3).
- A constitutive promoter driving expression of ZmSRTF18-del (ALT5).
- A constitutive promoter driving expression of ZmSRTF18-del with a heterologous nuclear localization signal.
- A constitutive promoter driving expression of a pearl millet homolog with the same truncation as in ZmSRTF18-del.
- A constitutive promoter driving expression of a sorghum homolog with the same truncation as in ZmSRTF18-del.
- A strong green tissue mesophyll cell promoter driving expression of ZmSRTF18-del.
- Root preferred promoters driving expression of ZmSRTF18-del.
- An ear preferred promoter driving expression of ZmSRTF18-del.
- A construct designed to reduce expression of ZmSRTF18 by an RNAi strategy.

EXAMPLE 10. Summary of ZmSRTF18 N-terminal truncations, and their proximity to the AP2 domain.

DREB transcription factors have a highly conserved 58 amino acid DNA binding domain known as the AP2 domain (SEQ ID NO: 10 and Fig. 3). Several ZmSRTF18 N-terminal truncations of differing lengths were made. The number of deleted amino acids, and the number of amino acids retained N terminal to the AP2 domain, are summarized for the truncated proteins in Table 3. Using the ZmSRTF18 (SPL VAR4) amino acid sequence (SEQ ID NO: 6 and Fig 3) as a reference, truncated proteins with deletions ranging from 18 to 71 amino acids were made. These truncated proteins retained from 12 to 65 amino acids N-terminal to the AP2 domain. The truncated proteins that retained from 12 to 52 amino acids N-terminal to the AP2

domain all lacked strong, specific binding to the DRE promoter element in gel shift assays in the presence of 1 mM DTT, as described in Example 3. The least severely truncated protein, which retained 65 amino acids N-terminal to the AP2 domain, strongly and specifically bound the DRE promoter element, as did the ZmSRTF18 (SPL VAR4) protein. The most severely truncated protein, ZmSRTF18-del, increased hybrid maize yield as described in Example 1, and showed no deleterious effects on growth when expressed constitutively, as described in Example 5,. ZmSRTF18-del (ALT2) also showed no deleterious effects on growth in maize inbreds when expressed constitutively, as described in Example 5. Thus, the N-terminal truncations may help to avoid deleterious effects on growth, while still allowing yield increases despite the diminished promoter binding observed with purified protein in vitro. Similar N-terminal truncations may be made in other DREB transcription factors. As specific examples, we made truncations at the identical location as the ZmSRTF18-del truncation in DREB2 proteins from sorghum (SEQ ID NO: 24) and pearl millet (SEQ ID NO: 25).

Table 3. Description of ZmSRTF18 N-terminal truncations.

Name of truncated protein And SEQ ID NO.	Number of amino acids deleted, using ZmSRTF18 (SPL VAR4) as reference	Number of amino acids retained N-terminal to the AP2 domain	Additional nonnative methionine added to serve as a start methionine	Strong, specific binding to DRE promoter element in vitro
ZmSRTF18-del SEQ ID NO: 2	71	12	0 (methionine already present)	no
ZmSRTF18-del (ALT3) SEQ ID NO: 4	69	14	1	no
ZmSRTF18-del (ALT2) SEQ ID NO: 3	55	28	1	no
ZmSRTF18-del (ALT5) SEQ ID NO: 26	31	52	1	no
ZmSRTF18-del (ALT6) SEQ ID NO: 27	18	65	1	yes
ZmSRTF18 (SPL VAR4) SEQ ID NO: 6	0	83	0 (methionine already present)	yes

WE CLAIM:

1. A method of improving abiotic stress tolerance of a plant, the method comprising transforming said plant with a construct comprising a promoter operably linked to a polynucleotide encoding a truncated DREB transcription factor.
2. The method of Claim 1, wherein the truncated DREB transcription factor lacks a functional N-terminal CBF domain.
3. The method of Claim 1, wherein the truncation affects at least one nuclear localization signal present in the DREB transcription factor prior to truncation.
4. The method of Claim 1, wherein the truncated DREB transcription factor lacks a functional N-terminal CBF domain or lacks a functional nuclear localization signal present in the DREB transcription factor prior to truncation.
5. The method of Claim 1, wherein the truncated DREB transcription factor lacks both a functional N-terminal CBF domain and at least one nuclear localization signal present in the DREB transcription factor prior to truncation.
6. The method of Claim 1, wherein the truncated DREB transcription factor lacks both a function N-terminal CBF domain and all nuclear localization signals present in the DREB transcription factor prior to truncation.
7. The method of Claim 1 wherein the polynucleotide encodes a polypeptide which is a truncation or variant of ZmSRTF18 (SEQ ID NO: 1) or ZmDREB2A (SEQ ID NO: 8).
8. The method of Claim 7 wherein the sequence of the encoded polypeptide is SEQ ID NO: 2, 3, 4, 5, 6, 19, 20, 21, 26, or 28.
9. The method of any of Claims 1-8, wherein the promoter drives constitutive expression.
10. The method of any of Claims 1-8, wherein the promoter drives tissue-preferred expression.
11. The method of any of Claims 1-8, wherein the plant produces increased seed yield, relative to a control.
12. The method of any of Claims 1-8, wherein the plant is maize, wheat, rice, or sorghum.
13. The method of Claim 11, wherein seed yield is increased under conditions of abiotic stress.
14. The method of Claim 13, wherein abiotic stress includes high salt concentrations.
15. The method of Claim 13, wherein abiotic stress includes chilling or freezing.
16. The method of Claim 13, wherein abiotic stress includes reduced water availability at or about the time of anthesis or the time of grain fill.

17. The method of Claim 13, wherein abiotic stress includes reduced nitrogen availability.

18. The method of Claim 1, wherein the truncation comprises an N-terminal deletion that retains at least 12 but fewer than 65 amino acids N-terminal to the AP2 domain.

19. The method of Claim 18, wherein the truncation retains 12, 14, 28, or 52 amino acids N-terminal to the AP2 domain.

20. The method of Claim 1, wherein the truncated DREB transcription factor has the polypeptide sequence of SEQ ID NO: 2, 3, 4, 19, 20, 21, 24, 25, 26, or 28.

21. A method of reducing pleiotropy resulting from ectopic expression of a DREB transcription factor, comprising expression of a truncated version of said transcription factor, wherein the truncation results in at least one of the following characteristics: (i) loss of a functional nuclear localization signal; (ii) loss of a functional CBF domain; (iii) altered binding properties of the transcription factor.

22. The method of Claim 21 wherein the truncated version lacks at least one nuclear localization signal which is present in the non-truncated DREB transcription factor.

23. The method of Claim 21 wherein the truncation deletes the first exon of the polypeptide.

24. The method of Claim 21 wherein the truncation deletes the first 2 exons of the polypeptide.

25. A recombinant polynucleotide encoding a truncation or variant of a DREB transcription factor from maize, wherein overexpression of said polynucleotide in a maize plant increases grain yield and the maize plant does not exhibit pleiotropic effects.

26. A recombinant polynucleotide encoding a truncation or variant of a DREB transcription factor from maize, wherein overexpression of said polynucleotide in a maize plant increases grain yield under conditions of drought or reduced nitrogen availability.

27. A method of improving abiotic stress tolerance of a plant, the method comprising transforming said plant with a construct comprising a promoter operably linked to a polynucleotide encoding a DREB transcription factor lacking a functional N-terminal CBF domain or a functional nuclear localization signal.

28. The method of Claim 1 wherein the polynucleotide encodes a polypeptide which is a truncation or variant of ZmSRTF18 (SEQ ID NO: 1) or ZmDREB2A (SEQ ID NO: 8).

29. The method of Claim 1 wherein the sequence of the encoded polypeptide is SEQ ID NO: 2, 3, 4, 5, 6, 19, 20, 21, 26, or 28.

30. The method of Claim 13, wherein the plant is maize, wheat, rice, or sorghum.

31. The method of Claim 12, wherein seed yield is increased under conditions of abiotic stress.

32. The method of Claim 12, wherein abiotic stress includes high salt concentrations.

33. The method of Claim 12, wherein abiotic stress includes chilling or freezing.

34. The method of Claim 12, wherein abiotic stress includes reduced water availability at or about the time of anthesis or the time of grain fill.

35. The method of Claim 12, wherein abiotic stress includes reduced nitrogen availability.

36. The method of Claim 2, wherein the truncation results in loss of one or more nuclear localization signals.

37. The method of Claim 1 wherein the polynucleotide is a homolog of ZmSRTF18.

38. The method of Claim 37 wherein the polynucleotide is isolated from sorghum or pearl millet.

39. The method of Claim 37 wherein the polynucleotide is at least 90% identical to SEQ ID NO: 24 or 25.

40. The method of Claim 39 wherein the sequence of the polynucleotide is SEQ ID NO: 24 or 25.

41. The method of Claim 37 wherein the polynucleotide lacks at least one nuclear localization signal present in the non-truncated homolog.

42. A method of improving stress tolerance of a maize plant, the method comprising expressing a polynucleotide encoding a polypeptide comprising a truncated version of ZmSRTF18 (SEQ ID NO: 2), where the truncation results in the loss of a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

43. A method of improving stress tolerance of a maize plant, the method comprising engineering a site-directed change in an endogenous genomic DNA encoding a polypeptide that is at least 85% identical to ZmSRTF18 (SEQ ID NO: 2), wherein the site-directed change results in the loss of a functional N-terminal CBF domain or a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

44. A plant comprising a polynucleotide encoding a polypeptide comprising a truncated version of ZmSRTF18 (SEQ ID NO: 2), where the truncation results in the loss of a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

45. A maize plant comprising a polynucleotide encoding a polypeptide comprising a truncated version of ZmSRTF18 (SEQ ID NO: 2), where the truncation results in the loss of a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

46. A seed comprising a polynucleotide encoding a polypeptide comprising a truncated version of ZmSRTF18 (SEQ ID NO: 2), where the truncation results in the loss of a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

47. A plant cell comprising a polynucleotide encoding a polypeptide comprising a truncated version of ZmSRTF18 (SEQ ID NO: 2), where the truncation results in the loss of a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

48. A plant comprising a site-directed change in an endogenous genomic DNA encoding a polypeptide that is at least 85% identical to ZmSRTF18 (SEQ ID NO: 2), wherein the site-directed change results in the loss of a functional N-terminal CBF domain or a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

49. A maize plant comprising a site-directed change in an endogenous genomic DNA encoding a polypeptide that is at least 85% identical to ZmSRTF18 (SEQ ID NO: 2), wherein the site-directed change results in the loss of a functional N-terminal CBF domain or a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

50. A seed comprising a site-directed change in an endogenous genomic DNA encoding a polypeptide that is at least 85% identical to ZmSRTF18 (SEQ ID NO: 2), wherein the site-directed change results in the loss of a functional N-terminal CBF domain or a functional nuclear localization signal of the polypeptide and

wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

51. A plant cell comprising a site-directed change in an endogenous genomic DNA encoding a polypeptide that is at least 85% identical to ZmSRTF18 (SEQ ID NO: 2), wherein the site-directed change results in the loss of a functional N-terminal CBF domain or a functional nuclear localization signal of the polypeptide and wherein the maize plant does not exhibit significant pleiotropic phenotype when grown in normal or drought conditions.

52. The method of Claim 21, wherein pleiotropy is exhibited as one or more characteristics selected from the group consisting of poor emergence, stunted growth, delayed flowering, asynchronous flowering, reduced photosynthetic rate, elongated internodes, compressed internodes, altered tillering, and reduced root growth, relative to a control.

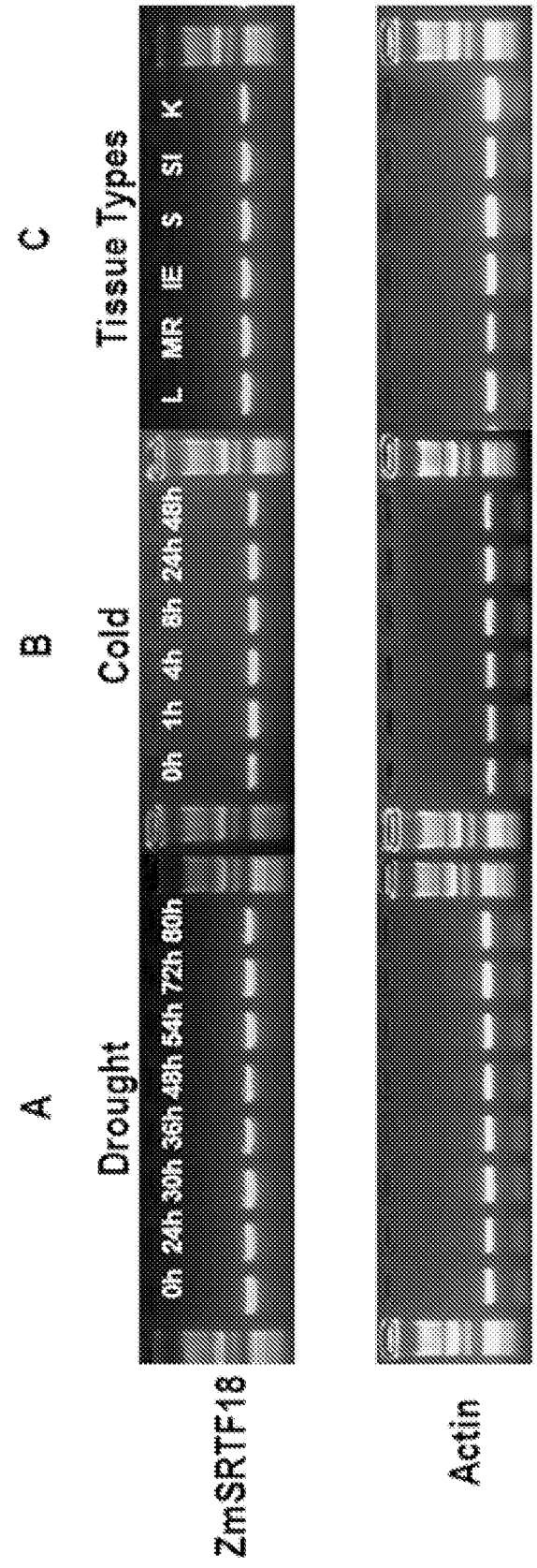

A
Drought
0h 24h 30h 36h 48h 54h 72h 80h
B
Cold
0h 1h 4h 8h 24h 48h
C
Tissue Types
L MR IE S SI K
ZmSRTF18
Actin

FIG. 1

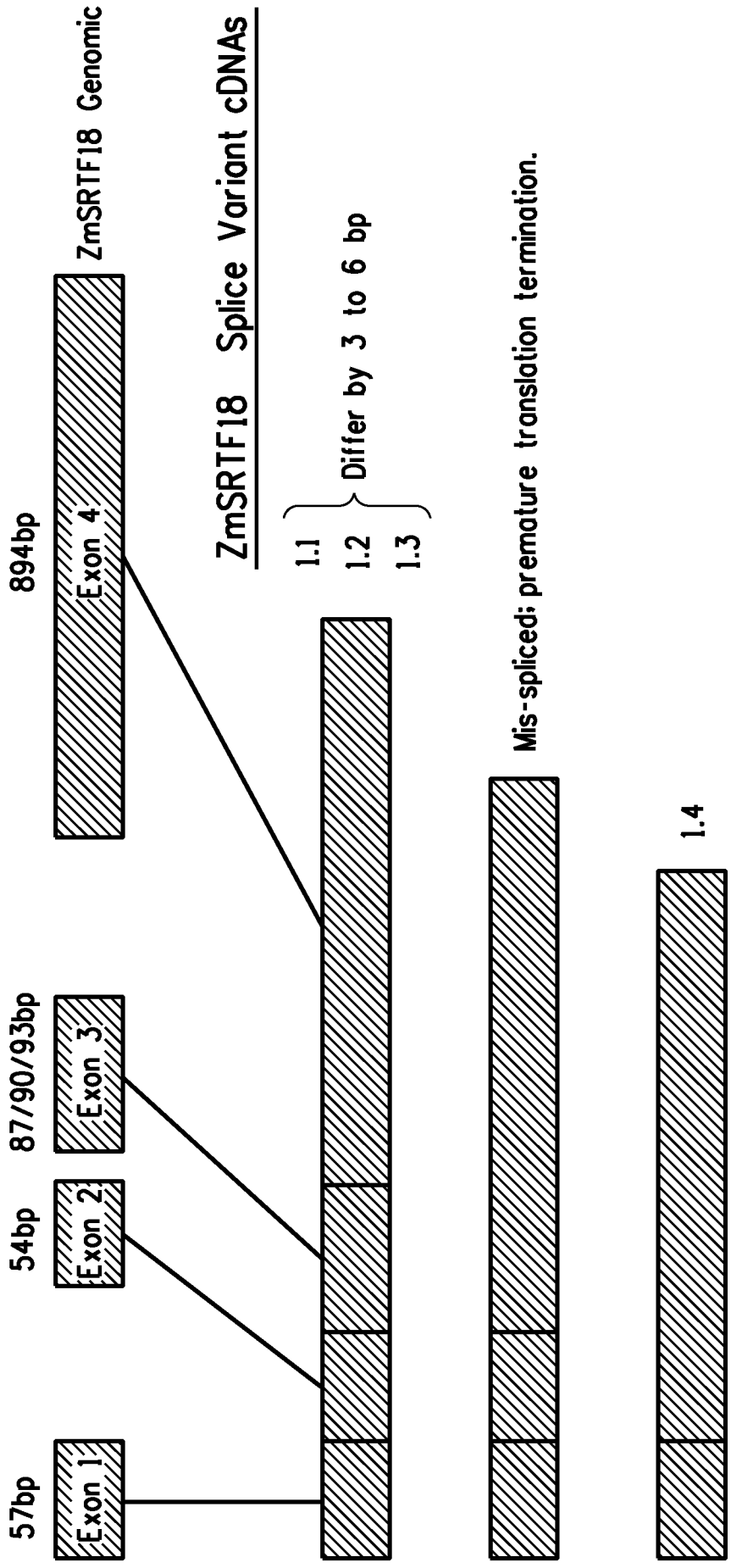

ZmSRTF18 Genomic
894bp
Exon 4
57bp
Exon 1
54bp
Exon 2
87/90/93bp
Exon 3
ZmSRTF18 Splice Variant cDNAs
1.1
1.2
1.3
Differ by 3 to 6 bp
Mis-spliced: premature translation termination.
1.4
Splice Variants of ZmSRTF18

FIG. 2

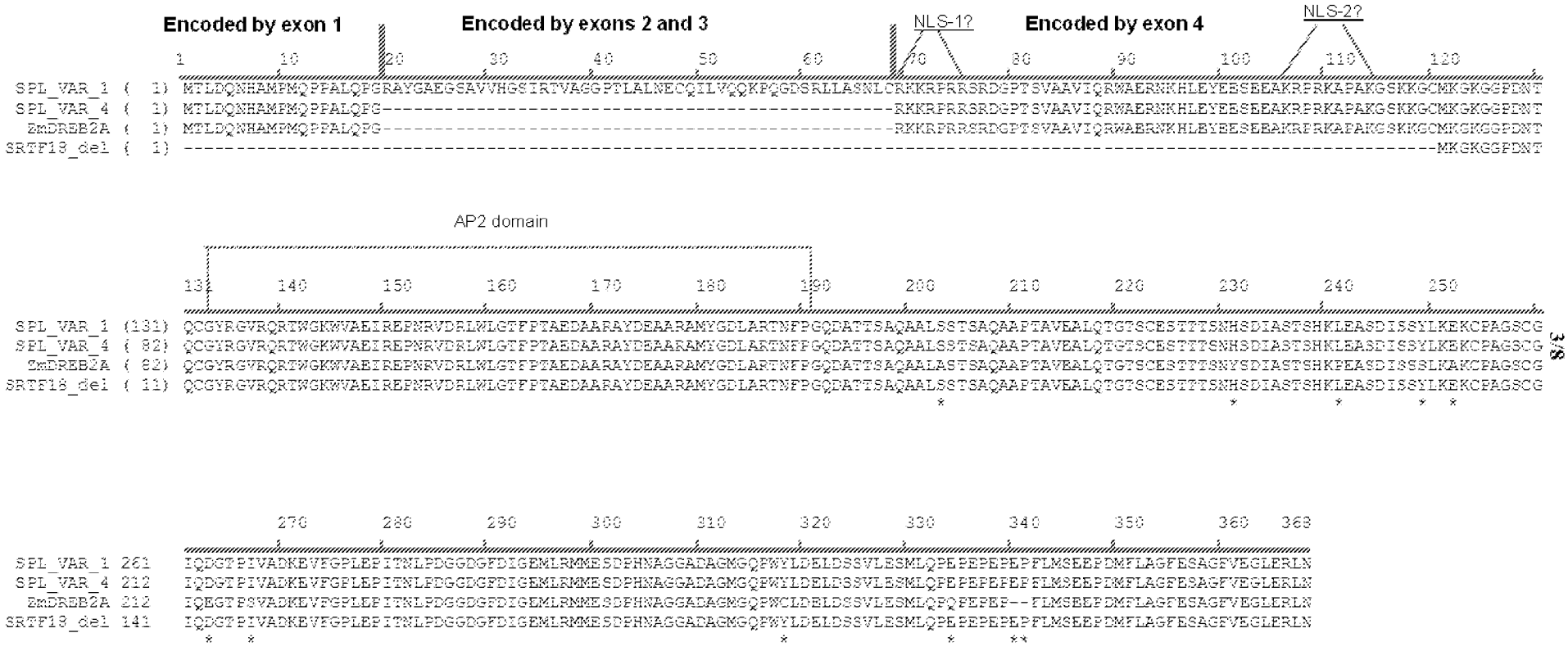

Encoded by exon 1
Encoded by exons 2 and 3
NLS-1?
Encoded by exon 4
NLS-2?
AP2 domain
SPL_VAR_1 (1) MTLDQNHAMPMQPPALQPGRAYGAEGSAVVHGSIRTVAGGPTLALNECQILVQQKPQGDSRLLASNLCRKKRPRRSRDGPTSVAAVIQRWAERNKHLEYEESEEAKRPRKAPAKGSKKGCMKGKGGPDNT
SPL_VAR_4 (1) MTLDQNHAMPMQPPALQPG----------RKKRPRRSRDGPTSVAAVIQRWAERNKHLEYEESEEAKRPRKAPAKGSKKGCMKGKGGPDNT
ZmDREB2A (1) MTLDQNHAMPMQPPALQPG----------RKKRPRRSRDGPTSVAAVIQRWAERNKHLEYEESEEAKRPRKAPAKGSKKGCMKGKGGPDNT
SRTF18_del (1) ----------MKGKGGPDNT
SPL_VAR_1 (131) QCGYRGVRQRTWGKWVAEIREPNRVDRLWLGTFPTAEDAARAYDEAARAMYGDLARTNFPGQDATTSAQAALSSTSAQAAPTAVEALQTGTSCESTTTSNHSDIASTSHKLEASDISSYLKEKCPAGSCG
SPL_VAR_4 (82) QCGYRGVRQRTWGKWVAEIREPNRVDRLWLGTFPTAEDAARAYDEAARAMYGDLARTNFPGQDATTSAQAALSSTSAQAAPTAVEALQTGTSCESTTTSNHSDIASTSHKLEASDISSYLKEKCPAGSCG
ZmDREB2A (82) QCGYRGVRQRTWGKWVAEIREPNRVDRLWLGTFPTAEDAARAYDEAARAMYGDLARTNFPGQDATTSAQAALASTSAQAAPTAVEALQTGTSCESTTTSNYSDIASTSHKFEASDISSSLKAKCPAGSCG
SRTF18_del (11) QCGYRGVRQRTWGKWVAEIREPNRVDRLWLGTFPTAEDAARAYDEAARAMYGDLARTNFPGQDATTSAQAALSSTSAQAAPTAVEALQTGTSCESTTTSNHSDIASTSHKLEASDISSYLKEKCPAGSCG
SPL_VAR_1 261 IQDGTPIVADKEVFGPLEPITNLPDGGDGFDIGEMLRMMESDPHNAGGADAGMGQPWYLDELDSSVLESMLQPEPEPEPEPFLMSEEPDMFLAGFESAGFVEGLERLN
SPL_VAR_4 212 IQDGTPIVADKEVFGPLEPITNLPDGGDGFDIGEMLRMMESDPHNAGGADAGMGQPWYLDELDSSVLESMLQPEPEPEPEPFLMSEEPIMFLAGFESAGFVEGLERLN
ZmDREB2A 212 IQEGTPSVADKEVFGPLEPITNLPDGGDGFDIGEMLRMMESDPHNAGGADAGMGQPWCLDELDSSVLESMLQPQPEPEP--FLMSEEPDMFLAGFESAGFVEGLERLN
SRTF18_del 141 IQDGTPIVADKEVFGPLEPITNLPDGGDGFDIGEMLRMMESDPHNAGGADAGMGQPWYLDELDSSVLESMLQPEPEPEPEPFLMSEEPDMFLAGFESAGFVEGLERLN

FIG 3

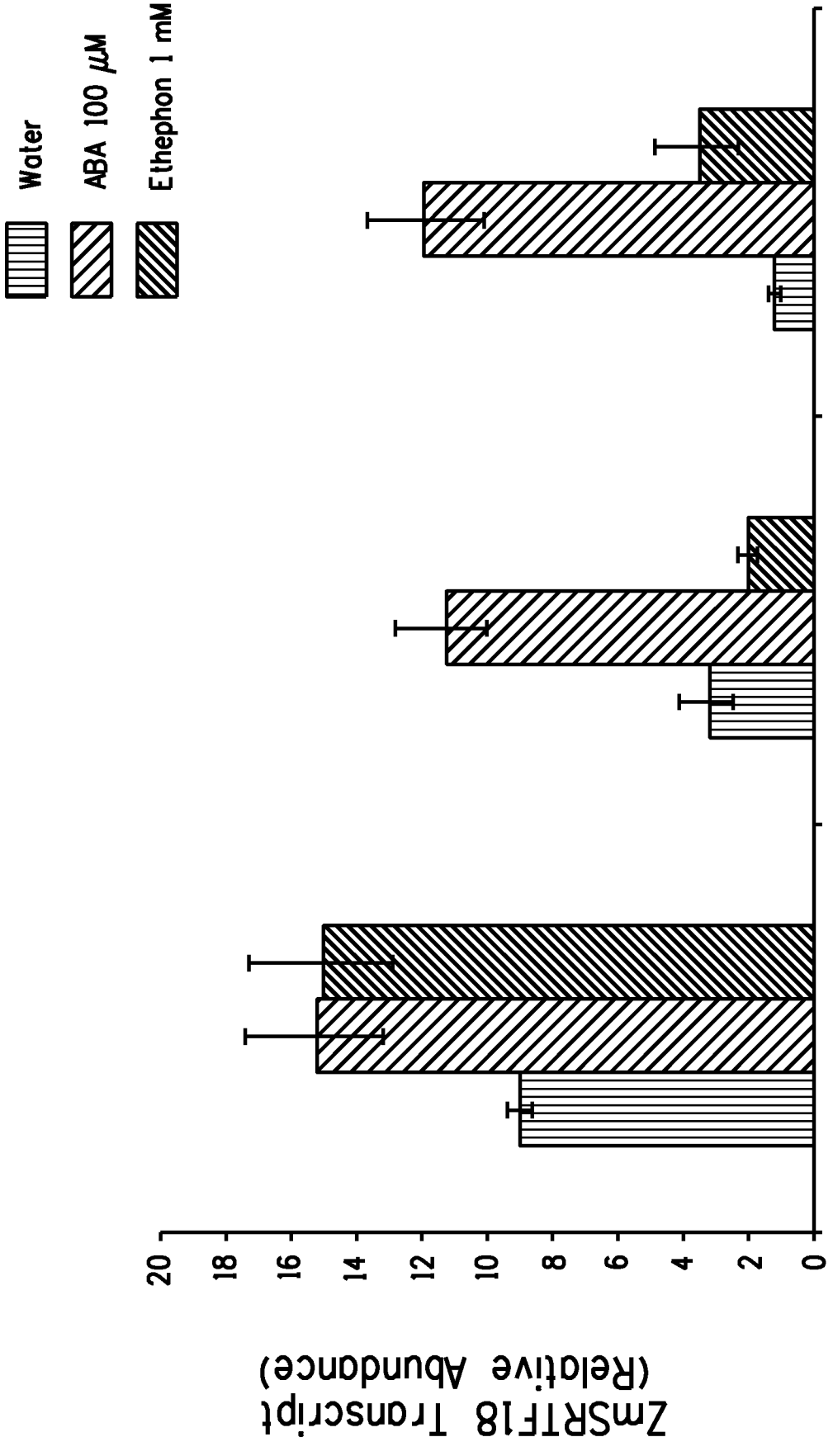

Water
ABA 100 μM
Ethephon 1 mM
ZmSRTF18 Transcript
(Relative Abundance)
0
2
4
6
8
10
12
14
16
18
20
3
6
24
Hours

FIG. 4

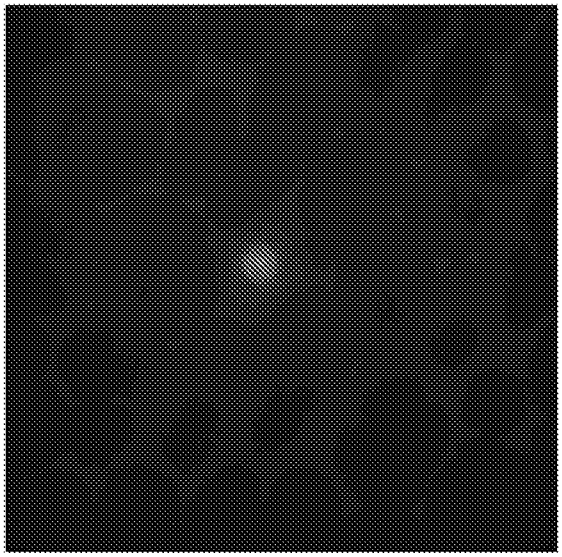


FIG. 5C

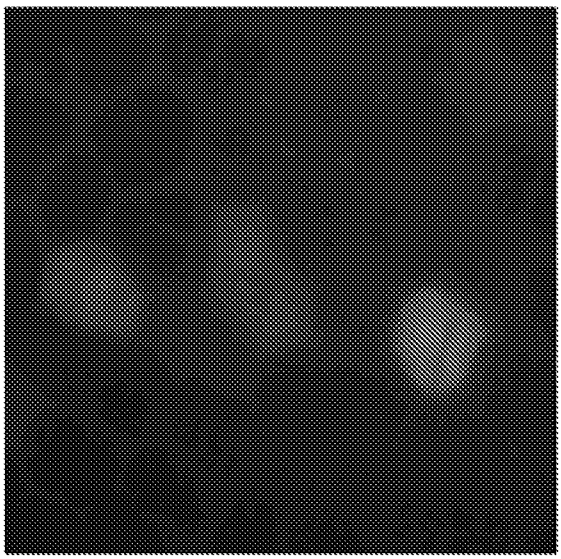


FIG. 5B

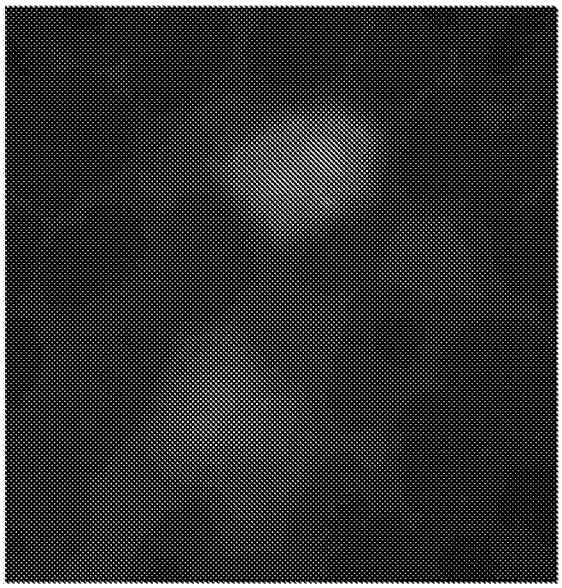


FIG. 5A

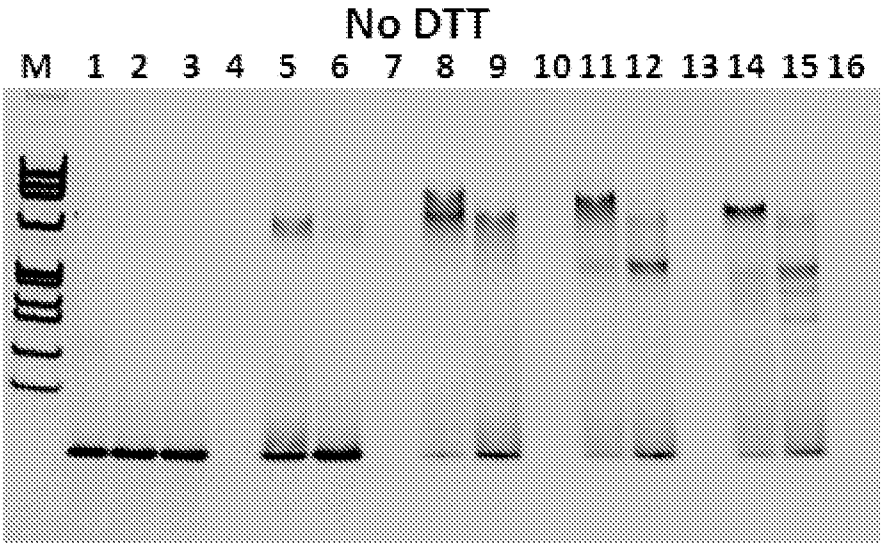

No DTT
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

FIG. 6A

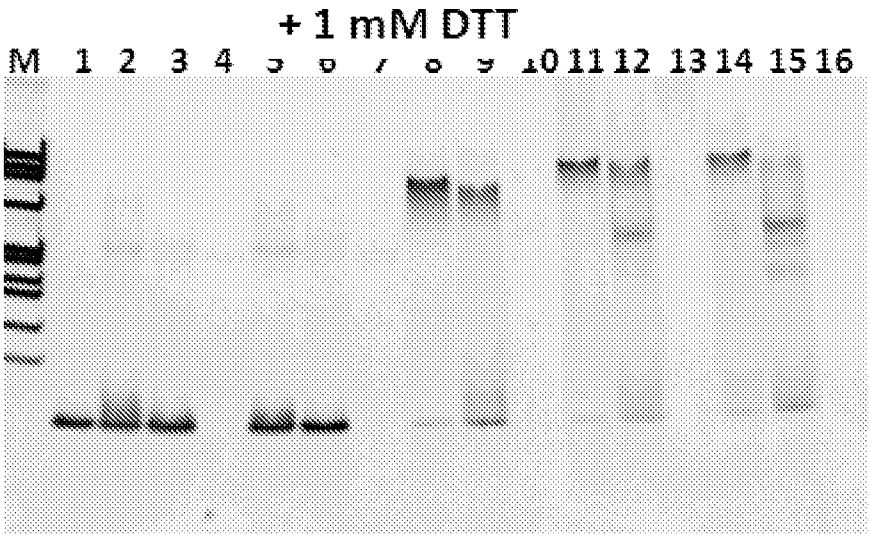

+ 1 mM DTT
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

FIG. 6B

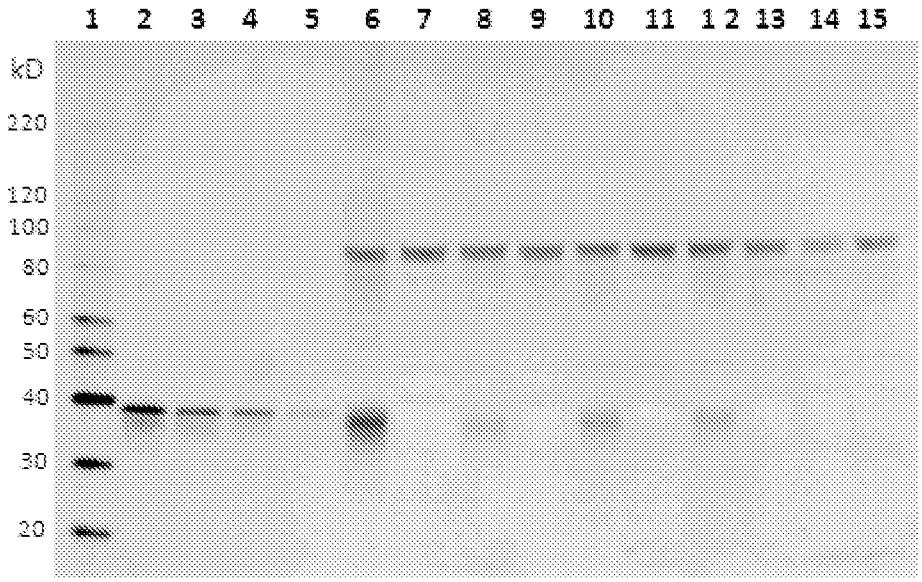


FIG. 7

												# native amino acids 5' to AP2 domain:
SEQ2_ZmSRTF18-del	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	MKGKGGPDNT	QCGYRGVRQ				12
SEQ24_SbDREB2A-del	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	MRGKGGPDNT	QCGYRGVRQ				12
SEQ25_PgDREB2A-del	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	MKGKGGPENT	HCGYRGVRQ				12
SEQ4_ZmSRTF18-del(ALT3)	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~(M)GC	MKGKGGPDNT	QCGYRGVRQ			14
SEQ3_ZmSRTF18-del(ALT2)	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~~~~~~~~~~~~	~(M)AKRPRK	APAKGSKKGC	MKGKGGPDNT	QCGYRGVRQ				28
SEQ26_ZmSRTF18-del(ALT5)	~~~~~~~~~~~~	~~~~~~~~(M)	TSVAAVIQRW	AERNKHLEYE	ESEEAKRPRK	APAKGSKKGC	MKGKGGPDNT	QCGYRGVRQ				52
SEQ27_ZmSRTF18-del(ALT6)	~~~~(M)GRK	KRPRRSRDGP	TSVAAVIQRW	AERNKHLEYE	ESEEAKRPRK	APAKGSKKGC	MKGKGGPDNT	QCGYRGVRQ				65

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2014/025862

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/415 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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 5 May 2009 (2009-05-05), "SubName: Full=Dehydration responsive element binding protein 2 isoform b; SubName: Full=Putative AP2/EREBP transcription factor superfamily protein isoform 1; SubName: Full=Putative AP2/EREBP transcription factor superfamily protein isoform 2; SubName: Full=Putative AP2/EREBP transcription factor su", XP002726589, retrieved from EBI accession no. UNIPROT:C0P2R8 Database accession no. C0P2R8 sequence -/--	25,26, 44-51

[X] Further documents are listed in the continuation of Box C.

[X] 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	Date of mailing of the international search report
3 July 2014	17/07/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Krüger, Julia

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	& Carol Soderlund ET AL: "Sequencing, Mapping, and Analysis of 27,455 Maize Full-Length cDNAs", PLoS Genetics, vol. 5, no. 11, 20 November 2009 (2009-11-20), page e1000740, XP055066952, ISSN: 1553-7390, DOI: 10.1371/journal.pgen.1000740	
X	Y. SAKUMA ET AL: "Functional Analysis of an Arabidopsis Transcription Factor, DREB2A, Involved in Drought-Responsive Gene Expression", THE PLANT CELL ONLINE, vol. 18, no. 5, 1 May 2006 (2006-05-01), pages 1292-1309, XP055126136, ISSN: 1040-4651, DOI: 10.1105/tpc.105.035881 abstract; figures 1,3	1-3,9
X	WO 2006/006236 A1 (JAPAN INTERNAT RES CT FOR AGRI [JP]; SHINOZAKI KAZUKO [JP]; SAKUMA YO) 19 January 2006 (2006-01-19) paragraphs [0197], [0198], [0203]; claims 6, 8, 18; example 1	1-3,9
X	FENG QIN ET AL: "Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in Zea mays L", THE PLANT JOURNAL, vol. 50, no. 1, 1 April 2007 (2007-04-01), pages 54-69, XP055007577, ISSN: 0960-7412, DOI: 10.1111/j.1365-313X.2007.03034.x abstract; figures 1, 3 page 55, column 1	1-52
X	US 2009/307794 A1 (SHINOZAKI KAZUKO [JP] ET AL) 10 December 2009 (2009-12-10) examples 4-6; sequence 2	1-52
A	C. LATA ET AL: "Role of DREBs in regulation of abiotic stress responses in plants", JOURNAL OF EXPERIMENTAL BOTANY, vol. 62, no. 14, 1 October 2011 (2011-10-01), pages 4731-4748, XP055126333, ISSN: 0022-0957, DOI: 10.1093/jxb/err210 the whole document page 4738 - page 4739, column 1	1-52

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2014/025862

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2006006236 A1	19-01-2006	BR PI0414161 A	31-10-2006
		CN 1950503 A	18-04-2007
		JP 4706050 B2	22-06-2011
		JP 2008505603 A	28-02-2008
		WO 2006006236 A1	19-01-2006
US 2009307794 A1	10-12-2009	BR PI0616222 A2	14-06-2011
		CA 2620766 A1	22-03-2007
		JP 4771259 B2	14-09-2011
		US 2009307794 A1	10-12-2009
		WO 2007032111 A1	22-03-2007