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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 September 2012 (27.09.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/127373 A1

(51) International Patent Classification:
C12N 15/113 (2010.01) *C12N 15/82* (2006.01)
C12N 9/12 (2006.01) *A01H 1/00* (2006.01)

(21) International Application Number:
PCT/IB2012/051239

(22) International Filing Date:
15 March 2012 (15.03.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/453,970 18 March 2011 (18.03.2011) US

(71) Applicant *(for all designated States except US)*: **BASF PLANT SCIENCE COMPANY GMBH** [DE/DE]; 67056 Ludwigshafen (DE).

(71) Applicant *(for MN only)*: **BASF (CHINA) COMPANY LIMITED** [CN/CN]; 300 Jiangxinsha Road, Shanghai 200137 (CN).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KUHN, Josef Martin** [DE/DE]; Faberstrasse 5, 67063 Ludwigshafen (DE). **SCHULTHEISS, Holger** [DE/DE]; Am alten Bach 30, 67459 Böhl-Iggelheim (DE). **HARTIG, Julia Verena** [DE/DE]; Kohlplatz 23, 67433 Neustadt (DE). **LOYALL, Linda Patricia** [DE/DE]; Maxburgstrasse 24, 67117 Limburgerhof (DE). **DUWENIG, Elke** [DE/DE]; Eschkopfstr. 9, 67117 Limburgerhof (DE). **SARIC, Marc** [DE/DE]; Posthof 3, 69254 Malsch (DE). **STUIVER, Maarten Hendrik** [NL/DE]; An der Althart 31, 67433 Neustadt (DE). **OSWALD, Oliver** [DE/DE]; Nibelungenstr. 158, 64686 Lautertal (DE).

(74) Agent: **POPP, Andreas**; BASF SE, GVX/B - C006, 67056 Ludwigshafen (DE).

(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, MD, 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).

Published:

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

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

— *with sequence listing part of description (Rule 5.2(a))*

WO 2012/127373 A1

(54) **Title**: PROMOTERS FOR REGULATING EXPRESSION IN PLANTS

(57) **Abstract**: Isolated nucleic acid molecules capable of regulating expression in plants, as well as expression [cassettes, vectors and transgenic plants comprising the same are provided.

Promoters for regulating expression in plants

FIELD OF THE INVENTION

The present invention relates to isolated nucleic acid molecules capable of regulating expression in plants as well as expression cassettes, vectors and transgenic plants comprising the same.

BACKGROUND OF THE INVENTION

Manipulation of plants to alter and/or improve phenotypic characteristics (such as productivity, quality or pest resistance) requires the expression of heterologous genes in plant tissues. Such genetic manipulation relies on the availability of a means to drive and to control gene expression as required. For example, genetic manipulation relies on the availability and use of suitable promoters which are effective in plants and which regulate gene expression so as to give the desired effect(s) in the transgenic plant. For numerous applications in plant biotechnology a tissue-specific expression profile is advantageous, since beneficial effects of expression in one tissue may have disadvantages in others. For example, promoters driving expression in the plant epidermis, such as epidermis-preferential or epidermis-specific promoters are useful for expressing genes that prevent pathogens such as fungi or bacteria from infecting a plant through the epidermis. It is advantageous to have the choice of a variety of different promoters so that the most suitable promoter may be selected for a particular gene, construct, cell, tissue, plant or environment. Moreover, the increasing interest in transforming plants with multiple plant transcription units (PTU) and the potential problems associated with using common regulatory molecules for these purposes merit having a variety of promoter molecules available.

There is, therefore, a constant need in the art for the identification of novel molecules that can be used for expression of selected transgenes in economically important plants. It is thus an objective of the present invention to provide new and alternative expression cassettes for expression of transgenes in various tissues of plants, for example in the epidermis. The objective is solved by the present invention.

SUMMARY OF THE INVENTION

A first embodiment of the invention relates to an isolated nucleic acid molecule capable of regulating expression in plants selected from the list comprising

i) a molecule described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, and

ii) a fragment of at least 250 consecutive bases of a molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48,

49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89 and

iii) a nucleotide molecule of at least 250 consecutive bases with a sequence identity of at least 60% to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, and

iv) a nucleotide molecule with a sequence identity of at least 60% to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, and

v) a nucleotide molecule of at least 250 bases capable of hybridizing under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, or the complement thereof;

vi) a nucleotide molecule of at least 250 bases capable of hybridizing under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C to a nucleic acid comprising 250 or more consecutive nucleotides of a transcription regulating nucleotide molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, or the complement thereof;

vii) a nucleotide molecule which is the complement or reverse complement of any of the previously mentioned nucleotide sequences under i) to vi).

In another embodiment the isolated nucleic acid molecules having a sequence as defined above under i), ii), iii), iv) v), vi) and vii) are capable for driving constitutive expression in plants, or expression in plant epidermis or mesophyll. In another embodiment, the expression derived from the isolated nucleic acid molecules as defined above under i), ii), iii), iv) v), vi) and vii) is inducible in plant epidermis and/or mesophyll by pathogen infection, for example by infection with a fungus.

In a preferred embodiment, the isolated nucleic acid molecules capable of regulating expression in plants having a sequence as defined above under ii) comprise a minimal promoter, preferably the minimal promoter of the respective isolated nucleic acid molecule.

The isolated nucleic acid molecule may be obtained or is obtainable from plant genomic DNA from a gene (e.g., from plant genomic DNA) encoding a polypeptide comprising an amino acid sequence which has at least 80% amino acid sequence homology to a polypeptide selected from the group described by SEQ ID NO: 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 and 255.

The isolated nucleic acid molecule may be obtained or is obtainable from plant genomic DNA from a gene which has at least 80% sequence identity to a nucleic acid molecule selected from the group described by SEQ ID NO: 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 and 254.

In one embodiment the isolated nucleic acid molecules having a sequence as specified above are capable of modifying transcription in a plant, or part thereof, for example in a plant cell. More specifically, the isolated nucleic acid molecules having a sequence as specified above are capable of modifying transcription constitutively, in epidermis and or mesophyll of a plant or a plant cell derived from such tissue.

It is also an embodiment of the invention at hand that the isolated nucleic acid molecules as defined above having the sequences specified under ii), iii), iv) v), vi) and vii) have substantially the same transcription regulating activity as the corresponding transcription regulating nucleotide molecule described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18 they are derived from.

Preferably, the isolated nucleic acid molecule is selected from the group of molecules described by SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89 or any homolog or fragment thereof.

More preferably the transcription regulating nucleotide molecule is selected from the group of molecules consisting of:

i) the molecule described by SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89, and

ii) a fragment of at least 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably at least 500 consecutive bases, most preferably at least 750 consecutive bases of a molecule

described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89 and

iii) a nucleotide molecule of at least 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably at least 500 consecutive bases, most preferably at least 750 consecutive bases with a sequence identity of at least 60%, 65% or 70%, preferably at least 75%, 80% or 85%, more preferably at least 90% or 95%, even more preferably at least 96% or 97%, most preferably at least 98% or 99% to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, and

iv) a nucleotide molecule having a sequence identity of at least 60%, 65% or 70%, preferably at least 75%, 80% or 85%, more preferably at least 90% or 95%, even more preferably at least 96% or 97%, most preferably at least 98% or 99% to an isolated nucleic acid molecule capable of regulating expression in plants described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, and

v) a nucleotide molecule of at least 250 bases, preferably at least 300 bases, more preferably at least 400 bases, even more preferably at least 500 bases, most preferably at least 750 bases capable of hybridizing preferably under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C, more desirably in 7% SDS, 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 65°C to an isolated nucleic acid molecule capable of regulating expression in plants described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, or the complement thereof;

vi) a nucleotide molecule of at least 250 bases, preferably at least 300 bases, more preferably at least 400 bases, even more preferably at least 500 bases, most preferably at least 750 bases capable of hybridizing preferably under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C, more desirably in 7% SDS, 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 65°C to a nucleic acid comprising at least 250 preferably at least 300, more preferably at least 400, even more

preferably at least 500, most preferably at least 750 consecutive nucleotides of an isolated nucleic acid molecule capable of regulating expression in plants described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, or the complement thereof;

vii) an isolated nucleic acid molecule capable of regulating expression in plants which is the complement or reverse complement of any of the previously mentioned nucleic acid molecules under i) to vi).

In a preferred embodiment, the isolated nucleic acid molecules capable of regulating expression in plants having a sequence as defined above under ii) comprise a minimal promoter, preferably the minimal promoter of the respective isolated nucleic acid molecule.

Another embodiment of the invention is an expression cassette for regulating expression in plants comprising

a) at least one isolated nucleic acid molecule capable of regulating expression in plants as defined above,

and functionally linked thereto

b) at least one nucleic acid molecule which is heterologous in relation to said transcription regulating nucleotide molecule which is to be expressed in a plant or part thereof.

The heterologous nucleotide molecule to be expressed in a plant or part thereof is preferably furthermore operably linked to introns, having expression enhancing effects, NEENAs (WO2011023537, WO2011023539), 5' and or 3'-untranslated regions, transcription termination and/or polyadenylation signals. 3'-untranslated regions are suitable to stabilize mRNA expression and structure. This can result in prolonged presence of the mRNA and thus enhanced expression levels. Termination and polyadenylation signals are suitable to stabilize mRNA expression (e.g., by stabilization of the RNA transcript and thereby the RNA level) to ensure constant mRNA transcript length and to prevent read-through transcription. Especially in multigene expression constructs this is an important feature. Furthermore correct termination of transcription is linked to re-initiation of transcription from the regulatory 5´nucleotide sequence resulting in enhanced expression levels. The above-mentioned signals can be any signal functional in plants and can for example be isolated from plant genes, plant virus genes or other plant pathogens. However, in a preferred embodiment the 3'-untranslated regions, transcription termination and polyadenylation signals are from the genes employed as the source for the promoters of this invention.

The transcription regulating molecules of the invention can be utilized to express any kind of nucleic acid molecule. For example, expression of the nucleic acid molecule can result in expression of a protein, or expression of an antisense RNA, sense or double-stranded RNA. Pref-

erably, expression of the nucleic acid molecule confers to the plant an agronomically valuable trait.

Another embodiment of the invention relates to a vector comprising an isolated nucleic acid molecule or an expression cassette of the invention. Yet another embodiment of the invention relates to a transgenic host cell or non-human organism comprising an expression cassette or a vector of the invention. Yet another embodiment of the invention relates to a transgenic plant or plant cell comprising an expression cassette or a vector of the invention. Preferably, said plant or plant cell is from a dicotyledonous plant, preferably of the family Fabacea, more preferably of the genus Glycine, most preferably the species Glycine max.

A further embodiment of the invention is a method for the production of an expression cassette as defined above or a vector as defined above comprising the steps of

a. providing an isolated nucleic acid molecule capable of regulating expression in plants as defined above and
b. functionally linking said isolated nucleic acid molecule to at least one nucleic acid molecule heterologous to said isolated nucleic acid molecule.

An additional embodiment of the invention is a method for the production of a transgenic plant comprising the steps of

a. providing an expression cassette as defined above or a vector as defined above and
b. transforming said expression cassette or vector into a plant part or plant cell and
c. regenerating a plant from said transformed plant part or plant cell.

An additional embodiment of the invention is a method for providing an expression cassette comprising an isolated nucleic acid molecule capable of regulating expression in plants comprising the steps of

a) isolating a first nucleic acid molecule from plant genomic DNA be using at least 15 consecutive bp, preferably at least 20 consecutive bp of a sequence described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 or 18 or 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 and 254 and
b) functionally linking the first nucleic acid molecule obtained in step a) to at least one additional nucleic acid molecule heterologous to said first nucleic acid molecule.

DEFINITIONS

It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, plant species or genera, constructs, and reagents described as such. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which will be limited only by the appended claims. It must be noted that as used herein and in the appended claims, the singular forms "a," "and," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a vector" is a reference to one or more vectors and includes equivalents thereof known to those skilled in the art, and so forth.

The term "about" is used herein to mean approximately, roughly, around, or in the region of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 20 percent, preferably 10 percent up or down (higher or lower).

As used herein, the word "or" means any one member of a particular list and also includes any combination of members of that list.

The term "gene" is used broadly to refer to any segment of nucleic acid associated with a biological function. Thus, genes include coding sequences and/or the regulatory molecules required for their expression. For example, gene refers to a nucleic acid fragment that expresses mRNA or functional RNA, or encodes a specific protein, and which includes regulatory molecules. Genes also include non-expressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

The term "native" or "wild type" gene refers to a gene that is present in the genome of an untransformed cell, i.e., a cell not having a known mutation.

A "marker gene" encodes a selectable or screenable trait.

The term "chimeric gene" refers to any gene that contains

1) DNA sequences, including regulatory and coding sequences, that are not functionally linked 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 molecules and coding sequences that are derived from different sources, or comprise regulatory molecules, 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 that is introduced by gene transfer.

An "oligonucleotide" corresponding to a nucleotide sequence of the invention, e.g., for use in probing or amplification reactions, may be about 30 or fewer nucleotides in length (e.g., 9, 12,

15, 18, 20, 21, 22, 23, or 24, or any number between 9 and 30). Generally specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16 to 24 nucleotides in length may be preferred. Those skilled in the art are well versed in the design of primers for use processes such as PCR. If required, probing can be done with entire restriction fragments of the gene disclosed herein which may be 100's or even 1000's of nucleotides in length.

The terms "polypeptide", "peptide", "oligopeptide", "gene product", "expression product" and "protein" are used interchangeably herein to refer to a polymer or oligomer of consecutive amino acid residues. As used herein, the term "amino acid sequence" or a "polypeptide sequence" refers to a list of abbreviations, letters, characters or words representing amino acid residues. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. The abbreviations used herein are conventional one letter codes for the amino acids: A, alanine; B, asparagine or aspartic acid; C, cysteine; D aspartic acid; E, glutamate, glutamic acid; F, phenylalanine; G, glycine; H histidine; I isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine ; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; Z, glutamine or glutamic acid (see L. Stryer, Biochemistry, 1988, W. H. Freeman and Company, New York. The letter "x" as used herein within an amino acid sequence can stand for any amino acid residue.

"Coding sequence" refers to a DNA or RNA molecule that codes for a specific amino acid molecule and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", i.e., lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a molecule of RNA which is contained in the primary transcript but which is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

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).

A "functional RNA" refers to an antisense RNA, microRNA, siRNA, ribozyme, or other RNA that is not translated.

The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA molecule. When the RNA transcript is a perfect complementary copy of the DNA molecule, it is referred to as the primary transcript or it may be a RNA molecule derived from 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.

"Isolated nucleic acid molecule capable of regulating expression", „transcription regulating nucleotide molecule", "regulatory molecule", or "suitable regulatory molecules", each refer to nucleotide molecules influencing the transcription, RNA processing or stability, or translation of the associated (or functionally linked) nucleotide molecules to be transcribed. The transcription regulating nucleotide molecule may have various localizations with respect to the nucleotide molecules to be transcribed. The transcription regulating nucleotide molecule may be located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of the molecule to be transcribed (e.g., a coding sequence). The transcription regulating nucleotide molecule may be selected from the group comprising enhancers, promoters, translation leader sequences, introns, 5'-untranslated sequences, 3'-untranslated sequences, and polyadenylation signal sequences. They include natural and synthetic molecules as well as molecules, which may be a combination of synthetic and natural molecules. As is noted above, the term "transcription regulating nucleotide molecule" is not limited to promoters. However, preferably a transcription regulating nucleotide molecule of the invention comprises at least one promoter molecule (*e.g.*, a molecule localized upstream of the transcription start of a gene capable to induce transcription of the downstream molecules). In one preferred embodiment the transcription regulating nucleotide molecule of the invention comprises the promoter molecule of the corresponding gene and – optionally and preferably – the native 5'-untranslated region of said gene. Furthermore, the 3'-untranslated region and/or the polyadenylation region of said gene may also be employed. As used herein, the term "cis-element" or "promoter motif" 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. The promoters of the present invention desirably contain cis-elements that can confer or modulate gene expression. 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 subsequent manipulation.

"5' non-coding sequence" or "5'-untranslated sequence" or "-region" refers to a sequence of a nucleotide molecule 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.

"3' non-coding sequence" or "3'-untranslated sequence" or "-region" refers to a sequence of a nucleotide molecule 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.

The term "translation leader 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 translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

"Signal peptide" refers to the amino terminal extension of a polypeptide, which is translated in conjunction with the polypeptide forming a precursor peptide and which is required for its entrance into the secretory pathway. The term "signal sequence" refers to a nucleotide sequence that encodes the signal peptide. The term “transit peptide” as used herein refers part of an expressed polypeptide (preferably to the amino terminal extension of a polypeptide), which is translated in conjunction with the polypeptide forming a precursor peptide and which is required for its entrance into a cell organelle (such as the plastids (e.g., chloroplasts) or mitochondria). The term "transit sequence" refers to a nucleotide sequence that encodes the transit peptide.

"Promoter" refers to a nucleotide molecule, 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 molecule that includes a minimal promoter plus regulatory elements that is capable of controlling the expression of a coding sequence or functional RNA. This type of promoter molecule consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA molecule, which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other upstream promoter elements bind sequence-specific DNA-binding proteins that mediate their effects. 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 sequences that are involved in the binding of protein factors, which control the effectiveness of transcription initiation in response to physiological or developmental conditions. A person skilled in the art is aware of methods for rendering a unidirectional to a bidirectional promoter and of methods to use the complement or reverse complement of a promoter sequence for creating a promoter having the same promoter specificity as the original sequence. Such methods are for example described for constitutive as well as inducible promoters by Xie et al. (2001) “Bidirectionalization of polar promoters in plants” nature biotechnology 19 pages 677 – 679. The authors describe that it is sufficient to add a

minimal promoter to the 5´ prime end of any given promoter to receive a promoter controlling expression in both directions with same promoter specificity.

The "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.

Promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation are referred to as "minimal or core promoters." In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription. A "minimal or core promoter" thus consists only of all basal elements needed for transcription initiation, e.g., a TATA box and/or an initiator.

"Constitutive expression" refers to expression using a constitutive or regulated promoter. "Conditional" and "regulated expression" refer to expression controlled by a regulated promoter.

"Constitutive promoter" refers to a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant. Each of the transcription-activating elements do not exhibit an absolute tissue-specificity, but mediate transcriptional activation in most plant parts at a level of at least 1% of the level reached in the part of the plant in which transcription is most active.

"Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes both tissue-specific and inducible promoters. It includes natural and synthetic molecules as well as molecules which may be a combination of synthetic and natural molecules. 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. New promoters of various types useful in plant cells are constantly being discovered, numerous examples may be found in the compilation by Okamuro et al. (1989). Typical regulated promoters useful in plants include but are not limited to safener-inducible promoters, promoters derived from the tetracycline-inducible system, promoters derived from salicylate-inducible systems, promoters derived from alcohol-inducible systems, promoters derived from glucocorticoid-inducible system, promoters derived from pathogen-inducible systems, and promoters derived from ecdysone-inducible systems.

"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 or seeds), specific tissues (such as epidermis, green tissue, 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 leaf expansion fruit ripening in developing seeds or fruit, in fully differentiated leaf, or at the onset of senescence.

"Inducible promoter" refers to those regulated promoters that can be turned on in one or more cell types or that cause increased expression upon an external stimulus, such as a chemical, light, hormone, stress, or a pathogen.

"Operably-linked" or "functionally linked" refers preferably to the association of nucleic acid molecules on single nucleic acid fragment so that the function of one is affected by the other. For example, a regulatory DNA molecule is said to be "operably linked to" or "associated with" a DNA molecule that codes for an RNA or a polypeptide if the two molecules are situated such that the regulatory DNA molecule affects expression of the coding DNA molecule (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory molecules in sense or antisense orientation.

"Expression" refers to the transcription and/or translation of an endogenous gene, ORF or portion thereof, or a transgene in plants. For example, in the case of antisense constructs, expression may refer to the transcription of the antisense DNA only. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

"Specific expression" is the expression of gene products, which is limited to one or a few plant tissues (spatial limitation) and/or to one or a few plant developmental stages (temporal limitation). It is acknowledged that hardly a true specificity exists: promoters seem to be preferably switch on in some tissues, while in other tissues there can be no or only little activity. This phenomenon is known as leaky expression. However, with specific expression in this invention is meant preferable expression in one or a few plant tissues.

The "expression pattern" of a promoter (with or without enhancer) is the pattern of expression levels, which shows where in the plant and in what developmental stage transcription is initiated by said promoter. 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. The level of expression of a promoter can be determined by measuring the `steady state` concentration of a standard transcribed reporter mRNA. This measurement is indirect since the concentration of the reporter mRNA is dependent not only on its synthesis rate, but also on the rate with which the mRNA is degraded. Therefore, the steady state level is the product of synthesis rates and degradation rates. The rate of degradation can however be considered to proceed at a fixed rate when the transcribed molecules are identical, and thus this value can serve as a measure of synthesis rates. When promoters are compared in this way techniques available to those skilled in the art are hybridization S1-RNAse analysis, northern blots and competitive RT-PCR. This list of techniques in no way represents all available techniques, but rather describes commonly used procedures used to analyze transcription activity and expression levels of mRNA. The analysis of transcription start points in practically all promoters has revealed that there is usually no single base at which transcription starts, but rather a more or less clustered set of initiation sites, each of which accounts for some start points of the mRNA. Since this distribution varies from promoter to promoter the sequences of the reporter mRNA in each of the populations would differ from each other. Since each mRNA spe-

cies is more or less prone to degradation, no single degradation rate can be expected for different reporter mRNAs. It has been shown for various eukaryotic promoter molecules that the sequence surrounding the initiation site (`initiator`) plays an important role in determining the level of RNA expression directed by that specific promoter. This includes also part of the transcribed sequences. The direct fusion of promoter to reporter molecules would therefore lead to suboptimal levels of transcription. A commonly used procedure to analyze expression patterns and levels is through determination of the `steady state` level of protein accumulation in a cell. Commonly used candidates for the reporter gene, known to those skilled in the art are beta-glucuronidase (GUS), chloramphenicol acetyl transferase (CAT) and proteins with fluorescent properties, such as green fluorescent protein (GFP) from Aequora victoria. In principle, however, many more proteins are suitable for this purpose, provided the protein does not interfere with essential plant functions. For quantification and determination of localization a number of tools are suited. Detection systems can readily be created or are available which are based on, e.g., immunochemical, enzymatic, fluorescent detection and quantification. Protein levels can be determined in plant tissue extracts or in intact tissue using in situ analysis of protein expression. Generally, individual transformed lines with one chimeric promoter reporter construct will vary in their levels of expression of the reporter gene. Also frequently observed is the phenomenon that such transformants do not express any detectable product (RNA or protein). The variability in expression is commonly ascribed to `position effects`, although the molecular mechanisms underlying this inactivity are usually not clear.

"Overexpression" refers to the level of expression in transgenic cells or organisms that exceeds levels of expression in normal or untransformed (non-transgenic) cells or organisms.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from an endogenous gene or a transgene.

"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 1996). Gene silencing includes virus-induced gene silencing (Ruiz et al. 1998).

The terms "heterologous DNA molecule", "exogenous DNA segment" or "heterologous nucleic acid," as used herein, each refer to a molecule that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA molecule. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides. A "homologous" DNA molecule is a DNA molecule that is naturally associated with a host cell into which it is introduced.

"Homologous to" in the context of nucleotide sequence identity refers to the similarity between the nucleotide sequences of two nucleic acid molecules or between the amino acid sequences of two protein molecules. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (as described in Haines and Higgins (eds.), Nucleic Acid Hybridization, IRL Press, Oxford, U.K.), or by the comparison of sequence similarity between two nucleic acids or proteins.

The term "substantially similar" or "similar" refers to nucleotide and amino acid sequences that represent functional and/or structural equivalents or orthologs of *Arabidopsis thaliana* or *Glycine max* sequences disclosed herein.

In its broadest sense, the term "substantially similar" or "similar" when used herein with respect to a nucleotide sequence means that the nucleotide sequence is part of a gene which encodes a polypeptide having substantially the same structure and function as a polypeptide encoded by a gene for the reference nucleotide sequence, e.g., the nucleotide sequence comprises a promoter from a gene that is the ortholog of the gene corresponding to the reference nucleotide sequence, as well as promoter sequences that are structurally related the promoter sequences particularly exemplified herein, i.e., the substantially similar promoter sequences hybridize to the complement of the promoter sequences exemplified herein under high or very high stringency conditions. For example, altered nucleotide sequences, which simply reflect the degeneracy of the genetic code but nonetheless encode amino acid sequences that are identical to a particular amino acid sequence are substantially similar to the particular sequences. The term "substantially similar" also includes nucleotide sequences wherein the sequence has been modified, for example, to optimize expression in particular cells, as well as nucleotide sequences encoding a variant polypeptide having one or more amino acid substitutions relative to the (unmodified) polypeptide encoded by the reference sequence, which substitution(s) does not alter the activity of the variant polypeptide relative to the unmodified polypeptide.

In its broadest sense, the term "substantially similar" or "similar" when used herein with respect to polypeptide or nucleic acids means that the polypeptide or nucleic acid has substantially the same structure and function as the reference polypeptide. In addition, amino acid sequences or nucleic acids that are substantially similar to a particular sequence are those wherein overall amino acid or nucleic acid identity is at least 90% or greater to the instant sequences. Modifications that result in equivalent nucleotide or amino acid sequences are well within the routine skill in the art. The percentage of amino acid or nucleic acid sequence identity between the substantially similar and the reference polypeptide or nucleic acid is at least 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, wherein the reference polypeptide is an polypeptide encoded by a gene with a promoter having any one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, a nucleotide sequence comprising an open reading frame comprised in SEQ ID NOs: 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 or 254, which encodes a polypeptide described by SEQ ID NOs: 221, 223,

225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 or 255. One indication that two polypeptides are substantially similar to each other, besides having substantially the same function, is that an agent, e.g., an antibody, which specifically binds to one of the polypeptides, also specifically binds to the other.

Sequence comparisons maybe carried out using a Smith-Waterman sequence alignment algorithm (see e.g., Waterman (1995)). The localS program, version 1.16, is preferably used with following parameters: match: 1, mismatch penalty: 0.33, open-gap penalty: 2, extended-gap penalty: 2.

Moreover, a nucleotide sequence that is "substantially similar" or "similar" to a reference nucleotide sequence is said to be "equivalent" to the reference nucleotide sequence. The skilled artisan recognizes that equivalent nucleotide sequences encompassed by this invention can also be defined by their ability to hybridize, under low, moderate and/or stringent conditions (e.g., 0.1 X SSC, 0.1% SDS, 65°C), with the nucleotide sequences that are within the literal scope of the instant claims.

What is meant by "substantially the same activity" or "the same activity" when used in reference to a polynucleotide fragment or a homolog is that the fragment or homolog has at least 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99% of the expression regulating activity of the full length polynucleotide.

"Target gene" refers to a gene on the replicon that expresses the desired target coding sequence, functional RNA, or protein. The target gene is not essential for replicon replication. Additionally, target genes may comprise native non-viral genes inserted into a non-native organism, or chimeric genes, and will be under the control of suitable regulatory sequences. Thus, the regulatory sequences in the target gene may come from any source, including the virus. Target genes may include coding sequences that are either heterologous or homologous to the genes of a particular plant to be transformed. However, target genes do not include native viral genes. Typical target genes include, but are not limited to genes encoding a structural protein, a seed storage protein, a protein that conveys herbicide resistance, and a protein that conveys insect resistance. Proteins encoded by target genes are known as "foreign proteins". The expression of a target gene in a plant will typically produce an altered plant trait.

The term "altered plant trait" means any phenotypic or genotypic change in a transgenic plant relative to the wild-type or non-transgenic plant host.

"Replication gene" refers to a gene encoding a viral replication protein. In addition to the ORF of the replication protein, the replication gene may also contain other overlapping or non-overlapping ORF(s), as are found in viral sequences in nature. While not essential for replication, these additional ORFs may enhance replication and/or viral DNA accumulation. Examples of such additional ORFs are AC3 and AL3 in ACMV and TGMV geminiviruses, respectively.

"Chimeric trans-acting replication gene" refers either to a replication gene in which the coding sequence of a replication protein is under the control of a regulated plant promoter other than that in the native viral replication gene, or a modified native viral replication gene, for example, in which a site specific sequence(s) is inserted in the 5' transcribed but untranslated region. Such chimeric genes also include insertion of the known sites of replication protein binding between the promoter and the transcription start site that attenuate transcription of viral replication protein gene.

"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.

The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to as "transgenic" cells, and organisms comprising transgenic cells are referred to as "transgenic organisms". Examples of methods of transformation of plants and plant cells include Agrobacterium-mediated transformation (De Blaere 1987) and particle bombardment technology (US 4,945,050). Whole plants may be regenerated from transgenic cells by methods well known to the skilled artisan (see, for example, Fromm 1990).
"Transformed," "transgenic," and "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 generally known in the art and are disclosed (Sambrook 1989; Innis 1995; Gelfand 1995; Innis & Gelfand 1999. Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially mismatched primers, and the like. For example, "transformed," "transformant," and "transgenic" plants or calli have been through the transformation process and contain a foreign gene integrated into their chromosome. The term "untransformed" refers to normal plants that have not been through the transformation process.

"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.

"Transient expression" refers to expression in cells in which a virus or a transgene is introduced by viral infection or by such methods as Agrobacterium-mediated transformation, electroporation, or biolistic bombardment, but not selected for its stable maintenance.

"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.

"Primary transformant" and "T0 generation" refer to transgenic plants that are of the same genetic generation as the tissue which 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.

Wild-type: The term "wild-type", "natural" or "natural origin" means with respect to an organism, polypeptide, or nucleic acid sequence, that said organism is naturally occurring or available in at least one naturally occurring organism which is not changed, mutated, or otherwise manipulated by man.

The terms "genome" or "genomic DNA" is referring to the heritable genetic information of a host organism. Said genomic DNA comprises the DNA of the nucleus (also referred to as chromosomal DNA) but also the DNA of the plastids (*e.g.,* chloroplasts) and other cellular organelles (*e.g.*, mitochondria). Preferably the terms genome or genomic DNA is referring to the chromosomal DNA of the nucleus.

The term "chromosomal DNA" or "chromosomal DNA-sequence" is to be understood as the genomic DNA of the cellular nucleus independent from the cell cycle status. Chromosomal DNA might therefore be organized in chromosomes or chromatids, they might be condensed or uncoiled. An insertion into the chromosomal DNA can be demonstrated and analyzed by various methods known in the art like *e.g.*, polymerase chain reaction (PCR) analysis, Southern blot analysis, fluorescence *in situ* hybridization (FISH), and *in situ* PCR.

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

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. In the context of the present invention, an "isolated" or "purified" DNA molecule or an "isolated" or "purified" polypeptide is a DNA molecule or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or polypeptide may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein or polypeptide having less than about 30%, 20%, 10%, 5%, (by dry weight) of contaminating protein. When the protein of the invention, or biologically active portion thereof, is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein of interest chemicals. The nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant (variant) forms. Such variants will continue to possess the desired activity, i.e., either promoter activity or the activity of the product encoded by the open reading frame of the non-variant nucleotide sequence.

The term "variant" or “homolog” with respect to a sequence (e.g., a polypeptide or nucleic acid sequence such as – for example – a transcription regulating nucleotide molecule of the invention) is intended to mean substantially similar sequences. For nucleotide sequences comprising an open reading frame, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical amino acid sequence of the native protein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis and for open reading frames, encode the native protein, as well as those that encode a polypeptide having amino acid substitutions relative to the native protein. Generally, nucleotide sequence variants of the invention will have at least 40, 50, 60, to 70%, e.g., preferably 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, generally at least 80%, e.g., 81%-84%, at least 85%, e.g., 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, to 98% and 99% nucleotide sequence identity to the native (wild type or endogenous) nucleotide sequence.

"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 a nucleotide sequence of interest, which is – optionally - operably linked to termination signals and/or other regulatory elements. An expression cassette may also comprise 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. The expression cassette may also be one, which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. An expression cassette may be assembled entirely extracellularly (e.g., by recombinant cloning techniques). However, an expression cassette may also be assembled using in part endogenous components. For example, an expression cassette may be obtained by placing (or inserting) a promoter sequence upstream of an endogenous sequence, which thereby becomes functionally linked and controlled by said promoter sequences. Likewise, a nucleic acid sequence to be expressed may be placed (or inserted) downstream of an endogenous promoter sequence thereby forming an expression cassette. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter, which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, the promoter can also be specific to a particular tissue or organ or stage of development. In a preferred embodiment, such expression cassettes will comprise the transcriptional initiation region of the invention linked to a nucleotide sequence of interest. Such an expression cassette is preferably provided with a plurality of restriction sites for insertion of the gene of interest to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes. The cassette will include 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. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such, as the octopine synthase and nopaline synthase termination regions and others described below (see also, Guerineau 1991; Proudfoot 1991; Sanfacon 1991; Mogen 1990; Munroe 1990; Ballas 1989; Joshi 1987).

"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 actinomy-

cetes and related species, bacteria and eukaryotic (e.g. higher plant, mammalian, yeast or fungal cells).

Preferably the nucleic acid in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell such as a microbial, e.g. bacterial, or plant cell. The vector may be a bi-functional expression vector which functions in multiple hosts. In the case of genomic DNA, this may contain its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell.

"Cloning vectors" typically contain one or a small number of restriction endonuclease recognition sites at which foreign DNA sequences can be inserted in a determinable fashion without loss of essential biological function of the vector, as well as a marker gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Marker genes typically include genes that provide tetracycline resistance, hygromycin resistance or ampicillin resistance.

A "transgenic plant" is a plant having one or more plant cells that contain an expression vector or recombinant expression construct.

"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.

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

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a fulllength cDNA or gene sequence or isolated nucleic acid sequence capable of regulating expression in plants, preferably the complete cDNA or gene sequence or isolated nucleic acid sequence capable of regulating expression in plants is the reference sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. In a preferred embodiment the comparison window defining the homology of sequence consists of the entire query sequence. Those of skill in the art understand that to avoid a high similarity to a reference se-

quence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

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

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described (Higgins 1988, 1989; Corpet 1988; Huang 1992; Pearson 1994). The ALIGN program is based on the algorithm of Myers and Miller, supra. The BLAST programs of Altschul et al., 1990, are based on the algorithm of Karlin and Altschul, supra. Multiple alignments (i.e. of more than 2 sequences) are preferably performed using the Clustal W algorithm (Thompson 1994; e.g., in the software VectorNTI™, version 9; Invitrogen Inc.) with the scoring matrix BLOSUM62MT2 with the default settings (gap opening penalty 15/19, gap extension penalty 6.66/0.05; gap separation penalty range 8; % identity for alignment delay 40; using residue specific gaps and hydrophilic residue gaps).

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul 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.

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 (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.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. 1997. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al., supra. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (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 wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989). See http://www.ncbi.nlm.nih.gov. Alignment may also be performed manually by inspection.

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

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-

conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif.).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, preferably the complete query or reference sequence as defined by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e) (i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 90%, 91%, 92%, 93%, or 94%, and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 90%, 95%, and most preferably at least 98%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions (see below). Generally, stringent 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. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 90%, 91%, 92%, 93%, or 94%, or even more preferably,

95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (1970). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution.

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. The reference sequences of the invention are defined by the sequences comprised in the sequence protocol, preferably SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 and 254, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 and 255. Preferably the reference sequence comprises the complete sequence as defined by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89 or 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 or 254 or 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 or 255, more preferably the reference sequence consists of the complete sequence as defined by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89 or 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 or 254 or 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 or 255.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "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.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridization are sequence dependent, and are different under different environmental parameters. 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. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, 1984:

$T_m = 81.5°C + 16.6 (\log_{10} M)+0.41 (\%GC) - 0.61 (\%\ form) – 500 / L$

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. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point I for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10°C lower than the thermal melting point I; moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point I; 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 I. 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. Generally, highly stringent 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.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2 X 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 1 X SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4 to 6 X SSC at 40°C for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, more preferably 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 and at least about 60°C for long robes (e.g., >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2 X (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 hy-

bridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent 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, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1 x SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1 X to 2 X SSC (20 X 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 0.5 X to 1 X SSC at 55 to 60°C.

The following are examples of sets of hybridization/wash conditions that may be used to clone orthologous 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 2 X 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 1 X 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 0.5 X 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 0.1 X 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 0.1 X SSC, 0.1% SDS at 65°C.

"DNA shuffling" is a method to introduce mutations or rearrangements, preferably randomly, in a DNA molecule or to generate exchanges of DNA sequences between two or more DNA molecules, preferably randomly. The DNA molecule resulting from DNA shuffling is a shuffled DNA molecule that is a non-naturally occurring DNA molecule derived from at least one template DNA molecule. The shuffled DNA preferably encodes a variant polypeptide modified with respect to the polypeptide encoded by the template DNA, and may have an altered biological activity with respect to the polypeptide encoded by the template DNA.

"Recombinant DNA molecule" is a combination of DNA sequences that are joined together using recombinant DNA technology and procedures used to join together DNA sequences as described, for example, in Sambrook et al., 1989.

The word "plant" refers to any plant, particularly to agronomically useful plants (e.g., seed plants), and "plant cell" is a structural and physiological unit of the plant, which comprises a cell wall but may also refer to a protoplast. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, a plant tissue, or a plant organ differentiated into a structure that is present at any stage of a plant's development. Such structures include one or more plant organs including, but are not limited to, fruit, shoot,

stem, leaf, flower petal, *etc.* Preferably, the term "plant" includes whole plants, shoot vegetative organs/structures (*e.g.* leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.* bracts, sepals, petals, stamens, carpels, anthers and ovules), seeds (including embryo, endosperm, and seed coat) and fruits (the mature ovary), plant tissues (*e.g.* vascular tissue, ground tissue, and the like) and cells (*e.g.* guard cells, egg cells, trichomes and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, and multicellular algae. It includes plants of a variety of ploidy levels, including aneuploid, polyploid, diploid, haploid and hemizygous. Included within the scope of the invention are all genera and species of higher and lower plants of the plant kingdom. Included are furthermore the mature plants, seed, shoots and seedlings, and parts, propagation material (for example seeds and fruit) and cultures, for example cell cultures, derived therefrom. Preferred are plants and plant materials of the following plant families: *Amaranthaceae, Brassicaceae, Carophyllaceae, Chenopodiaceae, Compositae, Cucurbitaceae, Labiatae, Leguminosae, Papilionoideae, Liliaceae, Linaceae, Malvaceae, Rosaceae, Saxifragaceae, Scrophulariaceae, Solanaceae, Tetragoniaceae.* Annual, perennial, monocotyledonous and dicotyledonous plants are preferred host organisms for the generation of transgenic plants. The use of the recombination system, or method according to the invention is furthermore advantageous in all ornamental plants, forestry, fruit, or ornamental trees, flowers, cut flowers, shrubs or turf. Said plant may include – but shall not be limited to - bryophytes such as, for example, Hepaticae (hepaticas) and Musci (mosses); pteridophytes such as ferns, horsetail and clubmosses; gymnosperms such as conifers, cycads, ginkgo and *Gnetaeae*; algae such as *Chlorophyceae, Phaeophpyceae, Rhodophyceae, Myxophyceae, Xanthophyceae, Bacillariophyceae* (diatoms) and *Euglenophyceae.* Plants for the purposes of the invention may comprise the families of the Rosaceae such as rose, Ericaceae such as rhododendrons and azaleas, *Euphorbiaceae* such as poinsettias and croton, *Caryophyllaceae* such as pinks, Solanaceae such as petunias, *Gesneriaceae* such as African violet, *Balsaminaceae* such as touch-me-not, Orchidaceae such as orchids, *Iridaceae* such as gladioli, iris, freesia and crocus, *Compositae* such as marigold, *Geraniaceae* such as geraniums, *Liliaceae* such as *Drachaena, Moraceae* such as ficus, *Araceae* such as philodendron and many others. The transgenic plants according to the invention are furthermore selected in particular from among dicotyledonous crop plants such as, for example, from the families of the Leguminosae such as pea, alfalfa and soybean; the family of the *Umbelliferae*, particularly the genus *Daucus* (very particularly the species *carota* (carrot)) and *Apium* (very particularly the species *graveolens* var. dulce (celery)) and many others; the family of the Solanaceae, particularly the genus *Lycopersicon*, very particularly the species *esculentum* (tomato) and the genus *Solanum*, very particularly the species *tuberosum* (potato) and *melongena* (aubergine), tobacco and many others; and the genus *Capsicum*, very particularly the species *annum* (pepper) and many others; the family of the *Leguminosae*, particularly the genus *Glycine*, very particularly the species *max* (soybean) and many others; and the family of the *Cruciferae*, particularly the genus *Brassica*, very particularly the species *napus* (oilseed rape), *campestris* (beet), *oleracea* cv Tastie (cabbage), *oleracea* cv Snowball Y (cauliflower) and *oleracea* cv Emperor (broccoli); and the genus *Arabidopsis*, very particularly the species *thaliana* and many others; the family of the Compositae, particularly the genus *Lactuca*, very particularly the spe-

cies *sativa* (lettuce) and many others. The transgenic plants according to the invention may be selected among monocotyledonous crop plants, such as, for example, cereals such as wheat, barley, sorghum and millet, rye, triticale, maize, rice or oats, and sugarcane. Further preferred are trees such as apple, pear, quince, plum, cherry, peach, nectarine, apricot, papaya, mango, and other woody species including coniferous and deciduous trees such as poplar, pine, sequoia, cedar, oak, *etc.* Especially preferred are *Arabidopsis thaliana, Nicotiana tabacum*, oilseed rape, soybean, corn (maize), wheat, *Linum usitatissimum* (linseed and flax), *Camelina sativa, Brassica juncea,* potato and tagetes.

"Significant increase" is an increase that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater.

"Significantly less" means that the decrease is larger than the margin of error inherent in the measurement technique, preferably a decrease by about 2-fold or greater.

DETAILED DESCRIPTION OF THE INVENTION

The present invention thus provides for isolated nucleic acid molecules comprising a plant nucleotide sequence that directs transcription in epidermis of an operably linked nucleic acid fragment in a plant or part thereof.

The present invention further provides for isolated nucleic acid molecules comprising a plant nucleotide sequence that directs transcription in epidermis of an operably linked nucleic acid fragment in a plant or part thereof upon induction by a pathogen, preferably a fungal pathogen.

The present invention further provides for isolated nucleic acid molecules comprising a plant nucleotide sequence that directs constitutive transcription of an operably linked nucleic acid fragment in a plant or part thereof.

The present invention further provides for isolated nucleic acid molecules comprising a plant nucleotide sequence that directs mesophyll specific or mesophyll preferable transcription of an operably linked nucleic acid fragment in a plant or part thereof.

The present invention further provides for isolated nucleic acid molecules comprising a plant nucleotide sequence that directs mesophyll and epidermis specific or mesophyll and epidermis preferable transcription of an operably linked nucleic acid fragment in a plant or part thereof upon induction by a pathogen, preferably a fungal pathogen.

In addition, the present invention provides transgenic expression cassettes for regulating expression in plant epidermis, inducible expression in plant epidermis and/or mesophyll or constitutive expression comprising

a) at least one transcription regulating nucleotide molecule derived from any of the *Glycine max* genes described by the GenBank *Glycine max* genome loci Glyma11g14950, Glyma14g06680, Glyma02g47670, Glyma14g02930, Glyma17g27610, Glyma13g44640, Glyma08g37270, Glyma04g40860.1, Glyma01g33070.2,

Glyma15g05820.1, Glyma01g42660.1, Glyma17g14320 or Glyma01g01510.1 or their orthologous genes
and functionally linked thereto

b) at least one nucleic acid molecule which is heterologous in relation to said transcription regulating nucleotide molecule.

In addition, the present invention provides transgenic expression cassettes for regulating expression in plant mesophyll comprising

a) at least one transcription regulating nucleotide molecule derived from any of the *Arabidopsis thaliana* genes described by the GenBank *Arabidopsis thaliana* genome loci At1g49750, At3g62410, At1g61520, At1g30380 or At1g65490 or their orthologous genes and functionally linked thereto

b) at least one nucleic acid molecule which is heterologous in relation to said transcription regulating nucleotide sequence.

"tissue-specific transcription" in the context of this invention means the transcription of a nucleic acid molecule by a transcription regulating nucleic acid molecule in a way that transcription of said nucleic acid molecule in said tissue contribute to more than 90%, preferably more than 95%, more preferably more than 99% of the entire quantity of the RNA transcribed from said nucleic acid molecule in the entire plant during any of its developmental stage. The transcription regulating nucleotide molecules specifically disclosed herein are considered to be tissue-specific transcription regulating nucleotide molecules.

"tissue-preferential transcription" in the context of this invention means the transcription of a nucleic acid molecule by a transcription regulating nucleic acid molecule in a way that transcription of said nucleic acid sequence in the said tissue contribute to more than 50%, preferably more than 70%, more preferably more than 80% of the entire quantity of the RNA transcribed from said nucleic acid sequence in the entire plant during any of its developmental stage.

"substantially the same transcription regulating activity" in the context of this invention means the transcription of a nucleic acid molecule by a transcription regulating nucleic acid molecule which is a fragment or a homolog or a variant of any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 or 18 has the same tissue specific or tissue preferential transcription as the transcription regulating nucleic acid molecule it is derived from and has at least 50% preferably at least 60%, more preferably at least 80% or 90%, even more preferably at least 95%, most preferably the same expression strength as the transcription regulating nucleic acid molecule it is derived from.

Preferably a transcription regulating nucleotide molecule of the invention comprises at least one promoter sequence of the respective gene (*e.g.*, a sequence localized upstream of the transcription start of the respective gene capable to induce transcription of the downstream sequences). Said transcription regulating nucleotide molecule may comprise the promoter sequence of said genes but may further comprise other elements such as the 5'-untranslated se-

quence, enhancer, introns etc. Preferably, said promoter sequence directs transcription of an operably linked nucleic acid segment in a plant epidermis or plant epidermis cell e.g., a linked plant DNA comprising an open reading frame for a structural or regulatory gene.

As specified above in the DEFINTION section, identities between nucleotide sequences are preferably measured by the BLASTN program using default parameters with a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For measuring identity between amino acid sequences, the BLASTP program is used with default parameters with a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989). The BLAST Program version 1.4.7 or later is used.

Preferably, such homolog or fragment of said isolated nucleotide sequence (e.g., the sequences specified under ii), iii), iv) v), vi) and vii) above) is capable to modify transcription in a plant cell or organism, more preferably said homolog or fragment (e.g., the sequences specified under ii), iii), iv) v) and vi) above) has substantially the same transcription regulating activity as the transcription regulating nucleotide molecule described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 or 18. Preferably, the homolog or fragment (e.g., the sequences specified under iv) or v) above) is hybridizing under stringent conditions (i.e. medium stringent, more preferably high stringent conditions) with the specified target sequence.

Preferably, the transcription regulating nucleotide molecule employed in the expression cassettes of the invention is selected from the group of molecules consisting of the molecules described by SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89 or any homolog or fragment thereof. More preferably the transcription regulating nucleotide molecule employed in the expression cassette of the invention is selected from the group of molecules consisting of

i) the molecule described by SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89, and

ii) a fragment of at least 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably at least 500 consecutive bases, most preferably at least 750 consecutive bases of a molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89 and

iii) a nucleotide molecule of at least 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably

at least 500 consecutive bases, most preferably at least 750 consecutive bases with a sequence identity of at least 60%, 65% or 70%, preferably at least 75%, 80% or 85%, more preferably at least 90% or 95%, even more preferably at least 96% or 97%, most preferably at least 98% or 99% to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 or 89, and

iv) a nucleotide molecule having a sequence identity of at least 60%, 65%, 70%, 75% or 80%, preferably at least 85%, more preferably at least 90% or 95%, even more preferably at least 96% or 97%, most preferably at least 98% or 99% to an isolated nucleic acid molecule capable of regulating expression in plants described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89, and

v) a nucleotide molecule of at least 250 bases, preferably at least 300 bases, more preferably at least 400 bases, even more preferably at least 500 bases, most preferably at least 750 bases capable of hybridizing preferably under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C, more desirably in 7% SDS, 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 65°C to an isolated nucleic acid molecule capable of regulating expression in plants described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89, or the complement thereof;

vi) a nucleotide molecule of at least 250 bases, preferably at least 300 bases, more preferably at least 400 bases, even more preferably at least 500 bases, most preferably at least 750 bases capable of hybridizing preferably under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C, more desirably in 7% SDS, 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 65°C to a nucleic acid comprising at least 250 preferably at least 300, more preferably at least 400, even more preferably at least 500, most preferably at least 750 consecutive nucleotides of an isolated nucleic acid molecule capable of regulating expression in plants described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89, or the complement thereof;

vii) an isolated nucleic acid molecule capable of regulating expression in plants which is the complement or reverse complement of any of the previously mentioned nucleic acid molecules under i) to vi).

Preferably, such homolog or fragment of the transcription regulating nucleotide molecule to be employed in the expression cassette of the invention (e.g., the sequences specified under ii), iii), iv) v), vi) and vii) above) is capable to modify transcription in a plant cell or organism, more preferably said homolog or fragment (e.g., the sequences specified under ii), iii), iv) v) and vi) above) has substantially the same transcription regulating activity as the transcription regulating nucleotide molecule described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18. Preferably, the homolog or fragment (e.g., the sequences specified under iii) above) has a sequence identity of at least 60% or 65%, preferably at least 70%, 75% or 80%, more preferably at least 90% or 95%, most preferably at least 97%, 98% or 99% to a sequence described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18. Preferably, the homologs or fragments (e.g., the sequences specified under iv) or vi) above) are hybridizing under stringent conditions (i.e. preferably medium stringent, more preferably high stringent conditions) with the specified target sequence.

In the applications US 61/419895 and EP 10193800.9 methods for the production of such homologs having the same expression pattern as the reference sequence as defined by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18 are described.

The homologs or fragments of the transcription regulating nucleotide molecule of the invention (e.g., the sequence described by any of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18) may be obtained or is obtainable from plant genomic DNA from a gene which is encoding an amino acid sequence having at least 90% amino acid sequence identity, more preferably at least 90% or 95%, most preferably at least 97% or 98% amino acid sequence identity, to a polypeptide as described by SEQ ID NO: 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 and 255.

The activity of a transcription regulating nucleotide molecule is considered equivalent if transcription is initiated in the same tissues as is by the reference molecule. Such expression profile is preferably demonstrated using reporter genes operably linked to said transcription regulating nucleotide sequence. Preferred reporter genes (Schenborn 1999) in this context are green fluorescence protein (GFP) (Chui 1996; Leffel 1997), chloramphenicol transferase, luciferase (Millar 1992), beta -glucuronidase or beta–galactosidase. Especially preferred is beta -glucuronidase (Jefferson 1987).

Beside this the transcription regulating activity of a functional equivalent homolog or fragment of the transcription regulating nucleotide molecule may vary from the activity of its parent sequence, especially with respect to expression level. The expression level may be higher or lower than the expression level of the parent sequence. Both derivations may be advantageous depending on the nucleic acid sequence of interest to be expressed. Preferred are such functional equivalent sequences, which – in comparison with its parent sequence – does, not deri-

vate from the expression level of said parent sequence by more than 50%, preferably 25%, more preferably 10% (as to be preferably judged by either mRNA expression or protein (e.g., reporter gene) expression). Furthermore preferred are equivalent sequences which demonstrate an increased expression in comparison to its parent sequence, preferably an increase by at least 50%, more preferably by at least 100%, most preferably by at least 500%.

Preferably a functional equivalent of the transcription regulating nucleotide molecule of the invention can be obtained or is obtainable from plant genomic DNA from a gene expressing a mRNA described by a cDNA comprising a sequence which is substantially similar and preferably has at least 90%, preferably at least 92% or 95%, more preferably at least 96% or 97%, most preferably at least 99% sequence identity to a sequence described by any SEQ ID NOs: 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 and 254. Preferably said transcription regulating nucleotide molecule exhibits promoter activity in the same tissue/s as the reference molecule as defined by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18.

Such functional equivalent of the transcription regulating nucleotide molecule may be obtained from other plant species by using the *Arabidopsis thaliana, or Glycine max* promoter sequences described herein as probes to screen for homologous structural genes in other plants by hybridization under low, moderate or stringent hybridization conditions. Regions of the promoter sequences of the present invention which are conserved among species could also be used as PCR primers to amplify a segment from a species other than *Arabidopsis thaliana, or Glycine max,* and that segment used as a hybridization probe (the latter approach permitting higher stringency screening) or in a transcription assay to determine promoter activity. Moreover, the promoter sequences could be employed to identify structurally related sequences in a database using computer algorithms.

More specifically, based on the nucleic acid sequences of the present invention, orthologs may be identified or isolated from the genome of any desired organism, preferably from another plant, according to well known techniques based on their sequence similarity to the *Arabidopsis thaliana, or Glycine max* nucleic acid sequences, e.g., hybridization, PCR or computer generated sequence comparisons. For example, all or a portion of a particular *Arabidopsis thaliana, or Glycine max* nucleic acid sequence is used as a probe that selectively hybridizes to other gene sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen source organism. Further, suitable genomic and cDNA libraries may be prepared from any cell or tissue of an organism. Such techniques include hybridization screening of plated DNA libraries (either plaques or colonies; see, e.g., Sambrook 1989) and amplification by PCR using oligonucleotide primers preferably corresponding to sequence domains conserved among related polypeptide or subsequences of the nucleotide sequences provided herein (see, e.g., Innis 1990). These methods are particularly well suited to the isolation of gene sequences from organisms closely related to the organism from which the probe sequence is derived. The application of these methods using the *Arabidopsis thaliana, or Glycine max* sequences as probes is well suited for the isolation of gene sequences from any source organism, preferably other plant species. In a PCR approach, oligonucleotide

primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any plant of interest. Methods for designing PCR primers and PCR cloning are generally known in the art.

In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen organism. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labeled with a detectable group such as ^{32}P, or any other detectable marker. Thus, for example, probes for hybridization can be made by labeling synthetic oligonucleotides based on the sequence of the invention. Methods for preparation of probes for hybridization and for construction of cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook et al. (1989). In general, sequences that hybridize to the sequences disclosed herein will have at least about 50% to 70% and even about 80% 85%, 90%, 95% to 98% or more identity with the disclosed sequences. That is, the sequence similarity of sequences may range, sharing at least about 50% to 70%, and even about 80%, 85%, 90%, 95% to 98% sequence similarity.

The nucleic acid molecules of the invention can also be identified by, for example, a search of known databases for genes encoding polypeptides having a specified amino acid sequence identity or DNA having a specified nucleotide sequence identity. Methods of alignment of sequences for comparison are well known in the art and are described hereinabove.

Hence, the isolated nucleic acid molecules of the invention include the orthologs of the *Arabidopsis thaliana, or Glycine max* sequences disclosed herein, i.e., the corresponding nucleotide sequences in organisms other than *Arabidopsis thaliana, or Glycine max* including, but not limited to, plants other than *Arabidopsis thaliana, or Glycine max*, preferably dicotyledonous plants, e.g., alfalfa, sunflower, rape seed, cotton, peanut, tobacco, or sugar beet, but also cereal plants such as corn, wheat, rye, turfgrass, sorghum, millet, sugarcane, barley and banana. An orthologous gene is a gene from a different species that encodes a product having the same or similar function, e.g., catalyzing the same reaction as a product encoded by a gene from a reference organism. Thus, an ortholog includes polypeptides having less than, e.g., 50% amino acid sequence identity, but which ortholog encodes a polypeptide having the same or similar function. Databases such GenBank may be employed to identify sequences related to the *Arabidopsis thaliana, or Glycine max* sequences, e.g., orthologs in other dicotyledonous plants. Alternatively, recombinant DNA techniques such as hybridization or PCR may be employed to identify sequences related to the *Arabidopsis thaliana, or Glycine max* sequences or to clone the equivalent sequences from different DNAs.

The transcription regulating nucleotide sequences of the invention or their functional equivalents can be obtained or isolated from any plant or non-plant source, or produced synthetically by purely chemical means. Preferred sources include, but are not limited to the plants defined in the DEFINITION section above.

Thus, another preferred embodiment of the invention relates to a method for identifying and/or isolating a transcription regulating nucleotide molecule characterized that said identification and/or isolation utilizes a nucleic acid molecule encoding a polypeptide comprising a sequence as described by SEQ ID NO: 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 and 255, or a part of said nucleic acid sequence. Preferred are nucleic acid molecules comprising nucleic acid sequences described by or comprising any of SEQ ID NO: 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 and 254 or parts thereof. "Part" in this context means a nucleic acid sequence of at least 15 consecutive nucleotides, preferably at least 25 consecutive nucleotides, more preferably at least 50 consecutive nucleotides.

The method for identification and/or isolation can be based on (but is not limited to) the methods described above such as polymerase chain reaction, hybridization or database screening. Preferably, this method of the invention is based on a polymerase chain reaction, wherein said nucleic acid sequence or its part is utilized as oligonucleotide primer. The person skilled in the art is aware of several methods to amplify and isolate the promoter of a gene starting from part of its coding sequence (such as, for example, part of a cDNA). Such methods may include but are not limited to method such as inverse PCR ("iPCR") or "thermal asymmetric interlaced PCR" ("TAIL PCR").

Thus, another embodiment of the invention relates to a method for providing or producing a transgenic expression cassette for heterologous expression in plants comprising the steps of:

I. isolating of a transcription regulating nucleotide molecule of a plant gene utilizing at least one nucleic acid sequence comprising any of SEQ ID NO: 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 and 254, or a part of at least 15, preferably at least 20 consecutive nucleotides of said nucleic acid sequence, and

II. functionally linking said transcription regulating nucleotide molecule to another nucleotide molecule of interest, which is heterologous in relation to said transcription regulating nucleotide molecule.

Still another embodiment of the invention relates to a method for providing a transgenic expression cassette for expression comprising the steps of:

I. isolating of a transcription regulating nucleotide molecule utilizing at least one nucleic acid molecule or a part thereof, wherein said molecule is encoding a protein comprising any of SEQ ID NO: 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253 and 255, or a part of at least 15 consecutive nucleotides thereof, and

II. functionally linking said transcription regulating nucleotide molecule to another nucleotide molecule of interest, which is heterologous in relation to said transcription regulating nucleotide molecule.

Preferably, the nucleic acid molecule employed for the isolation comprises at least 15 consecutive nucleotides, preferably at least 25 consecutive nucleotides, more preferably at least 50

consecutive nucleotides of a nucleic acid molecule comprising a sequence described by any of SEQ ID NO: 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 and 254. Preferably, the isolation of the transcription regulating nucleotide molecule is realized by a polymerase chain reaction utilizing said nucleic acid sequence as a primer. The operable linkage can be realized by standard cloning method known in the art such as ligation-mediated cloning or recombination-mediated cloning.

Preferably, the transcription regulating nucleotide sequences and promoters of the invention include a consecutive stretch of about 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably at least 500 consecutive bases, most preferably at least 750 consecutive bases, of any one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18, or the promoter orthologs thereof, which include the minimal promoter region.

Preferably, the transcription regulating nucleotide sequences and promoters of the invention include a consecutive stretch of about 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably at least 500 consecutive bases, most preferably at least 750 consecutive bases, of any one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18, or the promoter orthologs thereof, which include the minimal promoter region. In a particular embodiment of the invention said consecutive stretch of about 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably at least 500 consecutive bases, most preferably at least 750 consecutive bases, has at least 50% or 60%, preferably at least 70% or 80%, more preferably at least 90% even more preferably at least 95% or 97%, most preferably 98% or 99%, nucleic acid sequence identity with a corresponding consecutive stretch of about 250 consecutive bases, preferably at least 300 consecutive bases, more preferably at least 400 consecutive bases, even more preferably at least 500 consecutive bases, most preferably at least 750 consecutive bases, of any one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18 or the promoter orthologs thereof, which include the minimal promoter region. The above-defined stretch of contiguous nucleotides preferably comprises one or more promoter motifs selected from the group consisting of TATA box, GC-box, CAAT-box and a transcription start site.

The transcription regulating nucleotide sequences of the invention or their functional equivalents are capable of driving expression of a coding sequence in a target cell, particularly in a plant cell. The promoter sequences and methods disclosed herein are useful in regulating expression, respectively, of any heterologous nucleotide sequence in a host plant or part thereof in order to vary the phenotype of that plant. These promoters can be used with combinations of enhancer, upstream elements, and/or activating sequences from the 5' flanking regions of plant expressible structural genes. Similarly the upstream element can be used in combination with various plant promoter sequences.

The transcription regulating nucleotide sequences and promoters of the invention are useful to modify the phenotype of a plant. Various changes in the phenotype of a transgenic plant are

desirable, i.e., modifying the fatty acid composition in a plant, altering the amino acid content of a plant, altering a plant's pathogen defense mechanism, and the like. These results can be achieved by providing expression of heterologous products or increased expression of endogenous products in plants. Alternatively, the results can be achieved by providing for a reduction of expression of one or more endogenous products, particularly enzymes or cofactors in the plant. These changes result in an alteration in the phenotype of the transformed plant.

Generally, the transcription regulating nucleotide sequences and promoters of the invention may be employed to express a nucleic acid segment that is operably linked to said promoter such as, for example, an open reading frame, or a portion thereof, an anti-sense sequence, a sequence encoding for a double-stranded RNA sequence, or a transgene in plants.

An operable linkage may - for example – comprise an sequential arrangement of the transcription regulating nucleotide molecule of the invention (for example a sequence as described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89) with a nucleic acid sequence to be expressed, and – optionally – additional regulatory elements such as for example polyadenylation or transcription termination elements, enhancers, introns etc., in a way that the transcription regulating nucleotide molecule can fulfill its function in the process of expressing the nucleic acid sequence of interest under the appropriate conditions. The term "appropriate conditions" means preferably the presence of the expression cassette in a plant cell. Preferred are arrangements, in which the nucleic acid sequence of interest to be expressed is placed downstream (i.e., in 3'-direction) of the transcription regulating nucleotide molecule of the invention in a way, that both sequences are covalently linked. Optionally additional sequences may be inserted in-between the two sequences. Such sequences may be for example linker or multiple cloning sites. Furthermore, sequences can be inserted coding for parts of fusion proteins (in case a fusion protein of the protein encoded by the nucleic acid of interest is intended to be expressed). Preferably, the distance between the nucleic acid sequence of interest to be expressed and the transcription regulating nucleotide molecule of the invention is not more than 200 base pairs, preferably not more than 100 base pairs, more preferably no more than 50 base pairs.

An operable linkage in relation to any expression cassette or of the invention may be realized by various methods known in the art, comprising both *in vitro* and *in vivo* procedure. Thus, an expression cassette of the invention or an vector comprising such expression cassette may by realized using standard recombination and cloning techniques well known in the art (see e.g., Maniatis 1989; Silhavy 1984; Ausubel 1987).

An expression cassette may also be assembled by inserting a transcription regulating nucleotide molecule of the invention (for example a sequence as described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56,

57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89) into the plant genome. Such insertion will result in an operable linkage to a nucleic acid sequence of interest which as such already existed in the genome. By the insertion the nucleic acid of interest is expressed in a way due to the transcription regulating properties of the transcription regulating nucleotide sequence. The insertion may be directed or by chance. Preferably the insertion is directed and realized by for example homologous recombination. By this procedure a natural promoter may be exchanged against the transcription regulating nucleotide molecule of the invention, thereby modifying the expression profile of an endogenous gene. The transcription regulating nucleotide molecule may also be inserted in a way, that antisense mRNA of an endogenous gene is expressed, thereby inducing gene silencing.

Similar, a nucleic acid sequence of interest to be expressed may by inserted into a plant genome comprising the transcription regulating nucleotide molecule in its natural genomic environment (i.e. linked to its natural gene) in a way that the inserted sequence becomes operably linked to the transcription regulating nucleotide sequence, thereby forming an expression cassette of the invention.

The expression cassette may be employed for numerous expression purposes such as for example expression of a protein, or expression of an antisense RNA, sense or double-stranded RNA. Preferably, expression of the nucleic acid sequence confers to the plant an agronomically valuable trait.

The open reading frame to be linked to the transcription regulating nucleotide molecule of the invention may be obtained from an insect resistance gene, a disease resistance gene such as, for example, a bacterial disease resistance gene, a fungal disease resistance gene, a viral disease resistance gene, a nematode disease resistance gene, a herbicide resistance gene, a gene affecting grain composition or quality, a nutrient utilization gene, a mycotoxin reduction gene, a male sterility gene, a selectable marker gene, a screenable marker gene, a negative selectable marker, a positive selectable marker, a gene affecting plant agronomic characteristics, i.e., yield, standability, and the like, or an environment or stress resistance gene, i.e., one or more genes that confer herbicide resistance or tolerance, insect resistance or tolerance, disease resistance or tolerance (viral, bacterial, fungal, oomycete, or nematode), stress tolerance or resistance (as exemplified by resistance or tolerance to drought, heat, chilling, freezing, excessive moisture, salt stress, or oxidative stress), increased yields, food content and makeup, physical appearance, male sterility, drydown, standability, prolificacy, starch properties or quantity, oil quantity and quality, amino acid or protein composition, and the like. By "resistant" is meant a plant, which exhibits substantially no phenotypic changes as a consequence of agent administration, infection with a pathogen, or exposure to stress. By "tolerant" is meant a plant, which, although it may exhibit some phenotypic changes as a consequence of infection, does not have a substantially decreased reproductive capacity or substantially altered metabolism.

The expression regulating nucleotide sequences specified above may be optionally operably linked to other suitable regulatory sequences, e.g., a transcription terminator sequence, operator, repressor-binding site, transcription factor binding site and/or an enhancer.

The present invention further provides a recombinant vector containing the expression cassette of the invention, and host cells comprising the expression cassette or vector, e.g., comprising a plasmid. The expression cassette or vector may augment the genome of a transformed plant or may be maintained extra-chromosomally. The expression cassette or vector of the invention may be present in the nucleus, chloroplast, mitochondria and/or plastid of the cells of the plant. Preferably, the expression cassette or vector of the invention is comprised in the chromosomal DNA of the plant nucleus. The present invention also provides a transgenic plant prepared by this method, a seed from such a plant and progeny plants from such a plant including hybrids and inbreds. The expression cassette may be operatively linked to a structural gene, the open reading frame thereof, or a portion thereof. The expression cassette may further comprise a Ti plasmid and be contained in an Agrobacterium tumefaciens cell; it may be carried on a microparticle, wherein the microparticle is suitable for ballistic transformation of a plant cell; or it may be contained in a plant cell or protoplast. Further, the expression cassette or vector can be contained in a transformed plant or cells thereof, and the plant may be a dicot or a monocot. In particular, the plant may be a dicotyledonous plant. Preferred transgenic plants are transgenic maize, soybean, barley, alfalfa, sunflower, canola, soybean, cotton, peanut, sorghum, tobacco, sugarbeet, rice, wheat, rye, turfgrass, millet, sugarcane, tomato, or potato.

The invention also provides a method of plant breeding, e.g., to prepare a crossed fertile transgenic plant. The method comprises crossing a fertile transgenic plant comprising a particular expression cassette of the invention with itself or with a second plant, e.g., one lacking the particular expression cassette, to prepare the seed of a crossed fertile transgenic plant comprising the particular expression cassette. The seed is then planted to obtain a crossed fertile transgenic plant. The plant may be a monocot or a dicot. In a particular embodiment, the plant is a dicotyledonous plant. The crossed fertile transgenic plant may have the particular expression cassette inherited through a female parent or through a male parent. The second plant may be an inbred plant. The crossed fertile transgenic may be a hybrid. Also included within the present invention are seeds of any of these crossed fertile transgenic plants.

The transcription regulating nucleotide sequences of the invention further comprise sequences which are complementary to one (hereinafter "test" sequence) which hybridizes under stringent conditions with a nucleic acid molecule as described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88 and 89, as well as RNA which is transcribed from the nucleic acid molecule. When the hybridization is performed under stringent conditions, either the test or nucleic acid molecule of invention is preferably supported, e.g., on a membrane or DNA chip. Thus, either a denatured test or nucleic acid molecule of the invention is preferably first bound to a support and hybridization is effected for a specified period of time at a temperature of, e.g., between 55 and 70°C, in

double strength citrate buffered saline (SC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SC concentration. Depending upon the degree of stringency required such reduced concentration buffers are typically single strength SC containing 0.1% SDS, half strength SC containing 0.1% SDS and one-tenth strength SC containing 0.1% SDS. More preferably hybridization is carried out under high stringency conditions (as defined above).

Virtually any DNA composition may be used for delivery to recipient plant cells, e.g., dicotyledonous cells, to ultimately produce fertile transgenic plants in accordance with the present invention. For example, DNA segments or fragments in the form of vectors and plasmids, or linear DNA segments or fragments, in some instances containing only the DNA element to be expressed in the plant, and the like, may be employed. The construction of vectors, which may be employed in conjunction with the present invention, will be known to those of skill of the art in light of the present disclosure (see, e.g., Sambrook 1989; Gelvin 1990).

The nucleotide sequence of interest linked to one or more of the transcription regulating nucleotide sequences of the invention can, for example, code for a ribosomal RNA, an antisense RNA or any other type of RNA that is not translated into protein. In another preferred embodiment of the invention, said nucleotide sequence of interest is translated into a protein product. The transcription regulating nucleotide molecule and/or nucleotide sequence of interest linked thereto may be of homologous or heterologous origin with respect to the plant to be transformed. A recombinant DNA molecule useful for introduction into plant cells includes that which has been derived or isolated from any source, that may be subsequently characterized as to structure, size and/or function, chemically altered, and later introduced into plants. An example of a nucleotide sequence or segment of interest "derived" from a source, would be a nucleotide sequence or segment that is identified as a useful fragment within a given organism, and which is then chemically synthesized in essentially pure form. An example of such a nucleotide sequence or segment of interest "isolated" from a source, would be nucleotide sequence or segment that is excised or removed from said source by chemical means, e.g., by the use of restriction endonucleases, so that it can be further manipulated, e.g., amplified, for use in the invention, by the methodology of genetic engineering. Such a nucleotide sequence or segment is commonly referred to as "recombinant."

Therefore a useful nucleotide sequence, segment or fragment of interest includes completely synthetic DNA, semi-synthetic DNA, DNA isolated from biological sources, and DNA derived from introduced RNA. Generally, the introduced DNA is not originally resident in the plant genotype which is the recipient of the DNA, but it is within the scope of the invention to isolate a gene from a given plant genotype, and to subsequently introduce multiple copies of the gene into the same genotype, e.g., to enhance production of a given gene product such as a storage protein or a protein that confers tolerance or resistance to water deficit.

The introduced recombinant DNA molecule includes but is not limited to, DNA from plant genes, and non-plant genes such as those from bacteria, yeasts, animals or viruses. The introduced DNA can include modified genes, portions of genes, or chimeric genes, including genes from

the same or different genotype. The term "chimeric gene" or "chimeric DNA" is defined as a gene or DNA sequence or segment comprising at least two DNA sequences or segments from species which do not combine DNA under natural conditions, or which DNA sequences or segments are positioned or linked in a manner which does not normally occur in the native genome of untransformed plant.

The introduced recombinant DNA molecule used for transformation herein may be circular or linear, double-stranded or single-stranded. Generally, the DNA is in the form of chimeric DNA, such as plasmid DNA that can also contain coding regions flanked by regulatory sequences, which promote the expression of the recombinant DNA present in the resultant plant. Generally, the introduced recombinant DNA molecule will be relatively small, i.e., less than about 30 kb to minimize any susceptibility to physical, chemical, or enzymatic degradation which is known to increase as the size of the nucleotide molecule increases. As noted above, the number of proteins, RNA transcripts or mixtures thereof, which is introduced into the plant genome, is preferably preselected and defined, e.g., from one to about 5-10 such products of the introduced DNA may be formed.

Two principal methods for the control of expression are known, viz.: overexpression and underexpression. Overexpression can be achieved by insertion of one or more than one extra copy of the selected gene. It is, however, not unknown for plants or their progeny, originally transformed with one or more than one extra copy of a nucleotide sequence, to exhibit the effects of underexpression as well as overexpression. For underexpression there are two principle methods, which are commonly referred to in the art as "antisense downregulation" and "sense downregulation" (sense downregulation is also referred to as "cosuppression"). Generically these processes are referred to as "gene silencing". Both of these methods lead to an inhibition of expression of the target gene.

Obtaining sufficient levels of transgene expression in the appropriate plant tissues is an important aspect in the production of genetically engineered crops. Expression of heterologous DNA sequences in a plant host is dependent upon the presence of an operably linked promoter that is functional within the plant host. Choice of the promoter sequence will determine when and where within the organism the heterologous DNA sequence is expressed.

It is specifically contemplated by the inventors that one could mutagenize a promoter to potentially improve the utility of the elements for the expression of transgenes in plants. The mutagenesis of these elements can be carried out at random and the mutagenized promoter sequences screened for activity in a trial-by-error procedure. Alternatively, particular sequences which provide the promoter with desirable expression characteristics, or the promoter with expression enhancement activity, could be identified and these or similar sequences introduced into the sequences via mutation. It is further contemplated that one could mutagenize these sequences in order to enhance their expression of transgenes in a particular species.

The means for mutagenizing a DNA segment encoding a promoter sequence of the current invention are well known to those of skill in the art. As indicated, modifications to promoter or

other regulatory element may be made by random, or site-specific mutagenesis procedures. The promoter and other regulatory element may be modified by altering their structure through the addition or deletion of one or more nucleotides from the sequence which encodes the corresponding unmodified sequences.

Mutagenesis may be performed in accordance with any of the techniques known in the art, such as, and not limited to, synthesizing an oligonucleotide having one or more mutations within the sequence of a particular regulatory region. In particular, site-specific mutagenesis is a technique useful in the preparation of promoter mutants, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to about 75 nucleotides or more in length is preferred, with about 10 to about 25 or more residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector, which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phages are readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids also are routinely employed in site directed mutagenesis, which eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double stranded vector which includes within its sequence a DNA sequence which encodes the promoter. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as E. coli polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform or transfect appropriate cells, such as E. coli cells, and cells are selected which include recombinant vectors bearing the mutated sequence arrangement. Vector DNA can then be isolated from these cells and used for plant transformation. A genetic selection scheme was devised by Kunkel et al. (1987) to enrich for clones incorporating mutagenic oligonucleotides. Alternatively, the use of PCR with commercially available thermostable enzymes such as Taq polymerase may be used to incorporate a mutagenic oligonucleotide primer into an amplified DNA fragment that can then be cloned into an appropriate cloning or expression vector. The PCR-mediated mutagenesis procedures of Tomic et al. (1990) and Upender et al. (1995) provide two examples of such protocols. A PCR employing a thermostable ligase in

addition to a thermostable polymerase also may be used to incorporate a phosphorylated mutagenic oligonucleotide into an amplified DNA fragment that may then be cloned into an appropriate cloning or expression vector. The mutagenesis procedure described by Michael (1994) provides an example of one such protocol.

The preparation of sequence variants of the selected promoter-encoding DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting, as there are other ways in which sequence variants of DNA sequences may be obtained. For example, recombinant vectors encoding the desired promoter sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

As used herein; the term "oligonucleotide directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" also is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template-dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson and Rarnstad, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U.S. Pat. No. 4,237,224. A number of template-dependent processes are available to amplify the target sequences of interest present in a sample, such methods being well known in the art and specifically disclosed herein below.

Where a clone comprising a promoter has been isolated in accordance with the instant invention, one may wish to delimit the essential promoter regions within the clone. One efficient, targeted means for preparing mutagenizing promoters relies upon the identification of putative regulatory elements within the promoter sequence. This can be initiated by comparison with promoter sequences known to be expressed in similar tissue-specific or developmentally unique manner. Sequences, which are shared among promoters with similar expression patterns, are likely candidates for the binding of transcription factors and are thus likely elements that confer expression patterns. Confirmation of these putative regulatory elements can be achieved by deletion analysis of each putative regulatory region followed by functional analysis of each deletion construct by assay of a reporter gene, which is functionally attached to each construct. As such, once a starting promoter sequence is provided, any of a number of different deletion mutants of the starting promoter could be readily prepared.

Functionally equivalent fragments of a transcription regulating nucleotide molecule of the invention can also be obtained by removing or deleting non-essential sequences without deleting the essential one. Narrowing the transcription regulating nucleotide molecule to its essential, transcription mediating elements can be realized in vitro by trial-and-error deletion mutations, or in silico using promoter element search routines. Regions essential for promoter activity often

demonstrate clusters of certain, known promoter elements. Such analysis can be performed using available computer algorithms such as PLACE ("Plant Cis-acting Regulatory DNA Elements"; Higo 1999), the BIOBASE database "Transfac" (Biologische Datenbanken GmbH, Braunschweig; Wingender 2001) or the database PlantCARE (Lescot 2002).

A method for producing such regulatory nucleic acid molecules with mutated sequences regulating the same expression specificity as the parent sequence or reference sequence is for example defined in US 61/419895 and EP 10193800.9.

Preferably, functional equivalent fragments of one of the transcription regulating nucleotide sequences of the invention comprises at least 250 base pairs, preferably, at least 300 base pairs, more preferably at least 400 base pairs, even more preferably 500 base pairs, most preferably 750 base pairs of a transcription regulating nucleotide molecule as described by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 or 18, . More preferably this fragment is starting from the 3'-end of the indicated sequences.

Especially preferred are equivalent fragments of transcription regulating nucleotide sequences, which are obtained by deleting the region encoding the 5'-untranslated region of the mRNA, thus only providing the (untranscribed) promoter region. The 5'-untranslated region can be easily determined by methods known in the art (such as 5'-RACE analysis). Accordingly, some of the transcription regulating nucleotide sequences of the invention are equivalent fragments of other sequences.

An expression cassette of the invention may comprise further regulatory elements. The term in this context is to be understood in a broad meaning comprising all sequences, which may influence construction or function of the expression cassette. Regulatory elements may for example modify transcription and/or translation in prokaryotic or eukaryotic organism. In an preferred embodiment the expression cassette of the invention comprised downstream (in 3'-direction) of the nucleic acid sequence to be expressed a transcription termination sequence and – optionally additional regulatory elements – each operably liked to the nucleic acid sequence to be expressed (or the transcription regulating nucleotide sequence).

Additional regulatory elements may comprise additional promoter, minimal promoters, or promoter elements, which may modify the expression regulating properties. For example the expression may be made depending on certain stress factors such water stress, abscisin (Lam 1991) or heat stress (Schoffl 1989). Furthermore additional promoters or promoter elements may be employed, which may realize expression in other organisms (such as E.coli or Agrobacterium). Such regulatory elements can be found in the promoter sequences or bacteria such as amy and SPO2 or in the promoter sequences of yeast or fungal promoters (such as ADC1, MFa, AC, P-60, CYC1, GAPDH, TEF, rp28, and ADH).

Furthermore, it is contemplated that promoters combining elements from more than one promoter may be useful. For example, US 5,491,288 discloses combining a Cauliflower Mosaic Virus promoter with a histone promoter. Thus, the elements from the promoters disclosed herein may be combined with elements from other promoters. Promoters, which are useful for

plant transgene expression include those that are inducible, viral, synthetic, constitutive (Odell 1985), temporally regulated, spatially regulated, tissue-specific, and spatial-temporally regulated.

Where expression in specific tissues or organs is desired, tissue-specific promoters may be used. In contrast, where gene expression in response to a stimulus is desired, inducible promoters are the regulatory elements of choice. Such inducible promoters may additionally be tissue or organ specifically expressed and or induced. Where continuous expression is desired throughout the cells of a plant, constitutive promoters are utilized. Additional regulatory sequences upstream and/or downstream from the core promoter sequence may be included in expression constructs of transformation vectors to bring about varying levels of expression of heterologous nucleotide sequences in a transgenic plant.

A variety of 5' and 3' transcriptional regulatory sequences are available for use in the present invention. Transcriptional terminators are responsible for the termination of transcription and correct mRNA polyadenylation. The 3' nontranslated regulatory DNA sequence preferably includes from about 50 to about 1,000, more preferably about 100 to about 1,000, nucleotide base pairs and contains plant transcriptional and translational termination sequences. Appropriate transcriptional terminators and those which are known to function in plants include the CaMV 35S terminator, the tml terminator, the nopaline synthase terminator, the pea rbcS E9 terminator, the terminator for the T7 transcript from the octopine synthase gene of Agrobacterium tumefaciens, and the 3' end of the protease inhibitor I or II genes from potato or tomato, although other 3' elements known to those of skill in the art can also be employed. Alternatively, one also could use a gamma coixin, oleosin 3 or other terminator from the genus Coix.

Preferred 3' elements include those from the nopaline synthase gene of Agrobacterium tumefaciens (Bevan 1983), the terminator for the T7 transcript from the octopine synthase gene of Agrobacterium tumefaciens, and the 3' end of the protease inhibitor I or II genes from potato or tomato.

As the DNA sequence between the transcription initiation site and the start of the coding sequence, i.e., the untranslated leader sequence, can influence gene expression, one may also wish to employ a particular leader sequence. Preferred leader sequences are contemplated to include those, which include sequences, predicted to direct optimum expression of the attached gene, i.e., to include a preferred consensus leader sequence, which may increase or maintain mRNA stability and prevent inappropriate initiation of translation. The choice of such sequences will be known to those of skill in the art in light of the present disclosure. Sequences that are derived from genes that are highly expressed in plants will be most preferred.

Preferred regulatory elements also include the 5'-untranslated region, introns and the 3'-untranslated region of genes. Such sequences that have been found to enhance gene expression in transgenic plants include intron sequences (e.g., from Adh1, bronze1, actin1, actin 2 (WO 00/760067), or other plant introns as disclosed in WO2011/023537, WO2011/023539, WO2006/094976.

Additional preferred regulatory elements are enhancer sequences or polyadenylation sequences.

The heterologous nucleotide sequence to be expressed is preferably furthermore operably linked to 3'-untranslated regions, transcription termination and/or polyadenylation signal. 3'-untranslated regions are suitable to stabilize mRNA expression and structure. This can result in prolonged presence of the mRNA and thus enhanced expression levels. Termination and polyadenylation signals are suitable to stabilize mRNA expression, to ensure constant mRNA transcript length and to prevent read-through transcription. Especially in multigene expression constructs this is an important feature. Furthermore correct termination of transcription is linked to re-initiation of transcription from the regulatory 5´nucleotide sequence resulting in enhanced expression levels. The above-mentioned signals can be any signal functional in plants and can for example be isolated from plant genes, plant virus genes or other plant pathogens. However, in a preferred embodiment the 3'-untranslated regions, transcription termination and polyadenylation signals are from the genes employed as the source for the promoters of this invention.

Preferred polyadenylation sequences are those from plant genes or Agrobacterium T-DNA genes (such as for example the terminator sequences of the OCS (octopine synthase) or NOS (nopaline synthase) genes).
Examples of enhancers include elements from the CaMV 35S promoter, octopine synthase genes (Ellis el al., 1987), the rice actin I gene, the maize alcohol dehydrogenase gene (Callis 1987), the maize shrunken I gene (Vasil 1989), TMV Omega element (Gallie 1989) and promoters from non-plant eukaryotes (e.g. yeast; Ma 1988). Vectors for use in accordance with the present invention may be constructed to include the ocs enhancer element. This element was first identified as a 16 bp palindromic enhancer from the octopine synthase (ocs) gene of ultilane (Ellis 1987), and is present in at least 10 other promoters (Bouchez 1989). The use of an enhancer element, such as the ocs elements and particularly multiple copies of the element, will act to increase the level of transcription from adjacent promoters when applied in the context of plant transformation.

An expression cassette of the invention (or a vector derived thereof) may comprise additional functional elements, which are to be understood in the broad sense as all elements, which influence construction, propagation, or function of an expression cassette or a vector or a transgenic organism comprising them. Such functional elements may include origin of replications (to allow replication in bacteria; for the ORI of pBR322 or the P15A ori; Sambrook 1989), or elements required for Agrobacterium T-DNA transfer (such as for example the left and/or rights border of the T-DNA).

Ultimately, the most desirable DNA segments for introduction into, for example, a dicot genome, may be homologous genes or gene families which encode a desired trait (e.g., increased yield per acre, fungal resistance) and which are introduced under the control of novel promoters or enhancers, etc., or perhaps even homologous or tissue specific (e.g., root-, collar/sheath-, whorl-, stalk-, earshank-, kernel- or leaf-specific) promoters or control elements. Indeed, it is

envisioned that a particular use of the present invention will be the expression of a gene in the epidermis or mesophyll or inducible in the epidermis and/or mesophyll.

Additionally, vectors may be constructed and employed in the intracellular targeting of a specific gene product within the cells of a transgenic plant or in directing a protein to the extracellular environment. This will generally be achieved by joining a DNA sequence encoding a transit or signal peptide sequence to the coding sequence of a particular gene. The resultant transit or signal peptide will transport the protein to a particular intracellular or extracellular destination, respectively, and will then be post-translationally removed. Transit or signal peptides act by facilitating the transport of proteins through intracellular membranes, e.g., vacuole, vesicle, plastid and mitochondrial membranes, whereas signal peptides direct proteins through the extracellular membrane.

A particular example of such a use concerns the direction of a herbicide resistance gene, such as the EPSPS gene, to a particular organelle such as the chloroplast rather than to the cytoplasm. This is exemplified by the use of the rbcs transit peptide which confers plastid-specific targeting of proteins. In addition, it is proposed that it may be desirable to target certain genes responsible for male sterility to the mitochondria, or to target certain genes for resistance to phytopathogenic organisms to the extracellular spaces, or to target proteins to the vacuole.

By facilitating the transport of the protein into compartments inside and outside the cell, these sequences may increase the accumulation of gene product protecting them from proteolytic degradation. These sequences also allow for additional mRNA sequences from highly expressed genes to be attached to the coding sequence of the genes. Since mRNA being translated by ribosomes is more stable than naked mRNA, the presence of translatable mRNA in front of the gene may increase the overall stability of the mRNA transcript from the gene and thereby increase synthesis of the gene product. Since transit and signal sequences are usually post-translationally removed from the initial translation product, the use of these sequences allows for the addition of extra translated sequences that may not appear on the final polypeptide. Targeting of certain proteins may be desirable in order to enhance the stability of the protein (US 5,545,818).

It may be useful to target DNA itself within a cell. For example, it may be useful to target introduced DNA to the nucleus as this may increase the frequency of transformation. Within the nucleus itself it would be useful to target a gene in order to achieve site-specific integration. For example, it would be useful to have a gene introduced through transformation replace an existing gene in the cell. Other elements include those that can be regulated by endogenous or exogenous agents, e.g., by zinc finger proteins, including naturally occurring zinc finger proteins or chimeric zinc finger proteins (see, e.g., US 5,789,538, WO 99/48909; WO 99/45132; WO 98/53060; WO 98/53057; WO 98/53058; WO 00/23464; WO 95/19431; and WO 98/54311) or myb-like transcription factors. For example, a chimeric zinc finger protein may include amino acid sequences, which bind to a specific DNA sequence (the zinc finger) and amino acid sequences that activate (e.g., GAL 4 sequences) or repress the transcription of the sequences linked to the specific DNA sequence.

It is one of the objects of the present invention to provide recombinant DNA molecules comprising a nucleotide sequence according to the invention operably linked to a nucleotide segment of interest.

A nucleotide segment of interest is reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest changes, and as developing nations open up world markets, new crops and technologies will also emerge. In addition, as the understanding of agronomic traits and characteristics such as yield and heterosis increase, the choice of genes for transformation will change accordingly. General categories of nucleotides of interest include, for example, genes involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. More specific categories of transgenes, for example, include genes encoding important traits for agronomics, insect resistance, disease resistance, herbicide resistance, sterility, grain characteristics, and commercial products. Genes of interest include, generally, those involved in starch, oil, carbohydrate, or nutrient metabolism, as well as those affecting kernel size, sucrose loading, zinc finger proteins, see, e.g., US 5,789,538, WO 99/48909; WO 99/45132; WO 98/53060; WO 98/53057; WO 98/53058; WO 00/23464; WO 95/19431; and WO 98/54311, and the like.

One skilled in the art recognizes that the expression level and regulation of a transgene in a plant can vary significantly from line to line. Thus, one has to test several lines to find one with the desired expression level and regulation. Once a line is identified with the desired regulation specificity of a chimeric Cre transgene, it can be crossed with lines carrying different inactive replicons or inactive transgene for activation.

Other sequences, which may be linked to the gene of interest, which encodes a polypeptide, are those which can target to a specific organelle, e.g., to the mitochondria, nucleus, or plastid, within the plant cell. Targeting can be achieved by providing the polypeptide with an appropriate targeting peptide sequence, such as a secretory signal peptide (for secretion or cell wall or membrane targeting, a plastid transit peptide, a chloroplast transit peptide, e.g., the chlorophyll a/b binding protein, a mitochondrial target peptide, a vacuole targeting peptide, or a nuclear targeting peptide, and the like. For example, the small subunit of ribulose bisphosphate carboxylase transit peptide, the EPSPS transit peptide or the dihydrodipicolinic acid synthase transit peptide may be used, for examples of plastid organelle targeting sequences (see WO 00/12732). Plastids are a class of plant organelles derived from proplastids and include chloroplasts, leucoplasts, amyloplasts, and chromoplasts. The plastids are major sites of biosynthesis in plants. In addition to photosynthesis in the chloroplast, plastids are also sites of lipid biosynthesis, nitrate reduction to ammonium, and starch storage. And while plastids contain their own circular genome, most of the proteins localized to the plastids are encoded by the nuclear genome and are imported into the organelle from the cytoplasm.

Transgenes used with the present invention will often be genes that direct the expression of a particular protein or polypeptide product, but they may also be non-expressible DNA segments, e.g., transposons such as Ds that do no direct their own transposition. As used herein, an "ex-

pressible gene" is any gene that is capable of being transcribed into RNA (e.g., mRNA, antisense RNA, etc.) or translated into a protein, expressed as a trait of interest, or the like, etc., and is not limited to selectable, screenable or non-selectable marker genes. The invention also contemplates that, where both an expressible gene that is not necessarily a marker gene is employed in combination with a marker gene, one may employ the separate genes on either the same or different DNA segments for transformation. In the latter case, the different vectors are delivered concurrently to recipient cells to maximize cotransformation.

The choice of the particular DNA segments to be delivered to the recipient cells will often depend on the purpose of the transformation. One of the major purposes of transformation of crop plants is to add some commercially desirable, agronomically important traits to the plant. Such traits include, but are not limited to, herbicide resistance or tolerance; insect resistance or tolerance; disease resistance or tolerance (viral, bacterial, fungal, nematode); stress tolerance and/or resistance, as exemplified by resistance or tolerance to drought, heat, chilling, freezing, excessive moisture, salt stress; oxidative stress; increased yields; food content and makeup; physical appearance; male sterility; drydown; standability; prolificacy; starch properties; oil quantity and quality; and the like. One may desire to incorporate one or more genes conferring any such desirable trait or traits, such as, for example, a gene or genes encoding pathogen resistance.

In certain embodiments, the present invention contemplates the transformation of a recipient cell with more than one advantageous transgene. Two or more transgenes can be supplied in a single transformation event using either distinct transgene-encoding vectors, or using a single vector incorporating two or more gene coding sequences. For example, plasmids bearing the bar and aroA expression units in either convergent, divergent, or colinear orientation, are considered to be particularly useful. Further preferred combinations are those of an insect resistance gene, such as a Bt gene, along with a protease inhibitor gene such as pinII, or the use of bar in combination with either of the above genes. Of course, any two or more transgenes of any description, such as those conferring herbicide, insect, disease (viral, bacterial, fungal, nematode) or drought resistance, male sterility, drydown, standability, prolificacy, starch properties, oil quantity and quality, or those increasing yield or nutritional quality may be employed as desired.

EXAMPLES

Materials and General Methods

Unless indicated otherwise, chemicals and reagents in the Examples were obtained from Sigma Chemical Company (St. Louis, MO), restriction endonucleases were from New England Biolabs (Beverly, MA) or Roche (Indianapolis, IN), oligonucleotides were synthesized by MWG Biotech Inc. (High Point, NC), and other modifying enzymes or kits regarding biochemicals and molecular biological assays were from Clontech (Palo Alto, CA), Pharmacia Biotech (Piscataway, NJ), Promega Corporation (Madison, WI), or Stratagene (La Jolla, CA). Materials for cell culture media were obtained from Gibco/BRL (Gaithersburg, MD) or DIFCO (Detroit, MI). The cloning steps carried out for the purposes of the present invention, such as, for example, restriction cleavages, agarose gel electrophoresis, purification of DNA fragments, transfer of nucleic acids

to nitrocellulose and nylon membranes, linking DNA fragments, transformation of *E. coli* cells, growing bacteria, multiplying phages and sequence analysis of recombinant DNA, are carried out as described by Sambrook (1989). The sequencing of recombinant DNA molecules is carried out using ABI laser fluorescence DNA sequencer following the method of Sanger (Sanger 1977).

Example 1: Identification and validation of promoters from soybean putatively conferring constitutive expression or expression in epidermis of a plant

1.1 Identification of promoters putatively conferring constitutive expression or expression in epidermis

A soybean gene expression profiling analysis was carried out by a commercial supplier of AFLP comparative expression technology (Keygene N.V., P.O.Box 216, 6700 AE Wageningen, The Netherlands) using RNA samples from 34 soybean tissues generated by BASF (Table 1). AFLP bands were selected for constitutive expression in all tissues and expression in epidermis.

Table1: Soybean samples for AFLP screen

Sample #	Tissue	Source	Stage	Treatment
1	Leaf - Epidermis	unifoliolate leaf	V2-3	
2	Leaf - mesophyll	unifoliolate leaf	V2-3	
3	Leaf - Epidermis*	unifoliolate leaf	V2-3	mock 8 + 16h
4	Leaf - Epidermis*	unifoliolate leaf	V2-3	+ ASR 8 hai
5	Leaf - Epidermis*	unifoliolate leaf	V2-3	+ ASR 16 hai
6	Leaf - Epidermis*	unifoliolate leaf	V2-3	+ ASR 112 hai
7	Leaf - Mesophyll*	unifoliolate leaf	V2-3	mock 8 + 16h
8	Leaf - Mesophyll*	unifoliolate leaf	V2-3	+ ASR 16 hai
9	Leaf - Mesophyll*	unifoliolate leaf	V2-3	+ ASR 112 hai
10	Leaf	unifoliolate leaves	V2-3	
11	Leaf	trifoliolate leaves	V2-3	
12	Leaf	trifoliolate leaves	R1-2	
13	Leaf	trifoliolate leaves	R7	
14	Stem	complete	VC	
15	Stem	complete	V2-3	
16	Stem	complete	R2	
17	Shoot tip	complete	VC	
18	root	complete	VC	
19	root	complete	R2	
20	flowers	buds	R1	
21	fowers	complete	R2	
22	embryo	18-20 days	R5	
23	embryo	5-9 mm	R5	
24	embryo	complete	R6	

25	embryo	complete	R7	
26	embryo	complete	R8	
27	whole seeds	14 days	R4	
28	endosperm	14 days	R4 early	
29	endosperm	complete	R4 late	
30	endosperm	complete	R5	
31	endosperm	complete	R6	
32	siliques	- seeds	R3	
33	siliques	- seeds	R4	
34	siliques	- seeds	R6	

hai = hours after infection; ASR = Asian Soybean Rust

1.2 Identification of the genes corresponding to AFLP bands

Expressed Sequence Tag (EST) sequences of AFLP bands were used as query for BLASTN searching against a soybean sequence database. The corresponding genes are listed in table 2.

Table 2: Overview over corresponding genes for G. max promoters conferring constitutive expression and expression in the epidermis

Feature name	SEQ ID #
Glyma11g14950_gene	220
Glyma14g06680_gene	222
Glyma02g47670_gene	224
Glyma14g02930_gene	226
Glyma17g27610_gene	228

1.3 Confirmation of allele-specific expression pattern using quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR)

In order to confirm the native expression patterns of the identified soybean genes in an allele-specific manner, quantitative reverse transcription PCR (qRT-PCR) was performed using total RNA isolated from the same materials as were used for the AFLP expression profiling (Table 1).

Primers for qRT-PCR were designed based on the sequences of the isolated EST fragments using the Vector NTI software package (Invitrogen, Carlsbad, CA, USA). Primers were designed to distinguish individual alleles of the candidate gene in the tetraploid soybean genome. Primers for qRT-PCR are listed in table 3. The tubulin gene served as a control for normalization purposes.

Table 3: Primer sequences for qRT-PCR

Feature name	Primer	Sequence	SEQ ID #
Glyma11g14950	Loy1294	CCTTCATAGACCTGAATCAACACACCG	90
	Loy1296	GGAGGAGTCATGACTGTGTTGATTCC	91
Glyma14g06680	Loy1275	CGCCGGGTTTATGTGTC	92

	Loy1278	CCGGGGCTAAGTCTAAGTGT	93
Glyma02g47670	Loy1420	CGTCTGCTACAGCGTGTGGAAGGACGAGG	94
	Loy1421	GAGACGTGGCGGTGCTTCTTGCGGTAATC	95
Glyma14g02930	Loy1425	TGGAATCAAAGACAGGTAGACTGGC	96
	Loy1427	CTGCTTTCAGTGTAATGGTTTCCAGA	97
Glyma17g27610	Loy1429	TTGTCTGGTTTGGAAAGAAGAAAGTTGTGA	98
	Loy1430	CACACAGAGCACAAGGAATAGTGGCAAT	99
Tubulin	Loy1145	TGGGAATCCACTCAACGAAGT	100
	Loy1146	CCTGACAGCATCAGCCATGT	101

qRT-PCR was performed using QuantiTect Kit (Qiagen, Hilden, Germany) and SYBR Green qPCR Master Mix (Roche Diagnostics, Mannheim, Germany) in a Roche LightCycler (Roche Diagnostics, Mannheim, Germany). cDNA was synthesized using 800 ng of total RNA and 1 µl reverse transcriptase in a 20 µl volume. The cDNA was diluted with 60 µl of RNAse free water to a final volume of 80 µl. 4 µl of diluted cDNA were used in a 10 µl PCR reaction according to manufacturer's instruction. The thermocycling conditions were as follows: Denature at 95°C for 2 minutes, and run 45 cycles at 95°C for 10 seconds and 60°C for 20 seconds and 72°C for 20 seconds for amplification. After the final cycle of the amplification, the dissociation curve analysis was carried out to verify that the amplification occurred specifically and no primer dimer product was generated during the amplification process. The tubulin gene (primer sequences in table 3) was used as an endogenous reference gene to normalize the calculation using the Comparative Ct (Cycle of threshold) value method. The DeltaCt value was obtained by subtracting the Ct value of tubulin gene from the Ct value of the respective candidate gene, and the relative transcription quantity (expression level) of the candidate gene was expressed as $2^{-DeltaCt}$.

1.4 Identification of the promoter region

For promoter identification purposes, the sequence upstream of the start codon of the identified genes was defined as the respective promoter. To characterize these promoter regions, 5´RACE PCR analyses were performed using the primers listed in table 4.

Table 4: Primer sequences for 5´RACE PCR

Feature name	Primer	Sequence	SEQ ID #
Glyma11g14950	Loy1685	CATCGCCGATCAACCGTTCTGTG	102
Glyma14g06680	Loy1682	GACTTAGCCCCGGCGACTCCCATA	103
Glyma02g47670	Loy1649	TGATCAAACGCTCTGTAAACTTTCTTCACA	104
Glyma14g02930	Loy1665	GACTGAACTGGGGTTGAAGGTGAACACT	105
Glyma17g27610	Loy1672	CATCTTCTGGTGCCGAGGCAGGGAT	106

1.5 Isolation of the promoter region by PCR amplification

The promoter regions of the respective genes were isolated via genomic PCR using the following sequence specific primers (table 5):

Promoters putatively conferring constitutive expression or expression in the epidermis amplified with these primer pairs are listed in table 6.

In addition 5´-deletions of the promoters are made by using different 5´ primers in combination with the same 3´ oligonucleotide primer. The resulting promoters are indicated by their respective length in base pairs (cp. table 6) and the corresponding primer pairs for PCR are listed in table 5.

Promoters are bioinformatically analyzed with MatInspector Professional 8.0.4, Release August 2010 (Genomatix Software GmbH; Munich, Germany) for transcription factor binding sites. Regions free of core motifs are permutated and the resulting nucleic acid sequences are synthesized. The principles for generating permutated promoters that retain their respective tissue specificities are described in EP10193800.9 und US61/419895. The resulting sequences are indicated by the original identifier of the corresponding gene followed by "_perm" (cp. table 6).

Table 5: Primer sequences for PCR amplification of promoters putatively conferring constitutive expression or expression in the epidermis

Feature name	Primer	Sequence	SEQ ID #
Glyma11g14950_1939bp	Loy1436	TATATAGGTACCAAAGAGCCAAGTTGTTATTC	107
	Loy1437	TATATACCATGGTACTCACTCACACACAAAC	108
Glyma14g06680_1056bp	Loy1432	TATATAGGTACCATTTCCAACTCCTGACTGAGA	109
	Loy1433	TATATACCATGGTCTTTCTCCTCGCCTGGGA	110
Glyma02g47670_1753bp	Loy1490	TATATAGGTACCTCTCAATCAAGGCCTTTAT	111
	Loy1491	TATATACCATGGTTAATTAATTTCAATCTCTCCCTCTCTAT	112
Glyma14g02930_1688bp	Loy1492	TATATAGGTACCCGGTTATTCTTAATCCTTTTCA	113
	Loy1493	TATATACCATGGTTAATTAAGCTGTGTGTGACCACTGATG	114
Glyma17g27610_1889bp	Loy1494	TATATAGGTACCGATTCTAGATATTGAAGTTTGTGA	115
	Loy1495	TATATACCATGGTTAATTAATGTTGTGTTAACAAAGGGT	116
Glyma11g14950_500bp	Loy1436.500	TATATAGGTACCCGCGCTTTACACGGAGTTAGTGAA	117
	Loy1437	TATATACCATGGTACTCACTCACACACAAAC	108
Glyma11g14950_1000bp	Loy1436.1000	TATATAGGTACCAAGAAAAAAAACATATCGGAGGAGGA	118
	Loy1437	TATATACCATGGTACTCACTCACACACAAAC	108
Glyma11g14950_1500bp	Loy1436.1500	TATATAGGTACCAATTTCAATTTCTCACCTTTTTAATTGT	119
	Loy1437	TATATACCATGGTACTCACTCACACACAAAC	108
Glyma14g06680_500bp	Loy1432.500	TATATAGGTACCGGAGAAAGAAAAACTGTTGAC	120
	Loy1433	TATATACCATGGTCTTTCTCCTCGCCTGGGA	110
Glyma14g06680_700bp	Loy1432.700	TATATAGGTACCATTTATACCACATGTGGGAAGTATTG	121
	Loy1433	TATATACCATGGTCTTTCTCCTCGCCTGGGA	110
Glyma14g06680_1000bp	Loy1432.1000	TATATAGGTACCCATCTTTCACGCTACAAAACATTGGT	122
	Loy1433	TATATACCATGGTCTTTCTCCTCGCCTGGGA	110
Glyma02g47670_500bp	Loy1490.500	TATATAGGTACCCTGAAGATTACACCAGTAGTTAGT	123
	Loy1491	TATATACCATGGTTAATTAATTTCAATCTCTCCCTCTCTAT	112

Glyma02g47670_1000bp	Loy1490.1000	TATATAGGTACCTATGCCAGAATCAACAATGAAAC	124
	Loy1491	TATATACCATGGTTAATTAATTTCAATCTCTCCCTCTCTAT	112
Glyma02g47670_1500bp	Loy1490.1500	TATATAGGTACCAGCTAGGTAGCGGGTGGTGGTAGGA	125
	Loy1491	TATATACCATGGTTAATTAATTTCAATCTCTCCCTCTCTAT	112
Glyma14g02930_500bp	Loy1492.500	TATATAGGTACCCCACCGACCTTTTTTTTATATAAAAAAAATC	126
	Loy1493	TATATACCATGGTTAATTAAGCTGTGTGACCACTGATG	114
Glyma14g02930_1000bp	Loy1492.1000	TATATAGGTACCTTAAATTACATGAATAACGAAATTAAG	127
	Loy1493	TATATACCATGGTTAATTAAGCTGTGTGACCACTGATG	114
Glyma14g02930_1500bp	Loy1492.1500	TATATAGGTACCAAACAAAATTATCCATCTCACA	128
	Loy1493	TATATACCATGGTTAATTAAGCTGTGTGACCACTGATG	114
Glyma17g27610_500bp	Loy1494.500	TATATAGGTACCAATAAACATATTAATCAACTATGAAAC	129
	Loy1495	TATATACCATGGTTAATTAATGTTGTGTTAACAAAGGGT	116
Glyma17g27610_1000bp	Loy1494.1000	TATATAGGTACCAAACTCATTCCACATGGACTGTGGCCT	130
	Loy1495	TATATACCATGGTTAATTAATGTTGTGTTAACAAAGGGT	116
Glyma17g27610_1500bp	Loy1494.1500	TATATAGGTACCTTGATTAACAAAAGTTTTATAAATAAAC	131
	Loy1495	TATATACCATGGTTAATTAATGTTGTGTTAACAAAGGGT	116

Table 6: Overview over G. max promoters conferring constitutive expression and expression in the epidermis

Feature name	SEQ ID #
p-Glyma11g14950_1939bp	1
p-Glyma14g06680_1056bp	2
p-Glyma02g47670_1753bp	3
p-Glyma14g02930_1688bp	4
p-Glyma17g27610_1889bp	5
p-Glyma11g14950_1939bp_perm	19
p-Glyma11g14950_500bp	20
p-Glyma11g14950_1000bp	21
p-Glyma11g14950_1500bp	22
p-Glyma14g06680_1056bp_perm	23
p-Glyma14g06680_500bp	24
p-Glyma14g06680_700bp	25
p-Glyma14g06680_1000bp	26
p-Glyma02g47670_1753bp_perm	27
p-Glyma02g47670_500bp	28
p-Glyma02g47670_1000bp	29
p-Glyma02g47670_1500bp	30
p-Glyma14g02930_1688bp_perm	31
p-Glyma14g02930_500bp	32
p-Glyma14g02930_1000bp	33
p-Glyma14g02930_1500bp	34
p-Glyma17g27610_1889bp_perm	35
p-Glyma17g27610_500bp	36
p-Glyma17g27610_1000bp	37
p-Glyma17g27610_1500bp	38

1.6 Cloning of promoter elements into the context of a GUS reporter gene cassette

To facilitate sub-cloning, promoter elements were modified by the addition of KpnI+Acc65I restriction enzyme sites at their 5' end and PacI+NcoI sites at their 3'end.

Using the Multisite Gateway System (Invitrogen, Carlsbad, CA, USA), the promoter::reporter-gene cassettes were assembled into binary constructs for plant transformation. The respective Glycine max promoters (with the prefix p- denoting promoter) were used in the reporter gene construct, and betaglucoronidase coding sequence (GUS) was utilized as reporter protein for subsequent histo-chemical analysis.

An ENTR/A vector containing the betaglucoronidase coding sequence followed by the t-nos nopalin synthase transcriptional terminator (Genbank V00087) was generated. Glycine max promoters were cloned using the restriction enzyme sites (see above) added by PCR amplifica-

tion at either end. Positive pENTR/A clones underwent sequence analysis to ensure correctness.

The pENTR/B and pENTR/C did not contain any additional elements. By performing a site-specific recombination (LR-reaction), the created pENTR/A, pENTR/B and pENTR/C were combined with the pSUN destination vector (pSUN derivative) according to the manufacturers (Invitrogen, Carlsbad, CA, USA) Multisite Gateway manual. The reactions yielded binary vectors LJK291, LJK296, LJK303, LJK304, LJK305 (cp. table7) with the respective Glycine max promoter, the GUS coding sequence c-GUS and the t-nos terminator, with promoter molecules having the prefix p-, coding sequences having the prefix c-, and terminator molecules having the prefix t-.
Table 7 shows an overview over reporter gene constructs with promoter elements putatively conferring constitutive expression or expression in the epidermis.

Table 7: Overview over G. max reporter gene constructs with promoters conferring constitutive expression or expression in the epidermis

Feature name	SEQ ID #
LJK291	
LJK296	256
LJK303	257
LJK304	
LJK305	

1.7 Generation of transgenic soybean plants (amended protocol according to WO2005/121345; Olhoft et al., 2007).

Soybean seed germination, propagation, A. rhizogenes and axillary meristem explant preparation, and inoculations were done as previously described (WO2005/121345; Olhoft et al., 2007) with the exception that the LJK291, LJK296, LJK303, LJK304, LJK305 (cp. example 1.6) each contained a mutated AHAS gene driven by the parsley ubiquitin promoter PcUbi4-2, mediating tolerance to imidazolinone herbicides for selection.

1.8 Promoter evaluation in transgenic soybean

Expression patterns and levels driven by the constitutive promoters and promoters putatively conferring expression in the epidermis measured using GUS histo-chemical analysis following a protocol known in the art (Jefferson 1987). GUS expression was assayed in the following vegetative and reproductive tissue for the constitutive promoters at various developmental stages:

1) leaf surface
2) root
3) stem
4) stem section
5) meristem
6) petioles
7) flowers
8) bud

9) embryo
10) seedcoat
11) silique seed-pocket
12) silique end

Expression in the epidermis was assayed in leaf surface views and sections based on visual assessment of the GUS staining. LJK291 and LJK296 showed constitutive expression in all analyzed tissues. LJK303, LJK304 and LJK305 showed strong expression in the lower leaf epidermis and parts of the spongy layer of mesophyll as well as some background expression in reproductive stages of the plant and can therefore be considered as epidermis preferential promoters. The permutated promoters and the 5´deleted versions of the promoters show the same expression patterns as the original promoters they are derived from.
The results are indicated in table 8.

Table 8: Expression profiles of constitutive promoters and promoters conferring expression in the epidermis

Construct	Feature Name	Specificity	Events	leaf	leaf section				root	stem	Stem section	meristem	petioles	flower	bud	embryo	seedcoat	silique seed pocket	silique end	Induction yes/no
					upper epidermis	lower epidermis	spongy mesophyll	palisade mesophyll												
LJK291	p-Glyma11g14950_1939bp	constitutive	17	++	n.d.	n.d.	n.d.	n.d.	++	+++	+++	+	+	+	+	++	+++	+++	+++	no
LJK296	p-Glyma14g06680_1056bp	constitutive	15	+++	n.d.	n.d.	n.d.	n.d.	++	++	++	++	++	++	++	+++	+++	++	++	no
LJK303	p-Glyma02g47670_1753bp	epidermis	8	++	+	+++	+++	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	+	+	+	+	no
LJK304	p-Glyma14g02930_1688bp	epidermis	4	++	+	+++	+++	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	++	++	++	++	no
LJK305	p-Glyma17g27610_1889bp[	epidermis	5	+	+	++	++	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	+	+	+	+	no

0 no GUS staining; + minimal GUS staining; medium GUS staining; +++ strong GUS staining; n.d. no analysis

Example 2: Identification and validation of pathogen-inducible promoters from soybean

2.1 Identification of pathogen-inducible transcripts by AFLP

AFLP bands from table 1 were selected for pathogen-inducible expression in epidermis or both mesophyll and epidermis.

2.2 Identification of the genes corresponding to AFLP bands

Expressed Sequence Tag (EST) sequences of AFLP bands were used as query for BLASTN searching against a soybean sequence database. The corresponding genes are listed in table 9

2.3 Identification of pathogen-inducible transcripts by microarray

In addition to identification of ESTs by AFLP, a microarray experiment was performed in triplicate for samples 3 and 5-8 from table 1, which identified the following genes: Glyma01g33070.2; Glyma01g42660.1 (cp. Table 9).

Table 9: Overview over corresponding genes for G. max promoters conferring pathogen-inducible expression

Feature name	SEQ ID #
Glyma13g44640_gene	230
Glyma08g37270_gene	232
Glyma04g40860.1_gene	234
Glyma01g33070.2_gene	236
Glyma15g05820.1_gene	238
Glyma01g42660.1_gene	240
Glyma17g14320_gene	242
Glyma01g01510.1_gene	244

2.4 Confirmation of allele-specific expression pattern using quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR)

In order to confirm the native expression patterns of soybean genes in an allele-specific manner from both AFLP and microarray approaches, quantitative reverse transcription PCR (qRT-PCR) was performed using total RNA isolated from the same materials as were used for the AFLP expression profiling and the microarray experiment (cp. table 1).

Primers for qRT-PCR were designed based on the sequences of the isolated EST fragments or on microarray data using the Vector NTI software package (Invitrogen, Carlsbad, CA, USA). Primers were designed to distinguish individual alleles of the candidate gene in the tetraploid soybean genome. Primers for qRT-PCR are listed in table 10. The tubulin gene served as a control for normalization purposes (cp. table 3).

Table 10: Primer sequences for qRT-PCR

Feature name	Primer	Sequence	SEQ ID #
Glyma13g44640	Loy1405	GCCATGCCTCAGCTAACCGACAGATCA	132
	Loy1407	ACAGCCACTGCAGCAACCTGATACAAATGC	133
Glyma08g37270	Loy1335	GCTGGAGCTAGCGCTTATGCACGTCTC	134
	Loy1337	CAGCTGCAACCAATCCACTGATGTG	135
Glyma04g40860.1	Loy1306	AGTGGCTAGACCAGTTGAATTGCAACGA	136
	Loy1307	CATGCAAGCGAGGTGTTTACATTTTGCT	137
Glyma01g33070.2	Loy1456	AAATGGCTTTCGGAGTTTCCCTAGTGGCA	138
	Loy1457	GAACTGAAGCAAGAACAGCATTCCCCACAC	139
Glyma15g05820.1	Loy1382	TCACTATGCCCTCAAAACGGTGACG	140
	Loy1383	GCTTGATCAGATTGCAGAATTCCACGA	141
Glyma01g42660.1	Loy1443	CTCAGGCAGCGAACTTCAACATCACAAAT	142
	Loy1444	CCACGATTCGCCAGGGTTAAGCCTT	143
Glyma17g14320	Loy1330	AAAGAAGGTGGAATTGGAAGGGGC	144
	Loy1331	TTAACGTGGGTGATGGTGAGTGGC	145
Glyma01g01510.1	Loy1373	AACATTGTTTCAGGACATGCACACCG	146
	Loy1375	TGAAGTGGGGGTATTGATCAAGAGCCT	147

qRT-PCR was performed using QuantiTect Kit (Qiagen, Hilden, Germany) and SYBR Green qPCR Master Mix (Roche Diagnostics, Mannheim, Germany) in a Roche LightCycler (Roche Diagnostics, Mannheim, Germany). cDNA was synthesized using 800 ng of total RNA and 1 µl reverse transcriptase in a 20 µl volume. The cDNA was diluted with 60 µl of RNAse free water to a final volume of 80 µl. 4 µl of diluted cDNA were used in a 10 µl PCR reaction according to manufacturer's instruction. The thermocycling conditions were as follows: Denature at 95°C for 2 minutes, and run 45 cycles at 95°C for 10 seconds and 60°C for 20 seconds and 72°C for 20 seconds for amplification. After the final cycle of the amplification, the dissociation curve analysis was carried out to verify that the amplification occurred specifically and no primer dimer product was generated during the amplification process. The tubulin gene (primer sequences in table 2) was used as an endogenous reference gene to normalize the calculation using the Comparative Ct (Cycle of threshold) value method. The DeltaCt value was obtained by subtracting the Ct value of tubulin gene from the Ct value of the respective candidate gene, and the relative transcription quantity (expression level) of the candidate gene was expressed as $2^{-DeltaCt}$.

2.5 Identification of the promoter region

For promoter identification purposes, the sequence upstream of the start codon of the identified genes was defined as the respective promoters. To characterize these promoter regions, 5´RACE PCR analyses were performed using the primers listed in table 11.

Table 11: Primer sequences for 5´RACE PCR

Feature name	Primer	Sequence	SEQ ID #
Glyma13g44640	Loy1667	CAGGAATAGGAAGACTCCAACAAGAAGAGC	148
Glyma08g37270	Loy1657	GGCATTGAAGAGAGGGCGCAGAGGCTTG	149
Glyma04g40860.1	Loy1307	CATGCAAGCGAGGTGTTTACATTTTGCT	150
Glyma01g33070.2	Loy1765	AAGCAAGAACAGCATTCCCCACAC	151
Glyma15g05820.1	Loy1383	GCTTGATCAGATTGCAGAATTCCACGA	152
Glyma01g42660.1	Loy1663	GCTTTTCGTGACGGCCATTGTAAATT	153
Glyma17g14320	Loy1711	CAACAAAAACTGCAGAAAGTCCATCC	154
Glyma01g01510.1	Loy1766	ATCCAATAAGCTGCAGCAACCATACCAC	155

2.6 Isolation of the promoter region by PCR amplification

The promoter regions of the respective genes were isolated via genomic PCR using the following sequence specific primers (table 12).

Promoters putatively conferring pathogen-inducible expression amplified with these primer pairs are listed in table 13.

In addition 5´-deletions of the promoters are made by using different 5´ primers in combination with the same 3´ oligonucleotide primer. The resulting promoters are indicated by their respective length in base pairs (cp. table 13) and the corresponding primer pairs for PCR are listed in table 12.

Promoters are bioinformatically analyzed with MatInspector professional 8.0.4, Release August 2010 (Genomatix Software GmbH; Munich, Germany) for transcription factor binding sites. Regions free of core motifs are permutated and the resulting nucleic acid sequences are synthesized. The principles for generating permutated promoters that retain their respective tissue specificities are described in EP10193800.9 und US61/419895. The resulting sequences are indicated by the original identifier of the corresponding gene followed by "_perm" (cp. table 13).

Table 12: Primer sequences for PCR amplification of pathogen-inducible promoters

Feature name	Primer	Sequence	SEQ ID #
Glyma13g44640_1047bp	Loy1700	TATATAGGTACCGGGATGTTTATTTAAGGCATGGTCA	156
	Loy1501	TATATACCATGGTTAATTAACAAGGGAGTGGAA-TAACTT	157
Glyma08g37270_2043bp	Loy1712	TATATAGGTACC AGCTCATTACCTCAAATTTCCCTAC	158
	Loy1713	TATATACCATGG TTCTCGCACACACAGAACAGAGA	159
Glyma04g40860.1_1917bp	Loy1527	TATATAGGTACCTTTCTTAGATAAACATACGTACGTT	160
	Loy1528	TATATACCATGGTTAATTAATTCTAACAATACAA-AATCTGTATATG	161
Glyma01g33070.2_1921bp	Loy1525	TATATAGGTACCAATTGACAAGTTGATTGTTGTA	162
	Loy1526	TATATACCATGGTTAATTAAGGAAATTAACTGAAC-CAATTACT	163
Glyma15g05820.1_1393bp	Loy1548	TATATAGGTACCAAATTATAGGTGAAAAAATTC	164
	Loy1549	ATATATCCATGGTTTTGTGAGGAAATTAAAGG	165
Glyma01g42660.1_1948bp	Loy1518	TATATAGGTACCTGAGAGAGATGCCAATTTTA-CAAGCC	166
	Loy1519	TATATACCATGGTGTAAATTAATTGCCGTTCGTTA-AAGA	167
Glyma17g14320_1607bp	Loy1779	TATATAGGTACCTAAATAATTAATTTATTTCAAACACT	168
	Loy1482	TATATACCATGGTGTTGCGATATGAACGCAGAGA-GAGG	169
Glyma01g01510.1_2016bp	Loy1788	TATATAGGTACCGTTGAAGATTCACCACTTCTC	170
	Loy1544	TATATACCATGGTTAAAGAATTGCAAGAAGAAG-GAAG	171
Glyma13g44640_500bp	Loy1700.500	TATATAGGTACCATAATGTAGCGTTGAATGTACT	172
	Loy1501	TATATACCATGGTTAATTAACAAGGGAGTGGAA-TAACTT	157
Glyma13g44640_700bp	Loy1700.700	TATATAGGTACCAGTCACATACTGTTAACAATTATTC	173
	Loy1501	TATATACCATGGTTAATTAACAAGGGAGTGGAA-TAACTT	157
Glyma13g44640_1000bp	Loy1700.1000	TATATAGGTACCTAATTAATCACAAAGTGAAGAAC	174
	Loy1501	TATATACCATGGTTAATTAACAAGGGAGTGGAA-TAACTT	157
Glyma08g37270_500bp	Loy1712.500	TATATAGGTACCTTATGATTAGTATAAATCTATTG	175
	Loy1713	TATATACCATGGTTCTCGCACACACAGAACAGAGA	159
Glyma08g37270_1000bp	Loy1712.1000	TATATAGGTACCTAGATTTTTAAATATTTATAATAA-AATAATAAG	176
	Loy1713	TATATACCATGGTTCTCGCACACACAGAACAGAGA	159

Glyma08g37270_1500bp	Loy1712.1500	TATATAGGTACCTCATTAATTGAGTTATTTATATAA-AATG	177
	Loy1713	TATATACCATGGTTCTCGCACACACAGAACAGAGA	159
Glyma04g40860.1_500bp	Loy1527.500	TATATAGGTACCTAATATATAAGCGGAACTATACGGT	178
	Loy1528	TATATACCATGGTTAATTAATTCTAACAATACAA-AATCTGTATATG	161
Glyma04g40860.1_1000bp	Loy1527.1000	TATATAGGTACCGTTGATAAATAATTTTTTTATGAATAA	179
	Loy1528	TATATACCATGGTTAATTAATTCTAACAATACAA-AATCTGTATATG	161
Glyma04g40860.1_1500bp	Loy1527.1500	TATATAGGTACCGAAATATTTGATTCACAAGT	180
	Loy1528	TATATACCATGGTTAATTAATTCTAACAATACAAAATCTGTATATG	161
Glyma01g33070.2_500bp	Loy1525.500	TATATAGGTACCATTGAATTCACTAATTTTATATTTTATAATTTG	181
	Loy1526	TATATACCATGGTTAATTAAGGAAATTAACTGAACCAATTACT	163
Glyma01g33070.2_1000bp	Loy1525.1000	TATATAGGTACCCAACAGATTAAGATCTAGAATAAATAAAC	182
	Loy1526	TATATACCATGGTTAATTAAGGAAATTAACTGAACCAATTACT	163
Glyma01g33070.2_1500bp	Loy1525.1500	TATATAGGTACCTACTTATGAATTAAGCTTAGTTCTTGCA	183
	Loy1526	TATATACCATGGTTAATTAAGGAAATTAACTGAACCAATTACT	163
Glyma15g05820.1_500bp	Loy1548.500	TATATAGGTACCAATTTTTTTTTTTCAGTATTATTTCTATCT	184
	Loy1549	ATATATCCATGGTTTTGTGAGGAAATTAAAGG	165
Glyma15g05820.1_700bp	Loy1548.700	TATATAGGTACCTCATCACAATCGAAAAATTCCATC	185
	Loy1549	ATATATCCATGGTTTTGTGAGGAAATTAAAGG	165
Glyma15g05820.1_1000bp	Loy1548.1000	TATATAGGTACCGCGCGCGCTCTCTTAGAACTTTTTTTG	186
	Loy1549	ATATATCCATGGTTTTGTGAGGAAATTAAAGG	165
Glyma01g42660.1_500bp	Loy1518.500	TATATAGGTACCCATTTTCAACATTCAGAGTGGGT	187
	Loy1519	TATATACCATGGTGTAAATTAATTGCCGTTCGTTAAAGA	167
Glyma01g42660.1_1000bp	Loy1518.1000	TATATAGGTACCTTTTTTCACCCAATTAATTAGAGTATTTC	188
	Loy1519	TATATACCATGGTGTAAATTAATTGCCGTTCGTTAAAGA	167
Glyma01g42660.1_1500bp	Loy1518.1500	TATATAGGTACCGTTTTGCTATTGACTTTTGTTTTATTTCGT	189
	Loy1519	TATATACCATGGTGTAAATTAATTGCCGTTCGTTAAAGA	167
Glyma17g14320_500bp	Loy1779.500	TATATAGGTACCTTTTTTAATCTACTTTTTATTTGTTTAATC	190
	Loy1482	TATATACCATGGTGTTGCGATATGAACGCAGAGAGAGG	169

Glyma17g14320_1000bp	Loy1779.1000	TATATAGGTACCTTTAATTTGGAATAATTTTTTTTCTTCTC	191
	Loy1482	TATATACCATGGTGTTGCGATATGAACGCAGAGAGAGG	169
Glyma17g14320_1500bp	Loy1779.1500	TATATAGGTACCTTAGAGGAAAAATTTTGTCATCCATAA	192
	Loy1482	TATATACCATGGTGTTGCGATATGAACGCAGAGAGAGG	169
Glyma01g01510.1_500bp	Loy1788.500	TATATAGGTACCACATGGCAACATTTTTTTTTTATCTCT	193
	Loy1544	TATATACCATGGTTAAAGAATTGCAAAGAAGAAGGAAG	171
Glyma01g01510.1_1000bp	Loy1788.1000	TATATAGGTACCTATATATATATATATAATAAACTATATCT	194
	Loy1544	TATATACCATGGTTAAAGAATTGCAAAGAAGAAGGAAG	171
Glyma01g01510.1_1500bp	Loy1788.1500	TATATAGGTACCTACCTGTTCACTAGCTAGTTACAAAAATATATC	195
	Loy1544	TATATACCATGGTTAAAGAATTGCAAAGAAGAAGGAAG	171

Table 13: Overview over G. max promoters conferring pathogen-inducible expression

Feature name	SEQ ID #
p-Glyma13g44640_1047bp	6
p-Glyma08g37270_2043bp	7
p-Glyma04g40860.1_1917bp	8
p-Glyma01g33070.2_1921bp	9
pGlyma15g05820.1_1393bp	10
p-Glyma01g42660.1_1948bp	11
p-Glyma17g14320_1607bp	12
p-Glyma01g01510.1_2016bp	13
p-Glyma13g44640_1047bp_perm	39
p-Glyma13g44640_500bp	40
p-Glyma13g44640_700bp	41
p-Glyma13g44640_1000bp	42
p-Glyma08g37270_2043bp_perm	43
p-Glyma08g37270_500bp	44
p-Glyma08g37270_1000bp	45
p-Glyma08g37270_1500bp	46
p-Glyma04g40860.1_1917bp_perm	47
p-Glyma04g40860.1_500bp	48
p-Glyma04g40860.1_1000bp	49
p-Glyma04g40860.1_1500bp	50
p-Glyma01g33070.2_1921bp_perm	51
p-Glyma01g33070.2_500bp	52
p-Glyma01g33070.2_1000bp	53
p-Glyma01g33070.2_1500bp	54
p-Glyma15g05820.1_1393bp_perm	55
p-Glyma15g05820.1_500bp	56
p-Glyma15g05820.1_700bp	57
p-Glyma15g05820.1_1000bp	58
p-Glyma01g42660.1_1948bp_perm	59
p-Glyma01g42660.1_500bp	60
p-Glyma01g42660.1_1000bp	61
p-Glyma01g42660.1_1500bp	62
p-Glyma17g14320_1607bp_perm	63
p-Glyma17g14320_500bp	64
p-Glyma17g14320_1000bp	65
p-Glyma17g14320_1500bp	66
p-Glyma01g01510.1_2016bp_perm	67
p-Glyma01g01510.1_500bp	68
p-Glyma01g01510.1_1000bp	69
p-Glyma01g01510.1_1500bp	70

2.7 Cloning of promoter elements into the context of a GUS reporter gene cassette

To facilitate sub-cloning, promoter elements were modified by the addition of KpnI and Acc65I restriction enzyme sites at their 5' end and PacI andNcoI sites at their 3'end.

Using the Multisite Gateway System (Invitrogen, Carlsbad, CA, USA), the promoter::reporter-gene cassettes were assembled into binary constructs for plant transformation. The respective Glycine max promoters (with the prefix p- denoting promoter) were used in the reporter gene construct, and beta-glucoronidase coding sequence (GUS) was utilized as reporter protein for subsequent histo-chemical analysis.

An ENTR/A vector containing the beta-glucoronidase coding sequence followed by the t-nos nopalin synthase transcriptional terminator (Genbank V00087) was generated. Glycine max promoters were cloned using the restriction enzyme sites (see above) added by PCR amplification at either end. Positive pENTR/A clones underwent sequence analysis to ensure correctness.

The pENTR/B and pENTR/C did not contain any additional elements. By performing a site-specific recombination (LR-reaction), the created pENTR/A, pENTR/B and pENTR/C were combined with the pSUN destination vector (pSUN derivative) according to the manufacturers (Invitrogen, Carlsbad, CA, USA) Multisite Gateway manual. The reactions yielded binary vectors LJK306, LJK331, LJK334, LJK358, LJK360, LJK361, LJK363, LJK372 (cp. table 14) with the respective Glycine max promoter, the GUS coding sequence c-GUS and the t-nos terminator, with promoter molecules having the prefix p-, coding sequences having the prefix c-, and terminator molecules having the prefix t-.

Table 14 shows an overview over reporter gene constructs with promoter elements putatively conferring pathogen-inducible expression.

Table 14: Overview over G. max reporter gene constructs with promoters conferring pathogen-inducible expression

Feature name	SEQ ID #
LJK306	258
LJK331	
LJK334	
LJK358	
LJK360	259
LJK361	
LJK363	
LJK372	

2.8 Generation of transgenic soybean plants (amended protocol according to WO2005/121345; Olhoft et al., 2007).

Soybean seed germination, propagation, *A. rhizogenes* and axillary meristem explant preparation, and inoculations were done as previously described (WO2005/121345; Olhoft et al., 2007) with the exception that the constructs LJK306, LJK331, LJK334, LJK358, LJK360, LJK361,

LJK363, LJK372 (cp. example 2.7) each contained a mutated AHAS gene driven by the parsley ubiquitin promoter PcUbi4-2, mediating tolerance to imidazolinone herbicides for selection.

2.9 Promoter evaluation in transgenic soybean

Expression patterns and levels driven by the putatively pathogen-inducible promoters were measured using GUS histochemical analysis following a protocol known in the art (Jefferson 1987). Soybean transformation was conducted using an Agrobacterium-mediated transformation system.

The rust fungus is a wild isolate from Brazil. The plants were inoculated with P.pachyrhizi.

In order to obtain appropriate spore material for the inoculation, soy leaves which had been infected with rust 15-20 days ago, were taken 2-3 days before the inoculation and transferred to agar plates (1% agar in H_2O). The leaves were placed with their upper side onto the agar, which allowed the fungus to grow through the tissue and to produce very young spores. For the inoculation solution, the spores were knocked off the leaves and were added to a Tween-H_2O solution. The counting of spores was performed under a light microscope by means of a Thoma counting chamber. For the inoculation of the plants, the spore suspension was added into a compressed-air operated spray flask and applied uniformly onto the plants or the leaves until the leaf surface is well moisturized. For macroscopic assays a spore density of 1-5x10^5 spores/ml was used. For microscopy, a density of >5 x 10^5 spores / ml was used. The inoculated plants were placed for 24 hours in a darkened greenhouse chamber with an average of 22°C and >90% of air humidity. The following cultivation was performed in a chamber with an average of 25°C and 70% of air humidity for 48 hours.

GUS expression for pathogen-inducible promoters putatively conferring expression in the epidermis or in mesophyll and epidermis was assayed in leaf surface views and sections based on visual inspection. All promoters showed inducibility by soybean rust, either preferentially in the epidermis or in both epidermis and mesophyll. The permutated promoters and the 5´deleted versions of the promoters show the same expression patterns as the original promoters they are derived from.

The results are indicated in table 15.

Table 15: Expression profiles of pathogen-inducible promoters

Construct	Feature name	Specificity	events	leaf	leaf section				embryo	seedcoat	silique seed pocket	silique end	Induction yes/no
					upper epidermis	lower epidermis	spongy mesophyll	palisade mesophyll					
LJK306	p-Glyma13g44640_1047bp	epidermis induced	12	++	+	++	++	+	+	+	+	+	y
LJK331	p-Glyma08g37270_2043bp	mesophyll + epidermis induced	12	++	+	++	++	++	0	0	0	0	y
LJK334	p-Glyma04g40860.1_1917bp	mesophyll + epidermis induced	12	+++	++	++	++	++	0	0	0	0	y
LJK358	p-Glyma01g33070.2_1921bp	mesophyll + epidermis induced	17	+	+	+	+	+	0	0	0	0	y
LJK360	pGlyma15g05820.1_1393bp	mesophyll + epidermis induced	16	+++	++	++	+++	+++	+	+	+	+	y
LJK361	p-Glyma01g42660.1_1948bp	mesophyll + epidermis induced	12	+	0	+	+	+	0	0	0	0	y
LJK363	p-Glyma17g14320_1607bp	mesophyll + epidermis induced	12	+++	++	++	++	++	0	0	0	0	y
LJK372	p-Glyma01g01510.1_2016bp	mesophyll + epidermis induced	14	++	0	+	++	++	0	0	0	0	y

0 no GUS staining; + minimal GUS staining; medium GUS staining; +++ strong GUS staining; n.d. no analysis

Example 3: A. thaliana promoters putatively conferring expression in green tissue

3.1 Isolation of the promoter regions from of A. thaliana putatively conferring expression in green tissue of plants by PCR amplification

A. thaliana promoter elements putatively conferring expression in green tissue of plants were amplified by PCR using the primers listed in table 21. Preferentially the cloned region encompassed 1-2 kb upstream of the transcriptional start site or up to the stop codon of the previous open-reading frame. The corresponding genes are listed in table 22. The corresponding promoter sequences are listed in table 23.

In addition 5´-deletions of the promoters are made by using different 5´ primers in combination with the same 3´ oligonucleotide primer. The resulting promoters are indicated by their respective length in base pairs (cp. table 23) and the corresponding primer pairs for PCR are listed in table 21.

Promoters are bioinformatically analyzed with MatInspector professional 8.0.4, Release August 2010 (Genomatix Software GmbH; Munich, Germany) for transcription factor binding sites. Regions free of core motifs are permutated and the resulting nucleic acid sequences are synthesized. The principles for generating permutated promoters that retain their respective tissue specificities are described in EP10193800.9 und US61/419895. The resulting sequences are indicated by the original identifier of the corresponding gene followed by "_perm" (cp. table 23).

Table 21: Primer sequences for PCR amplification A. thaliana promoters putatively conferring expression in green tissue of plants

Gene ID	Primer	Sequence	SEQ ID #
At1g30380_1970bp	Loy1116	TATATACCCGGGTACCATGCAGAGGATCAAGAAGATTCTCC	196
	Loy1117	TATATACCATGGTTTCTTAGTTGATTCTACAAATCTTTTATTTTC	197
At1g49750_1922bp	Loy1124	TATATACCCGGGTACCGTGAAAGCAGTGAAGCCGTGA	198
	Loy1125	TATATACCATGGTGAGTTGATGAGATTTTGTGGTGAGT	199
At3g62410_332bp	Loy1128	TATATACCCGGGTACCGACATCTGTCTTGACTTTTCCTTAAACAGTGTGTG	200
	Loy1133	TATATACCATGGCTTTGGATGGAGAAGGTACACGGCAG	201
At1g61520_1970bp	Loy1120	TATATACCCGGGTACCGACGAACTCATGCTACTACTACAG	202
	Loy1121	TATATAACATGTTGAGCTCTTTCTCTGTTCCTCAACTCTTTTCT	203
At1g65490_1953bp	Loy1130	TATATACCCGGGTACCGAAACGAAACTGAACCGCCTCCTTT	204

	Loy1131	TATATAACATGTT-GAGCTCTTCGTTCTTCGTTGCGTTTTTGGTCATCG	205
At1g30380_500bp	Loy1116.500	TATATACCCGGGTACCTGACTAAT-TAAGCTCGAAAGTGTTCTTCA	206
	Loy1117	TATATACCATGGTTTCTTAGTTGATTCTA-CAAATCTTTTATTTTC	197
At1g30380_1000bp	Loy1116.1000	TATATACCCGGGTACCGGCTTTTGCGT-TAGGTTATATAACTCCA	207
	Loy1117	TATATACCATGGTTTCTTAGTTGATTCTA-CAAATCTTTTATTTTC	197
At1g30380_1500bp	Loy1116.1500	TATATACCCGGGTACCTGTAGCAA-GAATTGATCGATATGCTTTG	208
	Loy1117	TATATACCATGGTTTCTTAGTTGATTCTA-CAAATCTTTTATTTTC	197
At1g49750_500bp	Loy1124.500	TATATACCCGGGTACCGGACTCAATAAA-CAACTCAAAGATGA	209
	Loy1125	TATATACCATGGTGAGTTGATGA-GATTTTGTGGTGAGT	199
At1g49750_1000bp	Loy1124.1000	TATATACCCGGGTACCTCACT-GATGTTCTCTAATGAACGTTC	210
	Loy1125	TATATACCATGGTGAGTTGATGA-GATTTTGTGGTGAGT	199
At1g49750_1500bp	Loy1124.1500	TATATACCCGGGTACCAAGTGAAAATA-TAATATTCATACCTCTTG	211
	Loy1125	TATATACCATGGTGAGTTGATGA-GATTTTGTGGTGAGT	199
At3g62410_100bp	Loy1128.100	TATATACCCGGGTACCACGCACACTTCA-TATATCTTG	212
	Loy1133	TATATACCATGGCTTTGGATGGAGAAGG-TACACGGCAG	201
At3g62410_200bp	Loy1128.200	TATATACCCGGGTACCAAATTTTCAA-CATCGTACTGCTTCATAAAC	213
	Loy1133	TATATACCATGGCTTTGGATGGAGAAGG-TACACGGCAG	201
At1g61520_500bp	Loy1120.500	TATATACCCGGGTACCGTTGAATTGTTA-TATCAAAATTTGA	214
	Loy1121	TATATAACATGTT-GAGCTCTTTCTCTGTTCCT-CAACTCTTTTCT	203
At1g61520_1000bp	Loy1120.1000	TATATACCCGGGTACCTTTGGCTGAAT-CAGCTTCAGCAGA	215

	Loy1121	TATATAACATGTT-GAGCTCTTTCTCTGTTCCT-CAACTCTTTTCT	203
At1g61520_1500bp	Loy1120.1500	TATATACCCGGGTAC-CAATGGTTCTGTTGCTCCTAATGTAGA	216
	Loy1121	TATATAACATGTT-GAGCTCTTTCTCTGTTCCT-CAACTCTTTTCT	203
At1g65490_500bp	Loy1130.500	TATATACCCGGGTACCTTTTTGTAAA-CAATTTTTTGTGATATATAT	217
	Loy1131	TATATAACATGTT-GAGCTCTTCGTTCTTCGTTGCGTTTTTGGTCATCG	205
At1g65490_1000bp	Loy1130.1000	TATATACCCGGGTACCCAGAATTTTAAA-GACACACAAAGCA	218
	Loy1131	TATATAACATGTT-GAGCTCTTCGTTCTTCGTTGCGTTTTTGGTCATCG	205
At1g65490_1500bp	Loy1130.1500	TATATACCCGGGTACCACCGCTTAA-TATCGTATGATTAG	219
	Loy1131	TATATAACATGTT-GAGCTCTTCGTTCTTCGTTGCGTTTTTGGTCATCG	205

Table 22: Overview over corresponding genes for A. thaliana promoters conferring expression in green tissue of plants

Feature name	SEQ ID #
At1g30380_gene	246
At1g49750_gene	248
At3g62410_gene	250
At1g61520_gene	252
At1g65490_gene	254

Table 23: Overview over A. thaliana promoters conferring expression in green tissue of plants

Feature name	SEQ ID #
p-mes-At1g30380_1970bp	14
p-mes-At1g49750_1922bp	15
p-mes-At3g62410_332bp	16
p-photo-At1g61520_1970bp	17
p-mes-At1g65490_1953bp	18
p-mes-At1g30380_1970bp_perm	71
p-mes-At1g30380_500bp	72
p-mes-At1g30380_1000bp	73

p-mes-At1g30380_1500bp	74
p-mes-At1g49750_1922bp_perm	75
p-mes-At1g49750_500bp	76
p-mes-At1g49750_1000bp	77
p-mes-At1g49750_1500bp	78
p-mes-At3g62410_332bp_perm	79
p-mes-At3g62410_100bp	80
p-mes-At3g62410_200bp	81
p-photo-At1g61520_1970bp_perm	82
p-photo-At1g61520_500bp	83
p-photo-At1g61520_1000bp	84
p-photo-At1g61520_1500bp	85
p-mes-At1g65490_1953bp_perm	86
p-mes-At1g65490_500bp	87
p-mes-At1g65490_1000bp	88
p-mes-At1g65490_1500bp	89

3.2 Cloning of promoter elements into the context of a GFP reporter gene cassette

To facilitate sub-cloning the following restriction sites were added to the ends of the promoter element (table 24):

Table 24: Restriction enzyme sites for cloning A. thaliana promoters putatively conferring expression in green tissue of plants

Promoter	5´end	3´end
p-mes-At1g30380_1970bp	KpnI	NcoI
p-mes-At1g49750_1922bp	KpnI	NcoI
p-mes-At3g62410_332bp	KpnI	NcoI
p-photo-At1g61520_1970bp	KpnI	PciI
p-mes-At1g65490_1953bp	KpnI	PciI

Using the Multisite Gateway System (Invitrogen, Carlsbad, CA, USA), the promoter::reporter-gene cassettes were assembled into binary constructs for plant transformation. The respective Arabidopsis thaliana promoters (with the prefix p- denoting promoter) were used in the reporter gene construct, and Green Fluorescent Protein coding sequence (c-AcGFP1; Clontech Laboratories Inc., Mountain View, CA, USA) was utilized as reporter protein for subsequent fluorescence microscopic analysis.

An ENTR/A vector containing the Green Fluorescent protein coding sequence followed by the t-OCS agrobacterium terminator (Genbank DQ005456) was generated. Arabidopsis thaliana promoters were cloned using the restriction enzyme sites (see above) added by PCR amplification at either end. Positive pENTR/A clones underwent sequence analysis to ensure correctness.

The pENTR/B and pENTR/C did not contain any additional elements. By performing a site-specific recombination (LR-reaction), the created pENTR/A, pENTR/B and pENTR/C were combined with the pSUN destination vector (pSUN derivative) according to the manufacturers (Invitrogen, Carlsbad, CA, USA) Multisite Gateway manual. The reactions yielded binary vectors LJK186, LJK189, LJK190, LJK192 and LJK193 (cp. table 25); with the respective Arabidopsis thaliana promoter, the Green Fluorescent Protein coding sequence c-AcGFP1 and the t-OCS terminator, with promoter molecules having the prefix p-, coding sequences having the prefix c-, and terminator molecules having the prefix t-.

Table 25: GFP reporter gene constructs for A. thaliana promoters putatively conferring expression in green tissue of plants

Vector	promoter element used	SEQ ID #
LJK186	p-mes-At1g30380_1970bp	260
LJK189	p-mes-At1g49750_1922bp	
LJK190	p-mes-At3g62410_332bp	
LJK192	p-photo-At1g61520_1970bp	
LJK193	p-mes-At1g65490_1953bp	

3.3 Test of the A. thaliana promoters putatively conferring expression in green tissue of plants in transgenic soybean

Expression patterns and levels driven by the promoters putatively conferring expression in green tissue of plants were measured using GFP analysis. Analysis was performed with the Leica DM5000B microscope and the DFC490 camera with the following settings: Saturation 1.01; Gain 1; Exposure 2.1s; GFP-filter: L5; Excitation 480/40nm; Dichromatic mirror: 505nm; Suppression filter: 527/30nm.

The three promoter elements in vectors LJK 189, LJK 190 and LJK 192 showed medium to strong expression based on visual analysis exclusively in the mesophyll layer of leaves and can thus be rated mesophyll-specific. Promoter elements corresponding to LJK186 and LJK193 conferred preferential expression in the mesophyll layer of leaves as well as weak expression in the green tissue of the shoot and can thus be rated mesophyll-preferential. The permutated promoters and the 5´deleted versions of the promoters show the same expression patterns as the original promoters they are derived from.

The results are listed in table 26.

Table 26: Expression profiles of A. thaliana promoters conferring expression in the green tissue of plants

construct	Putative specificity	promoter	expression level					
			Leaf me-sophyll	Other leaf tissues	shoot (chlorophyll containing layer)	shoot (non-green)	root	flower
LJK186	mesophyll	p-mes-At1g30380_1970bp	+++	0	+	0	0	0
LJK189	mesophyll	p-mes-At1g49750_1922bp	++	0	0	0	0	0
LJK190	mesophyll	p-mes-At3g62410_332bp	++	0	0	0	0	0
LJK192	mesophyll	p-photo-At1g61520_1970bp	+++	0	0	0	0	0
LJK193	mesophyll	p-mes-At1g65490_1953bp	+++	0	+	0	0	0

0 no expression; + minimal expression; medium expression; +++ strong expression; n.d. no analysis

Table 27: Overview of Promoters of the Invention and corresponding homologs and fragments thereof

Original promoter		derivative promoters	
Seq ID NO	feature name	Seq ID NO	feature name
1	p-Glyma11g14950_1939bp		
		19	p-Glyma11g14950_1939bp_perm
		20	p-Glyma11g14950_500bp
		21	p-Glyma11g14950_1000bp
		22	p-Glyma11g14950_1500bp
2	p-Glyma14g06680_1056bp		
		23	p-Glyma14g06680_1056bp_perm
		24	p-Glyma14g06680_500bp
		25	p-Glyma14g06680_700bp
		26	p-Glyma14g06680_1000bp
3	p-Glyma02g47670_1753bp		
		27	p-Glyma02g47670_1753bp_perm
		28	p-Glyma02g47670_500bp
		29	p-Glyma02g47670_1000bp
		30	p-Glyma02g47670_1500bp
4	p-Glyma14g02930_1688bp		
		31	p-Glyma14g02930_1688bp_perm
		32	p-Glyma14g02930_500bp
		33	p-Glyma14g02930_1000bp
		34	p-Glyma14g02930_1500bp
5	p-Glyma17g27610_1889bp		
		35	p-Glyma17g27610_1889bp_perm
		36	p-Glyma17g27610_500bp
		37	p-Glyma17g27610_1000bp
		38	p-Glyma17g27610_1500bp
6	p-Glyma13g44640_1047bp		
		39	p-Glyma13g44640_1047bp_perm
		40	p-Glyma13g44640_500bp
		41	p-Glyma13g44640_700bp
		42	p-Glyma13g44640_1000bp
7	p-Glyma08g37270_2043bp		
		43	p-Glyma08g37270_2043bp_perm
		44	p-Glyma08g37270_500bp
		45	p-Glyma08g37270_1000bp
		46	p-Glyma08g37270_1500bp
8	p-Glyma04g40860.1_1917bp		
		47	p-Glyma04g40860.1_1917bp_perm

		48	p-Glyma04g40860.1_500bp
		49	p-Glyma04g40860.1_1000bp
		50	p-Glyma04g40860.1_1500bp
9	p-Glyma01g33070.2_1921bp	51	p-Glyma01g33070.2_1921bp_perm
		52	p-Glyma01g33070.2_500bp
		53	p-Glyma01g33070.2_1000bp
		54	p-Glyma01g33070.2_1500bp
10	pGlyma15g05820.1_1393bp	55	p-Glyma15g05820.1_1393bp_perm
		56	p-Glyma15g05820.1_500bp
		57	p-Glyma15g05820.1_700bp
		58	p-Glyma15g05820.1_1000bp
11	p-Glyma01g42660.1_1948bp	59	p-Glyma01g42660.1_1948bp_perm
		60	p-Glyma01g42660.1_500bp
		61	p-Glyma01g42660.1_1000bp
		62	p-Glyma01g42660.1_1500bp
12	p-Glyma17g14320_1607bp	63	p-Glyma17g14320_1607bp_perm
		64	p-Glyma17g14320_500bp
		65	p-Glyma17g14320_1000bp
		66	p-Glyma17g14320_1500bp
13	p-Glyma01g01510.1_2016bp	67	p-Glyma01g01510.1_2016bp_perm
		68	p-Glyma01g01510.1_500bp
		69	p-Glyma01g01510.1_1000bp
		70	p-Glyma01g01510.1_1500bp
14	p-mes-At1g30380_1970bp	71	p-mes-At1g30380_1970bp_perm
		72	p-mes-At1g30380_500bp
		73	p-mes-At1g30380_1000bp
		74	p-mes-At1g30380_1500bp
15	p-mes-At1g49750_1922bp	75	p-mes-At1g49750_1922bp_perm
		76	p-mes-At1g49750_500bp
		77	p-mes-At1g49750_1000bp
		78	p-mes-At1g49750_1500bp
16	p-mes-At3g62410_332bp	79	p-mes-At3g62410_332bp_perm
		80	p-mes-At3g62410_100bp
		81	p-mes-At3g62410_200bp
17	p-photo-At1g61520_1970bp	82	p-photo-At1g61520_1970bp_perm

		83	p-photo-At1g61520_500bp
		84	p-photo-At1g61520_1000bp
		85	p-photo-At1g61520_1500bp
18	p-mes-At1g65490_1953bp		
		86	p-mes-At1g65490_1953bp_perm
		87	p-mes-At1g65490_500bp
		88	p-mes-At1g65490_1000bp
		89	p-mes-At1g65490_1500bp

Claims:

1. An isolated nucleic acid molecule capable of regulating expression in plants selected from the list comprising
   i) a sequence described by SEQ ID NO: 3, 27, 28, 29 or 30 and
   ii) a fragment of at least 250 consecutive bases of a molecule described by any of SEQ ID NO: 3, 27, 28, 29 or 30 and
   iii) a nucleotide molecule with a sequence identity of at least 60% to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 3, 27, 28, 29 or 30, and
   iv) a nucleotide molecule of at least 250 consecutive bases with a sequence identity of at least 60% to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 3, 27, 28, 29, or 30, and
   v) a nucleotide molecule of at least 250 bases capable of hybridizing under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C to a transcription regulating nucleotide molecule described by any of SEQ ID NO: 3, 27, 28, 29 or 30, or the complement thereof;
   vi) a nucleotide molecule of at least 250 bases capable of hybridizing under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1 X SSC, 0.1% SDS at 50°C to a nucleic acid comprising 250 or more consecutive nucleotides of a transcription regulating nucleotide molecule described by any of SEQ ID NO: 3, 27, 28, 29 or 30, or the complement thereof;
   vii) a nucleotide molecule which is the complement or reverse complement of any of the previously mentioned nucleotide sequences under i) to vi).

2. The isolated nucleic acid molecule of Claim 1, wherein the isolated nucleic acid molecule is obtainable from plant genomic DNA from a gene encoding a polypeptide comprising an amino acid sequence which has at least 80% amino acid sequence homology to a polypeptide described by SEQ ID NO: 225.

3. The isolated nucleic acid molecule of Claim 1, wherein the isolated nucleic acid molecule is obtainable from plant genomic DNA from a gene which has at least 80% sequence identity to a nucleic acid molecule described by SEQ ID NO: 224.

4. The isolated nucleic acid molecule of any of claim 1 to 3, wherein the sequences specified under ii), iii), iv), v), vi) and vii) of Claim 1 are capable to modify transcription in a plant or part thereof.

5. The isolated nucleic acid molecule of any of claim 1 to 4, wherein the sequences specified under ii), iii), iv), v), vi) and vii) of Claim 1 have substantially the same transcription regulating activity as the transcription regulating nucleotide sequence described by SEQ ID NO: 3.

6. An expression cassette for regulating expression in plants comprising
   a) at least one isolated nucleic acid molecule capable of regulating expression in plants as defined in any of claim 1 to 5,

   and functionally linked thereto

   b) at least one nucleic acid molecule which is heterologous in relation to said isolated nucleic acid molecule capable of regulating expression in plants.

7. A vector comprising an isolated nucleic acid molecule of any of Claim 1 to 5 or an expression cassette of Claim 6.

8. A transgenic host cell or non-human organism comprising an isolated nucleic acid sequence of any of Claim 1 to 5, an expression cassette of Claim 6, or a vector of Claim 7.

9. A transgenic plant or plant cell comprising an isolated nucleic acid sequence of any of Claim 1 to 5, an expression cassette of Claim 6, or a vector of Claim 7.

10. The transgenic plant or plant cell of claim 9, wherein said plant or plant cell is from a dicotyledonous plant.

11. A use of an isolated nucleic acid molecule of any of claim 1 to 5, an expression cassette of Claim 6, or a vector of Claim 7 for regulating expression in a plant or plant cell.

12. A method for the production of an expression cassette of claim 6 or vector of claim 7 comprising the steps of
   a. providing an isolated nucleic acid molecule of any of claim 1 to 5 and
   b. functionally linking said isolated nucleic acid molecule of any of claim 1 to 5 to a nucleic acid molecule heterologous to said isolated nucleic acid molecule.

13. A method for the production of a transgenic plant comprising the steps of
   a. providing an expression cassette of claim 6 or a vector of claim 7 and
   b. transforming said expression cassette or vector into a plant part or plant cell and
   c. regenerating a plant from said transformed plant part or plant cell.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2012/051239

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

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)

IPC: C12N, A01H

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)

DATABASES: CPRSABS, CNABS, CJFD, CSCD, SIPONPL, CNKI, DWPI, SIPOABS, CPEA, ISI WEB OF KNOWLEDGE, CA, BA, MEDLINE, PUBMED, NCBI, EMBL, GENEBANK, DDBJ

SEARCH TERMS: SEQUENCE SEARCH ON THE SEQUENCES OF SEQ ID NO: 3, 27-30, 224, 225, GLYCINE MAX, SOY BEAN, PLANT, EXPRESS, EXPRESSION, PROMOTER, PROMOTOR, INCREASE, PROMOTE, ENHANCE, REGULATORY SEQUENCE, REGULATORY REGION

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN101914534A (UNIV NORTHEAST AGRICULTURAL) 15. Dec. 2010 (15.12.2010), see the whole document	1-13
A	Li, Pengli et al., Cloning and preliminary analysis of promoter of soybean receptor-like protein kinase. *Journal of Plant Physiology and Molecular Biology*, Vol. 32, No. 3, pages: 315-319, ISSN: 1671-3877, see the whole document	1-13

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

* Special categories of cited documents:

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

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

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

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

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

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
05 Jul 2012 (05.07.2012)	**26 Jul. 2012 (26.07.2012)**
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer LIAO, Yajing Telephone No. (86-10)62411074

Form PCT/ISA /210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2012/051239

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GenBank [online], database accession No.: BT096166.1 06 Aug. 2009 (06.08.2009), "Soybean clone JCVI-FLGm-15M12 unknown Mrna", see the sequence	1-13
A	GenBank [online], database accession No.: XP_002321177.1 04 Dec. 2009 (04.12.2009), "predicted protein [Populus trichocarpa]", see the sequence	1-13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2012/051239

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/113 (2010.01)i

C12N 9/12 (2006.01)i

C12N 15/82 (2006.01)i

A01H 1/00 (2006.01)i

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IB2012/051239

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN101914534A	15.12.2010	NONE	