

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 January 2012 (26.01.2012)

PCT

(10) International Publication Number
WO 2012/012412 A2

(51) International Patent Classification:

A01H 5/00 (2006.01) C12N 15/63 (2006.01)
C12N 15/11 (2006.01) C12N 15/82 (2006.01)

(21) International Application Number:

PCT/US2011/044516

(22) International Filing Date:

19 July 2011 (19.07.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/365,617 19 July 2010 (19.07.2010) US

(71) Applicant (for all designated States except US): **UNIVERSITY OF TENNESSEE RESEARCH FOUNDATION** [US/US]; UT Conference Center, Suite 211, 600 Henley Street, Knoxville, TN 37996-4122 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **STEWART, C., Neal** [US/US]; 1600 Wheaton Place, Knoxville, TN 37919 (US). **PENG, Yanhui** [CN/US]; 455-17 Bramblewood Lane, Knoxville, TN 37922 (US).

(74) Agents: **EISENSCHENK, Frank, C.** et al.; Saliwanchik, Lloyd & Eisenschenk, P.O. Box 142950, Gainesville, FL 32614-2950 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, RO, RS, RU, 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, 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:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

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

(54) Title: GLYPHOSATE-INDUCIBLE PROMOTER ITS USE

(57) Abstract: The subject application provides polynucleotides, compositions thereof and methods for regulating gene expression in a plant using a promoter that initiates transcription in an inducible manner. In a further aspect of the invention, methods for modulating expression of a gene product in a stably transformed plant comprising the steps of (a) transforming a plant cell with a DNA construct comprising the disclosed promoter or fragments thereof that are capable of driving the expression of an operably linked nucleic acid sequence operably linked to at least one nucleotide sequence; (b) growing the plant cell under plant growing conditions and (c) regenerating a stably transformed plant from the plant cell wherein the induced expression of the operably linked nucleotide sequence alters the phenotype of the plant.



WO 2012/012412 A2

GLYPHOSATE-INDUCIBLE PROMOTER AND ITS USE

CROSS-REFERENCE TO RELATED APPLICATION

5 This application claims the benefit of U.S. Provisional Application Serial No. 61/365,617, filed July 19, 2010, the disclosure of which is hereby incorporated by reference in its entirety, including all figures, tables and amino acid or nucleic acid sequences.

BACKGROUND OF THE INVENTION

Recent advances in plant genetic engineering have enabled the engineering of plants having improved characteristics or traits, such as disease resistance, insect resistance, herbicide resistance, enhanced stability or shelf-life of the ultimate consumer product obtained from the plants and improvement of the nutritional quality of the edible portions of the plant. Thus, one or more desired genes from a source different than the plant, but engineered to impart different or improved characteristics or qualities, can be incorporated into the plant's genome. One or more new genes can then be expressed in the plant cell to exhibit the desired phenotype such as a new trait or characteristic.

An inducible promoter is a promoter that is capable of directly or indirectly activating transcription of one or more DNA sequences or genes in response to an inducer. In the absence of an inducer, the DNA sequences or genes will not be transcribed. The inducer can be a chemical agent, such as a metabolite, growth regulator, herbicide or phenolic compound, or a physiological stress directly imposed upon the plant such as cold, heat, drought, flooding, salt or toxins. It is highly desirable to express genes using tightly regulated stress or chemically-inducible promoters.

5 Glyphosate has become the world's most widely-used herbicide for controlling weeds for a number of reasons, including its high efficacy, low cost, and because it is environmentally benign. Using glyphosate along with no-till cropping systems is considered to be a superior economic and environmental choice compared with other systems.^{1, 2} The widespread use of glyphosate, however, has exerted selection pressure on various species of weeds. In fact, agricultural weeds are becoming more difficult to control as they continue to rapidly evolve herbicide resistance.³ Horseweed (*Conyza canadensis*), which is in the Asteraceae family, was the first broadleaf weed to evolve glyphosate resistance,⁴ first occurring in Delaware in 2000. Resistant biotypes are found in 20 US states and several

countries on four continents. We have recently performed a phylogeographic study that gives evidence that horseweed has evolved glyphosate resistance independently in many locations in the USA.⁵ Resistant biotypes seem to abruptly appear and then spread within populations. This within-population spread of resistance is enabled by high seed production (each mature
5 plant can produce more than 200,000 wind-dispersed seeds) coupled with glyphosate treatment that kills non-adapted genotypes.⁶

Horseweed has several attractive features making it amenable for genomics research.^{7,8} From our own flow cytometry experiments, we estimate that horseweed has a genome size of about 335 Mb (unpublished data), which is approximately 2.5 times the size
10 of *Arabidopsis thaliana*. In fact, horseweed has the smallest known genome of all agricultural weeds.⁸ It is self-fertile, has high homozygosity, and is relatively easy to maintain in low-light growth rooms until plants bolt, when they quickly outgrow light rack spacing. Horseweed is a true diploid ($2n=18$), which simplifies sequence analysis compared to polyploid weed species. In addition, we have developed a plant transformation and
5 regeneration method⁹ that allows for overexpression or knockdown analysis of potential gene targets.

SUMMARY OF THE INVENTION

The subject application provides polynucleotides, compositions thereof and methods
10 for regulating gene expression in a plant. Polynucleotides disclosed herein comprise novel sequences for a promoter that initiates transcription in an inducible manner. Further embodiments of the invention comprise the nucleotide sequence of SEQ ID NO: 1 or fragments thereof that are capable of driving the expression of an operably linked nucleic acid sequence. Other polynucleotides disclosed herein provide nucleotide sequences having
5 at least 70% sequence identity to the sequence set forth in SEQ ID NO: 1. Polynucleotides complementary to such polynucleotides (polynucleotide sequences having at least 70% sequence identity to SEQ ID NO: 1) are also provided by the subject application.

Additionally, DNA constructs (sometimes referred to as nucleotide constructs) comprising a promoter, as disclosed herein, operably linked to a heterologous nucleotide
10 sequence of interest wherein said promoter is capable of driving expression of the operably linked heterologous nucleotide sequence in a plant cell are provided. Further aspects of the invention provide expression vectors and plants, seed or plant cells having stably incorporated into their genomes a DNA construct as disclosed herein.

Methods of selectively expressing a nucleotide sequence in a plant, comprising transforming a plant cell with a DNA construct, as disclosed herein, and optionally regenerating a transformed plant from said plant cell are also provided. The DNA construct comprises a promoter and a heterologous nucleotide sequence operably linked to said promoter, wherein said promoter initiates transcription of said nucleotide sequence in a plant cell in an inducible manner. Thus, the promoter disclosed herein is useful for controlling the expression of operably linked coding sequences in an inducible manner.

Downstream from and under the transcriptional initiation regulation of the promoter will be a sequence of interest that will provide for modification of the phenotype of the plant. Such modification caused by the sequence of interest includes modulating the production of an endogenous product, as to amount or relative distribution or the production of an exogenous expression product to provide for a novel function or product in the plant. For example, a heterologous nucleotide sequence that encodes a gene product that confers pathogen, herbicide, salt, cold, drought, or insect resistance to the plant. Other heterologous nucleotide sequences that encode a gene product that confers enhanced nutritional value can be operably linked to promoter sequences disclosed herein.

In a further aspect of the invention, methods for modulating expression of a gene product in a stably transformed plant comprising the steps of (a) transforming a plant cell with a DNA construct comprising the disclosed promoter or fragments thereof that are capable of driving the expression of an operably linked nucleic acid sequence operably linked to at least one nucleotide sequence; (b) growing the plant cell under plant growing conditions and (c) regenerating a stably transformed plant from the plant cell wherein the induced expression of the operably linked nucleotide sequence alters the phenotype of the plant.

BRIEF DESCRIPTION OF THE DRAWING

Figure 1. Frequency distribution of horseweed GS-FLX 454 sequence raw read lengths.

Figure 2. Characteristics of assembled horseweed GS-FLX 454 contigs; (a) length frequency distribution of assembled contigs; (b) average coverage frequency distribution of assembled contigs.

Figure 3. Summary of GO annotation of 454 unique sequences. Annotated sequences were classified into A, "Biological Process" B, "Molecular Function" and C, "Cellular Component" groups and 45 subgroups.

Figure 4. Expression levels of 17 ABC transporters genes in young leaves of glyphosate treated TN-R biotype horseweed plants relative to an internal control actin gene using real-time RT-PCR. Data are presented as mean $\pm$ SE of three technical replicates for each biotype-treatment combination (one pooled sample each).

Figure 5. Relative expression profiles (compared to expression level in TN-S control plants, SC) of 17 ABC transporter genes in young horseweed leaves from the following plant-treatment combinations: Tennessee-susceptible glyphosate-sprayed, SG; Tennessee-resistant untreated control, RC; and Tennessee-resistant glyphosate sprayed, RG. Data are presented as mean $\pm$ SE of three independent real-time RT-PCR analyses. Each RNA sample was isolated from leaves of six individual plants grown under the same conditions for each biotype and treatment and pooled to give one sample each.

Figure 6. Example of one unique sequence annotated by a similarity search of a custom plant protein database via NCBI Standalone Blast program.

Figure 7. Example of tabular annotation information of unique sequences. XML format BlastX results were parsed out with Query ID, hit accession number, annotated protein name, E-value, and score bits.

Figure 8. Number of contigs that have hits to the Arabidopsis protein database at various E-value thresholds.

Figures 9A and 9B. Gene structure of M11 (Fig. 9A). Figure 9B provides the sequence of M11 (see SEQ ID NO: 2); promoter underlined in Fig. 9B (SEQ ID NO: 1); ATG start codon in bold and double underlining.

Figures 10A and 10B. The promoter region of M11 were cloned into the pCR8/GW/TOPO vector, then subcloned into the *pMDC164* plant transformation vector upstream of the GUS reporter gene. The recombinant binary vectors were introduced into *Agrobacterium tumefaciens* GV3101 strain by frozen/thaw method. The constructs were transformed into young leaves of five week old tobacco via infiltration method. After infiltrated for two days, the leaves were treated with different amount Roundup WeatherMAX (0.108, 0.0108, 0.0054, 0.00108 kg/ha ae) or water as control. The transient expression of GUS reporter gene was observed after additional two (Figure 10A) or five days (Figure 10B).

Figure 11. Glyphosate concentration assay. Roundup WeatherMAX (540g/L) was diluted with water for 100, 1000, 10000, 100000 folds, respectively. Then, 2ml of each was

smear on the leaves of tobacco planted in 10cm*10cm plate with a brush. Same amount of water smear was as control. Tobacco plants treated with Roundup for one week.

Figures 12A and B. The ability of the promoter to drive expression of GUS was also examined in transgenic tobacco plants. The promoter (GUS as reporter gene) was induced by glyphosate (ROUNDUP) treatment in stable transgenic tobacco plants (multiple plants from the same independent transgene line). The dark areas in the leaf tips (panel B) clearly demonstrate GUS expression.

Figures 13A and B. The activity of the promoter (GFP as reporter gene) was induced by glyphosate (ROUNDUP) treatment in *Agrobacterium tumefaciens* infiltration tobacco plants. GFP expression is clearly evident in glyphosate treated leaves (light areas in panel B).

Figure 14. The relative activity of glyphosate (ROUNDUP) inducible promoters in stable transgenic tobacco plants (GFP as reporter gene) was assessed. GFP expression level was quantified with a fluorescence spectroscopy. the excitation peak was 490 nm and the emission peak was 509 nm.

Figure 15. The glyphosate (ROUNDUP) inducible promoters were cloned into FLP/*FRT* recombination system (GUS and GFP as reporter genes). In the new system, the capability of the inducible promoters will be amplified after excision and the reporter genes will be driven by the 35S promoter.

DETAILED DESCRIPTION OF THE INVENTION

The subject invention also provides isolated, recombinant, and/or purified polynucleotide sequences comprising:

- a) a polynucleotide sequence comprising SEQ ID NO: 1 or fragments thereof that are capable of driving the expression of an operably linked nucleic acid sequence;
- b) a polynucleotide sequence having at least about 70% to 99.99% identity to a polynucleotide sequence comprising SEQ ID NO: 1 or fragments thereof that are capable of driving the expression of an operably linked nucleic acid sequence;
- c) a polynucleotide that is fully complementary to the polynucleotides set forth in (a) or (b);
- d) a DNA construct comprising a polynucleotide sequence as set forth in (a), (b) or (c) operably linked to a heterologous nucleotide (polynucleotide) sequence;
- e) a host cell comprising a vector as set forth in (d);

f) a polynucleotide that hybridizes under low, intermediate or high stringency with a polynucleotide sequence as set forth in (a), (b) or (c); or

g) a probe comprising a polynucleotide according to (a), (b) or (c) and, optionally, a label or marker.

5 "Nucleotide sequence", "polynucleotide" or "nucleic acid" can be used interchangeably and are understood to mean, according to the present invention, either a double-stranded DNA, a single-stranded DNA or products of transcription of the said DNAs (*e.g.*, RNA molecules). It should also be understood that the present invention does not relate to genomic polynucleotide sequences in their natural environment or natural state. The
10 nucleic acid, polynucleotide, or nucleotide sequences of the invention can be isolated, purified (or partially purified), by separation methods including, but not limited to, ion-exchange chromatography, molecular size exclusion chromatography, or by genetic engineering methods such as amplification, subtractive hybridization, cloning, subcloning or chemical synthesis, or combinations of these genetic engineering methods.

5 A homologous polynucleotide or polypeptide sequence, for the purposes of the present invention, encompasses a sequence having a percentage identity with the polynucleotide or polypeptide sequences, set forth herein, of between at least (or at least about) 20.00% to 99.99% (inclusive). The aforementioned range of percent identity is to be taken as including, and providing written description and support for, any fractional
10 percentage, in intervals of 0.01%, between 20.00% and, up to, including 99.99%. These percentages are purely statistical and differences between two nucleic acid sequences can be distributed randomly and over the entire sequence length. For example, homologous sequences can exhibit a percent identity of 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,
15 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, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent with the sequences of the instant invention. Typically, the percent identity is calculated with reference to the full length, native, and/or naturally occurring polynucleotide. The terms "identical" or percent
20 "identity", in the context of two or more polynucleotide or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using a sequence comparison algorithm or by manual alignment and visual inspection. In certain aspects of the invention,

homologous sequences to SEQ ID NO: 1 have at least 70% sequence identity to SEQ ID NO: 1 over its full length (or over the full length of a given fragment of SEQ ID NO: 1).

Both protein and nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci. USA* 85(8):2444-2448; Altschul *et al.*, 1990, *J. Mol. Biol.* 215(3):403-410; Thompson *et al.*, 1994, *Nucleic Acids Res.* 22(2):4673-4680; Higgins *et al.*, 1996, *Methods Enzymol.* 266:383-402; Altschul *et al.*, 1990, *J. Mol. Biol.* 215(3):403-410; Altschul *et al.*, 1993, *Nature Genetics* 3:266-272). Sequence comparisons are, typically, conducted using default parameters provided by the vendor or using those parameters set forth in the above-identified references, which are hereby incorporated by reference in their entireties.

A “complementary” polynucleotide sequence, as used herein, generally refers to a sequence arising from the hydrogen bonding between a particular purine and a particular pyrimidine in double-stranded nucleic acid molecules (DNA-DNA, DNA-RNA, or RNA-RNA). The major specific pairings are guanine with cytosine and adenine with thymine or uracil. A “complementary” polynucleotide sequence may also be referred to as an “antisense” polynucleotide sequence or an “antisense sequence”. In various aspects of the invention, sequences are “fully complementary” to a reference sequence (e.g., SEQ ID NO: 1). The phrase “fully complementary” refers to sequences contain no mismatches in their base pairing.

Sequence homology and sequence identity can also be determined by hybridization studies under high stringency, intermediate stringency, and/or low stringency. Various degrees of stringency of hybridization can be employed. The more severe the conditions, the greater the complementarity that is required for duplex formation. Severity of conditions can be controlled by temperature, probe concentration, probe length, ionic strength, time, and the like. Preferably, hybridization is conducted under low, intermediate, or high stringency conditions by techniques well known in the art, as described, for example, in Keller, G.H., M.M. Manak [1987] *DNA Probes*, Stockton Press, New York, NY., pp. 169-170.

For example, hybridization of immobilized DNA on Southern blots with ³²P-labeled gene-specific probes can be performed by standard methods (Maniatis *et al.* [1982] *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York). In general, hybridization and subsequent washes can be carried out under intermediate to high

stringency conditions that allow for detection of target sequences with homology to the exemplified polynucleotide sequence. For double-stranded DNA gene probes, hybridization can be carried out overnight at 20-25° C below the melting temperature (T_m) of the DNA hybrid in 6X SSPE, 5X Denhardt's solution, 0.1% SDS, 0.1 mg/ml denatured DNA. The melting temperature is described by the following formula (Beltz *et al.* [1983] *Methods of Enzymology*, R. Wu, L. Grossman and K. Moldave [eds.] Academic Press, New York 100:266-285).

$T_m = 81.5^\circ\text{C} + 16.6 \log[\text{Na}^+] + 0.41(\%G+C) - 0.61(\%\text{formamide}) - 600/\text{length of duplex in base pairs.}$

Washes are typically carried out as follows:

- (1) twice at room temperature for 15 minutes in 1X SSPE, 0.1% SDS (low stringency wash);
- (2) once at $T_m - 20^\circ\text{C}$ for 15 minutes in 0.2X SSPE, 0.1% SDS (intermediate stringency wash).

For oligonucleotide probes, hybridization can be carried out overnight at 10-20°C below the melting temperature (T_m) of the hybrid in 6X SSPE, 5X Denhardt's solution, 0.1% SDS, 0.1 mg/ml denatured DNA. T_m for oligonucleotide probes can be determined by the following formula:

$T_m(^{\circ}\text{C}) = 2(\text{number T/A base pairs}) + 4(\text{number G/C base pairs})$ (Suggs *et al.* [1981] *ICN-UCLA Symp. Dev. Biol. Using Purified Genes*, D.D. Brown [ed.], Academic Press, New York, 23:683-693).

Washes can be carried out as follows:

- (1) twice at room temperature for 15 minutes 1X SSPE, 0.1% SDS (low stringency wash);
- (2) once at the hybridization temperature for 15 minutes in 1X SSPE, 0.1% SDS (intermediate stringency wash).

In general, salt and/or temperature can be altered to change stringency. With a labeled DNA fragment >70 or so bases in length, the following conditions can be used:

Low: 1 or 2X SSPE, room temperature

Low: 1 or 2X SSPE, 42°C

Intermediate: 0.2X or 1X SSPE, 65°C

High: 0.1X SSPE, 65°C.

By way of another non-limiting example, procedures using conditions of high stringency can also be performed as follows: Pre-hybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in pre-hybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 x 10⁶ cpm of ³²P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1X SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2X SSC and 0.1% SDS, or 0.5X SSC and 0.1% SDS, or 0.1X SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety.

Another non-limiting example of procedures using conditions of intermediate stringency are as follows: Filters containing DNA are pre-hybridized, and then hybridized at a temperature of 60°C in the presence of a 5X SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2X SSC at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of intermediate stringency which may be used are well known in the art and as cited in Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety.

Duplex formation and stability depend on substantial complementarity between the two strands of a hybrid and, as noted above, a certain degree of mismatch can be tolerated. Therefore, the probe sequences of the subject invention include mutations (both single and multiple), deletions, insertions of the described sequences, and combinations thereof, wherein said mutations, insertions and deletions permit formation of stable hybrids with the target

polynucleotide of interest. Mutations, insertions and deletions can be produced in a given polynucleotide sequence in many ways, and these methods are known to an ordinarily skilled artisan. Other methods may become known in the future.

It is also well known in the art that restriction enzymes can be used to obtain functional fragments of the subject DNA sequences. For example, *Bal31* exonuclease can be conveniently used for time-controlled limited digestion of DNA (commonly referred to as “erase-a-base” procedures). See, for example, Maniatis *et al.* [1982] *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York; Wei *et al.* [1983] *J. Biol. Chem.* 258:13006-13512.

The present invention further comprises fragments of the polynucleotide sequences of the instant invention. Representative fragments of the polynucleotide sequences according to the invention will be understood to mean any nucleotide fragment having at least 5 successive nucleotides, preferably at least 12 successive nucleotides, and still more preferably at least 15, 18, or at least 20 successive nucleotides of the sequence from which it is derived. The upper limit for such fragments is the total number of nucleotides found in the full-length sequence of SEQ ID NO: 1. The term “successive” can be interchanged with the term “consecutive” or the phrase “contiguous span”. Thus, in some embodiments, a polynucleotide fragment may be referred to as “a contiguous span of at least X nucleotides, wherein X is any integer value between 5 and 1507 (one nucleotide less than the total number of nucleotides found in the full-length sequence (SEQ ID NO: 1)).”

In some embodiments, the subject invention includes those fragments capable of hybridizing under various conditions of stringency conditions (*e.g.*, high or intermediate or low stringency) with a nucleotide sequence according to the invention; fragments that hybridize with a nucleotide sequence of the subject invention can be, optionally, labeled as set forth below.

Thus, the subject invention also provides detection probes (*e.g.*, fragments of the disclosed polynucleotide sequences) for hybridization with a target sequence or the amplicon generated from the target sequence. Such a detection probe will comprise a contiguous/consecutive span of at least 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides of SEQ ID NO: 1. Labeled probes or primers are labeled with a radioactive compound or with another type of label as set forth above (*e.g.*, 1) radioactive labels, 2) enzyme labels, 3) chemiluminescent labels, 4) fluorescent labels, or 5) magnetic labels). Alternatively, non-

labeled nucleotide sequences may be used directly as probes or primers; however, the sequences are generally labeled with a radioactive element (^{32}P , ^{35}S , ^3H , ^{125}I) or with a molecule such as biotin, acetylaminofluorene, digoxigenin, 5-bromo-deoxyuridine, or fluorescein to provide probes that can be used in numerous applications.

5 The promoter sequences disclosed herein are useful for expressing operably linked nucleotide sequences in an inducible manner. As disclosed herein, SEQ ID NO: 1 is a promoter that is induced by glyphosate. As used an “isolated” or “purified” nucleic acid molecule, or biologically active fragment 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. An “isolated” nucleic acid is essentially 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
5 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.

 The nucleic acid of SEQ ID NO: 1 is a promoter. The term “promoter” is intended to mean a regulatory region of DNA usually comprising a TATA box capable of directing RNA polymerase II to initiate RNA synthesis at the appropriate transcription initiation site for a particular coding sequence. A promoter may additionally comprise other recognition sequences generally positioned upstream or 5' to the TATA box, referred to as upstream promoter elements, which influence the transcription initiation rate. It is recognized that having identified the nucleotide sequences for the promoter regions disclosed herein, it is
5 within the state of the art to isolate and identify further regulatory elements in the 5' untranslated region upstream from the particular promoter regions identified herein. The promoter elements that enable inducible expression, can be identified, isolated, and used with core promoters to confer a preferred expression pattern. In this aspect of the invention, a “core promoter” is intended to mean a promoter without promoter elements generally found
 upstream and/or downstream of the core promoter (the minimal portion of the promoter required to properly initiate transcription which includes a Transcription Start Site (TSS) a binding site for RNA polymerase and general transcription factor binding sites.

The term “regulatory element” also refers to a sequence of DNA, usually, but not always, upstream (5') to the coding sequence of a structural gene, which includes sequences which control the expression of the coding region by providing the recognition for RNA polymerase and/or other factors required for transcription to start at a particular site. An example of a regulatory element that provides for the recognition for RNA polymerase or other transcriptional factors to ensure initiation at a particular site is a promoter element. A promoter element comprises a core promoter element, responsible for the initiation of transcription, as well as other regulatory elements (as discussed elsewhere in this application) that modify gene expression. It is to be understood that nucleotide sequences, located within introns, or 3' of the coding region sequence may also contribute to the regulation of expression of a coding region of interest. A regulatory element may also include those elements located downstream (3') to the site of transcription initiation, or within transcribed regions, or both. In the context of this disclosure, a post-transcriptional regulatory element may include elements that are active following transcription initiation, for example translational and transcriptional enhancers, translational and transcriptional repressors, and mRNA stability determinants.

The regulatory elements, or fragments thereof, may be operatively associated with heterologous regulatory elements or promoters in order to modulate the activity of the heterologous regulatory element. Such modulation includes enhancing or repressing transcriptional activity of the heterologous regulatory element, modulating post-transcriptional events, or both enhancing or repressing transcriptional activity of the heterologous regulatory element and modulating post-transcriptional events.

The promoter sequences disclosed herein, when assembled within a DNA construct such that the promoter is operably linked to a nucleotide sequence of interest, enable expression of the nucleotide sequence in the cells of a plant stably transformed with this DNA construct. The term “operably linked” is intended to mean that the transcription or translation of the heterologous nucleotide sequence is under the influence of the promoter sequence. “Operably linked” is also intended to mean the joining of two nucleotide sequences such that the coding sequence of each DNA fragment remain in the proper reading frame. In this manner, the nucleotide sequences for the promoters are provided in DNA constructs along with the nucleotide sequence of interest, typically a heterologous nucleotide sequence, for expression in the plant of interest. The term “heterologous nucleotide sequence” is intended to mean a sequence that is not naturally operably linked with the promoter sequence.

While this nucleotide sequence is heterologous to the promoter sequence, it may be homologous, or native; or heterologous, or foreign, to the plant host. It is recognized that the promoters may be used with their native coding sequences to increase or decrease expression, thereby resulting in a change in phenotype of the transformed plant after treatment with glyphosate.

Fragments and variants of the disclosed promoter sequences are also encompassed. A “fragment” is intended to mean a portion of the promoter sequence. Fragments of a promoter sequence may retain biological activity (the ability to drive expression of an operably linked nucleotide sequence in an inducible manner; this may also be referred to as a “biologically active portion” of the promoter). Thus, for example, less than the entire promoter sequence disclosed herein may be utilized to drive expression of an operably linked nucleotide sequence of interest, such as a nucleotide sequence encoding a heterologous protein.

Accordingly, a fragment of the promoter of SEQ ID NO: 1 can contain a biologically active portion of the promoter or it may be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed below. A biologically active portion of the promoter of SEQ ID NO: 1 can be prepared by isolating fragments of SEQ ID NO: 1 and assessing the activity of that fragment in causing the expression of an operably linked nucleic acid sequence (such as a reporter gene). Nucleic acid molecules that are fragments of a promoter nucleotide sequence comprise at least 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900 nucleotides or can be fragments that range from one nucleotide fewer (i.e., 911 nucleotides) than the number of nucleotides present in the full-length promoter nucleotide sequence disclosed herein, e.g., 912 nucleotides for SEQ ID NO:1 to fragments that are about 900 nucleotides shorter than the number of nucleotides present in the full-length promoter nucleotide sequence disclosed herein. Such fragments will usually comprise the TATA recognition sequence of the particular promoter sequence and can be obtained by use of restriction enzymes to cleave the naturally occurring promoter nucleotide sequence disclosed herein; by synthesizing a nucleotide sequence from the naturally occurring sequence of the promoter DNA sequence; or through the use of PCR technology. See particularly, Mullis et al. (1987) *Methods Enzymol.* 155:335-350, and Erlich, ed. (1989) *PCR Technology* (Stockton Press, New York). Variants of these promoter fragments, such as those resulting from site-directed mutagenesis and a procedure such as DNA “shuffling”, are also encompassed by the instant disclosure.

The term “variants” is intended to mean sequences having substantial similarity with a promoter sequence disclosed herein. For nucleotide sequences, naturally occurring 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 as outlined below. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis. Generally, variants of a particular nucleotide sequence will have at least 40%, 50%, 60%, 65%, 70%, generally at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, to 95%, 96%, 97%, 98%, 99% or more sequence identity to that particular nucleotide sequence (e.g., SEQ ID NO: 1) as determined by sequence alignment programs described elsewhere herein using default parameters. Biologically active variants are also encompassed. Biologically active variants include, for example, the native promoter sequence having one or more nucleotide substitutions, deletions, or insertions. Promoter activity may be measured by using techniques such as Northern blot analysis, reporter activity measurements taken from transcriptional fusions, and the like. See, for example, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), hereinafter “Sambrook,” herein incorporated by reference. Alternatively, levels of a reporter gene such as green fluorescent protein (GFP) or the like produced under the control of a promoter fragment or variant can be measured. See, for example, U.S. Patent No. 6,072,050, herein incorporated by reference. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel et al. (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein.

Variant promoter nucleotide sequences also encompass sequences derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different promoter sequences can be manipulated to create a new promoter possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined in vitro or in vivo. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Crameri et al. (1997) *Nature Biotech.* 15:436-438; Moore et al. (1997) *J. Mol.*

Biol. 272:336-347; Zhang et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Cramer et al. (1998) *Nature* 391:288-291; and U.S. Pat. Nos. 5,605,793 and 5,837,458.

The nucleotide sequences disclosed herein can be used to isolate corresponding sequences from other plants. In this manner, methods such as PCR, hybridization, and the like can be used to identify such sequences based on their sequence homology to the sequence set forth herein. Sequences isolated on the basis of sequence identity to SEQ ID NO: 1, or to fragments thereof, are encompassed by this disclosure.

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 and are disclosed in Sambrook, *supra*. See also Innis et al., eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis and Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); and Innis and Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). 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.

The promoter sequence disclosed herein, as well as variants and fragments thereof, are useful for genetic engineering of plants, e.g. for the production of a transformed or transgenic plant, to express a phenotype of interest. As used herein, the terms “transformed plant” and “transgenic plant” refer to a plant that comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome of a transgenic or transformed plant such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant DNA construct. It is to be understood that as used herein the term “transgenic” includes any cell, cell line, callus, tissue, plant part, or plant the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term “transgenic” as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

A transgenic “event” is produced by transformation of plant cells with a heterologous DNA construct, including a nucleic acid DNA construct that comprises a transgene of interest, the regeneration of a population of plants resulting from the insertion of the transgene into the genome of the plant, and selection of a particular plant characterized by insertion into a particular genome location. An event is characterized phenotypically by the expression of the transgene. At the genetic level, an event is part of the genetic makeup of a plant. The term “event” also refers to progeny produced by a sexual outcross between the transformant and another variety that include the heterologous DNA.

As used herein, the term “plant” includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, and progeny of same. Parts of transgenic plants are to be understood within the scope of the invention comprise, for example, plant cells, protoplasts, tissues, callus, embryos as well as flowers, stems, fruits, ovules, leaves, or roots originating in transgenic plants or their progeny previously transformed with a DNA molecule of the invention, and therefore consisting at least in part of transgenic cells. As used herein, the term “plant cell” includes, without limitation, seeds suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Monocotyledonous and dicotyledonous plants can be transformed with a promoter or DNA construct as disclosed herein.

The promoter sequences and methods disclosed herein are useful in regulating expression of any heterologous nucleotide sequence in a host plant. Thus, the heterologous nucleotide sequence operably linked to the promoters disclosed herein may be a structural gene encoding a protein of interest. Genes of interest are reflective of the commercial markets and interests of those involved in the development of the crop. General categories of genes of interest include, for example, those genes involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. More specific categories of transgenes, for example, include genes encoding proteins conferring resistance to abiotic stress, such as drought, flooding, temperature (heat or cold), salinity, and toxins such as pesticides and herbicides, or to biotic stress, such as attacks by fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms. Various changes in phenotype are of interest including modifying expression of a gene in a specific plant tissue, altering a plant's pathogen or insect defense mechanism, increasing the plant's tolerance to herbicides or nutrient content, altering tissue development to respond to environmental stress, and the like. The results can

be achieved by providing expression of heterologous 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, transporters, or cofactors, or affecting nutrients uptake in the plant. These changes result in a change in phenotype of the transformed plant.

It is recognized that any gene of interest can be operably linked to the promoter sequences disclosed herein and expressed in plant tissues. Thus, a DNA construct comprising a gene of interest, such as those described below, to create plants having a desired phenotype (e.g., disease, herbicide or insect resistance), to create heat or cold tolerance in a plant or to create or enhance resistance to drought or flood conditions in a plant. Accordingly, this disclosure encompasses methods that are directed to protecting plants against flooding, drought, heat, cold, fungal pathogens, bacteria, viruses, nematodes, insects, and the like. By “disease resistance” or “insect resistance” is intended that the plants avoid the harmful symptoms that are the outcome of the plant-pathogen interactions.

Disease resistance and insect resistance genes such as lysozymes, cecropins, maganins, or thionins for antibacterial protection, or the pathogenesis-related (PR) proteins such as glucanases and chitinases for anti-fungal protection, or *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, lectins, and glycosidases for controlling nematodes or insects are all examples of useful gene products. Pathogens include, but are not limited to, viruses or viroids, bacteria, insects, nematodes, fungi, and the like. Viruses include tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Nematodes include parasitic nematodes such as root knot, cyst, and lesion nematodes, etc.

Genes encoding disease resistance traits include detoxification genes, such as against fumonisin (U.S. Pat. No. 5,792,931) avirulence (avr) and disease resistance (R) genes (Jones et al. (1994) *Science* 266:789; Martin et al. (1993) *Science* 262:1432; Mindrinos et al. (1994) *Cell* 78:1089); and the like. Insect resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European corn borer, and the like. Such genes include, for example, *Bacillus thuringiensis* toxic protein genes (U.S. Pat. Nos. 5,366,892; 5,747,450; 5,737,514; 5,723,756; 5,593,881; and Geiser et al. (1986) *Gene* 48:109); lectins (Van Damme et al. (1994) *Plant Mol. Biol.* 24:825); and the like.

Herbicide resistance traits may be introduced into plants by genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene

containing mutations leading to such resistance, in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides that act to inhibit action of glutamine synthase, such as phosphinothricin or Basta® (glufosinate) (e.g., the bar gene), or other such genes known in the art. The bar gene encodes resistance to the herbicide Basta®, the nptII gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS gene encodes resistance to the herbicide chlorsulfuron. Glyphosate resistance is imparted by mutant 5-enolpyruvyl-3-phosphikimate synthase (EPSP) and aroA genes. See, for example, U.S. Pat. No. 4,940,835, which discloses the nucleotide sequence of a form of EPSPS which can confer glyphosate resistance. U.S. Pat. No. 5,627,061 also describes genes encoding EPSPS enzymes. See also U.S. Pat. Nos. 6,248,876; 6,040,497; 5,804,425; 5,633,435; 5,145,783; 4,971,908; 5,312,910; 5,188,642; 4,940,835; 5,866,775; 6,225,114; 6,130,366; 5,310,667; 4,535,060; 4,769,061; 5,633,448; 5,510,471; RE 36,449; RE 37,287; and 5,491,288; and international publications WO 97/04103; WO 97/04114; WO 00/66746; WO 01/66704; WO 00/66747 and WO 00/66748, which are incorporated herein by reference for this purpose. Glyphosate resistance is also imparted to plants that express a gene that encodes a glyphosate oxido-reductase enzyme as described more fully in U.S. Pat. Nos. 5,776,760 and 5,463,175, which are incorporated herein by reference for this purpose. In addition glyphosate resistance can be imparted to plants by the over-expression of genes encoding glyphosate N-acetyltransferase.

Commercial traits can also be encoded on a gene or genes that could increase for example, starch for ethanol production, or provide expression of proteins. Another important commercial use of transformed plants is the production of polymers and bioplastics such as described in U.S. Pat. No. 5,602,321. Genes such as β -ketothiolase, PHBase (polyhydroxybutyrate synthase), and acetoacetyl-CoA reductase (see Schubert et al. (1988) *J. Bacteriol.* 170:5837-5847) facilitate expression of polyhydroxyalkanoates (PHAs).

Agronomically important traits that affect quality of grain or various other plants (e.g., soybean or maize/corn), such as levels and types of oils, saturated and unsaturated, quality and quantity of essential amino acids, levels of cellulose, starch, and protein content can be genetically altered. Modifications include increasing content of oleic acid, saturated and unsaturated oils, increasing levels of lysine and sulfur, providing essential amino acids, and modifying starch. Hordothionin protein modifications in corn are described in U.S. Pat. Nos. 5,990,389; 5,885,801; 5,885,802 and 5,703,049; herein incorporated by reference. Another example is lysine and/or sulfur rich seed protein encoded by the soybean 2S albumin described in U.S. Pat. No. 5,850,016, and the chymotrypsin inhibitor from barley, Williamson

et al. (1987) *Eur. J. Biochem.* 165:99-106, the disclosures of which are herein incorporated by reference.

Exogenous products include plant enzymes and products as well as those from other sources including prokaryotes and other eukaryotes. Such products include enzymes, cofactors, hormones, and the like. Examples of other applicable genes and their associated phenotype include genes that confer viral resistance; genes that confer fungal resistance; genes that confer insect resistance; genes that promote yield improvement; and genes that provide for resistance to stress, such as dehydration resulting from heat and salinity, toxic metal or trace elements, or the like.

In one embodiment, DNA constructs will comprise a transcriptional initiation region comprising a promoter sequence, as disclosed herein, or variants or fragments thereof, operably linked to a heterologous nucleotide sequence whose expression is to be controlled by the promoter. Such a DNA construct is provided with a plurality of restriction sites for insertion of the nucleotide sequence to be under the transcriptional regulation of the regulatory regions. The DNA construct may additionally contain selectable marker genes.

Where appropriate, the heterologous nucleotide sequence whose expression is to be under the control of the promoter sequence disclosed herein may be optimized for increased expression in the transformed plant. That is, these nucleotide sequences can be synthesized using plant preferred codons for improved expression. Methods are available in the art for synthesizing plant-preferred nucleotide sequences. See, for example, U.S. Pat. Nos. 5,380,831 and 5,436,391, and Murray et al. (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

Reporter genes or selectable marker genes may be included in the DNA constructs. Examples of suitable reporter genes known in the art can be found in, for example, Jefferson et al. (1991) in *Plant Molecular Biology Manual*, ed. Gelvin et al. (Kluwer Academic Publishers), pp. 1-33; DeWet et al. (1987) *Mol. Cell. Biol.* 7:725-737; Goff et al. (1990) *EMBO J.* 9:2517-2522; Kain et al. (1995) *BioTechniques* 19:650-655; and Chiu et al. (1996) *Current Biology* 6:325- 330. Selectable marker genes for selection of transformed cells or tissues can include genes that confer antibiotic resistance or resistance to herbicides. Examples of suitable selectable marker genes include, but are not limited to, genes encoding resistance to chloramphenicol (Herrera Estrella et al. (1983) *EMBO J.* 2:987-992); methotrexate (Herrera Estrella et al. (1983) *Nature* 303:209-213; Meijer et al. (1991) *Plant Mol. Biol.* 16:807-820); hygromycin (Waldron et al. (1985) *Plant Mol. Biol.* 5:103-108;

Zhijian et al. (1995) *Plant Science* 108:219-227); streptomycin (Jones et al. (1987) *Mol. Gen. Genet.* 210:86-91); spectinomycin (Bretagne-Sagnard et al. (1996) *Transgenic Res.* 5:131-137); bleomycin (Hille et al. (1990) *Plant Mol. Biol.* 7:171-176); sulfonamide (Guerineau et al. (1990) *Plant Mol. Biol.* 15:127-136); bromoxynil (Stalker et al. (1988) *Science* 242:419-423); glyphosate (Shaw et al. (1986) *Science* 233:478-481); phosphinothricin (DeBlock et al. (1987) *EMBO J.* 6:2513-2518). Other genes that could serve utility in the recovery of transgenic events but might not be required in the final product would include, but are not limited to, examples such as GUS (b-glucuronidase; Jefferson (1987) *Plant Mol. Biol. Rep.* 5:387), GFP (green fluorescence protein; Chalfie et al. (1994) *Science* 263:802), luciferase (Riggs et al. (1987) *Nucleic Acids Res.* 15(19): 8115 and Luehrsen et al. (1992) *Methods Enzymol.* 216:397-414), and the maize genes encoding for anthocyanin production (Ludwig et al. (1990) *Science* 247:449).

The nucleic acid molecules disclosed herein are useful in methods of expressing a nucleotide sequence in a plant. This may be accomplished by transforming a plant cell of interest with a DNA construct comprising a promoter identified herein, operably linked to a heterologous nucleotide sequence, and regenerating a stably transformed plant from said plant cell. The plant can then be exposed to glyphosate that cause the promoter to drive expression of the heterologous nucleotide sequence.

Plant species suitable for transformation include, but are not limited to, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers.

Other plants suitable for transformation with a promoter as disclosed herein include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Additionally, monocots, such as maize, rice, barley, oats, wheat, sorghum, rye, sugarcane, pineapple, yams, onion, banana, coconut, and dates.

As used herein, “vector” refers to a DNA molecule such as a plasmid, cosmid, or bacterial phage for introducing a nucleotide construct, for example, a DNA construct, into a 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.

Various methods disclosed herein include introducing a nucleotide (DNA) construct into a plant. The term “introducing” is used herein to mean presenting to the plant the nucleotide construct in such a manner that the construct gains access to the interior of a cell of the plant. These methods do not depend on a particular method for introducing a nucleotide construct to a plant, only that the nucleotide construct gains access to the interior of at least one cell of the plant. Methods for introducing nucleotide constructs into plants are known in the art including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

By “stable transformation” is intended that the nucleotide construct introduced into a plant integrates into the genome of the plant and is capable of being inherited by progeny thereof. By “transient transformation” is intended that a nucleotide construct introduced into a plant does not integrate into the genome of the plant. The nucleotide constructs disclosed herein may be introduced into plants by contacting plants with a virus or viral nucleic acids. Generally, such methods involve incorporating a nucleotide construct within a viral DNA or RNA molecule. Methods for introducing nucleotide constructs into plants and expressing a

protein encoded therein, involving viral DNA or RNA molecules, are known in the art. See, for example, U.S. Pat. Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367, and 5,316,931; herein incorporated by reference.

Transformation protocols as well as protocols for introducing nucleotide sequences into plants may vary depending on the type of plant or plant cell, i.e., monocot or dicot, targeted for transformation. Suitable methods of introducing nucleotide sequences into plant cells and subsequent insertion into the plant genome include microinjection (Crossway et al. (1986) *Biotechniques* 4:320-334), electroporation (Riggs et al. (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, *Agrobacterium*-mediated transformation (U.S. Pat. Nos. 5,981,840 and 5,563,055), direct gene transfer (Paszkowski et al. (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, U.S. Pat. Nos. 4,945,050; 5,879,918; 5,886,244; 5,932,782; Tomes et al. (1995) in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); and McCabe et al. (1988) *Biotechnology* 6:923-926). Also see Weissinger et al. (1988) *Ann. Rev. Genet.* 22:421-477; Sanford et al. (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou et al. (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe et al. (1988) *Bio/Technology* 6:923-926 (soybean); Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh et al. (1998) *Theor. Appl. Genet.* 96:319-324 (soybean); Datta et al. (1990) *Biotechnology* 8:736-740 (rice); Klein et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein et al. (1988) *Biotechnology* 6:559-563 (maize); U.S. Pat. Nos. 5,240,855; 5,322,783 and 5,324,646; Klein et al. (1988) *Plant Physiol.* 91:440-444 (maize); Fromm et al. (1990) *Biotechnology* 8:833-839 (maize); Hooykaas-Van Slogteren et al. (1984) *Nature (London)* 311:763-764; U.S. Pat. No. 5,736,369 (cereals); Bytebier et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet et al. (1985) in *The Experimental Manipulation of Ovule Tissues*, ed. Chapman et al. (Longman, N.Y.), pp. 197-209 (pollen); Kaeppler et al. (1990) *Plant Cell Reports* 9:415-418 and Kaeppler et al. (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin et al. (1992) *Plant Cell* 4:1495-1505 (electroporation); Li et al. (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford (1995) *Annals of Botany* 75:407-413 (rice); Osjoda et al. (1996) *Nature Biotechnology* 14:745-750 (maize via *Agrobacterium tumefaciens*); all of which are herein incorporated by reference.

The cells that have been transformed may be grown into plants according to methods known in the art. See, for example, McCormick et al. (1986) *Plant Cell Reports* 5:81-84.

These plants may then be grown, and either pollinated with the same transformed plant variety or different varieties, and the resulting hybrid having a desired phenotypic characteristic. Two or more generations may be grown to ensure that the expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure that expression of the desired phenotypic characteristic has been achieved. Thus as used herein, "transformed seeds" refers to seeds that contain the nucleotide construct stably integrated into the plant genome.

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

Another aspect of the invention provides methods of identifying homologous promoters. In various embodiments of this aspect of the invention, nucleic acid sequences are compared against the disclosed promoter sequence as discussed above and sequences having at least 20% sequence identity over the entire length of SEQ ID NO: 1 can be selected for further evaluation. Thus, putative glyphosate inducible promoters can exhibit at least 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, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent identity with SEQ ID NO: 1 (over its full length). In certain embodiments, the putative promoters have at least 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent with SEQ ID NO: 1 (over its full length).

Determination of the percent identity between two sequences can be accomplished, for example, using a mathematical algorithm, such as the mathematical algorithm utilized for the comparison of two sequences disclosed in Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264, or modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873- 5877 or by visual alignment. Where similarity/identity searches are carried out using commercially available sequence alignment suites, default alignment parameters can be used. Nucleic acid sequences that are compared against the disclosed promoter sequence (SEQ ID NO: 1) can be mined from commercially available databases or isolated from various plant sources. Potential glyphosate inducible promoters identified using this aspect of the invention can be screened by operably linking the putative promoter sequence to a reporter or selectable marker as disclosed herein and using the construct to transform a plant or plant part. The transformed plant can then be treated with glyphosate and the ability of the putative promoter to drive expression of the reporter or selectable marker assessed by methods known in the art, including comparison of expression levels of a reporter or selectable marker between glyphosate treated plants and control plants (plants to which glyphosate is not applied). An increase in reporter or selectable marker expression in glyphosate treated plants is indicative of the inducibility of a promoter identified by this aspect of the invention. Of course other traits can be linked to a putative promoter identified according to this aspect of the invention (for example, a gene product that confers improved nutritional content or resistance to herbicides, salts, heat, cold, flood, drought, pathogens, or insects). Increases in reporter or selectable marker expression can be assessed by methods known in the art or as discussed in the Examples that follow. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

EXAMPLES

MATERIAL AND METHODS

Sample preparation for 454 sequencing

Horseweed plants were grown in potting media in a greenhouse at the University of Tennessee, Knoxville, TN, USA under a 16 hour photoperiod at ambient temperatures ($25\pm 2^{\circ}\text{C}$). Plants were watered and fertilized as necessary with Osmocote slow-release fertilizer. Young leaves and meristematic tissue were harvested at the rosette stage from

plants that were approximately 3 months old and 6 to 8 cm in diameter. Total RNA was isolated and pooled into one sample from the following three sample types represented by six plants each: untreated (water-sprayed) TN-susceptible horseweed (from Knoxville, TN, USA), and TN-resistant biotype from western Tennessee (Lauderdale County, TN, USA), and the latter biotype after 24 h glyphosate-sprayed with the field rate of Round-up Weathermax (0.84 kg/ha ae, Monsanto, St. Louis, MO, USA).²⁴ RNA extraction was done using TriReagent according to the manufacturer's protocol (MRC, Cincinnati, OH, USA). The pooled sample was used to generate double-stranded cDNA using SMART™ cDNA Library Construction Kit (Clontech, Mountain View, CA, USA). Normalization was performed using TRIMMER cDNA normalization kit (Evrogen, Moscow, Russia) to decrease the prevalence of abundant transcripts before sequencing. The cDNA sample was then fractionated into smaller pieces (300–500 bp). The ends of these fragments were subsequently polished by treating with DNA polymerase to fill in or remove any unpaired bases. The short A and B adaptors were then ligated on to each resulting fragment, which provided priming sequences for both emulsion PCR amplification and pyrosequencing, forming the basis of the single-stranded template library. Pyrosequencing using a Roche GS-FLX sequencer was performed by the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois as described previously.^{12,25} A preliminary titration run was followed with two bulk runs. The first bulk run was dedicated to horseweed cDNAs whereas in the second run, only half the plate was allocated to horseweed cDNAs. Raw reads of 454 data were submitted to the GenBank Short Read Archive (SRA) database, accession number is SRA010952 (which is hereby incorporated by reference in its entirety).

454 sequencing data trimming, assembling and annotation

The raw 454 sequences were processed according to standard protocols used in our waterhemp (*Amaranthus tuberculatus*) paper²⁶ and the waterhemp transcriptome data was used as a comparison in this study. Lucy²⁷ and EGAssembler (<http://egassembler.hgc.jp/>) were used to remove the low-quality sequences, end regions that are rich in ambiguous nucleotides, very short reads (<50 bp), poly (A/T) tails, SMART™ adaptors for cDNA synthesis, primers and potential contaminating vector sequences. The returned high-quality clean sequences were assembled using CAP3²⁸ and EGAssembler. All unique sequences (contigs and singletons) were annotated by similarity search (NCBI Standalone Blast program, <ftp://ftp.ncbi.nih.gov/blast/>) of three protein databases: Arabidopsis all proteins

database (AGIallAA.gz, 130,814 protein sequences; ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/other_datasets/12-18-07/), UniProtKB/Swiss-Prot annotated protein database (353,658 protein sequences, http://www.uniprot.org/downloads), and a custom protein database, in which included all green plants proteins from GenBank (677,422 protein sequences, ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/). The best five protein hits for each query sequences were parsed out to create annotated tables, which included available information, such as taxonomy, key words, protein function, accession number and/or gene ontology (GO) terms. The potential microRNAs in non-annotated sequences were scanned by searching the miRBase database (ftp://mirbase.org/pub/mirbase/CURRENT/). To help determine which sequences were of non-plant origin, contigs and singletons that had no hits found in custom plant proteins database were further searched using the Uniport database (Release 15.14).

Expression analysis of selected ABC transporter genes using real-time RT-PCR

Plants were grown, harvested, and total RNA extracted as described above. Four combinations of plant biotypes and treatments were made: TN-S and TN-R biotypes that were glyphosate-treated and untreated were compared for gene expression differences. Young leaves of six individual plants were used for total RNA extraction for each biotype-treatment combination. Therefore, the four combinations were represented by one sample each. The residual genomic DNA in the total RNA extract was removed by several treatments with RNase-free DNase I (Invitrogen, Carlsbad, CA, USA). First strand cDNA was synthesized using: 2 µg of total RNA, 0.5 µg oligo(dT)₁₈ and SuperScript® III reverse transcriptase, according to the manufacturer's instructions (Invitrogen) employing a Eppendorf MasterCycler (Eppendorf, Hamburg, GER). The cDNAs were diluted to 100 µl with sterile water of which 2 µl was used per real-time PCR sample. Real-time PCR was carried out in an ABI-7000 thermal cycling system using a real-time PCR Power Mix Kit (ABI, Foster City, CA, USA). The reaction mixture (25 µl) contained 2 µl of first strand cDNA, 0.5 µM of each of the forward and reverse primers and appropriate amounts of other components as recommended by the manufacturer (ABI). ABI-7000 thermal cycler was programmed as follows: 2 min at 95 °C for pre-denature; 40 cycles of 15 s at 94 °C, 15 s at 55 °C, 20 s at 72 °C. Data were collected during the extension step. The cDNA samples were tested by using three independent repetitions in the same condition. For control reactions, either no sample was added or RNA alone was added without reverse transcription to test if

the RNA sample was contaminated with genomic DNA. An actin-like housekeeping gene (contig9305, 916 bp) was used as a reference gene. The absolute expression level of this actin-like gene was relatively invariant (average ± 0.31 Ct value, within 10% variation) using equal amounts of cDNA samples from glyphosate treated plants in this study. Furthermore, abiotic stresses (salt, drought and cold; data not shown) did not perturb its expression. The relative expression of target genes to the actin control was calculated using the efficiency adjusted $\Delta\Delta\text{Ct}$ method as described by Yuan et al.²⁹ The oligonucleotide primers (Table 6) were designed with the Primer Express 2.0 software (ABI). To test the suitability of these primer sets, the specificity and identity of the RT-PCR products were monitored by a melting curve analysis (65–99°C, 5°C s⁻¹) of the reaction products, which can distinguish the gene-specific PCR products from the nonspecific PCR products. All primers were synthesized by Integrated DNA Technologies (IDT, Iowa City, IA, USA).

RESULTS AND DISCUSSION

Roche GS-FLX sequencing and assembly

Normalized cDNA was used to reduce oversampling of high abundance transcripts and obtain sufficient coverage of low abundance transcripts. Two sequencing runs (1.5 plates) plus a titration run yielded a total of 411,962 raw reads. The average length of each read was 233 bp (Table 7) with 79.2% distributed between 200 bp and 300 bp, and a total data size was 95.8 Mb (Fig. 1). The sequence yield was somewhat lower compared with genomic DNA 454 sequencing, but was higher than other *de novo* transcriptome sequencing projects for non-model plant species.^{30, 31} The difference resulted from shorter DNA fragments from the transcriptome preparation or other input effects compared with those data from genome studies. Compared with Sanger EST library sequencing methods, cDNA molecules needed to be fractionated into smaller pieces and size-scanned rather than fully cloned into vectors. Shotgun 454 sequences are located evenly across the cDNA of a given gene,³² which resulted in multiple fragments per single gene, requiring further analysis to assess their relationships.

Initial quality filtering of the 454 reads was performed at the machine level before base-calling. These sequences were subsequently trimmed as described in Materials and Methods. Ninety-four percent of sequences (379,152) passed the quality-control filter for assembly into unique sequences. A total of 363,471 high quality clean sequences resulted in 7.05 Mb representing 16,102 contigs. After assembly, 55% of contigs (8,817) were longer than 300 bp and 19.5% of contigs (3,145) were longer than 600 bp (Fig. 2). Of these contigs,

15,681 high-quality clean sequences (3.8%) remained as singletons (coverage depth = 1) with data size totaling 3.3 Mb. This resulted in 10.35 Mb of new horseweed transcriptome sequencing data representing 31,783 unique sequences (Table 1). Further quality-trimmed 2,016 unique transcripts (average length, 689 bp; total size, 1.39 Mb) obtained from horseweed cDNA libraries using traditional Sanger sequencing techniques⁵ were used to gauge the quality of the 454 sequencing and assembly.

The most challenging aspect of *de novo* assembly is obtaining abundant coverage of sequences. In this study, 95.9% of the high quality trimmed sequences were assembled into contigs with an average length of 438 bp. However, this average length was still shorter than the average length of Sanger ESTs. Given that the average coverage depth for each contig and each nucleotide position was ~22-fold and ~12-fold, respectively, this high coverage depth of contigs ensured the 454 sequences were likely more accurate than traditional Sanger sequences that rely on a single or very few reads.

Quality and performance of the 454 assembly

To test the quality and performance of the sequence assembly, we aligned contigs against themselves and the singletons using the NCBI Blastn program. 4,405 contigs (27.1%) had best Blast hits (i.e., had significantly similar sequences based on a bitscore > 45, E-value <0.0001 produced by the BlastN program) with >95% identity with other contigs and singletons, but in no case did these alignments extend over the entire length of either the Blast subjects or queries. These perfect match alignments averaged 92 bp and 73 bp in length for contig vs. contigs and contigs vs. singleton hits, respectively. Also, the average coverage of the match alignments were 18.5% and 11.5% of the length of the queried contigs in the cases of contigs vs. contigs and contigs vs. singletons, respectively. 2,768 of these contigs had Blastx hits (bitscore > 45) against the all green plants protein database, and only 285 (1.8%) of those Blastn-paired contigs had same best Blastx hits in the protein database (Table 2). Considering conserved motifs in different genes widely exist in the genome and different transcripts resulting from alternative splicing of single genes occurs frequently,^{33,34} our assembly appropriately partitioned these gene regions that produced high identity but short coverage alignments into different contigs.

To estimate the error rate of 454 sequencing and the quality of assembly, 2,016 high-quality trimmed Sanger-sequenced ESTs⁵ were aligned with the 454 contigs and singletons. Of these, 1,540 (76.4%) had strong Blast hits to 454 sequences (Table 3). Nucleotide alignments of Sanger vs. 454 sequences were 95.8% identical for all alignments, 97.3% for

those alignments involving 454 contigs and 99.3% for those alignments with bitscore>100. The average number of gaps for alignments involving 454 singletons was 0.22 and 7 per 1,000 aligned bases. The average number of gaps for alignments involving 454 contigs was 0.04 and 1.4 per 1,000 aligned bases, which was less than that for 454 singletons. This comparison might overestimate the real 454 sequencing error rates since they include base mismatches caused by polymorphisms, possible gaps created by alternative splicing, and alignments with end regions of Sanger sequences, which are known to have decreased accuracy. The horseweed genotypes between Sanger and 454 sequencing were not the same. However, these results indicated a sufficient coverage depth could efficiently reduce the error rate in 454 sequencing and it is reasonable to suggest that it could be more accurate than traditional Sanger sequences on the basis of depth of coverage.

Functional annotation of 454 unique sequences

One objective of this project was to assign hypothetical protein sequence and function to each EST. All unique sequences (contigs and singletons) were used as queries to search annotated protein databases and were assigned a gene description and/or a GO term (Figures 6 and 7). A number of factors, especially the E-value, affect the reliability of results when searching databases for similarities. The E-value is the probability, due to chance, that there is another alignment with a similarity greater than the given bitscore. In short, a lower E-value set translates to higher confidence in the search results. A total of 10,698 contigs had hits to the protein database with the E-value threshold set at <0.1, which was 1,438 (~16%) more than that with the E-value threshold set at <0.0001 (Figure 8). We sought to enlarge the database to allow maximal functional searching for gene discovery in this *de novo* transcriptome sequencing project. The number of contigs that had hits to different protein databases was counted based on the ‘best 5 hits’ of Blastx search (E-value <0.0001, score bits>45). The greatest yield of protein counts was obtained when searching the all green plant protein database, which hit 629 more contigs than searching the *Arabidopsis* protein database; about 20% more putative proteins were identified. Thus, a total 16,306 unique sequences were annotated. Of these, 13,708 (84.1%) were associated with Biological Process GO classification and were divided into 14 subgroups, 12,404 (76.1%) were associated with “cellular components” and were further divided into 16 subgroups, and 7,364 (45.2%) were associated with “molecular function” and were divided into 15 subgroups (Fig. 3).

Only 39.8% of singletons found hits in our custom protein database and could be annotated, while 61.5% of the contigs could be annotated, possibly the result of low coverage

depth and short average length of the singletons. The average length of annotated contigs was 526 bp with a 30.3 average coverage, while non-annotated contigs averaged 297 bp with only 9.6 coverage. Similarly, the average length of annotated singletons was 30 bp longer than that of non-annotated singletons. In 15,477 non-annotated unique sequences, ~2.8% of these (431) had hits in plant microRNA database, ~0.9% of them (134) had hits with non-plant proteins (Table 5).

The effectiveness of this horseweed 454 transcriptome sequencing for identifying gene candidates involved in herbicide resistance was estimated by comparing with waterhemp data²⁶. Eleven herbicide target-site genes/gene families and four non-target gene families were identified from unique horseweed and waterhemp sequences (Table 4). In ten gene families, we identified more resistance-gene candidates in waterhemp compared with horseweed. In the remaining five gene families, we identified more resistance gene candidates in horseweed than waterhemp. In short, some 430 unique sequences we identified in this study might be involved in the evolution of herbicide resistance. This finding demonstrates the enormous value of 454 transcriptome sequencing for gene discovery in an important weedy plant with scant sequence data. In the following section, we illustrate the utility of the horseweed transcriptome data by exploring a non-target glyphosate resistance hypothesis in this species.⁵ Specifically, the transcriptome data enabled expression analysis of ABC transporter genes based on real time RT-PCR experiments.

Expression analysis of ABC transporter-like genes

ABC transporters are transmembrane proteins that utilize the energy of ATP hydrolysis to transport a wide of variety substrates across extra- and intra-cellular membranes, including metabolic products, lipids and sterols, and drugs.^{35, 36} A number of ABC transporter genes were shown to be upregulated in our previous microarray analysis that suggested one or more might contribute to the glyphosate resistance in TN-R horseweed plants.⁵ One model for non-target resistance is glyphosate sequestration into vacuoles via active transport of glyphosate by ABC transporters;^{5,10,37,38} therefore overexpression of ABC transporters could account for the resistance. In fact, some gene families that might be involved in glyphosate resistance in horseweed³⁹ were found in the dataset, which included ABC transporters, glutathione S-transferases (GSTs), glycosyltransferases and P450s (Table 4). In the case of ABC transporters, we identified 67 unique sequences belonging to the subfamilies of multidrug resistance protein/multidrug resistance-associated protein (MRP) and pleiotropic drug resistance (PDR). Members of these subfamilies were shown to be up-regulated at a

high frequency by glyphosate in our previous heterologous microarray study.⁵ We therefore performed a preliminary gene-by-gene transcription analysis of 17 ABC-transporter genes (M1 to M11 from the MRP-like subfamily; P1 to P6 from the PDR-like subfamily). The most abundant transcript of these 17 ABC transporters, M11 (contig9470, 2120 bp of determined sequence), was found in glyphosate-treated TN-R horseweed plants and was 1.4 times higher than the expression of the actin housekeeping gene that we used as an internal control (Fig. 4). This AtMRP3-like ABC transporter was the highest up-regulated gene, with a fold change of 29.6, from our previous heterologous microarray study in horseweed.⁵ Also, the identity of M11 with the 70 bp *Arabidopsis* probe sequence was ~90%. All other ABC transporter genes had much lower absolute abundance: M1, M2, M3, M8, M9 and P4 were among those with moderate abundance levels, while the others can be classified as low abundance transcripts, but were still detectable by real-time RT-PCR (Fig. 4).

Compared with TN-S plants, M2 and P1 had lower expression levels in TN-R horseweed plants. M1 and M8 had the same expression level in both biotypes, while the remainder of ABC-transporters had higher expression levels in TN-R horseweed plants. The responses of these ABC transporter genes to 24-h glyphosate treatment varied as shown in Fig. 5. M1, M2, M8, M9, M10, M11, P4 and P5 were shown to be upregulated in both TN-S and TN-R biotypes. However, M1 and M2 had higher expression levels in TN-S plants. M9, M10 and M11 had higher expression levels in TN-R plants. The expression levels of M8, P4 and P5 were comparative between the two biotypes. M3, M6, M7 and P3 were shown to be upregulated in TN-S horseweed whereas there was almost no response in TN-R plants. M5 and P6 were shown to be upregulated in TN-S horseweed but downregulated in TN-R plants. P1 was downregulated in TN-S horseweed, whereas there was little response in TN-R plants. M4 and P2 had almost no response in both TN-S and TN-R biotype horseweed plants (Fig. 5). M6, M7, M10, M11 and P3 are more likely to be involved in the glyphosate resistance since their expression levels are always higher in resistant lines than in the susceptible lines, and we regard these as good preliminary target genes for further functional genomics studies.

The most interesting results were the strong responses of M10 and M11 transcription to glyphosate. M10 had a very low expression level, $\sim 6 \times 10^{-5}$ in TN-S and $\sim 1.2 \times 10^{-3}$ in TN-R plants, compared with the actin control gene. M10 was upregulated by nearly 300-fold in treated TN-S plants, and 16-fold in treated TN-R plants, compared with their untreated controls, but TN-R plants had the highest expression level (Fig 4). M11 was upregulated by 60-fold and 45-fold in treated TN-S plants and TN-R plants, respectively. Therefore, the

promoter of this gene could be used as a potential glyphosate sensor. Both M10 and M11 had a higher expression level in TN-R treated plants than TN-S treated plants. There are several features of M11 that are intriguing with regards to a potential non-target glyphosate resistance candidate. These features include its high levels of absolute transcription, up-regulation by glyphosate, which is also highest in resistant plants, and its putative tonoplast localization. Its orthologue in *Arabidopsis* is, tonoplast targeted.⁴⁰ Thus, M11 could play a very important role in glyphosate transport into vacuoles, thereby resulting in the glyphosate resistance in TN-R horseweed. One of our next steps is cloning full-length M11, as well as other candidates, and performing a functional analysis using overexpression analysis in susceptible- and/or knockdown analysis in resistant horseweed biotypes. Transcriptome sequencing has crucially enabled translational research in understanding herbicide resistance mechanisms.

We used young leaves and meristematic tissues from bulked samples of two horseweed biotypes, TN-S and TN-R with and without glyphosate treatment to carry out a large-scale transcriptome sequencing project using GS-FLX 454 sequencing, *de novo* assemblage, and functional annotation of the sequence data. We also used these data to design specific primers and measure the expression of potential candidate genes that might be involved in conferring non-target glyphosate-resistance in horseweed. Moreover, the data are sufficient to allow the design of microarray oligonucleotide probes and SNP discovery. The data also enable full-length cDNA cloning of non-target candidate genes via a RACE protocol for the next step in functional genomics research. Because of its very small genome size, horseweed could serve not only as a useful species for identifying non-target herbicide resistance mechanisms,^{10,39} but also be a good model for weed genomics.^{7,8} Sufficient data provided by this large-scale sequencing coupled with further application of multi genomic tools will improve our understanding of the genetic basis of weediness characteristics and the evolution mechanisms of herbicide resistance in weeds. In the long term, this research should be helpful in weed management and control.

Sequencing experiments yielded 411,962 raw reads, an average read length of 233 bp, and a total dataset of 95.8 Mb (NCBI Accession SRA010952). After trimming and quality control, we retained 379,152 high-quality sequences that were assembled into contigs. The assembly resulted in 31,783 unique transcripts, including 16,102 contigs and 15,681 singletons. The average coverage depth for each contig and each nucleotide position was 22-fold and 12-fold, respectively. A total of 16,306 unique sequences were annotated by

searching a custom plant protein database. The utility of the transcriptome data was demonstrated by further exploration of ABC transporters, which were previously hypothesized to play a role in non-target glyphosate resistance. Real-time RT-PCR primers were designed from the transcriptome data, which enabled assessing expression patterns of 17 ABC transporters from resistant and susceptible horseweed accessions from Tennessee with and without glyphosate treatment.

Potential contributions of transcriptomics research in weeds include better understanding of weediness characteristics, the identification of new molecular targets for improved weed control, and the gene discovery for transgenic crop improvement. However, the lack of good models and data hinder advances in genomics in weed biology. Although *Arabidopsis* is fully sequenced and commonly referred as “the weed”, it has few weediness characteristics and does not cause any significant economic loss in crops. Therefore, it is a poor model for weed genomics.^{7,8,41} Given the diversity in weedy species, no single species can encompass all weedy traits. Also, among weediness features, herbicide resistance is arguably the most critical trait affecting current and long-term control of weeds in agriculture. Several candidates have been suggested as potential weedy models,^{7,8} including herbicide-resistant horseweed and pigweeds, such as waterhemp. A 454 genomic DNA sequencing experiment on waterhemp produced 160,000 reads with an average read length of about 270 bp, yielding a total of about 43 Mb.⁴² Subsequently, a 454 transcriptome sequencing project on waterhemp from the same lab generated 483,225 raw reads with an average read length of 231 bp, yielding a total of 111.8 Mb,²⁶ which was compared to this study. Horseweed is one of the most attractive weeds for whole-genome sequencing because it has the smallest genome among 25 surveyed weeds most prevalent in the weed science literature.⁸ A GS-FLX genomic test run (half plate) with titanium reagents produced ~600,000 raw reads with an average read length of 403 bp, for a yield total of 248.7 Mb (Peng and Stewart, unpublished data), which included essentially the whole chloroplast genome (~150 Kb). Other weed genomics research has been reported recently. For example, 23,000 unique leafy spurge ESTs sequences⁴³ and about 9,000 unique cassava ESTs sequences⁴⁴ were obtained through traditional Sanger sequencing. More recently, the Roche GS-FLX 454 sequencing platform has been employed to sequence normalized cDNAs from 10 native and 10 invasive yellow star-thistle (*Centaurea solstitialis* L.) genotypes (K. Dlugosch, M. Barker, Z. Lai, and L. Rieseberg, unpublished data) in the Compositae Genome Project (http://compgenomics.ucdavis.edu/compositae_index.php). An average of 89,000 ~200 bp-

long reads and 32,000 unigenes were obtained per genotype. Compared with these examples, our large-scale sequencing of waterhemp and horseweed using the 454 platform yielded abundant data in a short period of time. This suggests that the development of weedy genomic research is largely being enabled by powerful next-generation sequencing platforms.

EXAMPLE 2

The promoter region of M11 was cloned into the pCR8/GW/TOPO vector, then subcloned into the *pMDC164* plant transformation vector upstream of the GUS reporter gene. The recombinant binary vectors were introduced into *Agrobacterium tumefaciens* GV3101 strain by freeze/thaw method. The constructs were transformed into young leaves of five week old tobacco via infiltration method. After being infiltrated for two days, the leaves were treated with different amounts Roundup WeatherMAX (0.108, 0.0108, 0.0054, 0.00108 kg/ha ae) or water as control. The transient expression of GUS reporter gene was observed after additional two (Figure 10A) or five days (Figure 10B).

A glyphosate concentration assay was also performed. Roundup WeatherMAX (540g/L) was diluted with water for 100, 1000, 10000, 100000 folds, respectively. Then, 2ml of each was smeared on the leaves of tobacco planted in 10cm*10cm plate with a brush. Same amount of water smeared was as control. Tobacco plants treated with Roundup for one week and the results are illustrated in Figure 11. The ability of the promoter to drive expression of GUS was also examined in transgenic tobacco plants. The promoter (GUS as reporter gene) was induced by RoundUp treatment in stable transgenic tobacco plants (multiple plants from the same independent transgene line). GUS expression is evident in Figure 12B (dark coloration of the leaves). Glyphosate induced expression of GFP was also observed in tobacco plants transformed by *A. tumefaciens* infiltration (Figs. 13A and B). The light colorations in panel B demonstrate the induction of GFP expression in glyphosate treated plants. GFP expression was also quantified using fluorescence spectroscopy (see Fig. 13).

REFERENCES

- 1 Dill GM, Cajacob CA and Padgett SR, Glyphosate-resistant crops: adoption, use and future consideration. *Pest Manag Sci* **64**:326-331(2008).
- 2 Duke SO and Powles SB, Glyphosate: a once-in-a-century herbicide. *Pest Manag Sci* **64**:319-325 (2008).
- 3 Heap I, The International Survey of Herbicide Resistant Weeds. Online Available www.weedscience.com (2010).
- 4 Van Gessel MJ, Glyphosate-resistant horseweed from Delaware. *Weed Sci* **49**:703-705 (2001).
- 5 Yuan JS, Good LG, Cao Y, Halfhill MD, Zhou X, Peng Y, Hu J, Rao MR, Heck GR, Larosa TJ *et al*: Functional genomics analysis of glyphosate resistance in *Conyza canadensis* (horseweed). *Weed Sci* in press (2010).
- 6 Weaver SE, The biology of Canadian weeds. 115. *Conyza canadensis*. *Can. J. Plant Sci* **81**:867-875 (2001).
- 7 Basu C, Halfhill MD, Mueller TC, and Stewart CN, Jr, Weed genomics: new tools to understand weed biology. *Trends Plant Sci* **9**:391-398 (2004).
- 8 Stewart CN, Tranel PJ, Horvath DP, Anderson JV, Rieseberg LH, Westwood JH, Mallory-Smith CA, Zapiola ML and Dlugosch KM, Evolution of weediness and invasiveness: charting the course for weed genomics. *Weed Sci* **57**: 451-462 (2009).
- 9 Halfhill MD, Good LL, Basu C, Burris J, Main CL, Mueller TC and Stewart CN, Transformation and segregation of GFP fluorescence and glyphosate resistance in horseweed (*Conyza canadensis*) hybrids. *Plant Cell Rep* **26**: 303-311 (2007).
- 10 Feng PCC, Tran M, Chiu T, Sammons RD, Heck GR and Cajacob CA, Investigations into glyphosate resistant horseweed (*Conyza canadensis*): retention, uptake, translocation, and metabolism. *Weed Sci.* **52**: 498-505 (2004).
- 11 Zelaya IA, Owen MDK, and VanGessel MJ, Inheritance of evolved glyphosate resistance in *Conyza canadensis* (L.) Cronq. *Theor. Appl. Genet* **110**:58-57 (2004).
- 12 Margulies M, Egholm E, Altman WE *et al.*, Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380 (2005).
- 13 Morozova O and Marra MA, Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**:255-264 (2008).
- 14 Rothberg JM and Leamon JH,. The development and impact of 454 sequencing. *Nat Biotechnol* **26**:1117-1124 (2008).

15 Emrich SJ, Barbazuk WB, Li L and Schnable PS, Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res* **17**:69-73 (2007).

16 Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders ACE, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, and Michael Snyder, Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**: 420-426 (2007).

17 Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M and Dougan G, High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat Genet* **40**: 987-993 (2008).

18 Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I and Marden JH, Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* **17**:1636-47 (2008).

19 Mao C, Evans C, Jensen RV and Sobral BW, Identification of new genes in *Sinorhizobium meliloti* using the Genome Sequencer FLX system. *BMC Microbiol* **8**:72 (2008).

20 Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L and Perrotta G, Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* **10**:399 (2009).

21 Droege M and Hill B, The genome sequencer FLXTM system-longer reads, more applications, straight forward bioinformatics and more complete datasets. *J Biotech* **136**: 3-10 (2008).

22 Nyren P, Pettersson B and Uhlen M, Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal. Biochem* **208**: 171-175 (1993).

23 Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, and Nyren P, Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem* **242**:84-89 (1996).

24 Mueller TC, Massey JH, Hayes RM, Main CL and Stewart CN Jr, Shikimate accumulates in both glyphosate-sensitive and glyphosate-resistant horseweed (*Conyza canadensis* L. Cronq.). *J Agr Food Chem* **51**: 680-684 (2003).

25 Dassanayake M, Haas JS, Bohnert HJ and Cheeseman JM, Shedding light on an extremophile lifestyle through transcriptomics. *New Phytol.* **183**:764-775 (2009).

26 Riggins CW, Peng Y, Stewart CN Jr and Tranel PJ, Characterization of waterhemp transcriptome using 454 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Management Science (accepted)*

27 Chou HH and Holmes MH, DNA sequencing quality trimming and vector removal. *Bioinformatics* **17**: 1093-1104 (2001)

28 Huang X and Madan A, CAP3: A DNA sequence assembly program. *Genome Res* **9**:868-877 (1999).

29 Yuan JS, Wang D and Stewart CN, Statistical methods for efficiency adjusted real-time PCR analysis. *Biotechnology Journal*. **3**:112-123 (2008).

30 Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L and Perrotta G, Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*. **10**:399 (2009).

31 Wang W, Wang Y, Zhang Q, Qi Y and Guo D, Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics*. **10**:465 (2009).

32 Weber AP, Weber KL, Carr K, Wilkerson C and Ohlrogge JB, Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiol* **144**:32–42 (2007).

33 Black DL, Mechanisms of alternative pre-messenger RNA splicing. *Ann Rev of Biochem* **72**: 291–336 (2003).

34 Yuan Y, Chung JD, Fu X, Johnson VE, Ranjan P, Booth SL, Harding SA and Tsaia CJ, Alternative splicing and gene duplication differentially shaped the regulation of isochorismate synthase in *Populus* and *Arabidopsis*. *Proc Natl Acad Sci U S A*. **106**:22020-22025 (2009)

35 Rea PA, Plant ATP-binding cassette transporters. *Ann Rev Plant Biol* **58**:347-375 (2007).

36 Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, Geisler M, Klein M, Kolukisaoglu U, Lee Y, Martinoia E, Murphy A, Rea PA, Samuels L, Schulz B, Spalding EJ, Yazaki K and Theodoulou FL, Plant ABC proteins – a unified nomenclature and updated inventory. *Trends Plant Sci* **13**: 151-159 (2008).

37 Shaner DL, The role of translocation as a mechanism of resistance to glyphosate. *Weed Sci* **57**:118–123 (2009).

38 Ge X, d'Avignon DA, Ackerman JJH and Sammons RD, Rapid vacuolar sequestration: the horseweed glyphosate resistance mechanism. *Pest Management Sci.* Online first. DOI: 10.1002/ps.1911 (2010).

39 Yuan JS, Tranel PJ, and Stewart CN Jr, Non-target site herbicide resistance: a family business. *Trends Plant Sci* **12**:6-13 (2007).

40 Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P and Lilley KS, Mapping the Arabidopsis organelle proteome. *Proc Natl Acad Sci U S A.* **103**:6518-6523 (2006).

41 Gressel J, Arabidopsis is not a weed, and mostly not a good model for weed genomics; there is no good model for weed genomics, in *Weedy and Invasive Plant Genomics*, ed. by Stewart CN, Jr., Wiley-Blackwell, Ames, Iowa, pp.25-32 (2009).

42 Lee, RM, Thimmapuram, J, Thinglum, KA, Gong, G, Hernandez, AG, Wright, CL, Kim, RW, Mikel, M and Tranel, PJ, Sampling the waterhemp (*Amaranthus tuberculatus*) genome using pyrosequencing technology. *Weed Sci.* **57**: 463-469 (2009).

43 Anderson JV, Horvath DP, and Chao WS, Foley ME., Hernandez AG, Thimmapuram J, Liu L, Gong GL, Band M, Kim R, and. Mikel MA, Characterization of an EST database for the perennial weed leafy spurge: an important resource for weed biology research. *Weed Sci* **55**:193–203 (2007).

44 Lokko Y, Anderson JV, Rudd S, Raji AAJ, Horvath D, Mikel MA, Kim R, Liu L, Hernandez A, Dixon AGO and Ingelbrecht IL, Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Rep* **26**:1605-1618 (2007).

Table 1. Summary of 454 sequencing, data trimming, assemblage, and annotation.

	Sequences (number; % of total)	Nucleotides (number; % of total)
Raw reads	411,962 (100%)	95,822,385 (100%)
After quality control and trimming	379,152 (92.04%)	88,013,743 (91.85%)
Sequences assembled to contigs	363,471 (88.23%)	84,692,110 (88.38%)
Assembled contigs	16,102	7,054,041 (7.36%)
Singletons	15,681 (3.81%)	3,321,633 (3.47%)
Annotated unique sequences	16,306	6,478,624

Table 2. Summary Blast data for of assembled FLX-454 contigs against themselves and all singletons. All Blast results refer to hits with bitscores greater than or equal to 45, E-value <0.0001 and alignments with greater than or equal to 95%.

Total assembled horseweed FLX-454 contigs	16,102
Number of contigs that had best Blast hits with other contigs and singletons	4,405 (27.1%)
Average length of alignments as contigs vs. contigs	92 bp
Average length of alignments as contigs vs. singletons	73 bp
Average coverage of alignments to queried contig as contigs vs. contigs	18.5%
Average coverage of alignments to queried contig as contigs vs. singletons	11.5%
Number of Blastn-paired contigs had Blastx hits against protein database	2,768 (17.2%)
Number of Blastn-paired contigs had identical best Blastx hits	285 (1.8%)

Table 3. Summary Blast data for assembled Sanger sequences against FLX-454 contigs and singletons. All Blast results refer to hits with bitscores greater than or equal to 45. Alignment lengths refer to nucleotides.

Number of Sanger sequences	2,016
Number of Sanger sequences with at least one Blast hit against 454 sequences	1,540
Percent Sanger sequences with a Blast hit against 454 sequence	76.4%
Mean percent identity of Sanger vs. all 454 contig Blast hit alignments	95.8%
Mean percent identity of Sanger vs. all 454 Blast hit alignment	97.3%
Mean percent identity of Sanger vs. 454 Blast hit alignment (bitscore>100)	99.3%
Mean number of gaps within Sanger vs. all 454 contig Blast hit alignments	0.04
Median number of gaps within Sanger vs. all 454 contig Blast hit alignments	0
Mean number of gaps within Sanger vs. all 454 singleton Blast hit alignments	0.22
Median number of gaps within Sanger vs. 454 singleton Blast hit alignments	0

Table 4. Comparison of the number of hits (contigs + singletons) to herbicide target-site genes and gene families and non-target gene families from transcriptome 454 sequencing of horseweed and waterhemp.

Herbicide target gene family	Horseweed	Waterhemp
Acetolactate synthase	6	2
D1 protein (plastidic gene)	4	2
Tubulin	29	33
Protoporphyrinogen oxidase	2	8
Phytoene desaturase	5	1
Glutamine synthetase	9	7
1-deoxy-D-xylulose-5-phosphate synthase	6	1
4-hydroxyphenylpyruvate dioxygenase	1	2
Acetyl-CoA carboxylase	6	8
Dihydropterolate synthase	1	2
5-enolpyruvylshikimate-3-phosphate synthase	2	3
Non-target gene family		
Glutathione S-transferase	7	22
Cytochrome P450 monooxygenases	125	191
Glycosyltransferases	76	84
ABC transporter genes	151	192

Table 5. Summary data for horseweed-unique annotated and non-annotated sequences.

Average length of annotated contigs	526 bp
Average coverage of annotated contigs	30.3-fold
Average length of non-annotated contigs	297 bp
Average coverage of non-annotated contigs	9.6-fold
Average length of annotated singletons	230 bp
Average length of non-annotated singletons	199 bp
Number of non-annotated unique sequences have hits to microRNAs	431
Number of non-annotated unique sequences have hits to non-plant proteins	135

Table 6. Specific primers list for real-time RT-PCR analysis of 17 ABC-transporter genes.			
Genes	Sequences	Primers	PCR product
actin	Contig9305	Forward: 5'-GTGGTTCAACTATGTTCCCTG-3' (SEQ ID NO:3) Reverse: 5'-CTTAGAAGCATTTCCTGTGG-3' (SEQ ID NO:4)	228 bp
M1	Contig4024	Forward: 5'-GGCTTATCTTTCAGCATTTC-3' (SEQ ID NO:5) Reverse: 5'-TTGTATCAGGTCATCCGTTC-3' (SEQ ID NO:6)	476 bp
M2	Contig375 Contig15669	Forward: 5'-TGTGATTCCCAGTGCTCCTT-3' (SEQ ID NO:7) Reverse: 5'-ATCGACCCATTAGTCAACCC-3' (SEQ ID NO:8)	167 bp
M3	Contig9365	Forward: 5'-TTGACCCCATTTGGTGAACAT-3' (SEQ ID NO:9) Reverse: 5'-TCAGCATCACTCGTCCTAAA-3' (SEQ ID NO:10)	174 bp
M4	Contig14889	Forward: 5'-GTGACAGTTCCTTGCCATTT-3' (SEQ ID NO:11) Reverse: 5'-GCACCACTATTCCCATCTCC-3' (SEQ ID NO:12)	204 bp
M5	Contig9009	Forward: 5'-TGATGTGGAGTCTGAGCGTG-3' (SEQ ID NO:13) Reverse: 5'-CGATCTTTATTATTTACGGT-3' (SEQ ID NO:14)	241 bp
M6	Contig15034	Forward: 5'-TAGTCGGTAAAAGTGGGTGT-3' (SEQ ID NO:15) Reverse: 5'-ATATTGTCACGAATAGTTCC-3' (SEQ ID NO:16)	190 bp
M7	Contig6777	Forward: 5'-TCATGCGTGTTGGATACCGT-3' (SEQ ID NO:17) Reverse: 5'-ACCAAGAAGTGAGGCGATGC-3' (SEQ ID NO:18)	245 bp
M8	Contig11627	Forward: 5'-ATCAGTGAGTGACGAAGGAG-3' (SEQ ID NO:19) Reverse: 5'-TTGTAGAATGGCATCAGTGG-3' (SEQ ID NO:20)	145 bp
M9	Contig531	Forward: 5'-GAGATAGGGGAAAGAGGTGT-3' (SEQ ID NO:21) Reverse: 5'-CATTAGTTGATTGTAAGGTGC-3' (SEQ ID NO:22)	285 bp

Table 6. Specific primers list for real-time RT-PCR analysis of 17 ABC-transporter genes.			
Genes	Sequences	Primers	PCR product
M10	Contig6841	Forward: 5'-TTGGCTCAACTTCGTGGTATCGGG-3' (SEQ ID NO:23) Reverse: 5'-CCAAGAAATTCCAAGCGGAACCCT-3' (SEQ ID NO:24)	253 bp
M11	Contig9470	Forward: 5'-ATGCTGTCTTCTTTTACCTTTGC-3' (SEQ ID NO:25) Reverse: 5'-CGACTTCCCACTACCAGTTCTTC-3' (SEQ ID NO:26)	393 bp
P1	Contig13034	Forward: 5'-AGGCATTGAAGTGTCATCCG-3' (SEQ ID NO:27) Reverse: 5'-CAAACATCTTTCTGGCGTCT-3' (SEQ ID NO:28)	531 bp
P2	Contig2514	Forward: 5'-ACGCCTTTGGACAAGTTATG-3' (SEQ ID NO:29) Reverse: 5'-AGCAGCGATGTTGTGGTTAG-3' (SEQ ID NO:30)	206 bp
P3	Contig75	Forward: 5'-GGGTGGGAGGGTCACATACA-3' (SEQ ID NO:31) Reverse: 5'-AAGACAGCGGCATACATAGC-3' (SEQ ID NO:32)	483 bp
P4	Contig14213	Forward: 5'-CAAAATTCTGCTGGTATCTGTA-3' (SEQ ID NO:33) Reverse: 5'-AAAGTATCGGTCCAAGTAGCCT-3' (SEQ ID NO:34)	311 bp
P5	Contig9365	Forward: 5'-GCTATCCAAAACGCATCATCAG-3' (SEQ ID NO:35) Reverse: 5'-AGCGACAACGCCAAGGAAATCA-3' (SEQ ID NO:36)	543 bp
P6	Contig2	Forward: 5'-CGACGACCGACTGCTGTTGCTT-3' (SEQ ID NO:37) Reverse: 5'-GACCGTTGACACCAGGTAATCC-3' (SEQ ID NO:38)	392 bp

Table 7. Summary of numbers of reads and nucleotides by 454 sequencing runs.

	Run A	Run B	Titration run	Total count
N raw reads	253,537	145,994	12,431	411,962
Mean length	235 bp	230 bp	228 bp	233 bp
N nucleotides	59,497,520	33,497,855	2,827,010	95,822,385

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising:
 - a) a nucleotide sequence comprising the sequence set forth in SEQ ID NO:1 or a complement thereof;
 - b) a nucleotide sequence comprising a fragment of the sequence set forth in SEQ ID NO:1, wherein said sequence initiates transcription in a plant cell; and
 - c) a nucleotide sequence comprising a sequence having at least 70% sequence identity to the sequence set forth in SEQ ID NO:1 or a fragment thereof, wherein said sequence initiates transcription in the plant cell.
2. A DNA construct comprising a nucleotide sequence according to claim 1 operably linked to a heterologous nucleotide sequence of interest.
3. A vector comprising the DNA construct of claim 2.
4. A plant cell having stably incorporated into its genome the DNA construct of claim 2.
5. The plant cell of claim 4, wherein said plant cell is from a monocot.
6. The plant cell of claim 4, wherein said plant cell is from a dicot.
7. A plant having stably incorporated into its genome the DNA construct of claim 2.
8. The plant of claim 7, wherein said plant is a monocot.
9. The plant of claim 7, wherein said plant is a dicot.
10. A transgenic seed of the plant of claim 7, wherein the seed comprises the DNA construct.

11. The plant of claim 8, wherein the heterologous nucleotide sequence of interest encodes a gene product that is a reporter or selectable marker, confers improved nutritional content or resistance to herbicides, salts, heat, cold, flood, drought, pathogens, or insects.

12. A method for expressing a nucleotide sequence in a plant, said method comprising introducing into a plant a DNA construct, said DNA construct comprising a promoter and operably linked to said promoter a heterologous nucleotide sequence of interest, wherein said promoter comprises a nucleotide sequence selected from the group consisting of:

a) a nucleotide sequence comprising the sequence set forth in SEQ ID NO:1 or a complement thereof;

b) a nucleotide sequence comprising a fragment of the sequence set forth in SEQ ID NO:1, wherein said sequence initiates transcription in a plant cell; and

c) a nucleotide sequence comprising a sequence having at least 70% sequence identity to the sequence set forth in SEQ ID NO:1 or a fragment thereof, wherein said sequence initiates transcription in the plant cell.

13. The method of claim 12, wherein said plant is a dicot.

14. The method of claim 12, wherein said plant is a monocot.

15. The method of claim 14, wherein the heterologous nucleotide sequence encodes a gene product that confers improved nutritional content or resistance to herbicides, salts, heat, cold, flood, drought, pathogens, or insects or is a reporter or selectable marker.

16. The method according to any one of claims 12-15, further comprising the application of glyphosate to said plant.

17. A method for introducing a nucleotide sequence into a plant cell comprising introducing into a plant cell a DNA construct comprising a promoter operably linked to a heterologous nucleotide sequence of interest, wherein said promoter comprises a nucleotide sequence selected from the group consisting of:

a) a nucleotide sequence comprising the sequence set forth in SEQ ID NO:1 or a complement thereof;

b) a nucleotide sequence comprising a fragment of the sequence set forth in SEQ ID NO:1, wherein said sequence initiates transcription in a plant cell; and

c) a nucleotide sequence comprising a sequence having at least 70% sequence identity to the sequence set forth in SEQ ID NO:1 or a fragment thereof, wherein said sequence initiates transcription in the plant cell.

18. The method of claim 17, wherein said plant cell is from a monocot.

19. The method of claim 17, wherein said plant cell is from a dicot.

20. A method for selectively expressing a nucleotide sequence in a plant cell comprising introducing into a plant cell a DNA construct, and regenerating a transformed plant from said plant cell and exposing said plant to glyphosate, said DNA construct comprising a promoter and a heterologous nucleotide sequence operably linked to said promoter, wherein said promoter comprises a nucleotide sequence selected from the group consisting of:

a) a nucleotide sequence comprising the sequence set forth in SEQ ID NO:1 or a complement thereof;

b) a nucleotide sequence comprising a fragment of the sequence set forth in SEQ ID NO:1, wherein said sequence initiates transcription in a plant cell; and

c) a nucleotide sequence comprising a sequence having at least 70% sequence identity to the sequence set forth in SEQ ID NO:1 or a fragment thereof, wherein said sequence initiates transcription in the plant cell.

21. A method of identifying a putative glyphosate inducible promoter comprising aligning nucleic acid sequences with SEQ ID NO: 1 and selecting those sequences having at least 50% sequence identity to SEQ ID NO: 1.

22. The method according to claim 21, further comprising transforming a plant with a DNA construct, said DNA construct comprising said putative promoter operably linked to a heterologous nucleotide sequence that confers a) improved nutritional content; b)

resistance to herbicides, salts, heat, cold, flood, drought, pathogens, or insects; or c) is a reporter or selectable marker.

23. The method according to claim 22, further comprising the testing said putative promoter for glyphosate inducibility, said testing comprising comparing the expression of said heterologous sequence between one or more glyphosate treated plant and one or more plant not treated with glyphosate (control plants) and selecting those plants expressing said heterologous nucleotide sequence in amounts or levels that exceed the amounts or levels of said heterologous sequence expressed in plants to which glyphosate was not applied (control plants).

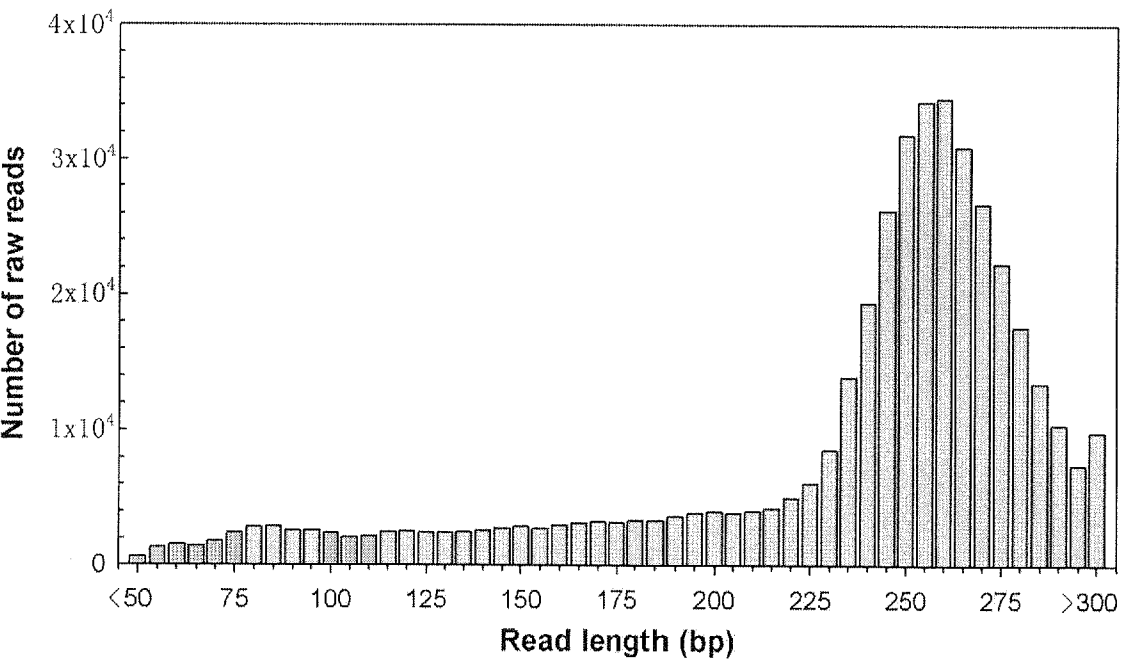


FIG. 1

FIG. 2A

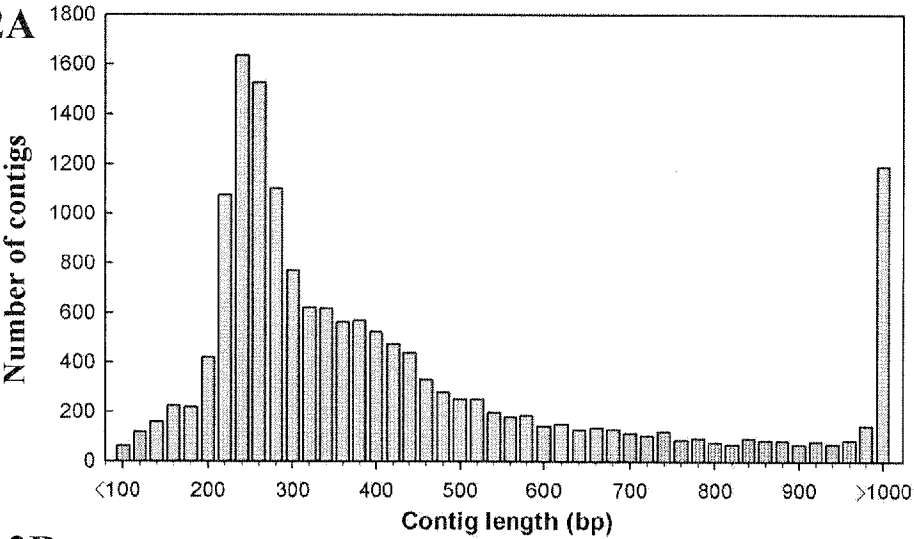
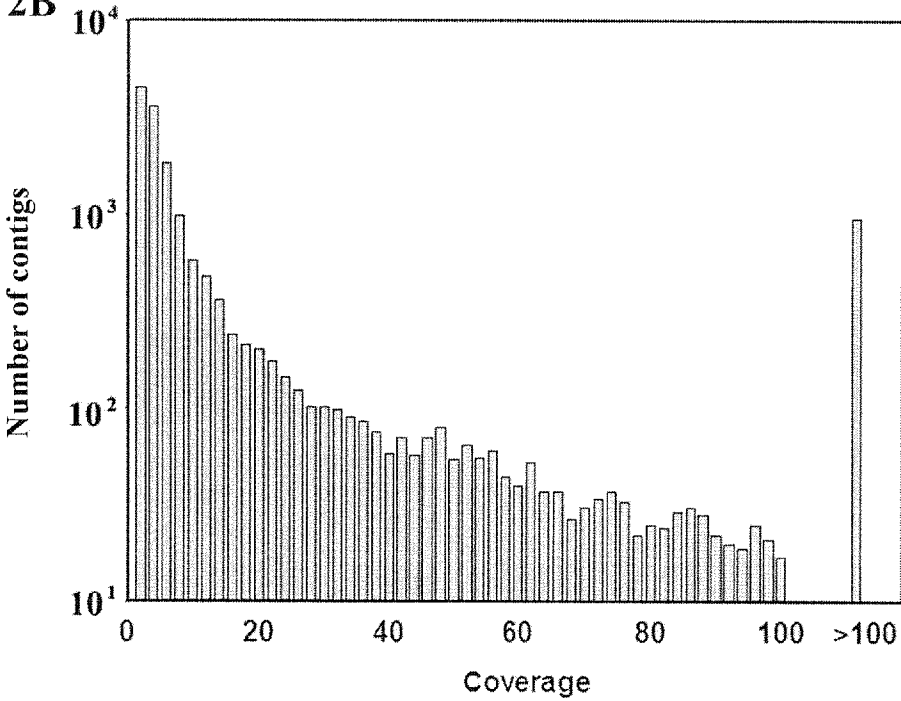


FIG. 2B



3/17

FIG. 3A

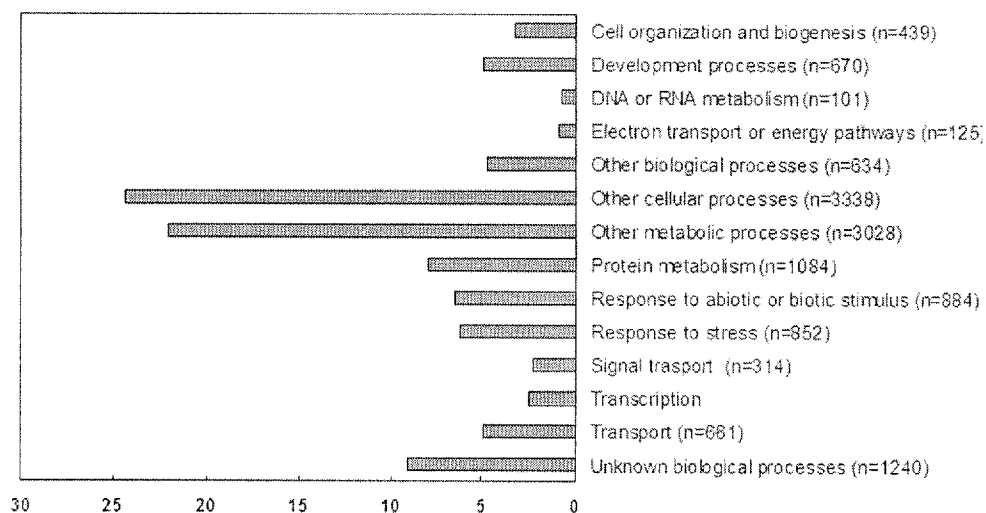


FIG. 3B

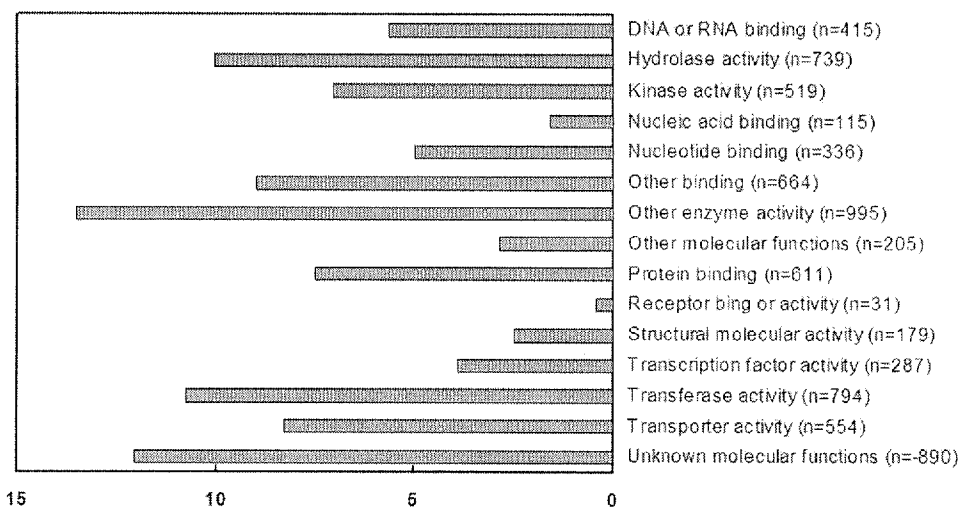
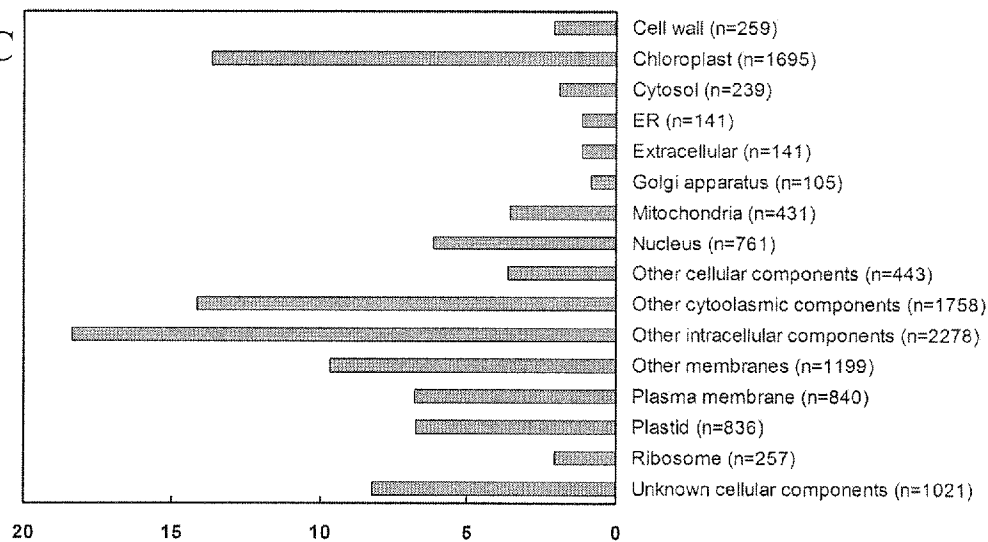


FIG. 3C



4/17

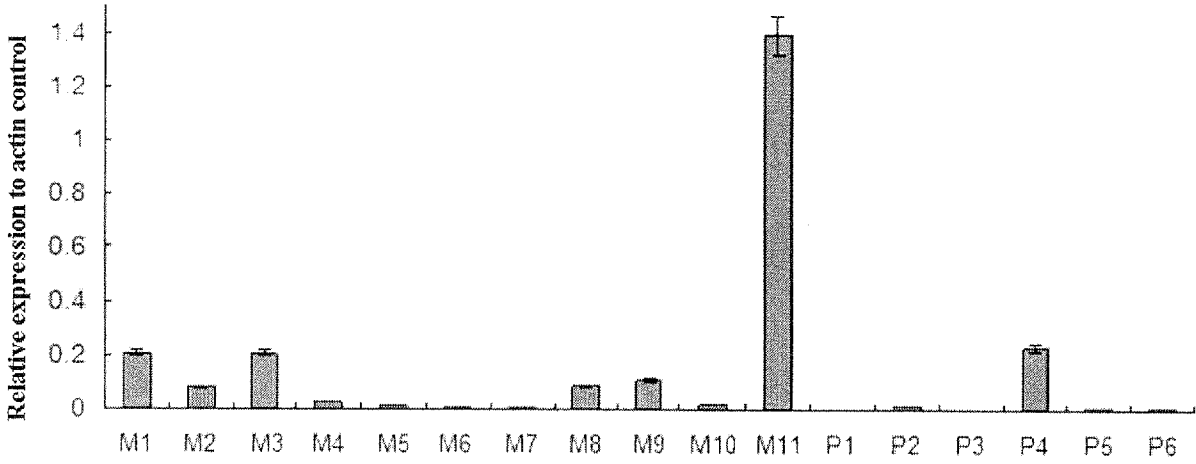


FIG. 4

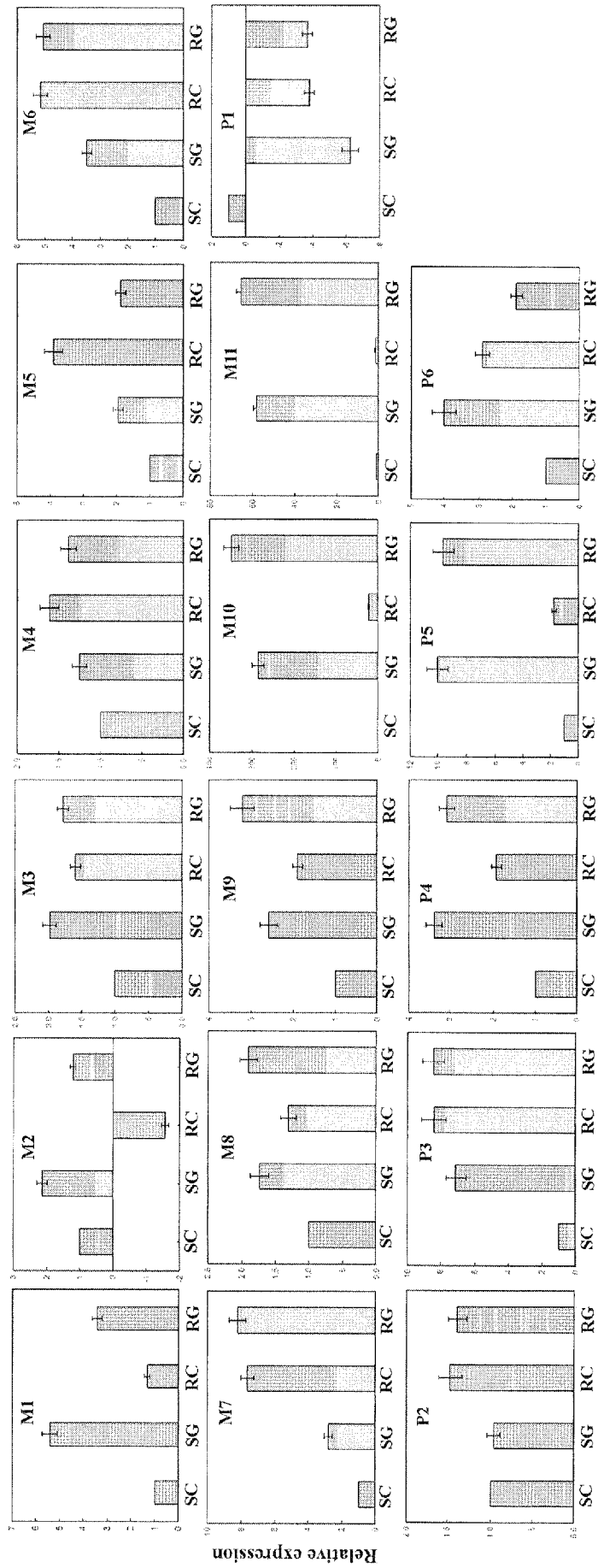


FIG. 5

6/17

BLASTX 2.2.18 Query= Contig19 (1511 letters)

	Score	E
	(bits)	Value
Sequences producing significant alignments:		
gb AAL68972.1 calmodulin-like-domain protein kinase CPK2 [Cucur...	757	0.0
gb AAL68971.1 phloem calmodulin-like-domain protein kinase PCPK...	755	0.0
dbj BAF57913.1 calcium-dependent protein kinases [Solanum tuber...	748	0.0
gb AAC49405.1 calcium dependent protein kinase	748	0.0
dbj BAF57914.1 calcium-dependent protein kinases [Solanum tuber...	747	0.0

>gb AAL68972.1 calmodulin-like-domain protein kinase CPK2 [Cucurbita maxima]

Length = 558

Score = 757 bits (1955), Expect = 0.0

Identities = 369/397 (92%), Positives = 387/397 (97%)

Frame = +2

Query: 2 AYEDLLYVHIVMELCNGGELFDRIIERGQYSEKAAELTKIIVGVVEACHSLGVMHRDLK 181

AYED LYVHIVMELC-GGELFDRII-RG YSEKAAELT+IIVGVVE CHSLGVMHRDLK

Sbjct: 162 AYEDSLYVHIVMELCSGGELFDRIIQRGHYSEKAAELTRIIVGVVETCHSLGVMHRDLK 221

Query: 182 PENFLLINKDDDFSLKAIDFGLSVFFKPGQVFTDVVGSPYYVAPEVLLKHYGPEADVWTA 361

PENFLL+NKDDDFSLKAIDFGLSVFFKPGQ+FTDVVGSPYYVAPEVLLKHYGP ADVWTA

Sbjct: 222 PENFLLVNKDDDFSLKAIDFGLSVFFKPGQIFTDVVVGSPYYVAPEVLLKHYGPAADVWTA 281

Query: 362 GVILYILLSGVPPFWAETQQGIFDAVLKGYIDFESDPWPLISDSAKDLIRKMLCSRPSDR 541

GVILYILLSGVPPFWAETQQGIFDAVLKG+IDF+SDPWPLISDSAKDLIRKMLCSRPSDR

Sbjct: 282 GVILYILLSGVPPFWAETQQGIFDAVLKGHIDFSDPWPLISDSAKDLIRKMLCSRPSDR 341

Query: 542 LTAHEVLCHPWICENG VAPDRALDPAVL SRLKQFSAMNKLKKMALRVIAESLSEEEIAGL 721

LTAHEVLCHPWICENG VAPDRALDPAVL SRLKQFSAMNKLKKMALRVIAESLSEEEIAGL

Sbjct: 342 LTAHEVLCHPWICENG VAPDRALDPAVL SRLKQFSAMNKLKKMALRVIAESLSEEEIAGL 401

Query: 902 IHLNKLEREEHLVAAFQYFDKDGSGYITVDELQQACADHNMTDFLVEDIIEVDQDNDGR 1081

IHLNKLEREEHLVAAF+YFDKDGSGYITVDELQQACA+HNMTD +EDII+EVDQDNDGR

Sbjct: 462 IHLNKLEREEHLVAAFQYFDKDGSGYITVDELQQACAEHNMTDAYLEDIIEVDQDNDGR 521

Query: 1082 IDYGEFVAMMTKGNAGVGRRTMRNSLNISMRDAPGAL 1192

IDY EFVAMM KGNAG+GRRTMRNSLN+SMRD PGAL

Sbjct: 522 IDYSEFVAMMQGNAGIGRRTMRNSLNLSMRDGPAL 558

FIG. 6

7/17

A		B		C		D		E	
Query ID	Query ID	Hits Accession	Annotation			E-value	Score bits		
1	Contig2	BAD07483	PDR-type ABC transporter 1 [Nicotiana tabacum]			2.70E-97	354.755		
2		BAD07484	PDR-type ABC transporter 2 [Nicotiana tabacum]			6.02E-97	353.599		
3		Q949G3	Pleiotropic drug resistance protein 1 (MpPDR1)			4.31E-86	347.436		
4		CAN63107	hypothetical protein [Vitis vinifera]			1.64E-94	345.51		
5		BAD05827	putative PDR-like ABC transporter [Oryza sativa japonica cultivar-group]			3.65E-94	344.354		
6	Contig3	CAA64173	soluble-starch-synthase [Solanum tuberosum]			5.00E-88	324.324		
7		CAH60282	unnamed protein product [Solanum tuberosum]			5.00E-88	324.324		
8		CAL49843	unnamed protein product [Solanum tuberosum]			5.00E-88	324.324		
9		CAL64687	unnamed protein product [Solanum tuberosum]			5.00E-88	324.324		
10		CAA65065	glycogen (starch) synthase [Solanum tuberosum]			1.90E-87	322.398		
11		AAS07184	putative peroxisomal membrane protein [Oryza sativa japonica cultivar-group]			6.43E-68	256.529		
12	Contig30	NP_001050563	Os03g053800 [Oryza sativa japonica cultivar-group]			6.43E-68	256.529		
13		AAK60317	At1g2870.F14G24_14 [Arabidopsis thaliana]			1.58E-66	251.906		
14		AAG52277	unknown protein [Arabidopsis thaliana]			1.58E-66	251.906		
15		AAP40491	unknown protein [Arabidopsis thaliana]			2.70E-66	251.136		
16		CAO65226	unnamed protein product [Vitis vinifera]			3.31E-09	61.2326		
17	Contig33	CAA93442	methionine sulfoxide reductase [Fragaria x ananassa]			2.14E-08	58.5362		
18		P54152	Peptide methionine sulfoxide reductase (Protein-methionine-S-oxide reductase)			2.14E-08	58.5362		
19		P54153	Peptide methionine sulfoxide reductase (Protein-methionine-S-oxide reductase)			2.37E-07	55.0694		
20		AAF19789	methionine sulfoxide reductase [Lactuca sativa]			2.01E-06	51.9878		
21		CAN166709	hypothetical protein [Vitis vinifera]			4.72E-29	126.716		
22	Contig50	CAN73906	hypothetical protein [Vitis vinifera]			1.16E-27	122.094		
23		CAN66040	hypothetical protein [Vitis vinifera]			2.19E-26	117.857		
24		AAF02838	Similar to serine/threonine kinases [Arabidopsis thaliana]			9.20E-25	112.464		
25		CAO43421	unnamed protein product [Vitis vinifera]			1.20E-24	112.079		
26		CAN76686	hypothetical protein [Vitis vinifera]			1.55E-32	139.813		
27	Contig51	AAL49790	unknown protein [Arabidopsis thaliana]			3.57E-29	128.642		
28		BAB85647	inflorescence and root apices receptor-like kinase [Arabidopsis thaliana]			3.57E-29	128.642		
29		NP_191196	leucine-rich repeat transmembrane protein kinase			3.57E-29	128.642		
30		AAQ42766	At5g01890.T20L15_160 [Arabidopsis thaliana]			1.36E-28	126.716		
31		AAK92723	unknown protein [Arabidopsis thaliana]			9.87E-19	92.4337		
32	Contig52	BAB02262	unnamed protein product [Arabidopsis thaliana]			9.87E-19	92.4337		
33		BAC84850	unknown protein [Oryza sativa japonica cultivar-group]			1.13E-14	78.9518		
34		EAZ04564	hypothetical protein Osl_025896 [Oryza sativa indica cultivar-group]			1.13E-14	78.9518		
35									

FIG. 7

8/17

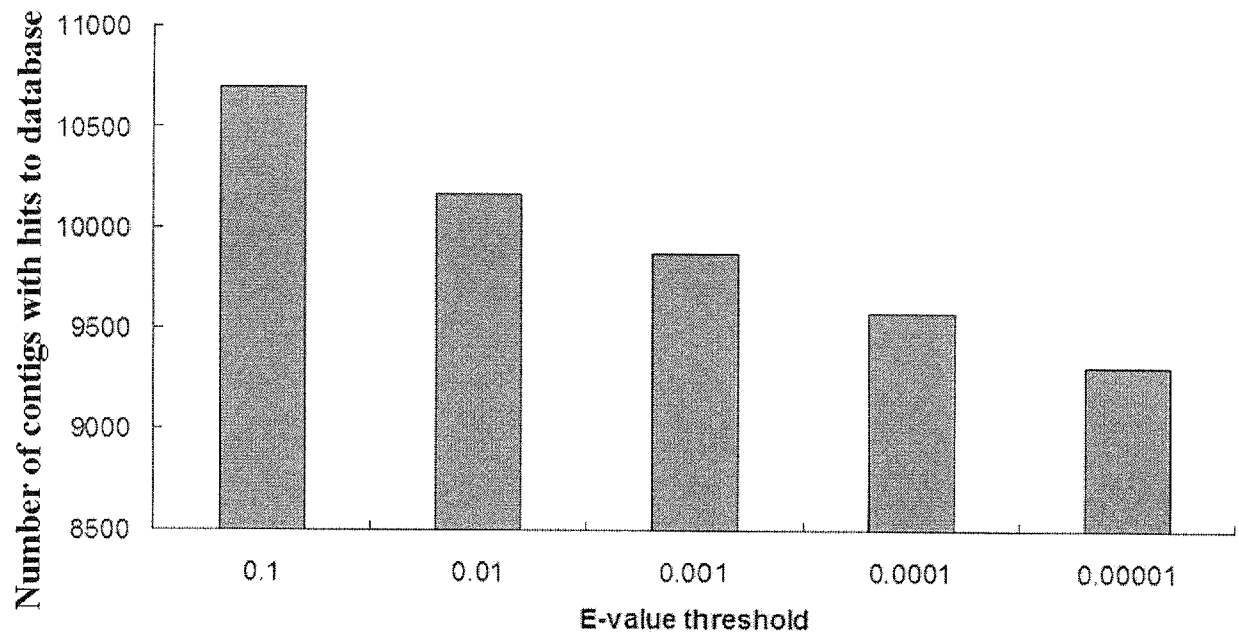


FIG. 8

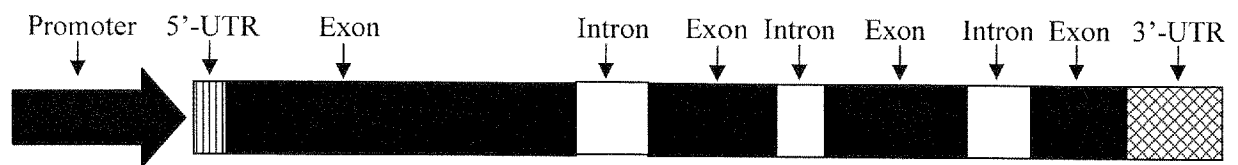


FIG. 9A

9/17

FIG. 9B

TTATTTTTCTTTTTGTTGATGGATTGTATTTTTTTTAAATTTTAAATGTGTGGATTATTTGCTCGCAACAGAGAAT
CTATCTTACGTCATTTTAAATCGGGTCCGCGTTATAGAGCCCCCTCACCTCAATACCCAGTAGAAGGAGAACTC
CCAACTAAATCGCCCGAAAACACTATATTTAATTGGGATAAACTTCTGCCCCCTCTGCAGGACTCAAACCTACTT
CACTGGAAGATCTTATTGTGAAATTTTTTAAATGTTCTTTTTATGTTTCGAACTCGAGACTTCCCGCTTATAAA
GCTGGTGTTCATAGTAAGTATAATAATTTATTTTTTATGTATATGTACGGCTCATTGCGAAATTTGGGAGTTGCA
ACAACCTTGCATACAGTTTTCTAACTTAGCACATCAAAGTTTTGTAAACATGTGAACGGTGAACCTATTTTTTTT
TTTATAGGGCTGTAACTTAAATTTTCCCAAAATTCCTAGAAAAATTTTTGTTTTATTTTATTTAACATATAG
AATCTATCAGAATTACAAAAATTATATATTTAAGTCCCCAAATCGATATCAAACACTACGAAATTGTCTCAAAC
GTTGTTGTCTCCGTCTAGCTACTTATTCTTGGCAACCACGTAAATTTTCCAAAAGAACCCATTTTTGACTATCTC
CACTAAATTTGCATATTTTTATGGGGGAAAAAAAATGTAGTATAAAACCAACATATCCATAAACCCCAAAACCC
CCCATAAATCATCATATCATTCACTACTTAAATTTTTTAGTGTGAGTTTGAGAGTGAGTGATATCAAGAAAA
TATATAAACAGAACACATCTTTCAATACTCATTTAACCATTTTCGTCACAAATAATAAATATTCTTAAAAAATT
TCAAGATTTGATATGTTTTACAATAATTTGATAATGATGAAAAACCTTGGTACTAATTTCTTGTAGACCCTATT
TTCTTACATAGGTGTGTGGTTTTTAATACAGTTTTTGTCTTGGTTTGGCTTATTAGTCTCATGGGTTTGGAAAACT
TTTAAAAAGGTTGATACAAGAAATTTGGTTTCAAAACAAAGTTTTAGGGGTGTGTTTTATAAAACAACTTTGGTT
TGTTGTGTTTTCTATCACTCTTTAATCTTGTGTTATGTTTCTTGAACAATTTCTATTGGTATAGAAATGGTTGG
AGTGATGGAAAAATAGTGACCTTTTGCATGTTGTTTTAGGAACCTTGTGTTGGTTATGTGTTTCTGTTTATTTG
CATACGCTTGTGTTTGAATTTGTCGACACAATCTAGAAAATACCCATTAGTGTTAAGGGCTTGGTGGGTGTTTTTC
TTTACATTTTCGTGTTATTCGTTGCCATAGACTAGTTTACTATAGAAGAACCCATGATTTCGCTTATATGTTG
TTTGTACCGATTCTGTGTCTAGTTTGATCGGTTGTTTTGTTTCTTTGTTTGTGGGTTTTCCCATAGACAGAGGAA
GAAATTCGAAATCTTGAAGAACCCCTTTTGAATTCGGGGTTCTAGAGTAGTTAAGAGTGAACAGCATCAACT
TATGAAATGCAAGTTTCTTTAGCTTGTGACATTTTCTTGGATGAGTCCATTATTATGCAAAAGGCAATAAGAAA
CCGTTAGACCTTGACGACGTCCCTCAACTCGCGAATATAGATAGTGTGAGAGGAGTGTTTCTGTTTTGTTAAAT
AAAGTCGAGTCTTTTAAACAATGGGAATAACCAATCACTACATTTGGTCTCACGAAAGCGTTATTCTACATAGTG
TGGAAAGAAGTAGTGATCACAGGTTTTTTCGCGATTGTATCAACATTGACTAGTTTTGTTGGTCTTACCTCATA
GAATCATTTGTTCAATACCTGAATGGAACCCGAGATTATGAACATCAAGGTGCCGTTTTGGTGGCTGCTTTTTTC
ATATCTAAGCTTATCGGGTGTTTTGCACAACGCCATTGGTATTTCAAGCTTCAACAACTGGCATTAGAGCCAGA
TCAGCTATTGTTGCCATGATTTACCAAAAAGGTTTGACCATTTCTGGCCAGTCAAGCAGGGGAACCTCAAGCGGT
GAAATCATCAATTTTATGGCTGTGATGACGAACGAATCGGCGACTATGCTTGGTATATGCACGATTTTGGCTA
GTACTTCTGCAAGTAGGCATTGCATTAGCACTTTTATATAAAAAATTTAGGGATAGCTGCAGTTGCTTCATTGATT
GCTACGATTGTGGTATTGGTGGCGAATTATCCACTTGGGCATATACAAGAAAAGCTTCAAGATGAGTTGATGAAA
TCTAAAGATAAAAGGATGAAATCAACATCCGAAATCTTGAGAAACATGAGGATTCTTAAACTTCAAGGTTGGGAA
ATGAAAGTTTTTCCAAAAATCATCAAACTTAGAGAATGAAGAAGAAAGTGCAATTGAAGAAATATATGTATACT
TTGTCATGATTTCTTTTATTTCTGGGGTGCTCCAATTATCGTAGCTGTGGCCACTTTCGCTACGTGTTTGT
GTTGGCACCCCTCTTGAATCAGGAAAAGTACTCTCTGCACCTTGCAACATTTAAGATACTTCAAGAGCCTATTTAC
AATCTTCTGACTCGATATCTGTGTTTTTCCAACTAAAGTCTCACTTGATCGCATTGCTACATATCTTCGGCTC
AGTGACATAGACTCCAGTGCTATAGATAAGGTTACACGAGGTAGTTCTGATACTGCTATTGAGATCATAAATGGG
AATTTTGCATGGGATTTGAGTGATCTTCAACCCCAACGTTGAAAGATATAAACATTAGGGTAAATCATGGGATG
CGGGTTGCTGTTTGTGGGACTGTTGGCTCGGGGAAGTCTAGTTTACTTTCTTGTATCTTGGGTGAGGTATCGAAG
ATATCAGGAAGTGTGAGAATTGAGGGATCAAAGGCATATGTAGCTCAATCGCCATGGATACAGAGCGGAAAGATT
GAAGATAATATCTTGTGTTGGTAGGGAGATGGATAGAGAGAGGTATGAGAATGTTTTAGAAGCATGTTCTTTGAAG
AAAGATCTTGAATTTCTTTTCAATTTGGTGATCAAACCTGTTATGGTGAGAGAGGGATTAACTTGAGTGGTGGTCAGA
AGCAGAGGATACAGATTGCTCGGGCTCTCTATCAAGATGCGGATATTTATCTTTTTGATGACCCGTTTCACTGCTG
TTGATGCTCATACAGGAAGTCATCTATTCAAAGTAAGTGAATTTTGCTTGTGATTTGCTCCGTGAAGTTTCACT
GTTTAAACGCATCCAAAAGTTTATCTGTATTAACCTGAATCTGCATTTGGTGTTTTTAATAACAGGAATGCATGT
TACAATTTCTGGACTCAAAAACAGTGATTTACATAACACATCAAGTTGAGTTCTTACCTGCTGCAGATCTTATCC
TGGTGAGCAATCAAACCTTTAAATCATTAGGCGATTTCGTAGAATCAAATTAACAAATATGGAATCTTGCATT
GTGTTTAGTTATAGAAAACCTTTTTTTTTTAGTTTTTCACTTTTCACTTTCTCAAAAAAATTTGAAAAACAATTTCCAA
AGATGATATCTGATGTGAAAGATGAGTCCTACAATGATCTGTTTTTGGTCTTACTCATATGATGATTATCTGTTT
CTCGTTGGATTTTCAATTTGAATTTGTACATGATATTAACATGACAAACAGGTTTTGAGAGATGGAAGGATCACA
CAAGCCGGAAGTACAATGATATTTCTCAATTCAGGAAGTGACTTTATGGAGCTTGTAGGTGCACACAAAGAAGCG
TTATCTGCCATTGATTCCATGGGGACAAACATGCAGGAACAAACAGATAGTTCCAAAAAGAGCACAAATGATAGT
CAAAATGGTAAAACAGATGACATATCAGGTACAAAAGCACAGCTTGTTCAAGAAGAAGAACGAGAGAAAGGCAAA

10/17

FIG. 9B Cont.

GTAGGTTTCTCAGTCTACTGGAACACATAACAACGGCTTACGGCGGGGCTCTTGCACCATTAATAGTTTTGGCA
CAAGTATTGTTCCAAGTACTTCAAATTGGAAGTAATTATTGGATGGCTTGGGCATCTCCAGTGTCTGCAAGTGAT
CCAGCCCCGGTTACAGGCTCAACCCTGATCCTGGTTTATGTAGTTTTAGCAGCTGGATGTGCATTGTGTATACTC
GCAAGAGGTCTGCTTCTTGCAACTGTTGCATATAAAGCAGCCACTATTCTCTTCCACAAAATGCACCTATCCATT
TTCCGTTTCGCCCATGTCTTTCTTTGACTCTACTCCAAGTGACGAATACTAAATAGAGTAAGCAAACCTCTTCTC
ATCCTATCCTTTTAATGAAACGGGTTGATTTGGGCTAGGTTTAATATTACATACTTGTAAAGACCAACTTGACAAA
GTTTCACATTTTCCAAACCTTGTAGAAAAAGTATGCGCAAAATCACTAACTTAGTCTTGATaCCCTCACGCTCT
TGCACCATTAATAGTTTTTGGCACAAAGTATTGTTCCAAGTACTTCAAATTGGAAGTAATTATTGGATGGCTTGGGC
ATCTCCAGTGTCTGCAAGTGATCCAGCCCCGGTTACAGGCTCAACCCTGATCCTGGTTTATGTAGTTTTAGCAGC
TGGATGTGCATTGTGTATACTCGCAAGAGGTCTGCTTCTTGCAACTGTTGCATATAAAGCAGCCACTATTCTCTT
CCACAAAATGCACCTATCCATTTTCCGTTTCGCCCATGTCTTTCTTTGACTCTACTCCAAGTGGACGAATACATAA
TAGAGCCTCTACAGACCAAAGTGGGTGGACATGCAAAATCCATACCAAGTTGGATCATTGTATTGCTCGTATCAT
TCAACTTCTTGGCATCATTGCAGTTATGTGCACAAATGTGCTTGGCAAGTGATCATAAATATTTATTCCTGTGCGTGG
AATGTGCATCTGGTTGCAGCAATATTACCTGCCTTCAGCACGAGAAATGGCACGGCTAGTTGGCGTTTGTAAAGG
TCCAGTGATACAGAATTTTGCTGAAACAATATCAGGGTCAACAACCATTAGAAGTTTTGATCAACAAGGCAGATT
CCAGGACACAAACCTGAAATTGAATGATGATTTTCGCAAGGCCAAAATTTTCATGCTGCTGCGGCTATGGAATGGTT
AGGCATACGTTTGGATATGCTGTCTTCTTTTACCTTTGCTGCATTTCTAATTTTCTTAATTTCTATCCCAGAAGG
AACTATAGATCCAAGTATCGCGGGCTTGGCTGCTACTTACGGTCTTACTTTGAACATGTTACAAGGATGGGTAGT
ATGGACTTTAACCAACCTTGAAAAACAAAATTAATTTCTGTTGAAAGGATATTTTCAAGTATTCATCTATCCCCAGCGA
ACCTCCTCTAGTTATAGAATCTAATAGGCCTGATGATCAGTGGCCGTCACAGGGAGAAGTTGATATCCGTAACCT
GCAGGTACTTTAATCTATTTACTTATCTTGCATAATTTATGTAGCAATGTAGCATGGAATCTGTCAAGATAAATT
CTAGCAAGTGACCCTATGAGACCTTTCCATGAGCTGTTTCATGGGCTCCATAACTTACCTCGCCTCAAACCTGAGCC
GTGTTTAATGTTCTTACATTTTTTATGGGTTTGCCGTGTTCCCATTTGTTTCATATAAATATAGGGTCATCTGTGT
ACTATGTTCTAGATGAATATGAATCTAATGGTTCTTTTTGGCTTTTAAAGGTTTCGGTATGCACCACATATGCCA
CTTGTGTTGCGAGGCCTTACGTGCAATTTCAAAGGGGGAAAGAAAACCTGGGATTGTGGGAAGAACTGGTAGTGGG
AAGTCGACTCTGATACAAACACTCTTCCGCCTTGTGAGCCAGCTTCTGGACAAATTTCAATAGATGGAATTAAT
ATATCCACAATAGGACTTCATGATCTTCGTTCAAGATTGAGTATAATTCCTCAAGATCCAACCATGTTTGAAGGA
ACTATTTCGAAGCAATCTGGATCCACTTGAAGAGTATACAGATGATAAAATTTGGGAGGCTCTCGATAAGTGTCAA
CTTGGAGATGAAGTAAGGAGCAAGAAAGGGAAGCTCGATTACCCAGTTACCGAGAATGGAGAAAACCTGGAGCGTG
GGTCAAAGGCAGCTGGTGTGTCTCGGGCGTGTAATACTCAAAAAAAGCAAAGTCCTGGTACTTGATGAAGCCACT
GCATCAGTCGACACGGCAACTGATGGAATGATTACAGCAACGCTCACGAAACATTTTACAGATTCAACTGTGATA
ATGATTGCGCATCGTATCACCTCTGTACTTGATAGTGACATGGTTTTAGTTCTAGAACAAGGTCTGATTGATGAA
TATGATTCTCCAACAAAGTTGCTGGAAGACAAATCATCTTCATTTGCTAAGCTTGTGCGGAGTATAGTATGAGA
TCGAGTTCCAGTTATGAAAACCTTAGCAACAGCTTAGTATGTTGGTGTAAAGATTGGTGCTTGATGATCTGATCTT
GATTGCTCAAATGAGAATATAGACATAGAAAGGTAAGTTTGTTTTGATTGTGCGGAGTTTGAGCTCTCATGGACT
GCATCTAGCTTGTGCAAGGAGCTTGGTGCCACAGGATCGTACTCTATGTACTACCATAACCTTGGCATATCTCC
ACTGTTTTCTAAATTTTTGATGTGTGTAAGCATAGCTATTTTCTCTTGTAAAGTTGTCTAAACTATCTAAACA
CTGCTTAAACTTAAAGATCATTTGTTGATCAAGCTTCTAGTTATTGATTATGTTGTATTCTCTTTGCGTATCAT
ATATATCATGCGGTTTTAGGTGGATTATGAGAAAATGAATCAAGAAAGTTCTTTCATTCCATC

11/17

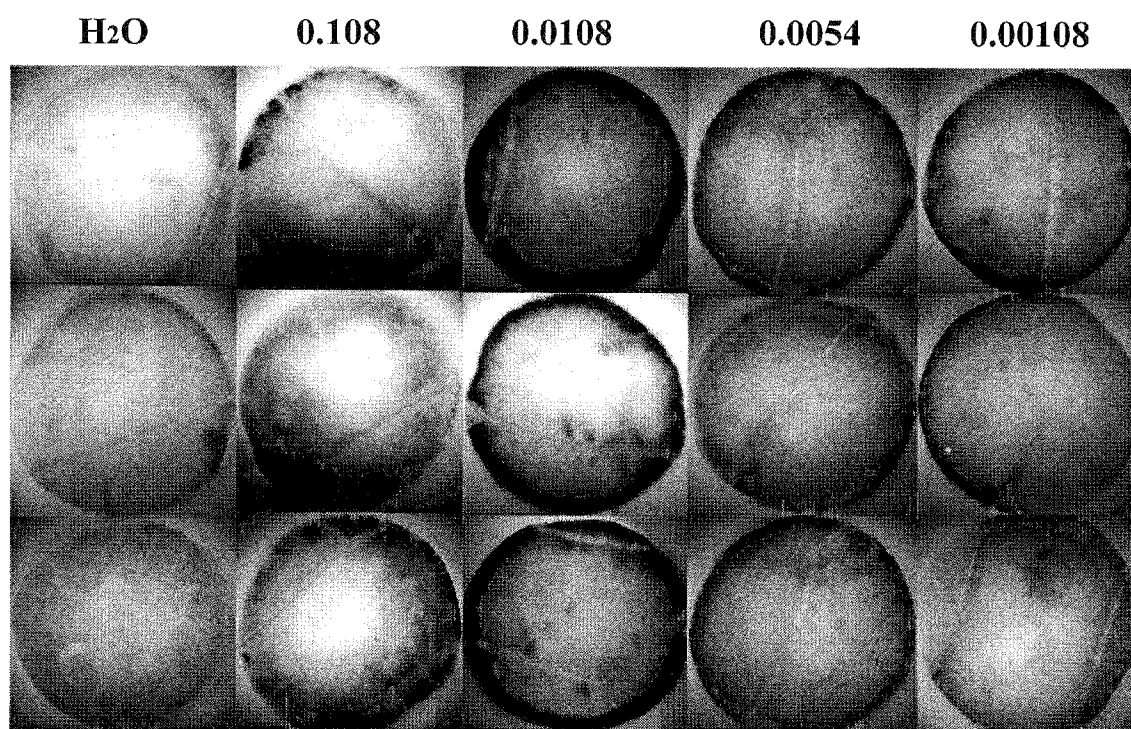


FIG. 10A

12/17

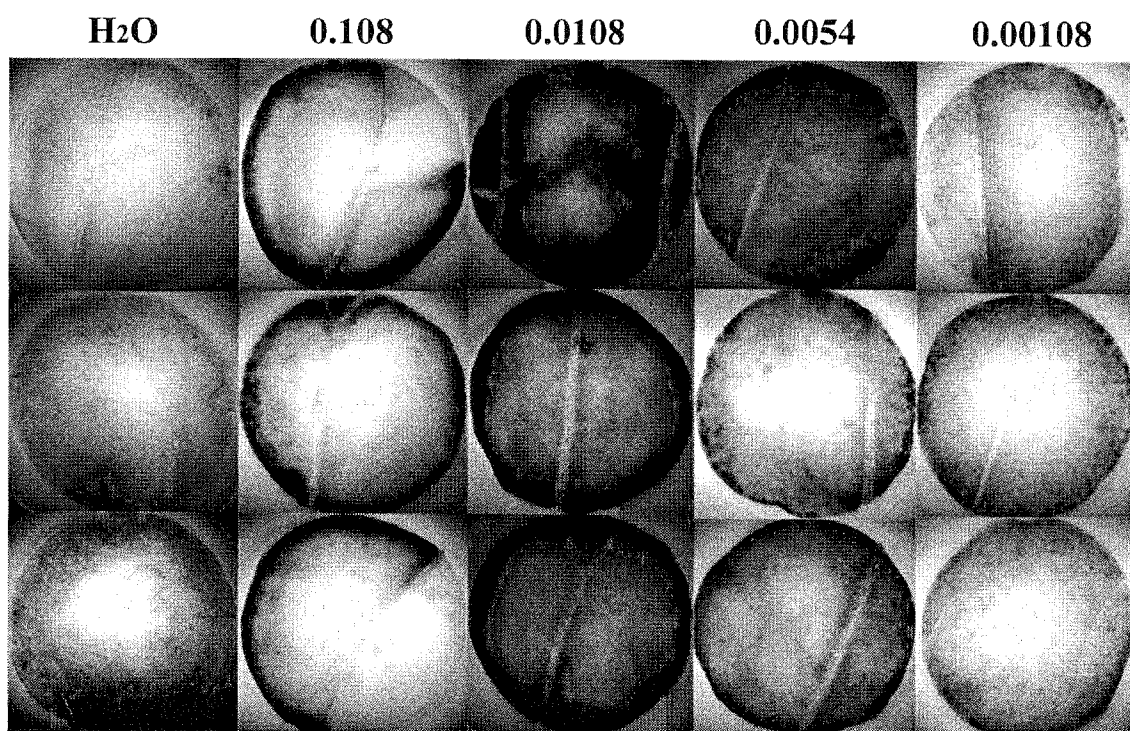


FIG. 10B

13/17

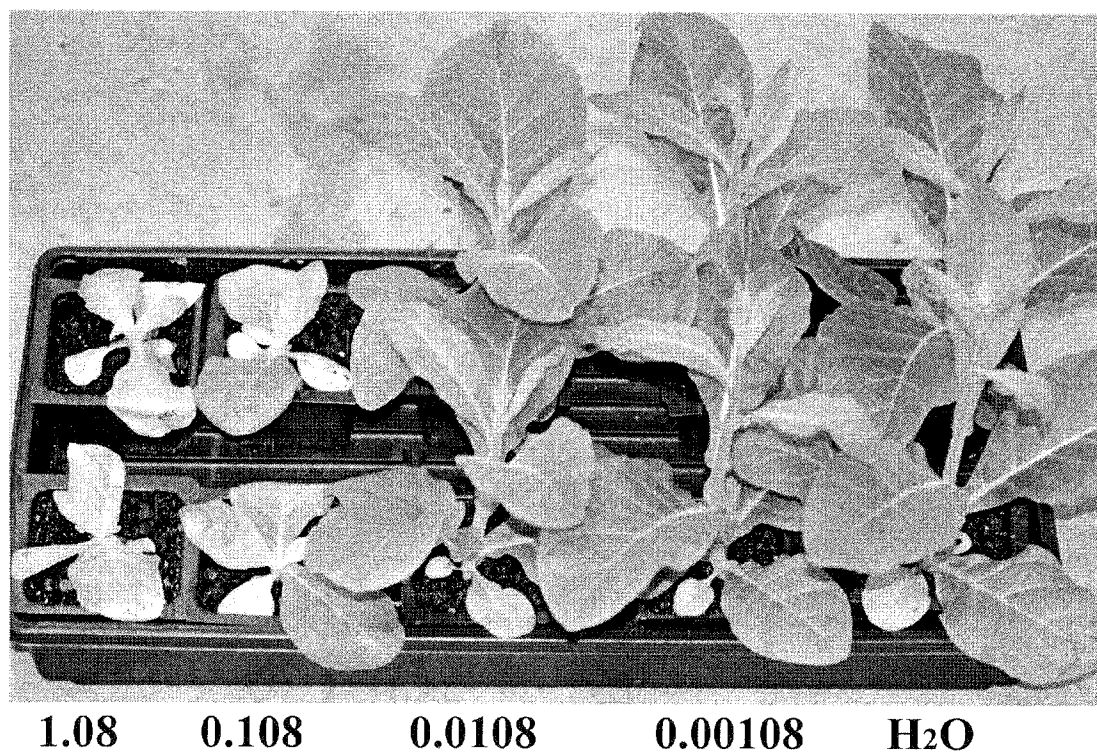


FIG. 11

14/17

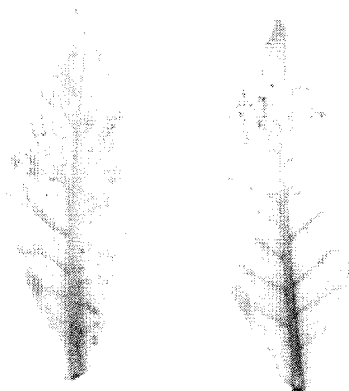
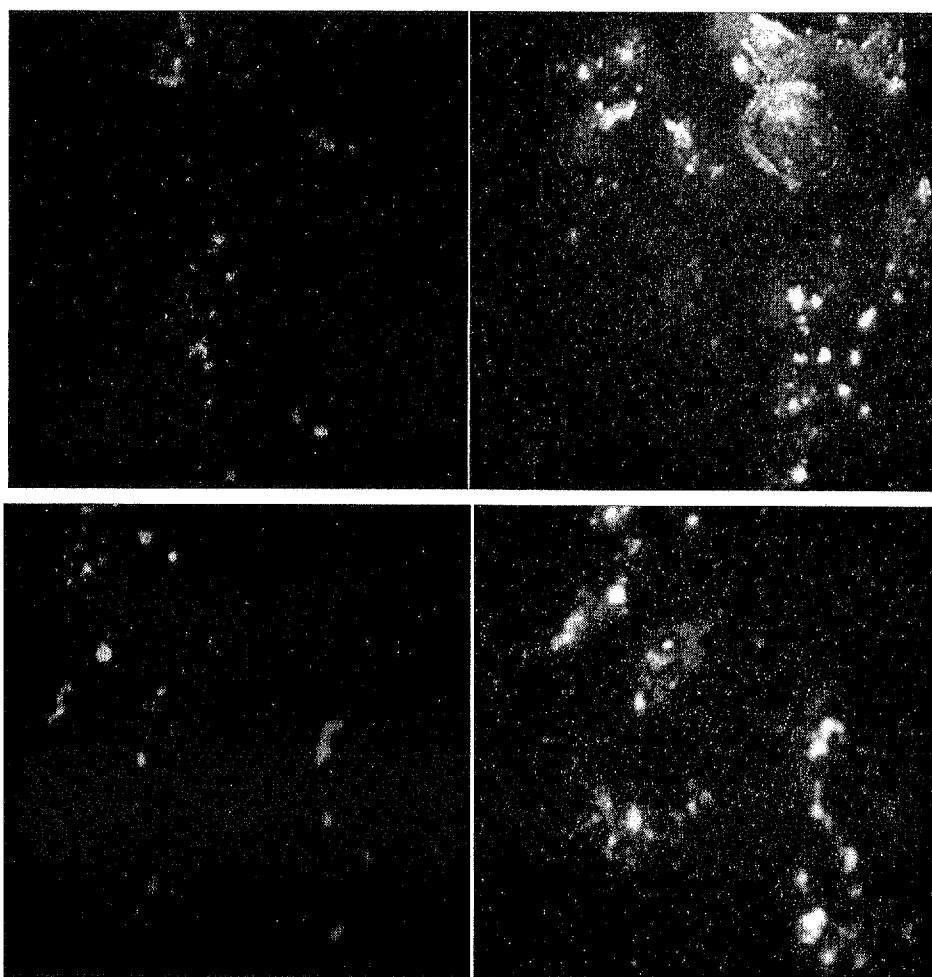


FIG. 12A Without Roundup



FIG. 12B Treated with Roundup

15/17



–Round up
FIG. 13A

+Round up
FIG. 13B

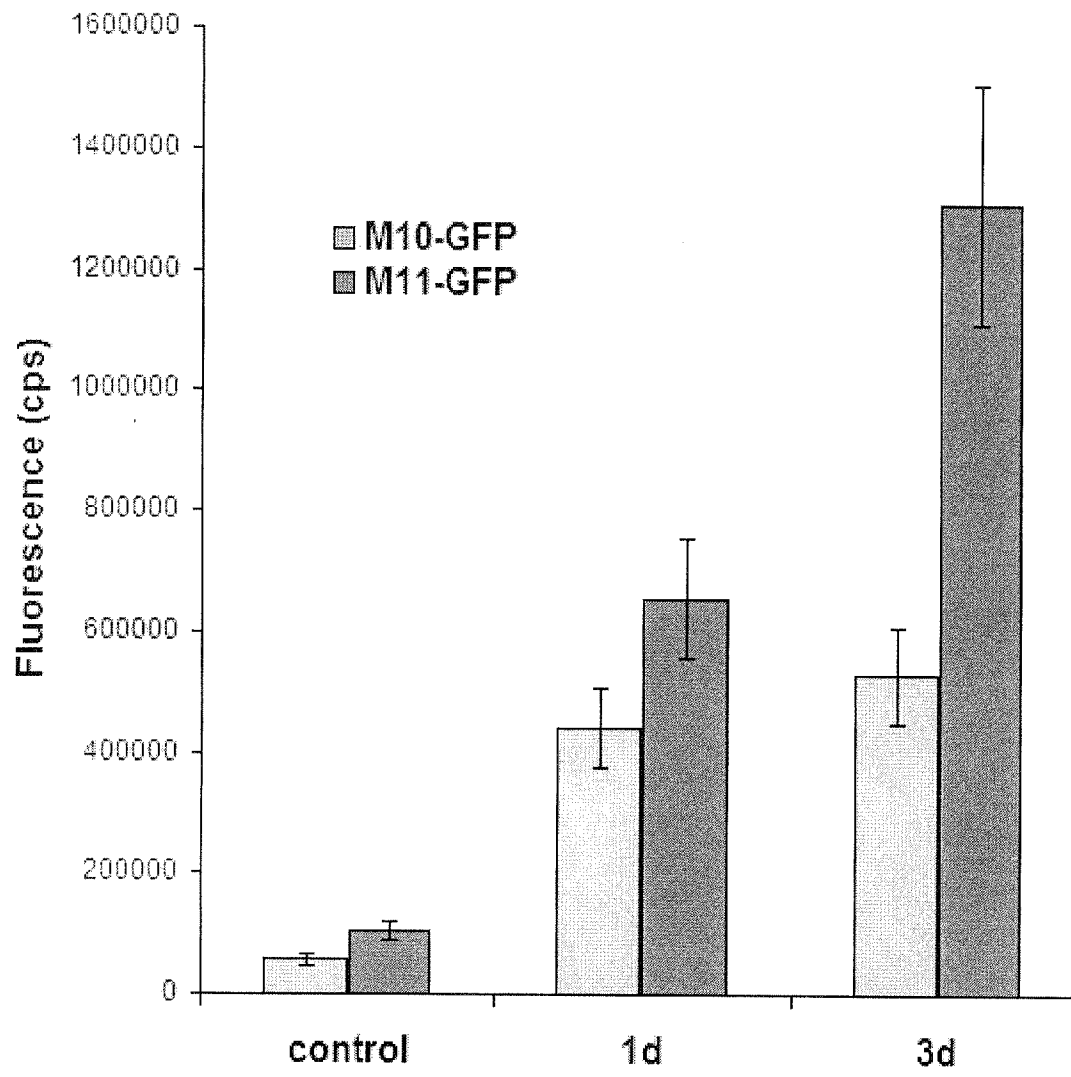


FIG. 14

17/17

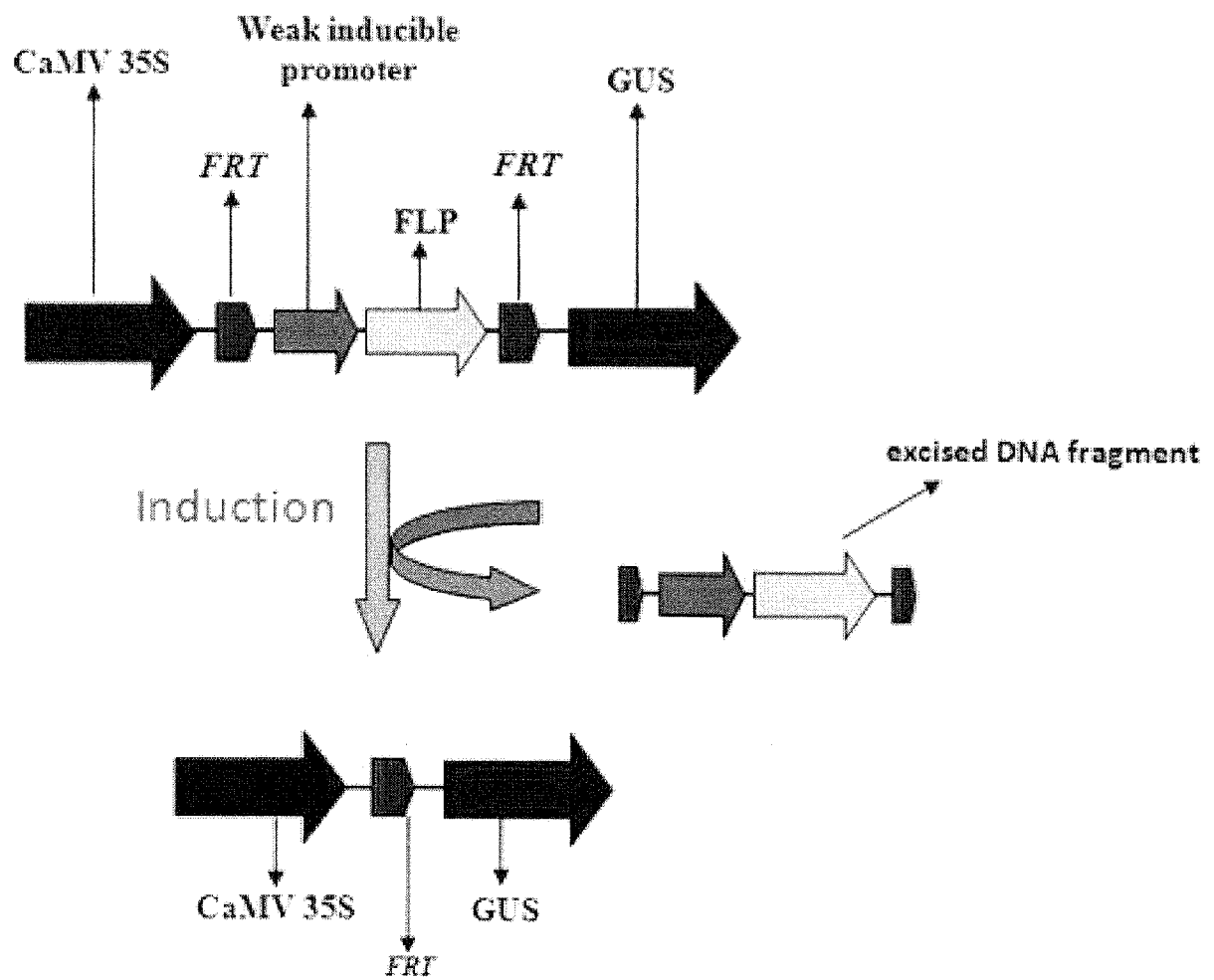


FIG. 15