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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 January 2013 (10.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/004668 A1

(51) International Patent Classification:
C12N 15/82 (2006.01)

(21) International Application Number:
PCT/EP2012/062855

(22) International Filing Date:
2 July 2012 (02.07.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/504,421 5 July 2011 (05.07.2011) US

(71) Applicant *(for all designated States except US)*: **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NUCCIO, Michael** [US/US]; 3054 East Cornwallis Road, Durham, North Carolina 27709 (US). **RICHMOND, Anthony** [US/US]; 4504 Baymar Drive, Apt. #106, Raleigh Durham, North Carolina 27612 (US).

(74) Agent: **RADKOV, Stoyan**; Syngenta International AG, WRO1004.6.22, Schwarzwaldallee 215, CH-4058 Basel (CH).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

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

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2013/004668 A1

(54) Title: ROOT PREFERRED PROMOTERS DERIVED FROM RICE AND METHODS OF USE

(57) Abstract: Highly active root-preferred expression cassettes are desirable tools for improving drought tolerance in maize, rice, and other monocots. Two rice genes with strong root-preferred expression profiles were identified by querying data from rice expression profiling experiments. The gene structure and regulatory sequence for each were used to design and build two novel root-preferred expression cassettes.

ROOT PREFERRED PROMOTERS DERIVED FROM RICE AND METHODS OF USE

FIELD OF THE INVENTION

The presently disclosed subject matter is closely related to the fields of molecular biology and biotechnology. The presently disclosed subject matter is additionally closely related to the field of plant biotechnology.

BACKGROUND

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

The identification of genes that are important for crop development is thus an ongoing effort in the agricultural community. Additional information can also be derived by analyzing the genomes of important plants. For example, the identification of transcriptional regulatory elements that can direct the expression of linked nucleotide sequences can also lead to the ability to manipulate the plant genome to express polypeptides of interest in desirable spatial and/or temporal manners.

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

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

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

SUMMARY

[0001] It is the object of the present invention to provide a isolated nucleic acid sequence capable of driving expression in a plant cell, wherein the isolated nucleic acid sequence comprises a 5'-untranslated region, a first exon, a first intron, and a portion of a second exon of a gene represented by a sequence selected from the group consisting of SEQ ID NOs: 31, 32, and 35-64.

[0002] It is also the object of the present invention to provide a nucleic acid sequence comprising SEQ ID NO: 1 or SEQ ID NO: 2, wherein the sequences are capable of driving expression in a plant cell. The plant cell can be a monocot or dicot plant cell.

[0003] It is also an object of the present invention to provide a nucleic acid sequence comprising SEQ ID NO: 1 or SEQ ID NO: 2, wherein the sequences are capable of driving expression in a plant root cell. The plant root cell may be a monocot root cell, and may further be a maize root cell.

[0004] It is a further object of the present invention to provide an isolated nucleic acid sequence, comprising SEQ ID NO: 1 and SEQ ID NO: 2, wherein the sequences are capable of driving expression of a single, two genes, or at least two genes in a plant cell.

[0005] In another aspect, the present invention also relates to an expression cassette comprising SEQ ID NO: 1 or SEQ ID NO: 2 operably linked to a nucleotide sequence of interest. In yet another aspect, the expression cassette is comprised within an expression vector.

[0006] In another aspect, the present invention relates to a method of expressing a heterologous gene, comprising constructing an expression cassette comprising SEQ ID NO: 1 or SEQ ID NO: 2, wherein the expression cassette is functional in a plant or plant cell, and creating a plant or plant cell comprising the expression cassette, wherein the heterologous gene is expressed. In yet another aspect, the plant or plant cell is a monocot, and may further be maize. In still another aspect, the invention relates to a plant, plant cell, or plant tissue or portion thereof made by the method of expressing a heterologous gene. In further yet another aspect, the invention relates to progeny of the plant, plant cell, or plant tissue or portion thereof made by the method of expressing a heterologous gene. In still yet another aspect, the invention relates to seed derived from the progeny of the plant, plant cell, or plant tissue or portion thereof made by the method of expressing a heterologous gene.

[0007] In another aspect, the present invention relates to a method of expressing a heterologous genes, comprising constructing an expression cassette comprising SEQ ID NO: 1 and SEQ ID NO: 2, wherein the expression cassette is functional in a plant or plant cell, and creating a plant or plant cell comprising the expression cassette, wherein the heterologous genes are expressed. In yet another aspect, the plant or plant cell is a monocot, and may further be maize. In still another aspect, the invention relates to a plant, plant cell, or plant tissue or portion thereof made by the method of expressing a heterologous gene or genes. In further yet another aspect, the invention relates to progeny of the plant, plant cell, or plant tissue or portion thereof made by the method of expressing a heterologous gene or genes. In still yet another aspect, the invention relates to seed derived from the progeny of the plant, plant cell, or plant tissue or portion thereof made by the method of expressing a heterologous gene or genes.

[0008] In another aspect, the present invention relates to a plant, plant cell, or plant tissue or portion thereof comprising an expression cassette comprising the promoter set forth in either SEQ ID NO: 1 or SEQ ID NO: 2. In yet another aspect, the plant, plant cell, or plant tissue or portion thereof is a monocot and further may be maize.

[0009] In yet another aspect, the present invention relates to a method of expressing a gene of interest in plant roots, comprising planting seed comprising SEQ ID NO: 1 or

SEQ ID NO: 2, growing the seed into a plant, and harvesting seed of the plant; wherein the plant expresses in the roots a the gene of interest operably linked to SEQ ID NO: 1 or SEQ ID NO: 2.

[0010] In another aspect, the present invention relates to a plant, plant cell, or plant tissue or portion thereof comprising an expression cassette comprising the promoters set forth in SEQ ID NO: 1 and SEQ ID NO: 2. In yet another aspect, the plant, plant cell, or plant tissue or portion thereof is a monocot and further may be maize.

[0011] In yet another aspect, the present invention relates to a method of expressing a gene or genes of interest in plant roots, comprising planting seed comprising SEQ ID NO: 1 and SEQ ID NO: 2, growing the seed into a plant, and harvesting seed of the plant; wherein the plant expresses in the roots a the gene of interest operably linked to SEQ ID NO: 1 and a gene of interest operably linked to SEQ ID NO: 2.

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

[0012] SEQ ID NO: 1 is the engineered promoter derived from Os1035893_at (also referred to as prOs1035893).

[0013] SEQ ID NO: 2 is the engineered promoter derived from Os1053629_at (also referred to as prOs1053629).

[0014] SEQ ID NO: 3 is the primer 629-P1.

[0015] SEQ ID NO: 4 is the primer 629-P2.

[0016] SEQ ID NO: 5 is the primer 629-P3.

[0017] SEQ ID NO: 6 is the primer 629-P4.

[0018] SEQ ID NO: 7 is the mutagenesis primer 629-m1.

[0019] SEQ ID NO: 8 is the mutagenesis primer 629-m2.

[0020] SEQ ID NO: 9 is the mutagenesis primer 629-m1.1.

[0021] SEQ ID NO: 10 is pNOV6901.

[0022] SEQ ID NO: 11 is the assembly vector 15672.

[0023] SEQ ID NO: 12 is vector 15289.

[0024] SEQ ID NO: 13 is the binary vector 15673; (comprises Os1053629-expression cassette).

[0025] SEQ ID NO: 14 is the binary vector 15706; (comprises Os1035893-expression cassette).

[0026] SEQ ID NO: 15 is the primer 893-P1.

[0027] SEQ ID NO: 16 is the primer 893-P2.

[0028] SEQ ID NO: 17 is the primer 893-P3.

[0029] SEQ ID NO: 18 is the primer 893-P4.

[0030] SEQ ID NO: 19 is the mutagenesis primer 893-m1.

[0031] SEQ ID NO: 20 is the mutagenesis primer 893-m2.

[0032] SEQ ID NO: 21 is the mutagenesis primer 893-m3.

[0033] SEQ ID NO: 22 is the mutagenesis primer 893-m4.

[0034] SEQ ID NO: 23 is the mutagenesis primer 893-m5.

[0035] SEQ ID NO: 24 is the mutagenesis primer 893-m7.

[0036] SEQ ID NO: 25 is the mutagenesis primer 893-m8.

[0037] SEQ ID NO: 26 is the mutagenesis primer 893-m9.

[0038] SEQ ID NO: 27 is the mutagenesis primer 893-m10.

[0039] SEQ ID NO: 28 is the mutagenesis primer 893-m11.

[0040] SEQ ID NO: 29 is the mutagenesis primer 893-m6.

[0041] SEQ ID NO: 30 is the assembly vector 15705.

[0042] SEQ ID NO: 31 is the consensus sequence for Os1053629_at from rice Affymetrix Chip probes.

[0043] SEQ ID NO: 32 is the consensus sequence for Os1035893_at from rice Affymetrix Chip probes.

[0044] SEQ ID NO: 33 is the engineered terminator derived from Os1035893_at (also referred to as tOs1035893).

[0045] SEQ ID NO: 34 is the engineered terminator derived from Os1053629_at (also referred to as tOs1053629).

DEFINITIONS

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

Following long-standing patent law convention, the terms “a”, “an”, and “the” refer to “one or more” when used in this application, including the claims. For example, the phrase "a cell" refers to one or more cells, including, for example, tissues and organs unless the context in which the term appears clearly excludes such an interpretation. Similarly, the phrase "at

least one", when employed herein to refer to an entity, refers to, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more of that entity, including but not limited to whole number values between 1 and 100 and greater than 100 as would be understood by one of ordinary skill in the art with respect to the context in which the phrase appears.

Unless otherwise indicated, all numbers expressing quantities of ingredients, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about". The term “about”, as used herein when referring to a measurable value such as an amount of mass, weight, time, volume, concentration or percentage is meant to encompass variations of in some embodiments ±20%, in some embodiments ±10%, in some embodiments ±5%, in some embodiments ±1%, in some embodiments ±0.5%, and in some embodiments ±0.1 % from the specified amount, as such variations are appropriate to perform the disclosed methods. Accordingly, unless indicated to the contrary, the numerical parameters set forth in this specification and attached claims are approximations that can vary depending upon the desired properties sought to be obtained by the presently disclosed subject matter.

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

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

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

As used herein, the phrase “consisting essentially of” limits the scope of the related disclosure or claim to the specified materials and/or steps, plus those that do not materially affect the basic and novel characteristic(s) of the disclosed and/or claimed subject matter. For example, a nucleic acid molecule of the presently disclosed subject matter can “consist essentially of” a promoter, a reporter gene coding sequence, and a transcriptional terminator.

It is noted, however, that additional nucleotides that are not specifically recited in the corresponding SEQ ID NOs. can also be present, provided that the additional nucleotides do not materially alter the activity of any of the promoter, the reporter gene coding sequence, and the transcriptional terminator.

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

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

[0046] "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from a target gene, including an endogenous gene, a transgene, or a gene from another organism, e.g., an insect.

[0047] "Cis-element" refers to a cis-acting transcriptional regulatory element that confers an aspect of the overall control of gene expression. A cis-element may function to bind transcription factors, trans-acting protein factors that regulate transcription. Some cis-

elements bind more than one transcription factor, and transcription factors may interact with different affinities with more than one cis-element. Cis-elements can be identified by a number of techniques, including deletion analysis, i.e., deleting one or more nucleotides from the 5′ end or internal to a promoter; DNA binding protein analysis using DNase I footprinting, methylation interference, electrophoresis mobility-shift assays, in vivo genomic footprinting by ligation-mediated PCR, and other conventional assays; or by DNA sequence similarity analysis with known cis-element motifs by conventional DNA sequence comparison methods. The fine structure of a cis-element can be further studied by mutagenesis (or substitution) of one or more nucleotides or by other conventional methods. Cis-elements can be obtained by chemical synthesis or by isolation from promoters that include such elements, and they can be synthesized with additional flanking nucleotides that contain useful restriction enzyme sites to facilitate subsequence manipulation.

[0048] "Chimeric" is used to indicate that a DNA sequence, such as a vector or a gene, is comprised of two or more DNA sequences of distinct origin that are fused together by recombinant DNA techniques resulting in a DNA sequence, which does not occur naturally.

[0049] "Chromosomally-integrated" refers to the integration of a foreign gene or DNA construct into the host DNA by covalent bonds. Where genes are not "chromosomally integrated" they may be "transiently expressed." Transient expression of a gene refers to the expression of a gene that is not integrated into the host chromosome but functions independently, either as part of an autonomously replicating plasmid or expression cassette, for example, or as part of another biological system such as a virus.

[0050] As used herein, the phrases "coding sequence" and "open reading frame" (ORF) are used interchangeably and refer to a nucleotide sequence that is transcribed into RNA, such as but not limited to mRNA, rRNA, tRNA, snRNA, sense RNA, or antisense RNA. In some embodiments, the RNA is then translated *in vivo* or *in vitro* to produce a polypeptide. In some embodiments, an ORF is a coding sequence of a reporter gene (*e.g.*, cellobiohydrolase I (CBHI); β-glucuronidase (GUS)). Thus, in some embodiments an "ORF" refers to a nucleotide sequence between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides (*i.e.*, a "codon") in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

[0051] As used herein, the term "intron" refers to an intervening section of DNA which occurs almost exclusively within eukaryotic genes, but which is not translated to amino acid sequences in the gene product as a consequence of being removed from a primary RNA transcript (sometimes referred to as a "heterogenous nuclear RNA", or "hnRNA") through the process of splicing. Splicing removes introns, thereby connecting exons into a final messenger RNA (mRNA) form than can be translated into a polypeptide. For purposes of the presently disclosed subject matter, the definition of the term "intron" includes modifications to the nucleotide sequence of an intron derived from a target gene, provided the modified intron does not significantly reduce the activity of its associated 5' regulatory sequence.

[0052] "Exon" refers to a section of DNA which carries the coding sequence for a protein or part of it. Exons are separated by intervening, non- coding sequences (introns). For purposes of the present invention, the definition of the term "exon" includes modifications to the nucleotide sequence of an exon derived from a target gene, provided the modified exon does not significantly reduce the activity of its associated 5' regulatory sequence.

[0053] "Constitutive promoter" refers to a promoter that is able to express the gene that it controls in all or nearly all of the plant tissues during all or nearly all developmental-stages of the plant, thereby generating "constitutive expression" of the gene.

[0054] "Co-suppression" and "sense suppression" refer to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially identical transgene or endogenous genes.

[0055] "Contiguous" is used herein to mean nucleic acid sequences that are immediately preceding or following one another.

[0056] "Expression" refers to the transcription and stable accumulation of mRNA. Expression may also refer to the production of protein.

[0057] "Expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette

comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components.

[0058] As used herein, the term "fragment" refers to a sequence that comprises a subset of another sequence. When used in the context of a nucleic acid or amino acid sequence, the terms "fragment" and "subsequence" are used interchangeably. A fragment of a nucleotide sequence can be any number of nucleotides that is less than that found in another nucleotide sequence, and thus includes, but is not limited to, the sequences of an exon or intron, a promoter, an enhancer, an origin of replication, a 5' or 3' untranslated region, a coding region, and a polypeptide binding domain. It is understood that a fragment or subsequence can also comprise less than the entirety of a nucleotide sequence, for example, a portion of an exon or intron, promoter, enhancer, etc. Similarly, a fragment or subsequence of an amino acid sequence can be any number of residues that is less than that found in a naturally occurring polypeptide, and thus includes, but is not limited to, domains, features, repeats, etc. Also similarly, it is understood that a fragment or subsequence of an amino acid sequence need not comprise the entirety of the amino acid sequence of the domain, feature, repeat, etc. A fragment can also be a "functional fragment", in which the fragment retains a specific biological function of the nucleotide sequence or amino acid sequence of interest. For example, a functional fragment of a transcription factor can include, but is not limited to, a DNA binding domain, a transactivating domain, or both. Similarly, a functional fragment of a receptor tyrosine kinase can include, but is not limited to, a ligand binding domain, a kinase domain, an ATP binding domain, and combinations thereof.

[0059] The "expression pattern" of a promoter (with or without an enhancer) is the pattern of expression level that shows where in the plant and in what developmental stage the promoter initiates transcription. Expression patterns of a set of promoters are said to be complementary when the expression pattern of one promoter shows little overlap with the expression pattern of the other promoter.

[0060] "Gene" refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. The term "Native gene" refers to a gene as found in nature. The term "chimeric gene" refers to any gene that contains 1) DNA sequences, including regulatory and coding sequences, that are not found together in nature, or 2) sequences encoding parts of proteins not naturally adjoined, or 3) parts of promoters that are not naturally adjoined. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or

comprise regulatory sequences and coding sequences derived from the same source, but arranged in a manner different from that found in nature. A "transgene" refers to a gene that has been introduced into the genome by transformation and is stably maintained. Transgenes may include, for example, genes that are either heterologous or homologous to the genes of a particular plant to be transformed. Additionally, transgenes may comprise native genes inserted into a non-native organism, or chimeric genes. The term "endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism but one that is introduced into the organism by gene transfer. Furthermore, a "gene" is a defined region that is located within a genome and that, besides the aforementioned coding nucleic acid sequence, comprises other, primarily regulatory, nucleic acid sequences responsible for the control of the expression, that is to say the transcription and translation, of the coding portion. A gene may also comprise other 5′ and 3′ untranslated sequences and termination sequences. Further elements that may be present are, for example, introns.

[0061] Gene silencing" refers to homology-dependent suppression of viral genes, transgenes, or endogenous nuclear genes. Gene silencing may be transcriptional, when the suppression is due to decreased transcription of the affected genes, or post-transcriptional, when the suppression is due to increased turnover (degradation) of RNA species homologous to the affected genes. (English, *et al.*, 1996, Plant Cell 8:179-1881). Gene silencing includes virus-induced gene silencing (Ruiz *et al.*, 1998, Plant Cell 10:937-946).

[0062] "Genetically stable" and "heritable" refer to chromosomally-integrated genetic elements that are stably maintained in the plant and stably inherited by progeny through successive generations.

[0063] "Heterologous DNA Sequence" is a DNA sequence not naturally associated with a host cell into which it is introduced, including non-naturally occurring multiple copies of a naturally occurring DNA sequence.

[0064] "Inducible promoter" refers to those regulated promoters that can be turned on in one or more cell types by an external stimulus, such as and by way of example only a chemical, light, hormone, stress, or a pathogen.

[0065] "Insecticidal" is defined as a toxic biological activity capable of controlling insects, preferably by killing them.

[0066] "5' non-coding sequence" refers to a nucleotide sequence located 5' (upstream) to the coding sequence. It is present in the fully processed mRNA upstream of the initiation

codon and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. (Turner *et al.*, 1995, Molecular Biotechnology, 3:225).

[0067] "3' non-coding sequence" refers to nucleotide sequences located 3' (downstream) to a coding sequence and include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht *et al.* (1989, Plant Cell, 1:671-680).

[0068] The term "nucleic acid" refers to a polynucleotide of high molecular weight which can be single-stranded or double-stranded, composed of monomers (nucleotides) containing a sugar, phosphate and a base which is either a purine or pyrimidine. A "nucleic acid fragment" is a fraction of a given nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. A "genome" is the entire body of genetic material contained in each cell of an organism. The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. An "isolated nucleic acid fragment" refers to a polymer of ribonucleotides (RNA) or deoxyribonucleotides (DNA) that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

[0069] As used herein, the phrase "percent identical"," in the context of two nucleic acid sequences, refers to two or more sequences or subsequences that have in some embodiments at least 60% (e.g., 60, 63, 65, 67, or 69%), in some embodiments at least 70% (e.g., 70, 73, 75, 77, or 79%), in some embodiments at least 80% (e.g., 80, 83, 85, 86, 87, 88, or 89%), in some embodiments at least 90% (e.g., 90, 91, 92, 93, 94, 95, 96, 97, or 98%), and in some embodiments at least 99% nucleotide identity, respectively, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. The percent identity exists in some embodiments over a region of the sequences that is at least about 50 nucleotides in length, in some embodiments over a region of at least about 100 nucleotides in length, in some embodiments over a region of at least about 250 nucleotides in length, in some embodiments over a region of at least about 500

nucleotides in length, in some embodiments over a region of at least about 1000 nucleotides in length, and in some embodiments, the percent identity exists over at least about 1500 residues. In some embodiments, the percent identity exists over the entire length of one or both of the sequences (e.g., any of SEQ ID NOs: 1-2).

[0070] The terms "open reading frame" and "ORF" refer to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides ('codon') in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

[0071] "Operably-linked" and "Operatively-linked" refer to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably-linked with a coding sequence or functional RNA when it is capable of affecting the expression of that coding sequence or functional RNA (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences in sense or antisense orientation can be operably-linked to regulatory sequences.

[0072] "Overexpression" refers to the level of expression in transgenic organisms that exceeds levels of expression in normal or untransformed organisms.

[0073] "Plant tissue" includes differentiated and undifferentiated tissues or plants, including but not limited to roots, stems, shoots, leaves, pollen, seeds, tumor tissue and various forms of cells and culture such as single cells, protoplast, embryos, and callus tissue. The plant tissue may be in plants or in organ, tissue or cell culture.

[0074] "Preferred expression" is the expression of gene products that are preferably expressed at a higher level in one or a few plant tissues (spatial limitation) and/or to one or a few plant developmental stages (temporal limitation) while in other tissues/developmental stages there is a relatively low level of expression.

[0075] "Primary transformant" and "T0 generation" refer to transgenic plants that are of the same genetic generation as the tissue that was initially transformed (i.e., not having gone through meiosis and fertilization since transformation). "Secondary transformants" and the "T1, T2, T3, etc. generations" refer to transgenic plants derived from primary transformants through one or more meiotic and fertilization cycles. They may be derived by self-fertilization of primary or secondary transformants or crosses of primary or secondary transformants with other transformed or untransformed plants.

[0076] The terms "protein," "peptide" and "polypeptide" are used interchangeably herein.

[0077] "Promoter" refers to a nucleotide sequence, usually upstream (5′) to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. "Promoter" includes a minimal promoter that is a short DNA sequence comprised of a TATA-box and other sequences that serve to specify the site of transcription initiation, to which regulatory elements are added for control of expression. "Promoter" also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements that is capable of controlling the expression of a coding sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also contain DNA segments that are involved in the binding of protein factors that control the effectiveness of transcription initiation in response to physiological or environmental conditions. "Promoter regulatory sequences" or "regulatory sequences" can comprise proximal and more distal upstream elements and/or downstream elements. Promoter regulatory sequences influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence. Promoter regulatory sequences include enhancers, untranslated leader sequences, introns, exons, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences that can be a combination of synthetic and natural sequences. An "enhancer" is a nucleotide sequence that can stimulate promoter activity and can be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. The primary sequence can be present on either strand of a double-stranded DNA molecule, and is capable of functioning even when placed either upstream or downstream from the promoter. The meaning of the term "promoter" includes "promoter regulatory sequences." Exemplary promoters of the presently disclosed subject matter include nucleotide sequences that are in some embodiments at least 70%, in some embodiments at least 80%, in some embodiments at least 90%, in some embodiments at least 91%, in some embodiments at least 92%, in some embodiments at least 93%, in some embodiments at least 94%, in some embodiments at least 95%, in some embodiments at least 96%, in some embodiments at least 97%, in some embodiments at least 98%, in some embodiments at least 99% identical, or that are 100% identical to any of SEQ ID NOs: 1-2. In some embodiments, the percent identity is

calculated over 100 nucleotides, 200 nucleotides, 300 nucleotides, 500 nucleotides, 750 nucleotides, 1000 nucleotides, 1500 nucleotides, or the full length of any of SEQ ID NOs: 1-2.

[0078] "Reference sequence" as used herein is defined as a sequence that is 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 fragment of a full-length cDNA or gene sequence, or the full-length cDNA or gene sequence.

[0079] "Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and include both tissue-specific and inducible promoters. It includes natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Inducible promoters are regulated by external stimuli.

[0080] The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived by posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA. A "functional RNA" refers to an antisense RNA, ribozyme, or other RNA that is not translated, but participates in a reaction or process as a RNA.

[0081] A "selectable marker gene" refers to a gene whose expression in a plant cell gives the cell a selective advantage. The selective advantage possessed by the cells transformed with the selectable marker gene may be due to their ability to grow in presence of a negative selective agent, such as an antibiotic or a herbicide, compared to the ability to grow of non-transformed cells. The selective advantage possessed by the transformed cells may also be due to their enhanced capacity, relative to non-transformed cells, to utilize an added compound as a nutrient, growth factor or energy source. A selective advantage possessed by a transformed cell may also be due to the loss of a previously possessed gene in what is called "negative selection". In the negative selection example, a compound is added that is toxic only to cells that did not lose a specific gene (a negative selectable marker gene) present in the parent cell (typically a transgene).

[0082] "Specific expression" is the expression of gene products that is limited to one or a few plant tissues (spatial limitation) and/or to one or a few plant developmental stages (temporal limitation).

[0083] Substantially identical: the phrase "substantially identical," in the context of two nucleic acid or protein sequences, refers to two or more sequences or subsequences that have at least 60%, preferably 80%, more preferably 90, even more preferably 95%, and most preferably at least 99% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In an especially preferred embodiment, the sequences are substantially identical over the entire length of the coding regions. Furthermore, substantially identical nucleic acid or protein sequences perform substantially the same function.

[0084] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters. 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.

[0085] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, 1981, Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman & Wunsch, 1970, J. Mol. Biol. 48: 443, by the search for similarity method of Pearson & Lipman, 1988, Proc. Nat'l. Acad. Sci. 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally, Ausubel *et al., infra)*.

[0086] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul

et al., 1990, J. Mol. Biol. 215: 403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. The BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix *(see* Henikoff & Henikoff, 1989, Proc. Natl. Acad. Sci. 89: 10915).

[0087] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences *(see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0088] For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default

parameters or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

[0089] Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent hybridization conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.,* total cellular) of DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

[0090] "Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *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. Generally, high stringency hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Typically, under high stringency conditions a probe will hybridize to its target subsequence, but to no other sequences.

[0091] The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very high stringency conditions are selected to be equal to the T_m for a particular probe. An example of high stringency hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of very high stringency wash conditions is 0.1 5M NaCl at 72°C for about 15 minutes. An example of high stringency wash conditions is a 0.2x SSC

wash at 65°C for 15 minutes (*see,* Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, *e.g.,* more than 100 nucleotides, is 4-6x SSC at 40°C for 15 minutes. For short probes (*e.g.,* about 10 to 50 nucleotides), high stringency conditions typically involve salt concentrations of less than about 1.0 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 typically at least about 30°C. High stringency conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under high stringency conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

[0092] 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, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 0% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0. 1 X SSC at 60 to 65°C.

[0093] The following are examples of sets of hybridization/wash conditions that may be used to clone homologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium

dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

[0094] 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 Anal. Biochem. 138:267-284 (1984); TM 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 TM is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T 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 >90% identity are sought, the T_m can be decreased 10°C. Generally, high stringency conditions are selected to be about 19°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, very high stringency 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, 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 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 (1993) Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, Part 1, Chapter 2 (Elsevier, New York); and Ausubel et al., eds. (1995) Current Protocols in Molecular Biology, Chapter 2 (Greene Publishing and Wiley - Interscience, New York). See Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2^{nd} ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

[0095] A further indication that two nucleic acid sequences or proteins are substantially identical is that the protein encoded by the first nucleic acid is immunologically cross reactive with, or specifically binds to, the protein encoded by the second nucleic acid. Thus, a protein is typically substantially identical to a second protein, for example, where the two proteins differ only by conservative substitutions.

[0096] "Tissue-specific promoter" refers to regulated promoters that are not expressed in all plant cells but only in one or more cell types in specific organs (such as leaves, roots or seeds), specific tissues (such as embryo or cotyledon), or specific cell types (such as leaf parenchyma or seed storage cells). These also include promoters that are temporally regulated, such as in early or late embryogenesis, during fruit ripening in developing seeds or fruit, in fully differentiated leaf, or at the onset of senescence.

[0097] "Transactivating gene" refers to a gene encoding a transactivating protein. It can encode a transcription factor. It can be a natural gene, for example, a plant transcriptional activator, or a chimeric gene, for example, when plant regulatory sequences are operably-linked to the open reading frame of a transcription factor from another organism. "Transactivating genes" may be chromosomally integrated or transiently expressed. "*Trans*-activation" refers to switching on of gene by the expression of another (regulatory) gene in *trans*.

[0098] A "transcriptional cassette" will comprise in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, or may be derived from another source.

[0099] The "transcription initiation site" is the position surrounding the first nucleotide that is part of the transcribed sequence, which is also defined as position +1. With respect to this site all other sequences of the gene and its controlling regions are numbered. Downstream sequences (i.e. further protein encoding sequences in the 3' direction) are denominated positive, while upstream sequences (mostly of the controlling regions in the 5' direction) are denominated negative. Furthermore, the "transcription initiation site" is the first nucleotide of a transcribed DNA sequence where RNA polymerase (DNA-Directed RNA Polymerase) begins synthesizing the RNA transcript. The "transcription initiation site" is also referred to as the "transcription start site" or TSS.

[00100] The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. "Transiently

transformed" refers to cells in which transgenes and foreign DNA have been introduced (for example, by such methods as *Agrobacterium*-mediated transformation or biolistic bombardment), but not selected for stable maintenance. "Stably transformed" refers to cells that have been selected and regenerated on a selection media following transformation.

[00101] "Transformed / transgenic / recombinant" refer to a host organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A "non-transformed", "non-transgenic", or "non-recombinant" host refers to a wild-type organism, e.g., a bacterium or plant, which does not contain the heterologous nucleic acid molecule.

[00102] The term "translational enhancer sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translational enhancer sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

[00103] "Vector" is defined to include, inter alia, any plasmid, cosmid, phage or *Agrobacterium* binary vector in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g. autonomous replicating plasmid with an origin of replication). Specifically included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g. higher plant, mammalian, yeast or fungal cells).

[00104] "Visible marker" refers to a gene whose expression does not confer an advantage to a transformed cell but can be made detectable or visible. Examples of visible markers include but are not limited to β-glucuronidase (GUS), luciferase (LUC) and green fluorescent protein (GFP).

[00105] "Wild-type" refers to the normal gene, virus, or organism found in nature without any known mutation.

DETAILED DESCRIPTION

[00106] The promoter nucleotide sequences and methods disclosed herein are useful in regulating expression of any heterologous nucleic acid sequences in a host plant in order to alter the phenotype of a plant.

[00107] Various changes in phenotype are of interest including, but not limited to, modifying the fatty acid composition in a plant, altering the amino acid composition of a plant, altering a plant's pathogen defense system, altering plant response to the environment 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.

[00108] 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 characteristics and traits such as yield and heterosis increase, the choice of genes for transformation will change accordingly. Categories of transgenes, also known as heterologous genes, for example, include, but are not limited to, genes encoding important agronomic traits, insect resistance, disease resistance, herbicide resistance, sterility, grain or seed characteristics, abiotic stress tolerance and commercial products. Genes of interest include, generally, those involved in oil, starch, carbohydrate, or nutrient metabolism as well as those affecting seed size, plant development, plant growth regulation, and yield improvement. Plant development and growth regulation also refer to the development and growth regulation of various parts of a plant, such as the flower, seed, root, leaf, and shoot.

[00109] Other commercially desirable traits are genes and proteins conferring cold, heat, salt, and drought resistance.

[00110] Disease and/or insect resistance genes may encode resistance to pests that have great yield drag such as for example, anthracnose, soybean mosaic virus, soybean cyst nematode, root-knot nematode, brown leaf spot, Downy mildew, purple seed stain, seed

decay, and seedling diseases commonly caused by the fungi *Pythium* sp., *Phytophthora* sp., *Rhizoctonia* sp., *Diaporthe* sp. Bacterial blight caused by the bacterium *Pseudomonas syringae* pv. *Glycinea*. Genes conferring insect resistance include, for example, *Bacillus thuringiensis* toxic protein genes (U.S. Pat. Nos. 5,366,892; 5,747,450; 5,737,514; 5,723,756; 5,593,881; and Geiser et al (1986) Gene 48:109); lectins (Van Damme et al. (1994) Plant Mol. Biol. 24:825); vegetative insecticidal proteins (VIP3C, U.S. Pat. No. 7,378,493); and the like.

[00111] 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). The ALS-gene mutants encode resistance to the herbicide chlorosulfuron. Glyphosate acetyl transferase (GAT) is an N-acetyltransferase from Bacillus licheniformis that was optimized by gene shuffling for acetylation of the broad spectrum herbicide, glyphosate, forming the basis of a novel mechanism of glyphosate tolerance in transgenic plants (Castle et al. (2004) Science 304, 1151-1154). Other herbicide resistance traits include, but are not limited to, EPSPS (U.S. Pat. No. 6,248,076), Bar (U.S. Pat. No. 6,025,545), and HPPD (U.S. Pat. No. 7,312,379).

[00112] The present invention includes the transformation of a recipient cell with at least one advantageous transgene. Two or more transgenes can be supplied in a single transformation event using either distinct transgene-encoding vectors, or a single vector incorporating two or more gene coding sequences. Any two or more transgenes of any description, such as those conferring herbicide tolerance, insect resistance, disease (viral, bacterial, fungal, and nematode) resistance or drought resistance, oil quantity and quality, or those increasing yield or nutritional quality may be employed as desired.

[00113] The isolated promoter sequence of the present invention can be modified to provide a range of root-preferred expression levels of the heterologous nucleotide sequence. Thus, less than the entire promoter regions may be utilized and the ability to drive expression of the coding sequence retained. However, it is recognized that expression levels of the mRNA may be decreased with deletions of portions of the promoter sequences. Therefore, fragments of SEQ ID NO: 1 or 2 which are 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1 or 2, respectively, may still function as exemplified by this description.

[00114] Embraced by the present invention are also functional equivalents of the promoters of the present invention, i.e. nucleotide sequences that hybridize under stringent conditions to any one of SEQ ID NO:1 or SEQ ID NO: 2. A stringent hybridization can be performed at a temperature of 65°C, preferably 60°C and most preferably 55°C in double strength (2X) citrate buffered saline (SSC) containing 0.1 % SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SSC concentration. Such reduced concentration buffers are typically one tenth strength SSC (0.1 X SSC) containing 0.1 % SDS, preferably 0.2X SSC containing 0.1 % SSC and most preferably half strength SSC (0.5X SSC) containing 0.1 % SDS.

Example 1. Identification of root-preferred genes in rice

[00115] In order to identify rice candidate genes appropriate for expression cassette development, rice expression profiling data were analyzed for probes which demonstrated strong expression in root tissue and little or no expression in leaf tissue. Here expression refers to the normalized signal measured on an Affymetrix™ GeneChip. The query, summarized in Table 1, focused on expression in leaf and root tissue. The probes are ranked based on the ratio of the root to leaf signals. Analysis of the data identified Os1035893_at (SEQ ID NO: 32) and Os1053629_at (SEQ ID NO: 31) as target candidate probes for expression cassette development. Both showed very strong signals in root tissue and weak signals in leaf tissue (signals below 100 are considered background). The signal ratio between roots and leaf for SEQ ID NO: 32 and SEQ ID NO: 33 was 55-fold and 42-fold, respectively.

Table 1.

Relative signal strength of 15 leaf and 15 root samples under various nitrogen levels: vegetative developmental stage 28 days after planting				
SEQ ID NO:	Gene Description	Leaf	Root	Root/ Leaf
32	Motifs Copper_Blue Type-1 copper (blue) proteins	227.04	12400.87	54.62
33	No Description	300.54	12684.68	42.21

[00116] An electronic northern (see, for example, Toufighi, et al.) was generated to further explore the expression pattern of SEQ ID NO: 32 and SEQ ID NO: 33. Table 2 reports the relative expression of identified root-preferred and root specific probes. Data are sorted by Booting Stage root score. The data in Table 2 show that both SEQ ID NO: 32 and SEQ ID NO: 33 have strong or high signals in all root samples and low or weak

signals in leaf and reproductive tissue samples. SEQ ID NO: 32 also shows a strong signal in two stem samples. Taken together, the expression profiling data identify the genes represented by SEQ ID NO: 32 and SEQ ID NO: 33 as ideal candidates for root-preferred expression cassette development.

[00117] Table 2.

SEQ ID NO:	germinating seedling (root)	germinating seedling	3-4 leaf arial	tillering stage (root)	tillering stage (leaf)	tillering stage (arial)	Stage panicle 1-3 cm	stage panicle 4-7 cm	Stage panicle 8-14 cm	Booting Stage panicle 15-20 cm
32	5.78	2.14	1.57	66.41	2.18	2.15	2.16	2.25	2.91	2.04
33	5.51	1.92	1.90	36.63	2.25	2.32	1.81	1.65	2.52	1.81

[00118] Table 2, continued.

SEQ ID NO:	Booting Stage root	Booting Stage leaf	Booting stage arial	panicle emergence-root	panicle emergence-stem	panicle emergence-panicle	Seed milk stage	Seed -soft dough	Seed hard dough	inflorescence- no seeds
32	82.09	1.81	1.90	98.00	2.26	2.44	2.19	2.13	1.97	5.31
33	47.11	2.00	1.65	58.92	1.67	2.00	1.76	1.57	1.73	1.77

[00119] Table 2, continued.

SEQ ID NO:	maturation stem	maturation root	maturation leaf	embryo	endosperm	seed coat	Senescence-stem	Senescence	aleurone	seed day 0 post anthesis
32	51.94	101.01	2.62	2.29	1.67	2.57	18.61	2.12	2.22	2.25
33	1.83	68.41	1.87	1.66	1.90	1.74	1.49	1.88	1.66	1.58

[00120] Table 2, continued.

SEQ ID NO:	seed day 2 post anthesis	seed day 4 post anthesis	seed day 7 post anthesis	seed day 9post anthesis	seed day 11 post anthesis	seed day 14 post anthesis	seed day 17 post anthesis	seed day 19 post anthesis
32	2.14	2.20	1.73	1.95	1.93	1.81	2.09	2.12
33	1.72	2.05	1.89	1.69	1.71	1.64	1.97	1.74

Example 2. Identification and annotation of the Os1035893_at and Os1053629_at

[00121] Expression cassette development began with the identification of cDNA and gDNA represented by Os1035893_at (SEQ ID NO: 32) and Os1053629_at (SEQ ID NO: 33). The complete transcript for each gene was assembled, then the corresponding gDNA

sequence was identified. gDNA was sought that included the transcript plus 2-3 kb of 5'- and 3'-flanking sequence. This was accomplished by querying GenBank. The Os1053629_at transcript is represented by sequence AP005707 (GenBank). The Os1053629_at gene does not contain introns.

[00122] The gDNA sequence represented by Os1035893_at was identified in GenBank (AC135561). Only the sequence in close proximity to the Os1035893_at gene was reported. The Os1035893_at gene has one intron.

[00123] Next, the translation start and stop codons for both the Os1053629_at and Os1035893_at genes were identified. This was done by identifying the largest open reading frame in each transcript. Their location helps define boundaries for the 5'- and 3'-regulatory sequence.

Example 3. Design and construction of the Os1053629_at-based expression cassettes

[00124] The Os1053629_at expression cassette was designed to capture the regulatory sequence residing outside the protein coding sequence. The 1682 base pair (bp) promoter (SEQ ID NO: 2) includes a 97 bp 5'-UTR that terminates at an *Nco*I restriction endonuclease site. The *Nco*I site also encodes the translation start codon for trait protein expression. The promoter is designed to terminate in a maize-optimized Kozak sequence. The terminator tOs1053629 (SEQ ID NO: 34) initiates with sequence immediately downstream of the translation stop codon and extends for 897 bp. It includes, in addition to an engineered *Sac*I site, 361 bp of 3'-UTR.

[00125] The Os1053629 promoter and terminus were amplified from rice gDNA template in a 50 µL EXTENSOR DNA polymerase reaction kit, supplied by ABgene®, containing 100 ng gDNA, 5 µL 10X Extensor Buffer, 2.0 µL 10 mM dNTP mix, 10.0 µL Q solution, 5 Units Extensor polymerase and 1.0 µL each of 20 µM primers 629-P1 (SEQ ID NO: 3; 5'-ctcgagggaacccttcaatatgtgaggaa-3') and 629-P2 (SEQ ID NO: 4; 5'-ataccatggtggatcctgatgattattacttagcttgg-3') for the 1.68 kb promoter or 629-P3 (SEQ ID NO: 5; 5'-gagctcgatcaatcgatcacgtcgt-3') and 629-P4 (SEQ ID NO: 6; 5'-atatcccgggcggtccggaatttagcggttgacctaggt-3') for the 0.9 kb terminator. The thermocycling program was 95°C for 2 minutes followed by 10 cycles of 95°C for 30 seconds, 50°C for 1 minute and 68°C for 4 minutes, then 30 cycles of 95°C for 30 seconds and 68°C for 5 minutes. The final extension step was 68°C for 15 minutes. The reaction products were gel-purified on 1% TAE agarose and the DNA was extracted using QIAPREP® DNA extraction method according to the manufacturer's instructions. The DNA products were

concentrated by ethanol precipitation, recovered in 10 μL d^2H_2O and cloned into a pCR-TOPO vector, provided by Invitrogen®. Both products were completely sequenced. GENEAMP™ PCR System 9700 thermal cyclers, provided by Applied Biosystems®, were used to carry out the DNA sequencing reactions according to the manufacturer's instructions. 3730xl DNA Analyzer, also provided by Applied Biosystems®, was used to read the sequenced product according to the manufacturer's instructions.

[00126] Each cloned sequence required some modification. This was done using QUIKCHANGE™ Multi Site Directed Mutagenesis kit supplied by Stratagene®. The basic protocol included 1 μL DNA template, 2.5 μL 10X reaction buffer, 1μL dNTP mix, 0.75 μL Quik solution, 1 μL QuikChange enzyme and 1μL each of 20 μM phosphorylated oligonucleotide (noted in Table 3) in a 25 μL reaction. The thermocycling program was 95°C for 5 minutes followed by 35 cycles of 95°C for 1 minute, 50°C for 1 minute and 65°C for 13 minutes. The reaction products were completely sequenced. GENEAMP™ PCR System 9700 thermal cyclers, provided by Applied Biosystems®, were used to carry out the DNA sequencing reactions according to the manufacturer's instructions. 3730xl DNA Analyzer, also provided by Applied Biosystems®, was used to read the sequenced product according to the manufacturer's instructions.

[00127] Table 3.

SEQ ID NO:	Primer Name	Sequence 3'-5'	Component
7	629-m1	gcccttctcgagggacccttcaatatgtgag	promoter
8	629-m2	agtcggctatacgaacatcacccccctc	promoter
9	629-m1.1	gaataatttagtaacatttaaaatatcaaattag	terminus

[00128] The Os1053629 promoter was ligated to pNOV6901 (SEQ ID NO: 10) as an *Xho*I/*Nco*I fragment, then the Os1053629 terminus was ligated as a *Sac*I/*Xma*I fragment to create the Os1053629 assembly vector. This became vector 15672 (SEQ ID NO: 11). The Os1053629 expression cassette, with the β-glucuronidase (GUS) coding sequence was excised from 15672 as a *San*DI/*Rsr*II fragment and ligated to the *Rsr*II site of the agrobacterium binary vector 15289 (SEQ ID NO: 12), thereby creating vector 15673 (SEQ ID NO: 13).

Example 4. Design and construction of the Os1035893_at-based expression cassettes

[00129] The Os1035893_at expression cassette was designed to capture all the regulatory sequence residing upstream and downstream of the protein coding sequence. In order to capture the intron, the protein coding sequence in the first exon was silenced by altering

the methionine codons. The 2214 bp promoter (SEQ ID NO: 1) includes a 237 bp 5'-UTR that terminates at an *Nco*I restriction endonuclease site. The *Nco*I site also encodes the translation start codon for trait protein expression. The promoter is designed to terminate in a maize-optimized Kozak sequence. The tOs1035893 terminator (SEQ ID NO: 33) initiates with sequence immediately downstream of the translation stop codon and extends for 1041 bp. It includes, in addition to an engineered *Sac*I site, 330 bp of 3'-UTR.

[00130] The Os1035893_at promoter and terminus were amplified from rice gDNA template in a 50 µL Extensor (ABgene) DNA polymerase reaction containing 100 ng gDNA, 5 µL 10X Extensor Buffer, 2.0 µL 10 mM dNTP mix, 10.0 µL Q solution, 5 Units Extensor polymerase and 1.0 µL each of 20 µM primers 893-P1 (SEQ ID NO: 15; 5'-ctcgagggacccagcgtgcctcctcctct-3') and 893-P2 (SEQ ID NO: 16; 5'-ataccatggtggatccggtttgtagttgaaaactgcatgc-3') for the 2.2 kb promoter or 893-P3 (SEQ ID NO: 17; 5'-gagctctgaaactccggcgacagatc-3') and 893-P4 (SEQ ID NO: 18; 5'-atatcccgggcggtccgccactccctaccagcactgt-3') for the 1.0 kb terminator. The thermocycling program was 95°C for 2 minutes followed by 10 cycles of 95°C for 30 seconds, 50°C for 1 minute and 68°C for 4 minutes, then 30 cycles of 95°C for 30 seconds and 68°C for 5 minutes. The final extension step was 68°C for 15 minutes. The reaction products were gel-purified on 1% TAE agarose and the DNA was extracted using QIAPREP DNA extraction method. The DNA products were concentrated by ethanol precipitation, recovered in 10 µL d^2H_2O and cloned into a pCR-TOPO vector, provided by Invitrogen®. Both products were completely sequenced. GENEAMP™ PCR System 9700 thermal cyclers, provided by Applied Biosystems®, were used to carry out the DNA sequencing reactions according to the manufacturer's instructions. 3730xl DNA Analyzer, also provided by Applied Biosystems®, was used to read the sequenced product according to the manufacturer's instructions.

[00131] Each cloned sequence required some modification. This was done using Stratagene's QuikChange Multi Site Directed Mutagenesis kit. The basic protocol included 1 µL DNA template, 2.5 µL 10X reaction buffer, 1µL dNTP mix, 0.75 µL Quik solution, 1 µL QuikChange enzyme and 1µL each of 20 µM phosphorylated oligonucleotide (noted in Table 4) in a 25 µL reaction. The thermocycling program was 95°C for 5 minutes followed by 35 cycles of 95°C for 1 minute, 50°C for 1 minute and 65°C for 13 minutes. The reaction products were completely sequenced. GENEAMP™ PCR System 9700 thermal cyclers, provided by Applied Biosystems®, were used to carry out the DNA sequencing reactions according to the manufacturer's instructions. 3730xl

DNA Analyzer, also provided by Applied Biosystems®, was used to read the sequenced product according to the manufacturer's instructions.

[00132] Table 4.

SEQ ID NO:	Primer Name	Sequence 3'-5'	Component
19	893-m1	gcatagggccatcgggaagtacagtg	promoter
20	893-m2	cctaaaattgtgcgcatggttaacttctcg	promoter
21	893-m3	cgccggctaggcagccgtccgatggaagt	promoter
22	893-m4	tggctgaaacacaagttgttggtgact	promoter
23	893-m5	tgatagctgggctgttggaaaggtcttcg	promoter
24	893-m7	atggtactgaaatcaatagatgcaagcaac	promoter
25	893-m8	gtttacaacctgcgattgaaggggttgta	promoter
26	893-m9	gtggcttcacccgatggaataattaaccac	promoter
27	893-m10	caaaatcagcacgtatgcactaacacaagc	promoter
28	893-m11	cttaaatttttttgaataagacgaaccgtca	promoter
29	893-m6	ctatgtactggagtggctgacactgg	terminus

[00133] The Os1035893_at promoter was ligated to pNOV6901 (SEQ ID NO: 10) as an *Xho*I/*Nco*I fragment, then the Os1035893_at terminus was ligated as a *Sac*I/*Xma*I fragment to create the Os1035893_at assembly vector 15705 (SEQ ID NO: 30). The Os1035893_at expression cassette (SEQ ID NO: 14) , with the β-glucuronidase (GUS) coding sequence was excised from 15705 (SEQ ID NO: 30) as a *San*DI/*Rsr*II fragment and ligated to the *Rsr*II site of the agrobacterium binary vector 15289 (SEQ ID NO: 12). This created vector 15706 (SEQ ID NO: 14).

[00134] Other plasmid vectors comprising the instant recombinant expression constructs can be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host cells. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene.

[00135] Methods for transforming monocots, primarily by use of *Agrobacterium tumefaciens*, and obtaining transgenic plants have been published for monocots (U.S. Pat. No. 6,037,522), wheat (Cheng et al., Plant Cell Rep. 15:971-980 (1997), and in particular maize (U.S. pat. No. 6,051,409). Methods for transforming dicots, primarily by use of *Agrobacterium tumefaciens*, and obtaining transgenic plants have been published, among others, for cotton (U.S. Pat. No. 5,004,863, U.S. Pat. No. 5,159,135); soybean (U.S. Pat. No. 5,569,834, U.S. Pat. No. 5,416,011); *Brassica* (U.S. Pat. No. 5,463,174); and peanut (Cheng et al., Plant Cell Rep. 15:653-657 (1996), McKently et al., Plant Cell Rep.

14:699-703 (1995)). For a review of other commonly used methods of plant transformation see Newell, C. A., Mol. Biotechnol. 16:53-65 (2000).

[00136] There are a variety of methods for the regeneration of plants from plant tissues. The particular method of regeneration will depend on the starting plant tissue and the particular plant species to be regenerated. The regeneration, development and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art (Weissbach and Weissbach, Eds.; In Methods for Plant Molecular Biology; Academic Press, Inc.: San Diego, Calif., 1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development or through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

[00137] In addition to the above discussed procedures, practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant DNA fragments and recombinant expression constructs and the screening and isolating of clones, (see for example, Sambrook, J. et al., In Molecular Cloning: A Laboratory Manual; 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1989; Maliga et al., In Methods in Plant Molecular Biology; Cold Spring Harbor Press, 1995; Birren et al., In Genome Analysis: Detecting Genes, 1; Cold Spring Harbor: New York, 1998; Birren et al., In Genome Analysis: Analyzing DNA, 2; Cold Spring Harbor: New York, 1998; Clark, Ed., In Plant Molecular Biology: A Laboratory Manual; Springer: New York, 1997).

[00138] The skilled artisan will also recognize that different independent transformation events will result in different levels of expression of the chimeric genes (Jones et al., EMBO J. 4:2411-2418 (1985); De Almeida et al., Mol. Gen. Genetics 218:78-86 (1989)). Thus, multiple events must be screened in order to obtain lines displaying the desired expression level. Such screening may be accomplished by Northern analysis of mRNA

expression or qRT-PCR, Western analysis of protein expression, or phenotypic analysis. Also of interest are seeds obtained from transformed plants displaying the desired gene expression profile.

[00139] Transformation and selection can be accomplished using methods well-known to those skilled in the art including, but not limited to, the methods described herein.

Example 5. Analysis of root-preferred promoters in T0 maize

[00140] The agrobacterium binary vectors carrying the Os1053629 (15673; SEQ ID NO: 13) and Os1035893 (15706; SEQ ID NO: 14) expression cassettes fused to the plant reporter gene β-glucuronidase (GUS) were transformed into maize. A total of 11 low-medium copy events were produced for construct 15673 and 16 events were produced for construct 15706. Some higher copy number T0 plants were analyzed for GUS expression by histochemical assay. Root and leaf tissue were collected from each plant and vacuum-infiltrated with the histochemical reagent, then incubated in darkness overnight at 37°C. The tissue was cleared with ethanol and analyzed for presence of the blue histochemical precipitate. The data are summarized in Table 5.

[00141] Table 5.

SEQ ID NO:	Event Number	Insert Copy Number	Root Tissue Staining*	Leaf Tissue Staining*
13	5	>2	medium	light
13	8	>2	medium	light
14	5	2	strong	light
14	6	2	light	absent
14	21	1	light	absent

*absent=no detectable blue precipitate, light=faint blue precipitate, medium=dark blue precipitate, strong=dark blue precipitate in tissue and staining media

[00142] The histochemical analysis shows that both the Os1053629and Os1035893 expression cassettes function in maize. Both expression cassettes show a root-preferred expression pattern.

Example 6. Analysis of root-preferred promoters in T1 maize

[00143] T1/F1 seedlings for several events comprising the Os1053629 expression cassette, as comprised in SEQ ID NO: 13 were germinated in axenic culture. Seedlings were assayed for GUS enzyme activity, GUS protein and localization of GUS activity. Data are summarized in Table 6. It's clear from the T1 data that this expression cassette is not

root-preferred. Rather, the data indicate that the SEQ ID NO: 2 promoter causes strong GUS expression in leaf tissue. On average the amount of GUS protein/GUS enzyme activity is 4-6 times higher in leaf samples compared with root samples.

[00144] Table 6. Characterization of GUS expression in select prOs1053629 (15673) T1/F1 maize events germinated in axenic culture.

Plant ID	Generation	GUS Raw Copy Number	GUS histochemical stain		GUS activity (pmol/min/mg protein)		GUS ELISA (ng/mg protein)	
			leaf	root	leaf	root	leaf	root
A1	T1	2.05	strong	strong	91212	23676	18149	2831
A-2	T1	2.25	medium	medium	39501	12224	4116	1280
A-3	T1	2.12	strong	strong	90094	18550	9272	2183
A-4	T1	2.38	strong	strong	56549	18838	3933	1527
A-5	T1	1.95	strong	strong	72003	7166	11152	1989
A-6	T1	1.59	strong	strong	43112	25265	16721	6822
A-7	T1	1.72	strong	medium	31421	12395	3218	1932
A-8	T1	1.86	medium	light	24229	2878	3811	414
B-1	T1	1.93	medium	light	24421	2500	2532	327
B-2	T1	1.91	medium	light	31531	4777	4164	1271
B-3	T1	2.4	medium	medium	43675	4841	6988	473
B-4	T1	2.24	medium	medium	23530	699	2046	114
B-5	T1	1.69	medium	light	14986	1498	4040	187
B-6	T1	1.6	medium	medium	25941	5542	4025	1027
B-7	T1	1.83	medium	light	23106	2364	3566	328
B-8	T1	1.92	medium	medium	16689	3768	2888	356
B-9	T1	2	medium	medium	13120	4242	1434	530
B-10	T1	2.1	medium	medium	24601	8096	2472	816
B-11	T1	2.07	medium	medium	32779	5745	7147	637
C-1	T1	0	absent	absent	381	57	undetectable	undetectable
C-2	T1	0	absent	absent	59	109	undetectable	undetectable
C-3	T1	0	absent	absent	64	87	undetectable	undetectable
C-4	T1	0	absent	absent	194	78	undetectable	undetectable
C-5	T1	0	absent	absent	22	39	undetectable	undetectable
C-6	T1	0	absent	absent	9	33	undetectable	undetectable
C-7	T1	0	absent	absent	126	29	undetectable	undetectable
D-1	F1	0	absent	absent	35	88	undetectable	undetectable
D-2	F1	0	absent	absent	159	199	undetectable	undetectable
D-3	F1	0	absent	absent	0	209	undetectable	undetectable
D-4	F1	0	absent	absent	12	52	undetectable	undetectable

[00145] As shown in Table 6, seed from four Os1053629-GUS events were germinated and grown in axenic culture on MS basic media for 4 weeks at temperature 26/18^{0}C (day/night) and 12 hours photoperiod. Leaf and root tissue were sampled for GUS enzyme activity (MUG assay) and ELISA assay. Data are the mean of three replicate assays. Seedings were then subject to histochemical localization of the GUS enzyme. Tissues were frozen at -20°C for 2 hours, thawed, then infiltrated with the histochemical reagent. Tissues were incubated at 37°C overnight then cleared with ethanol. The scoring is identical to that of Table 5.

[00146] To ensure the results are not due to the culture conditions several 15673 T1/F1 seedlings were germinated in the greenhouse. Seed from four Os1053629-GUS events were germinated and grown under standard greenhouse conditions for two weeks. Leaf and root tissue were sampled for GUS enzyme activity (MUG assay) and ELISA assay. Roots were carefully cleaned of soil by rinsing with tap water. Data are the mean of three replicate assays. Data are summarized in Table 7. These data support the axenic culture data. Histochemical staining (data not shown) confirm the leaf preferred expression pattern.

[00147] Table 7. Characterization of GUS expression in select prOs1053629 (15673) T1/F1 maize events germinated in the greenhouse.

Plant ID	Generation	GUS activity (pmol/min/mg protein)		GUS ELISA (ng/mg protein)	
		Leaf	Root	Leaf	Root
E-9	T1	91185	21261	2085	634
E-7	T1	66253	32434	1926	1751
E-8	T1	123038	28309	2502	724
E-49	T1	84413	64908	2608	2978
F-59	T1	91835	9000	3596	389
F-57	T1	2089	8	96	undetectable
F-69	T1	1	236	undetectable	undetectable
F-38	T1	14559	4170	473	199
G-30	T1	89	187	undetectable	undetectable
G-3	T1	undetectable	117	undetectable	undetectable
G-31	T1	24774	2861	1063	414
G-39	T1	61348	5828	2736	467
G-44	T1	91251	32254	3759	2242
G-12	T1	36427	16574	1525	361
G-8	T1	36933	17310	1683	671
G-49	T1	31	202	undetectable	undetectable

H-21	F1	undetectable	14	undetectable	undetectable
H-48	F1	undetectable	456	undetectable	undetectable
H-14	F1	663	84	undetectable	undetectable
H-45	F1	67308	10414	3175	512

[00148] T1/F1 seedlings for several events comprising the expression cassette SEQ ID NO: 14 (Os1035893) were germinated in axenic culture. Seedlings were assayed for GUS enzyme activity, GUS protein and localization of GUS activity. Data are summarized in Table 8, below. It is clear from the T1 data, below, that this expression cassette shows stronger expression in root tissue. On average the amount of GUS protein/GUS enzyme activity is 8-10 times higher in root samples compared with leaf samples. The histochemical staining (data not shown) support this conclusion.

[00149] Table 8. Characterization of GUS expression in select prOs1035893 (15706) T1/F1 maize events germinated in axenic culture.

PLANT ID	Generation	GUS histochemical stain		MUG activity (pmol/min/mg protein)		GUS ELISA (ng/mg protein)	
		leaf	root	leaf	root	leaf	root
I-1	T1	absent	absent	undetectable	undetectable	13.4	undetectable
I-2	T1	absent	absent	undetectable	undetectable	6.56	15.2
I-3	T1	light	strong	880 ± 45.1	13000 ± 45.2	317	2020
I-4	T1	absent	absent	555 ± 82.3	111 ± 24.5	123	61.7
I-5	T1	absent	medium	321 ± 40.7	16000 ± 541	88.9	973
I-6	T1	light	strong	1070 ± 103	10200 ± 16.2	276	383
I-7	T1	light	strong	3300 ± 98.9	15800 ± 38.6	840	3140
I-8	T1	light	strong	1680 ± 35.8	9750 ± 843	390	1500
I-9	T1	light	medium	791 ± 64.7	4430 ± 136	186	1480
I-10	T1	light	strong	1530 ± 42.7	15300 ± 273	343	1110
J-1	F1	absent	medium	undetectable	1800 ± 95.8	17.2	760
J-2	F1	light	medium	684 ± 8.5	7260 ± 86.5	132	1500
J-3	F1	absent	absent	undetectable	62.0 ± 40.9	undetectable	42.9
J-4	F1	absent	absent	undetectable	76.2 ± 93.4	9.35	25.4
J-5	F1	absent	absent	undetectable	undetectable	7.6	23.8
J-6	F1	light	light	357 ± 2.0	2100 ± 19.8	98.9	789
J-7	F1	absent	absent	undetectable	undetectable	undetectable	undetectable
J-8	F1	absent	absent	undetectable	undetectable	undetectable	undetectable
J-9	F1	absent	absent	undetectable	undetectable	undetectable	undetectable
J-10	F1	light	medium	638 ± 66.9	9860 ± 333	88.1	1310
J-11	F1	light	light	519 ± 40.3	4090 ± 259	68.8	604
J-12	F1	light	strong	1010 ± 33.8	16400 ± 162	138	1890
J-13	F1	light	strong	826 ± 32	25500 ± 952	91.4	3070
J-14	F1	absent	medium	undetectable	8720 ± 182	undetectable	980
J-15	F1	light	medium	839 ± 50.5	10300 ± 277	129	1390
J-16	F1	light	strong	1650 ± 83.7	8840 ± 138.0	179	1370

J-17	F1	light	strong	917 ± 81.8	25600 ± 796	138	1770
K-1	T1	light	medium	1120 ± 67.5	6930 ± 442	105	719
K-3	T1	light	light	2480 ± 122	3450 ± 322	169	464
K-4	T1	light	medium	649 ± 31.4	6590 ± 786	72.2	902
K-5	T1	light	strong	1120 ± 53.8	15200 ± 581	93.9	1130
K-6	T1	light	medium	923 ± 53.6	10200 ± 128	115	1250
L-1	F1	light	light	252 ± 32.8	2420 ± 29.7	67.3	487
L-2	F1	light	medium	0.6 ± 35	1860 ± 79.8	10.8	265

[00150] As shown in Table 8, seed from four Os1035893-GUS events were germinated and grown in axenic culture on MS basic media for 4 weeks at temperature 26/18^{0}C (day/night) and 12 hours photoperiod. Leaf and root tissue were sampled for GUS enzyme activity (MUG assay) and ELISA assay. Data are the mean of three replicate assays. Seedings were then subject to histochemical localization of the GUS enzyme. Tissues were frozen at -20°C for 2 hours, thawed, then infiltrated with the histochemical reagent. Tissues were incubated at 37°C overnight then cleared with ethanol. The scoring is identical to that of Table 5.

[00151] In view of the results provided here, the present invention relates to a nucleic acid sequence, preferably an isolated nucleic acid sequence, capable of driving expression in a plant cell, wherein the isolated nucleic acid sequence comprises a 5'-untranslated region, a first exon, a first intron, and a portion of a second exon of a gene represented by a sequence selected from the group consisting of SEQ ID NOs: 31, 32, and 35-64.

[00152] The invention also relates to a nucleic acid sequence, preferably an isolated nucleic acid sequence, selected from the group consisting of prOs1035893 SEQ ID NO: 1 and prOs1053629 SEQ ID NO: 2, wherein the nucleic acid sequence, preferably an isolated nucleic acid sequence, is capable of driving expression in a plant cell. In another embodiment, the plant cell is a plant root cell. In yet another embodiment, the plant root cell is a monocot root cell. In still yet another embodiment, the monocot root cell is a maize root cell.

[00153] The present invention also relates to an expression cassette comprising a nucleic acid sequence, preferably an isolated nucleic acid sequence, selected from the group consisting of prOs1035893 SEQ ID NO: 1 and prOs1053629 SEQ ID NO: 2, wherein the nucleic acid sequence is capable of driving expression in a plant cell, and wherein the nucleic acid sequence is operably linked to a nucleotide sequence of interest. The present invention further relates to an expression vector comprising the expression cassette above.

[00154] The present invention also provides for a method of expressing a heterologous gene comprising: (a) constructing an expression cassette comprising a promoter selected from the group comprising SEQ ID NO: 1 and SEQ ID NO: 2, wherein the expression cassette is functional in a plant, plant cell, or plant tissue; and (b) creating a plant, plant cell, or plant tissue or portion thereof comprising the expression cassette, wherein the heterologous gene is expressed. In another embodiment, the present invention also provides for the method above, wherein the plant, plant cell, or plant tissue or portion thereof is a monocot. In yet another embodiment, the plant, plant cell, or plant tissue or portion thereof is maize.

[00155] The present invention also provides for a plant, plant cell, or plant tissue or portion thereof made by the method of expressing a heterologous gene comprising: (a) constructing an expression cassette comprising a promoter selected from the group comprising SEQ ID NO: 1 and SEQ ID NO: 2, wherein the expression cassette is functional in a plant, plant cell, or plant tissue; and (b) creating a plant, plant cell, or plant tissue or portion thereof comprising the expression cassette, wherein the heterologous gene is expressed. In another embodiment, the present invention also provides for the progeny of the above plant, plant cell, or plant tissue or portion thereof. In yet another embodiment, the present invention provides for seed derived from the above progeny, wherein the seed comprise an isolated nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 and SEQ ID NO: 2.

[00156] The present invention also provides for a plant, plant cell, or plant tissue or portion thereof comprising an expression cassette comprising a promoter selected from the group comprising SEQ ID NO: 1 and SEQ ID NO: 2. In another embodiment, the plant, plant cell, or plant tissue is a monocot. In still yet another embodiment, the plant, plant cell, or plant tissue is maize.

[00157] The present invention also provides for a method of expressing a gene of interest in plant roots by: (a) planting seed comprising an isolated nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 and SEQ ID NO: 2; (b) growing seed into a plant; and (c) harvesting seed of the plant; wherein roots of the plant of step (b) express a gene of interest operably linked to the isolated nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 and SEQ ID NO: 2.

[00158] The present invention also provides for grain produced from seed comprising a expression cassette comprising a promoter selected from the group comprising SEQ ID NO: 1 and SEQ ID NO: 2.

[00159] The present invention also provides for grain produced from seed comprising a expression cassette comprising the promoters SEQ ID NO: 1 and SEQ ID NO: 2

[00160]

REFERENCES

[00161] Altschul *et al.,* 1990, J. Mol. Biol. 215: 403-410

[00162] Ausubel et al., eds. (1995) Current Protocols in Molecular Biology, Chapter 2 (Greene Publishing and Wiley - Interscience, New York)

[00163] Birren et al., In Genome Analysis: Detecting Genes, 1; Cold Spring Harbor: New York, 1998

[00164] Birren et al., In Genome Analysis: Analyzing DNA, 2; Cold Spring Harbor: New York, 1998

[00165] Castle et al. (2004) Science 304, 1151-1154

[00166] Cheng et al., Plant Cell Rep. 15:653-657 (1996)

[00167] Cheng et al., Plant Cell Rep. 15:971-980 (1997)

[00168] Clark, Ed., In Plant Molecular Biology: A Laboratory Manual; Springer: New York, 1997

[00169] English, *et al.*, 1996, Plant Cell 8:179-1881

[00170] Geiser et al (1986) Gene 48:109

[00171] Henikoff & Henikoff, 1989, Proc. Natl. Acad. Sci. 89: 10915

[00172] Ingelbrecht *et al.* 1989, Plant Cell, 1:671-680

[00173] Jones et al., EMBO J. 4:2411-2418 (1985); De Almeida et al., Mol. Gen. Genetics 218:78-86 (1989)

[00174] Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)

[00175] Maliga et al., In Methods in Plant Molecular Biology; Cold Spring Harbor Press, 1995

[00176] McKently et al., Plant Cell Rep. 14:699-703 (1995)

[00177] Meinkoth and Wahl Anal. Biochem. 138:267-284 (1984)

[00178] Needleman & Wunsch, 1970, J. Mol. Biol. 48: 443

[00179] Newell, C. A., Mol. Biotechnol. 16:53-65 (2000)

[00180] Pearson & Lipman, 1988, Proc. Nat'l. Acad. Sci. 85: 2444

[00181] Ruiz *et al.*, 1998, Plant Cell 10:937-946

[00182] Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2nd ed., Cold Spring Harbor Laboratory Press, Plainview, New York)

[00183] Smith & Waterman, 1981, Adv. Appl. Math. 2: 482

[00184] Tijssen (1993) *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

[00185] Toufighi K, Brady SM, Austin R, Ly E, Provart NJ. The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *The Plant Journal* (2005) 43: 153-163

[00186] Turner *et al.*, 1995, Molecular Biotechnology, 3:225

[00187] Van Damme et al. (1994) Plant Mol. Biol. 24:825

[00188] Weissbach and Weissbach, Eds.; In Methods for Plant Molecular Biology; Academic Press, Inc.: San Diego, Calif., 1988

[00189] U.S. Pat. No. 5,004,863

[00190] U.S. Pat. No. 5,159,135

[00191] U.S. Pat. No. 5,366,892

[00192] U.S. Pat. No. 5,416,011

[00193] U.S. Pat. No. 5,463,174

[00194] U.S. Pat. No. 5,569,834

[00195] U.S. Pat. No. 5,593,881

[00196] U.S. Pat. No. 5,723,756

[00197] U.S. Pat. No. 5,737,514

[00198] U.S. Pat. No. 5,747,450

[00199] U.S. Pat. No. 6,025,545

[00200] U.S. Pat. No. 6,037,522

[00201] U.S. pat. No. 6,051,409

[00202] U.S. Pat. No. 6,248,076

[00203] U.S. Pat. No. 7,312,379

[00204] U.S. Pat. No. 7,378,493

What is claimed is:

1. An isolated promoter comprising a nucleic acid sequence with a sequence identity to any of SEQ ID NOs: 1-2 of greater than 98% over at least 100.
2. An isolated transcription terminator comprising a nucleic acid sequence with a sequence identity to any of SEQ ID NOs: 33-34 of greater than 98% over at least 100.
3. A nucleic acid molecule comprising a promoter operably linked to a heterologous nucleotide sequence, wherein the promoter is selected from the group consisting of:

   (a) a nucleic acid sequence comprising a fragment of at least about 100 consecutive nucleotides of any of SEQ ID NOs: 1-2, wherein the fragment has promoter activity;

   (b) a nucleic acid sequence with a sequence identity of greater than about 95% to one of SEQ ID NOs: 1-2 over the full length of the one of SEQ ID NOs: 1-2, wherein the nucleic acid sequence has promoter activity; and

   (c) the nucleic acid sequence of any of SEQ ID NOs: 1-2.
4. The nucleic acid sequence of claim 1, wherein the sequence is capable of driving expression in a plant root cell.
5. The nucleic acid sequence of claim 1, wherein the sequence is capable of driving expression in a monocot root cell.
6. The nucleic acid sequence of claim 1, wherein the sequence is capable of driving expression in a maize root cell.
7. An expression cassette comprising the nucleic acid sequence of claim 1, wherein the nucleic acid sequence is operably linked to a nucleotide sequence of interest.
8. An expression vector comprising the expression cassette of claim 7.
9. A plant, comprising the promoter set forth in claim 1.
10. Grain derived from the plant of claim 9.
11. A method of expressing a heterologous gene comprising:
    a) constructing an expression cassette comprising a promoter selected from the group comprising SEQ ID NO: 1 and SEQ ID NO: 2, wherein the expression cassette is functional in a plant, plant cell, or plant tissue; and
    b) creating a plant, plant cell, or plant tissue or portion thereof comprising the expression cassette, wherein the heterologous gene is expressed.
12. The method of claim 11, wherein the plant, plant cell, or plant tissue or portion thereof is a monocot.

13. The method of claim 9, wherein the plant, plant cell, or plant tissue or portion thereof is maize.

14. A plant, plant cell, or plant tissue or portion thereof made by the method of claim 11.

15. Progeny of the plant, plant cell, or plant tissue or portion thereof, of claim 14.

16. Seed derived from the progeny of claim 15.

17. A plant, plant cell, or plant tissue or portion thereof comprising an expression cassette according to any one of claims 7 or 8.

18. The plant, plant cell, or plant tissue or portion thereof of claim 17, wherein the plant, plant cell, or plant tissue is a monocot.

19. The plant, plant cell, or plant tissue or portion thereof of claim 18, wherein the plant, plant cell, or plant tissue is maize.

20. A method of expressing a gene of interest in plant roots by

a) planting seed of claim 16;

b) growing seed into a plant; and

c) harvesting seed of the plant;

wherein roots of the plant of step (b) express the gene of interest operably linked to a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 and SEQ ID NO: 2.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2012/062855

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

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

   a. (means)

   [ ] on paper

   [X] in electronic form

   b. (time)

   [ ] in the international application as filed

   [X] together with the international application in electronic form

   [ ] subsequently to this Authority for the purpose of search

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

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2012/062855

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1, 3-20(all partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/062855

A. CLASSIFICATION OF SUBJECT MATTER
INV. 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)
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, BIOSIS, Sequence Search, EMBASE, MEDLINE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE Geneseq [Online] 7 February 2008 (2008-02-07), "Rice genomic promoter sequence SEQ ID NO: 38514.", XP002683405, retrieved from EBI accession no. GSN:AOB93579 Database accession no. AOB93579 the whole document	1,3-20
X	-& US 2007/020621 A1 (BOUKHAROV ANDREY A [US] ET AL) 25 January 2007 (2007-01-25) abstract; paragraphs [0004], [0008], [0009], [0016]-[0020], [0060], [0117]-[0121]; Examples 4 and 7; claims 1, 6, 8-13, 17, 18, 23-29, 33-36	1,3-20

-/--

[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
14 September 2012	04/12/2012

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	

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE Geneseq [Online] 2 June 2005 (2005-06-02), "Rice stress-regulated promoter SEQ ID NO:13818.", XP002683406, retrieved from EBI accession no. GSN:ACL35255 Database accession no. ACL35255 the whole document	1,3-20
X	-& WO 03/008540 A2 (SYNGENTA PARTICIPATIONS AG [CH]; KREPS JOEL [US]; BRIGGS STEVEN P [US]) 30 January 2003 (2003-01-30) claims 48 and 68; page 7, lines 22-26; page 9, lines 22-28; page 10, lines 4-6 and 14-34; page 12, lines 4-15; page 14, lines 24-28; Seq. ID No. 13818	1,3-20
A	DATABASE EMBL [Online] 24 October 2006 (2006-10-24), "Oryza sativa (indica cultivar-group) cDNA clone:OSIGCRN124L23, full insert sequence.", XP002683407, retrieved from EBI accession no. EM_PL:CT835675 Database accession no. CT835675 the whole document	1,3-20
A	US 2008/227639 A1 (WU WEI [US] ET AL) 18 September 2008 (2008-09-18) paragraphs [0011]-[0014], [0017], [0056], [0057], [0081]-[0085]; Tables 1 and 3; Examples 7 and 8	1,3-20
A	US 2006/101541 A1 (FLASINSKI STANISLAW [US] ET AL) 11 May 2006 (2006-05-11) paragraphs [0006], [0031], [0033]; Figure 1; Examples 1-4; Tables 1-4	1,3-20
A	US 2011/160444 A1 (ZHAO YU [CN] ET AL) 30 June 2011 (2011-06-30) abstract; paragraphs [0006], [0007]; Figures 3 and 7, Example 7; claims 3 and 5	1,3-20
A	EP 1 375 668 A1 (NAT INST OF AGROBIO SCIENCES [JP]) 2 January 2004 (2004-01-02) abstract; paragraphs [0007]-[0012], [0015]-[0026]; Examples 1 and 2	1,3-20

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2012/062855

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2007020621 A1	25-01-2007	US 2007020621 A1	25-01-2007
		US 2007118921 A1	24-05-2007
WO 03008540 A2	30-01-2003	AU 2002341541 A1	03-03-2003
		EP 1402042 A2	31-03-2004
		US 2006075523 A1	06-04-2006
		WO 03008540 A2	30-01-2003
US 2008227639 A1	18-09-2008	US 2008227639 A1	18-09-2008
		WO 2009021041 A2	12-02-2009
US 2006101541 A1	11-05-2006	NONE	
US 2011160444 A1	30-06-2011	CN 101323853 A	17-12-2008
		US 2011160444 A1	30-06-2011
		WO 2009006782 A1	15-01-2009
EP 1375668 A1	02-01-2004	AU 2001244567 B2	27-04-2006
		CA 2437203 A1	03-10-2002
		CN 1494594 A	05-05-2004
		DE 60122785 T2	20-09-2007
		EP 1375668 A1	02-01-2004
		JP 3890410 B2	07-03-2007
		US 2004117876 A1	17-06-2004
		WO 02077248 A1	03-10-2002

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. claims: 1, 3-20(all partially)

   Promoter comprising a nucleic acid sequence with a sequence identity to Seq. ID No. 1 of greater than 98% over at least 100 nucleotides, corresponding expression cassettes, vectors, plants and seeds; methods of use thereof

   ---

2. claims: 1, 3-20(all partially)

   Promoter comprising a nucleic acid sequence with a sequence identity to Seq. ID No. 2 of greater than 98% over at least 100 nucleotides, corresponding expression cassettes, vectors, plants and seeds; methods of use thereof

   ---

3. claim: 2

   Transcriptional terminator with a sequence identity to any of Seq. ID NOs 33-34 of greater than 98% over at least 100 nucleotides

   ---