

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064794 A2

(51) International Patent Classification⁷: **C12N 15/29**,
G01N 33/68, A01H 5/00

(74) Agent: **BASTIAN, Werner**; Syngenta Participations AG,
Intellectual Property, P.O. Box, CH-4002 Basel (CH).

(21) International Application Number: PCT/EP02/00188

(22) International Filing Date: 10 January 2002 (10.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/260,967 11 January 2001 (11.01.2001) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LEVIN, Joshua, Zvi** [US/US]; Syngenta Biotechnology, Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709 (US). **BUDZISZEWSKI, Gregory, Joseph** [US/US]; Syngenta Biotechnology, Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709 (US). **WEGRICH, GLOVER, Lyn** [US/US]; 5446 Escover Lane, San Jose, CA 95118 (US). **AUX, George, W.** [US/US]; Syngenta Biotechnology, Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (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

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HERBICIDE TARGET GENES AND METHODS

(57) Abstract: The invention relates to genes isolated from *Arabidopsis* that code for proteins essential for seedling growth. The invention also includes the methods of using these proteins to discover new herbicides, based on the essentiality of these genes for normal growth and development. The invention is also applied to the development of herbicide tolerant plants, plant tissues, plant seeds, and plant cells.



WO 02/064794 A2

HERBICIDE TARGET GENES AND METHODS

The invention relates to genes isolated from *Arabidopsis* that code for proteins essential for seedling growth. The invention also includes the methods of using these proteins as herbicide targets, based on the essentiality of the genes for normal growth and development. The invention is also useful as a screening assay to identify inhibitors that are potential herbicides. The invention may also be applied to the development of herbicide tolerant plants, plant tissues, plant seeds, and plant cells.

The use of herbicides to control undesirable vegetation such as weeds in crop fields has become almost a universal practice. The herbicide market exceeds 15 billion dollars annually. Despite this extensive use, weed control remains a significant and costly problem for farmers. Effective use of herbicides requires sound management. For instance, the time and method of application and stage of weed plant development are critical to getting good weed control with herbicides. Since various weed species are resistant to herbicides, the production of effective new herbicides becomes increasingly important. Novel herbicides can now be discovered using high-throughput screens that implement recombinant DNA technology. Metabolic enzymes found to be essential to plant growth and development can be recombinantly produced through standard molecular biological techniques and utilized as herbicide targets in screens for novel inhibitors of the enzyme activity. The novel inhibitors discovered through such screens may then be used as herbicides to control undesirable vegetation.

Herbicides that exhibit greater potency, broader weed spectrum, and more rapid degradation in soil can also, unfortunately, have greater crop phytotoxicity. One solution applied to this problem has been to develop crops that are resistant or tolerant to herbicides. Crop hybrids or varieties tolerant to the herbicides allow for the use of the herbicides to kill weeds without attendant risk of damage to the crop. Development of tolerance can allow application of a herbicide to a crop where its use was previously precluded or limited (*e.g.* to pre-emergence use) due to sensitivity of the crop to the herbicide. For example, U.S. Patent No. 4,761,373 to Anderson *et al.* is directed to plants resistant to various imidazolinone or sulfonamide herbicides. An altered acetohydroxyacid synthase (AHAS) enzyme confers the resistance. U.S. Patent No. 4,975,374 to Goodman *et al.* relates to plant cells and plants containing a gene encoding a mutant glutamine synthetase (GS) resistant to inhibition by herbicides that

were known to inhibit GS, *e.g.* phosphinothricin and methionine sulfoximine. U.S. Patent No. 5,013,659 to Bedbrook *et al.* is directed to plants expressing a mutant acetolactate synthase that renders the plants resistant to inhibition by sulfonylurea herbicides. U.S. Patent No. 5,162,602 to Somers *et al.* discloses plants tolerant to inhibition by cyclohexanedione and
5 aryloxyphenoxypropanoic acid herbicides. The tolerance is conferred by an altered acetyl coenzyme A carboxylase (ACCase).

Notwithstanding the above described advancements, there remain persistent and ongoing problems with unwanted or detrimental vegetation growth (*e.g.* weeds). Furthermore, as the population continues to grow, there will be increasing food shortages. Therefore, there exists
10 a long felt, yet unfulfilled need, to find new, effective, and economic herbicides.

It is an object of the invention to provide effective and beneficial methods to identify novel herbicides. A feature of the invention is the identification of genes in *Arabidopsis*, herein referred to as the ET6497 gene, which encodes a protein with sequence similarity to
15 translation release factor (RF-1) (Craigén and Caskey (1987) *Biochimie* 69: 1031-1041; Craigén *et al.* (1990) *Mol. Micro.* 4: 861-865; Zhouravleva *et al.* (1995) *EMBO J.* 14: 4065-4072; Kisselev and Frolova (1995) *Biochem. Cell. Biol.* 73: 1079-1086; Stansfield *et al.* (1997) *Eur. J. Biochem.* 245: 557-563); the GT1773 gene, which encodes a protein that appears to be bi-functional and has sequence similarity to
20 diaminohydroxyphosphoribosylaminopyrimidine deaminases and 5-amino-6-(5-phosphoribosylamino) uracil reductases (Fuller and Mulks (1995) *J. Bacteriol.* 177: 7265-7270; Richter *et al.* (1997) *J. Bacteriol.* 179: 2022-2028; Begley *et al.* (1998) *Topics Curr. Chem.* 195: 93-141); the GT0992 gene, which encodes a protein with no known function, but may function as a transport protein (Jobling and Ritchie (1988) *Gene* 66: 245-258); the
25 ET6233 gene, which encodes a protein with similarity to the *Bacillus megaterium* *cbiX* protein required for cobyrinic acid biosynthesis (Raux *et al.* (1998) *Biochem. J.* 335: 159-166; Raux *et al.* (1998) *Biochem. J.* 335: 167-173; Raux *et al.* (1996) *J. Bacteriol.* 178: 753-767); the ET0763 gene, which encodes a protein with similarity to 30S ribosomal protein S1 (Sugita *et al.* (1995) *Mol. Gen. Genet.* 246: 142-147; Franzetti *et al.* (1992) *J. Biol. Chem.* 267: 19075-19081); the ET5848 gene which encodes a protein with no known function; the
30 GT5062 gene which encodes a protein with similarity to uracil phosphoribosyltransferase (also known as UMP pyrophosphorylase), (Martinussen *et al.* (1995) *J. Bacteriol.* 177: 271-274; Kern *et al.* (1990) *Gene* 88: 149-157); and the ET5036 gene which encodes a protein

with no known function. An important and unexpected feature of the invention is the discovery that each of these genes is essential for seedling growth and development. An advantage of the present invention is that the newly discovered essential genes containing novel herbicidal modes of action enable one skilled in the art to easily and rapidly identify novel herbicides.

One object of the present invention is to provide essential genes in plants for assay development for inhibitory compounds with herbicidal activity. Genetic results show that when either the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes are mutated in *Arabidopsis*, the resulting phenotype is seedling lethal in the homozygous state. This suggests a critical role for the gene products encoded by each of these genes.

Using Ac/Ds transposon mutagenesis, the inventors of the present invention have demonstrated that the activity encoded by the *Arabidopsis* ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes (herein referred to as ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity) is essential in *Arabidopsis* seedlings. This implies that chemicals that inhibit the function of any one of these proteins in plants are likely to have detrimental effects on plants and are potentially good herbicide candidates. The present invention therefore provides methods of using a purified protein encoded by any one of the gene sequences described below to identify inhibitors thereof, which can then be used as herbicides to suppress the growth of undesirable vegetation, e.g. in fields where crops are grown, particularly agronomically important crops such as maize and other cereal crops such as wheat, oats, rye, sorghum, rice, barley, millet, turf and forage grasses, and the like, as well as cotton, sugar cane, sugar beet, oilseed rape, and soybeans.

The present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the ET6497 gene. The nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:1, and the corresponding amino acid sequence is set forth in SEQ ID NO:2. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:17. Also, the present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the GT1773 gene. The nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:3, and the corresponding amino acid sequence is set forth in SEQ ID NO:4. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:18. Furthermore, the present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the GT0992 gene. The

nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:5, and the corresponding amino acid sequence is set forth in SEQ ID NO:6. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:19. Furthermore, the present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the ET6233 gene. The nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:7, and the corresponding amino acid sequence is set forth in SEQ ID NO:8. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:20. Furthermore, the present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the ET0763 gene. The nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:9, and the corresponding amino acid sequence is set forth in SEQ ID NO:10. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:21. Furthermore, the present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the ET5848 gene. The nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:11, and the corresponding amino acid sequence is set forth in SEQ ID NO:12. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:22. Furthermore, the present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the GT5062 gene. The nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:13, and the corresponding amino acid sequence is set forth in SEQ ID NO:14. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:23. Furthermore, the present invention discloses a nucleotide sequence derived from *Arabidopsis*, designated the ET5036 gene. The nucleotide sequence of the cDNA clone is set forth in SEQ ID NO:15, and the corresponding amino acid sequence is set forth in SEQ ID NO:16. The nucleotide sequence of the genomic DNA sequence is set forth in SEQ ID NO:24. The present invention also includes nucleotide sequences substantially similar to those set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, and SEQ ID NO:15. The present invention also encompasses plant proteins whose amino acid sequence are substantially similar to the amino acid sequences set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, and SEQ ID NO:16. Such proteins can be used in a screening assay to identify inhibitors that are potential herbicides.

In a preferred embodiment, the present invention relates to a method for identifying chemicals having the ability to inhibit ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity in plants preferably comprising the steps of: a) obtaining transgenic plants,

plant tissue, plant seeds or plant cells, preferably stably transformed, comprising a non-native nucleotide sequence encoding an enzyme having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, and capable of overexpressing an enzymatically active ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene product (either full length or truncated but still active), respectively; b) applying a chemical to the transgenic plants, plant cells, tissues or parts and to the isogenic non-transformed plants, plant cells, tissues or parts; c) determining the growth or viability of the transgenic and non-transformed plants, plant cells, tissues after application of the chemical; d) comparing the growth or viability of the transgenic and non-transformed plants, plant cells, tissues after application of the chemical; and e) selecting chemicals that suppress the viability or growth of the non-transgenic plants, plant cells, tissues or parts, without significantly suppressing the growth or viability or growth of the isogenic transgenic plants, plant cells, tissues or parts. In a preferred embodiment, the enzyme having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity is encoded by a nucleotide sequence derived from a plant, preferably a monocotyledonous or a dicotyledonous plant, preferably a dicotyledonous plant, preferably *Arabidopsis thaliana*, desirably identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively. In another embodiment, the enzyme having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity is encoded by a nucleotide sequence capable of encoding the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16, respectively. In yet another embodiment, the enzyme having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity has an amino acid sequence identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16, respectively.

The present invention further embodies plants, plant tissues, plant seeds, and plant cells that have modified ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity and that are therefore tolerant to inhibition by a herbicide at levels normally inhibitory to naturally occurring ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively. Herbicide tolerant plants encompassed by the invention include those that would otherwise be potential targets for normally inhibiting herbicides,

particularly the agronomically important crops mentioned above. According to this embodiment, plants, plant tissue, plant seeds, or plant cells are transformed, preferably stably transformed, with a recombinant DNA molecule comprising a suitable promoter functional in plants operatively linked to a nucleotide coding sequence that encodes a modified ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene that is tolerant to inhibition by a herbicide at a concentration that would normally inhibit the activity of wild-type, unmodified ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene product, respectively. Modified ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity may also be conferred upon a plant by increasing expression of wild-type herbicide-sensitive ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein by providing multiple copies of wild-type ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes, respectively, to the plant or by overexpression of wild-type ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes, respectively, under control of a stronger-than-wild-type promoter. The transgenic plants, plant tissue, plant seeds, or plant cells thus created are then selected by conventional selection techniques, whereby herbicide tolerant lines are isolated, characterized, and developed. Alternately, random or site-specific mutagenesis may be used to generate herbicide tolerant lines.

Therefore, the present invention provides a plant, plant cell, plant seed, or plant tissue transformed with a DNA molecule comprising a nucleotide sequence isolated from a plant that encodes an enzyme having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, wherein the DNA expresses the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, and wherein the DNA molecule confers upon the plant, plant cell, plant seed, or plant tissue tolerance to a herbicide in amounts that normally inhibits naturally occurring ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively. According to one example of this embodiment, the enzyme having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity is encoded by a nucleotide sequence identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, or has an amino acid sequence identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16, respectively.

The invention also provides a method for suppressing the growth of a plant comprising the step of applying to the plant a chemical that inhibits the naturally occurring ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity in the plant. In a related aspect, the present invention is directed to a method for selectively suppressing the growth of undesired vegetation in a field containing a crop of planted crop seeds or plants, comprising the steps of: (a) optionally planting herbicide tolerant crops or crop seeds, which are plants or plant seeds that are tolerant to a herbicide that inhibits the naturally occurring ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity; and (b) applying to the herbicide tolerant crops or crop seeds and the undesired vegetation in the field a herbicide in amounts that inhibit naturally occurring ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, wherein the herbicide suppresses the growth of the weeds without significantly suppressing the growth of the crops.

The invention thus provides:

An isolated DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15. In a preferred embodiment, the nucleotide sequence encodes an amino acid sequence substantially similar to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16. In another preferred embodiment, the nucleotide sequence is SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15. In yet another preferred embodiment, the nucleotide sequence encodes the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16. Preferably, the nucleotide sequence is a plant nucleotide sequence, which preferably encodes a polypeptide having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively.

The invention further provides:

A polypeptide comprising an amino acid sequence encoded by a nucleotide sequence substantially similar to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15. Preferably, the amino acid sequence is encoded by SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15. Preferably, the polypeptide

comprises an amino acid sequence substantially similar to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16, respectively. Preferably the amino acid sequence is SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16.

- 5 The amino acid sequence preferably has ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively. In another preferred embodiment, the amino acid sequence comprises at least 20 consecutive amino acid residues of the amino acid sequence encoded by SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively. Or, alternatively,
- 10 the amino acid sequence comprises at least 20 consecutive amino acid residues of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16, respectively.

The invention further provides:

- An expression cassette comprising a promoter operatively linked to a DNA molecule
- 15 according to the present invention, a recombinant vector comprising an expression cassette according to the present invention, wherein said vector is preferably capable of being stably transformed into a host cell, a host cell comprising a DNA molecule according to the present invention, wherein said DNA molecule is preferably expressible in the cell. The host cell is preferably selected from the group consisting of an insect cell, a yeast cell, a prokaryotic cell
- 20 and a plant cell. The invention further provides a plant or seed comprising a plant cell of the present invention, wherein the plant or seed is preferably tolerant to an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively.

The invention further provides:

- A process for making nucleotide sequences encoding gene products having altered ET6497,
- 25 GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, comprising: a) shuffling an unmodified nucleotide sequence of the present invention, b) expressing the resulting shuffled nucleotide sequences, and c) selecting for altered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, as compared to the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity,
- 30 respectively, of the gene product of said unmodified nucleotide sequence.

In a preferred embodiment, the unmodified nucleotide sequence is identical or substantially similar to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ

ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, or a homolog thereof. The present invention further provides a DNA molecule comprising a shuffled nucleotide sequence obtainable by the process described above, a DNA molecule comprising a shuffled nucleotide sequence produced by the process described above. Preferably, a shuffled nucleotide sequence obtained by the process described above has enhanced tolerance to an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively. The invention further provides an expression cassette comprising a promoter operatively linked to a DNA molecule comprising a shuffled nucleotide sequence a recombinant vector comprising such an expression cassette, wherein said vector is preferably capable of being stably transformed into a host cell, a host cell comprising such an expression cassette, wherein said nucleotide sequence is preferably expressible in said cell. A preferred host cell is selected from the group consisting of an insect cell, a yeast cell, a prokaryotic cell and a plant cell. The invention further provides a plant or seed comprising such plant cell, wherein the plant is preferably tolerant to an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively.

The invention further provides:

A method for selecting compounds that interact with the protein encoded by SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, comprising: a) expressing a DNA molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, or a sequence substantially similar to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, or a homolog thereof, to generate the corresponding protein, b) testing a compound suspected of having the ability to interact with the protein expressed in step (a), and c) selecting compounds that interact with the protein in step (b).

The invention further provides:

A process of identifying an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, comprising: a) introducing a DNA molecule comprising a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, and having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, or nucleotide sequences substantially similar thereto, or a

homolog thereof, into a plant cell, such that said sequence is functionally expressible at levels that are higher than wild-type expression levels, b) combining said plant cell with a compound to be tested for the ability to inhibit the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, under conditions conducive to such inhibition, c) measuring plant cell growth under the conditions of step (b),
5 d) comparing the growth of said plant cell with the growth of a plant cell having unaltered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, under identical conditions, and e) selecting said compound that inhibits plant cell growth in step (d).

10 The invention further comprises a compound having herbicidal activity identifiable according to the process described immediately above.

The invention further comprises:

A process of identifying compounds having herbicidal activity comprising:

a) combining a protein of the present invention and a compound to be tested for the ability to
15 interact with said protein, under conditions conducive to interaction, b) selecting a compound identified in step (a) that is capable of interacting with said protein, c) applying identified compound in step (b) to a plant to test for herbicidal activity, and d) selecting compounds having herbicidal activity.

The invention further comprises a compound having herbicidal activity identifiable according
20 to the process described immediately above.

The invention further comprises:

A method for suppressing the growth of a plant comprising, applying to said plant a compound that inhibits the activity of a polypeptide of the present invention in an amount sufficient to suppress the growth of said plant.

25 The invention further comprises:

A method for recombinantly expressing a protein having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity comprising introducing a nucleotide sequence encoding a protein having one of the above activities into a host cell and expressing the nucleotide sequence in the host cell. A preferred host cell is selected from the group
30 consisting of an insect cell, a yeast cell, a prokaryotic cell and a plant cell. A preferred prokaryotic cell is a bacterial cell, e.g. *E. coli*.

Other objects and advantages of the present invention will become apparent to those skilled in the art from a study of the following description of the invention and non-limiting examples.

DEFINITIONS

5 For clarity, certain terms used in the specification are defined and presented as follows:

Co-factor: natural reactant, such as an organic molecule or a metal ion, required in an enzyme-catalyzed reaction. A co-factor is e.g. NAD(P), riboflavin (including FAD and FMN), folate, molybdopterin, thiamin, biotin, lipoic acid, pantothenic acid and coenzyme A, S-adenosylmethionine, pyridoxal phosphate, ubiquinone, menaquinone. Optionally, a co-factor
10 can be regenerated and reused.

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

Enzyme activity: means herein the ability of an enzyme to catalyze the conversion of a
20 substrate into a product. A substrate for the enzyme comprises the natural substrate of the enzyme but also comprises analogues of the natural substrate, which can also be converted, by the enzyme into a product or into an analogue of a product. The activity of the enzyme is measured for example by determining the amount of product in the reaction after a certain period of time, or by determining the amount of substrate remaining in the reaction mixture
25 after a certain period of time. The activity of the enzyme is also measured by determining the amount of an unused co-factor of the reaction remaining in the reaction mixture after a certain period of time or by determining the amount of used co-factor in the reaction mixture after a certain period of time. The activity of the enzyme is also measured by determining the amount of a donor of free energy or energy-rich molecule (e.g. ATP, phosphoenolpyruvate, acetyl phosphate or phosphocreatine) remaining in the reaction mixture after a certain period
30 of time or by determining the amount of a used donor of free energy or energy-rich molecule (e.g. ADP, pyruvate, acetate or creatine) in the reaction mixture after a certain period of time.

“ET6497 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:2, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:1 or is substantially similar to SEQ ID NO:1.

5 “GT1773 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:4, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:3 or is substantially similar to SEQ ID NO:3.

10 “GT0992 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:6, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:5 or is substantially similar to SEQ ID NO:5.

15 “ET6233 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:8, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:7 or is substantially similar to SEQ ID NO:7.

20 “ET0763 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:10, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:9 or is substantially similar to SEQ ID NO:9.

“ET5848 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:12, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:11 or is substantially similar to SEQ ID NO:11.

25 “GT5062 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:14, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:13 or is substantially similar to SEQ ID NO:13.

30 “ET5036 Gene” as used herein refers to a DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:16, or a nucleotide sequence substantially similar thereto. Preferably, the nucleotide sequence is set forth in SEQ ID NO:15 or is substantially similar to SEQ ID NO:15.

Herbicide: a chemical substance used to kill or suppress the growth of plants, plant cells, plant seeds, or plant tissues.

Heterologous DNA Sequence: a DNA sequence not naturally associated with a host cell into which it is introduced, including non-naturally occurring multiple copies of a naturally occurring DNA sequence; and genetic constructs wherein an otherwise homologous DNA sequence is operatively linked to a non-native sequence.

Homologous DNA Sequence: a DNA sequence naturally associated with a host cell into which it is introduced.

Inhibitor: a chemical substance that causes abnormal growth, e.g., by inactivating the enzymatic activity of a protein such as a biosynthetic enzyme, receptor, signal transduction protein, structural gene product, or transport protein that is essential to the growth or survival of the plant. In the context of the instant invention, an inhibitor is a chemical substance that alters the enzymatic activity encoded by the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene from a plant. More generally, an inhibitor causes abnormal growth of a host cell by interacting with the gene product encoded by the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene.

Isogenic: plants which are genetically identical, except that they may differ by the presence or absence of a heterologous DNA sequence.

Isolated: in the context of the present invention, an isolated DNA molecule or an isolated enzyme is a DNA molecule or enzyme that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or enzyme may exist in a purified form or may exist in a non-native environment such as, for example, in a transgenic host cell.

Mature protein: protein which is normally targeted to a cellular organelle, such as a chloroplast, and from which the transit peptide has been removed.

Minimal Promoter: promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation. In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription.

Modified Enzyme Activity: enzyme activity different from that which naturally occurs in a plant (i.e. enzyme activity that occurs naturally in the absence of direct or indirect manipulation of such activity by man), which is tolerant to inhibitors that inhibit the naturally occurring enzyme activity.

Pre-protein: protein which is normally targeted to a cellular organelle, such as a chloroplast, and still comprising its transit peptide.

Significant Increase: an increase in enzymatic activity that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater of the activity of the wild-type enzyme in the presence of the inhibitor, more preferably an increase by about 5-fold or greater, and most preferably an increase by about 10-fold or greater.

Significantly less: means that the amount of a product of an enzymatic reaction is reduced by more than the margin of error inherent in the measurement technique, preferably a decrease by about 2-fold or greater of the activity of the wild-type enzyme in the absence of the inhibitor, more preferably an decrease by about 5-fold or greater, and most preferably an decrease by about 10-fold or greater.

In its broadest sense, the term "substantially similar", when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence, wherein the corresponding sequence encodes a polypeptide having substantially the same structure and function as the polypeptide encoded by the reference nucleotide sequence. Desirably the substantially similar nucleotide sequence encodes the polypeptide encoded by the reference nucleotide sequence. The term "substantially similar" is specifically intended to include nucleotide sequences wherein the sequence has been modified to optimize expression in particular cells. In the context of the "ET6497 gene", "substantially similar" refers to nucleotide sequences that encode a protein at least 60% identical, still more preferably at least 70% identical, still more preferably at least 80% identical, still more preferably at least 90% identical, still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:2; in the context of the "GT1773 gene", "substantially similar" refers to nucleotide sequences that encode a protein at least 46% identical, more preferably at least 55% identical, still more preferably at least 65% identical, still more preferably at least 75% identical, still more preferably at least 85% identical, still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:4; in the context of the "GT0992 gene", "substantially similar" refers to nucleotide sequences that encode a protein at least 43% identical, more preferably at least 55% identical, more preferably at least 65% identical, still more preferably at least 75% identical, still more preferably at least 85% identical, still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:6; in the context of the "ET6233 gene", "substantially similar" refers

to nucleotide sequences that encode a protein at least 38% identical, still more preferably at least 45% identical, still more preferably at least 55% identical, yet still more preferably at least 65% identical, still more preferably at least 75% identical, still more preferably at least 85% identical, still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:8; in the context of the "ET0763 gene", "substantially similar" refers to nucleotide sequences that encode a protein at least 34% identical, still more preferably at least 44% identical, still more preferably at least 55% identical, still more preferably at least 65% identical, still more preferably at least 75% identical, yet still more preferably at least 85% identical, still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:10; in the context of the "ET5848 gene", "substantially similar" refers to nucleotide sequences that encode a protein at least 52% identical, more preferably at least 65% identical, still more preferably at least 75% identical, still more preferably at least 85% identical, still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:12; in the context of the "GT5062 gene", "substantially similar" refers to nucleotide sequences that encode a protein at least 80% identical, more preferably at least 90% identical, more preferably at least 92% identical, still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:14; in the context of the "ET5036 gene", "substantially similar" refers to nucleotide sequences that encode a protein at least 50% identical, still more preferably at least 60% identical, still more preferably at least 65% identical, yet still more preferably at least 75% identical, yet still more preferably at least 85% identical, yet still more preferably at least 95% identical, yet still more preferably at least 99% identical to SEQ ID NO:16, wherein said protein sequence comparisons are conducted using GAP analysis as described below. A nucleotide sequence "substantially similar" to the reference nucleotide sequence hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

“Homologs of the ET6497 gene” include nucleotide sequences that encode an amino acid sequence that is at least 30% identical to SEQ ID NO:2, more preferably at least 34% identical, yet still more preferably at least 39% identical, yet still more preferably at least 43% identical to SEQ ID NO:2, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the ET6497 protein.

“Homologs of the GT1773 gene” include nucleotide sequences that encode an amino acid sequence that is at least 26% identical to SEQ ID NO:4, more preferably at least 32% identical, still more preferably at least 38% identical, still more preferably at least 43% identical to SEQ ID NO:4, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the GT1773 protein.

“Homologs of the GT0992 gene” include nucleotide sequences that encode an amino acid sequence that is at least 29% identical to SEQ ID NO:6, still more preferably at least 33% identical, yet still more preferably at least 37% identical, yet still more preferably at least 38% identical, yet still more preferably at least 39% identical to SEQ ID NO:6, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the GT0992 protein.

“Homologs of the ET6233 gene” include nucleotide sequences that encode an amino acid sequence that is at least 27% identical to SEQ ID NO:8, still more preferably at least 34% identical to SEQ ID NO:8, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the ET6233 protein.

“Homologs of the ET0763 gene” include nucleotide sequences that encode an amino acid sequence that is at least 24% identical to SEQ ID NO:10, more preferably at least 25% identical, yet still more preferably at least 26% identical to SEQ ID NO:10, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the ET0763 protein.

“Homologs of the ET5848 gene” include nucleotide sequences that encode an amino acid sequence that is at least 25% identical to SEQ ID NO:12, more preferably at least 29% identical, still more preferably at least 49% identical to SEQ ID NO:12, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the ET5848 protein.

“Homologs of the GT5062 gene” include nucleotide sequences that encode an amino acid sequence that is at least 29% identical to SEQ ID NO:14, more preferably at least 31% identical, yet still more preferably at least 41 % identical, yet still more preferably at least 54% identical to SEQ ID NO:14, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the GT5062 protein.

“Homologs of the ET5036 gene” include nucleotide sequences that encode an amino acid sequence that is at least 50% identical to SEQ ID NO:16, more preferably at least 60% identical, still more preferably at least 70% identical to SEQ ID NO:16, as measured, using the GAP parameters described below, wherein the amino acid sequence encoded by the homolog has the biological activity of the ET5036 protein.

The term "substantially similar", when used herein with respect to a protein, means a protein corresponding to a reference protein, wherein the protein has substantially the same structure and function as the reference protein, e.g. where only changes in amino acids sequence not affecting the polypeptide function occur. When used in the context of the "ET6497 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case SEQ ID NO:2) is at least 60%, more preferably at least 70%, still more preferably at least 80%, still more preferably at least 90%, still more preferably at least 95%, yet still more preferably at least 99%, as determined using default GAP analysis parameters with the University of Wisconsin GCG, SEQWEB application of GAP, based on the algorithm of Needleman and Wunsch (Needleman and Wunsch (1970) J Mol. Biol. 48: 443-453). In the context of the "GT1773 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case SEQ ID NO:4) is at least 46%, more preferably at least 55%, still more preferably at least 65%, still more preferably at least 75%, still more preferably at least 85%, still more preferably at least 95%, yet still more preferably at least 99%. In the context of the "GT0992 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case SEQ ID NO:6) is at least 43%, more preferably at least 55%, more preferably at least 65%, still more preferably at least 75%, still more preferably at least 85%, still more preferably at least 95%, yet still more preferably at least 99%. In the context of the "ET6233 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case

SEQ ID NO:8) is at least 38%, more preferably at least 45%, more preferably at least 55%, yet still more preferably at least 65%, yet still more preferably at least 75%, yet still more preferably at least 85%, yet still more preferably at least 95%, yet still more preferably at least 99%. In the context of the "ET0763 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case SEQ ID NO:10) is at least 34%, more preferably at least 44%, more preferably at least 55%, yet still more preferably at least 65%, still more preferably at least 75%, still more preferably at least 85%, still more preferably at least 95%, yet still more preferably at least 99%. In the context of the "ET5848 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case SEQ ID NO:12) is at least 52%, more preferably at least 65%, still more preferably at least 75%, still more preferably at least 85%, still more preferably at least 95%, yet still more preferably at least 99%. In the context of the "GT5062 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case SEQ ID NO:14) is at least 80%, more preferably at least 90%, still more preferably at least 92%, still more preferably at least 95%, yet still more preferably at least 99%. In the context of the "ET5036 gene", the percentage of identity between the substantially similar protein or amino acid sequence and the reference protein or amino acid sequence (in this case SEQ ID NO:16) is at least 50%, more preferably at least 60%, still more preferably at least 65%, still more preferably at least 75%, yet still more preferably at least 85%, yet still more preferably at least 95%, yet still more preferably at least 99% .

As used herein the term "ET6497 protein" refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:1.

"Homologs of the ET6497 protein" are amino acid sequences that are at least 30% identical to SEQ ID NO:2, more preferably at least 34% identical, yet still more preferably at least 39%, yet still more preferably at least 43% identical to SEQ ID NO:2, as measured using the GAP parameters described above, wherein the homologs of the ET6497 protein have the biological activity of the ET6497 protein.

As used herein the term "GT1773 protein" refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:3. "Homologs of the GT1773 protein" are amino acid sequences that are at least 26% identical to SEQ ID NO:4, more preferably at least 32% identical, still more preferably at least 38%

identical, yet still more preferably at least 43% identical to SEQ ID NO:4, as measured using the GAP parameters described above, wherein the homologs of the GT1773 protein have the biological activity of the GT1773 protein.

As used herein the term "GT0992 protein" refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:5.
5 "Homologs of the GT0992 protein" are amino acid sequences that are at least 29% identical to SEQ ID NO:6, still more preferably at least 33% identical, yet still more preferably at least 37% identical, yet still more preferably at least 38% identical, yet still more preferably at least 39% identical to SEQ ID NO:6, as measured using the GAP parameters described above,
10 wherein the homologs of the GT0992 protein have the biological activity of the GT0992 protein.

As used herein the term "ET6233 protein" refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:7.
"Homologs of the ET6233 protein" are amino acid sequences that are at least 27% identical to
15 SEQ ID NO:8, still more preferably at least 34% identical to SEQ ID NO:8, as measured using the GAP parameters described above, wherein the homologs of the ET6233 protein have the biological activity of the ET6233 protein.

As used herein the term "ET0763 protein" refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:9.
20 "Homologs of the ET0763 protein" are amino acid sequences that are at least 24% identical to SEQ ID NO:10, still more preferably at least 25% identical, yet still more preferably at least 26% identical to SEQ ID NO:10, as measured using the GAP parameters described above, wherein the homologs of the ET0763 protein have the biological activity of the ET0763 protein.

As used herein the term "ET5848 protein" refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:11.
"Homologs of the ET5848 protein" are amino acid sequences that are at least 25% identical to
25 SEQ ID NO:12, still more preferably at least 29% identical, yet still more preferably at least 49% identical to SEQ ID NO:12, as measured using the GAP parameters described above,
30 wherein the homologs of the ET5848 protein have the biological activity of the ET5848 protein.

As used herein the term "GT5062 protein" refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:13.

“Homologs of the GT5062 protein” are amino acid sequences that are at least 29% identical to SEQ ID NO:14, still more preferably at least 31% identical, yet still more preferably at least 41% identical, yet still more preferably at least 54% identical to SEQ ID NO:14, as measured using the GAP parameters described above, wherein the homologs of the GT5062 protein
5 have the biological activity of the GT5062 protein.

As used herein the term “ET5036 protein” refers to an amino acid sequence encoded by a DNA molecule comprising a nucleotide sequence substantially similar to SEQ ID NO:15. “Homologs of the ET5036 protein” are amino acid sequences that are at least 50% identical to SEQ ID NO:16, still more preferably at least 60% identical, yet still more preferably at least
10 70% identical to SEQ ID NO:16, as measured using the GAP parameters described above, wherein the homologs of the ET5036 protein have the biological activity of the ET5036 protein.

Substrate: a substrate is the molecule that an enzyme naturally recognizes and converts to a product in the biochemical pathway in which the enzyme naturally carries out its function, or
15 is a modified version of the molecule, which is also recognized by the enzyme and is converted by the enzyme to a product in an enzymatic reaction similar to the naturally-occurring reaction.

Tolerance: the ability to continue essentially normal growth or function (i.e. no more than 5% of herbicide tolerant plants show phytotoxicity) when exposed to an inhibitor or herbicide in
20 an amount sufficient to suppress the normal growth or function of native, unmodified plants.

Transformation: a process for introducing heterologous DNA into a cell, tissue, or plant. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof.

Transgenic: stably transformed with a recombinant DNA molecule that preferably comprises
25 a suitable promoter operatively linked to a DNA sequence of interest.

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

- SEQ ID NO:1 cDNA coding sequence for the *Arabidopsis* ET6497 gene
- SEQ ID NO:2 amino acid sequence encoded by the *Arabidopsis* ET6497 nucleotide
sequence shown in SEQ ID NO:1
- SEQ ID NO:3 cDNA coding sequence for the *Arabidopsis* GT1773 gene
- 5 SEQ ID NO:4 amino acid sequence encoded by the *Arabidopsis* GT1773 nucleotide
sequence shown in SEQ ID NO:3
- SEQ ID NO:5 cDNA coding sequence for the *Arabidopsis* GT0992 gene
- SEQ ID NO:6 amino acid sequence encoded by the *Arabidopsis* GT0992 nucleotide
sequence shown in SEQ ID NO:5
- 10 SEQ ID NO:7 cDNA coding sequence for the *Arabidopsis* ET6233 gene
- SEQ ID NO:8 amino acid sequence encoded by the *Arabidopsis* ET6233 nucleotide
sequence shown in SEQ ID NO:7
- SEQ ID NO:9 cDNA coding sequence for the *Arabidopsis* ET0763 gene
- SEQ ID NO:10 amino acid sequence encoded by the *Arabidopsis* ET0763 nucleotide
15 sequence shown in SEQ ID NO:9
- SEQ ID NO:11 cDNA coding sequence for the *Arabidopsis* ET5848 gene
- SEQ ID NO:12 amino acid sequence encoded by the *Arabidopsis* ET5848 nucleotide
sequence shown in SEQ ID NO:11
- SEQ ID NO:13 cDNA coding sequence for the *Arabidopsis* GT5062 gene
- 20 SEQ ID NO:14 amino acid sequence encoded by the *Arabidopsis* GT5062 nucleotide
sequence shown in SEQ ID NO:13
- SEQ ID NO:15 cDNA coding sequence for the *Arabidopsis* ET5036 gene
- SEQ ID NO:16 amino acid sequence encoded by the *Arabidopsis* ET5036 nucleotide
sequence shown in SEQ ID NO:15
- 25 SEQ ID NO:17 genomic sequence of the *Arabidopsis* ET6497 gene
- SEQ ID NO:18 genomic sequence of the *Arabidopsis* GT1773 gene
- SEQ ID NO:19 genomic sequence of the *Arabidopsis* GT0992 gene
- SEQ ID NO:20 genomic sequence of the *Arabidopsis* ET6233 gene
- SEQ ID NO:21 genomic sequence of the *Arabidopsis* ET0763 gene
- 30 SEQ ID NO:22 genomic sequence of the *Arabidopsis* ET5848 gene
- SEQ ID NO:23 genomic sequence of the *Arabidopsis* GT5062 gene
- SEQ ID NO:24 genomic sequence of the *Arabidopsis* ET5036 gene
- SEQ ID NO:25 oligonucleotide LWAD1

- SEQ ID NO:26 oligonucleotide CA51
 SEQ ID NO:27 oligonucleotide CA52
 SEQ ID NO:28 oligonucleotide CA53
 SEQ ID NO:29 oligonucleotide CA54
 5 SEQ ID NO:30 oligonucleotide CA55
 SEQ ID NO:31 oligonucleotide 5A
 SEQ ID NO:32 oligonucleotide 5B
 SEQ ID NO:33 oligonucleotide 5C
 SEQ ID NO:34 oligonucleotide 3A
 10 SEQ ID NO:35 oligonucleotide 3B
 SEQ ID NO:36 oligonucleotide 3C
 SEQ ID NO:37 second cDNA coding sequence for the *Arabidopsis* GT1773 gene
 SEQ ID NO:38 amino acid sequence encoded by the *Arabidopsis* GT1773 nucleotide
 sequence shown in SEQ ID NO:37

15

I. Essentiality of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062,
 and ET5036 Genes in *Arabidopsis* Demonstrated by Ac/Ds Transposon Mutagenesis

As shown in the examples below, the identification of a novel gene structure, as well as the
 essentiality of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or
 20 ET5036 genes for normal plant growth and development, have been demonstrated for the first
 time in *Arabidopsis* using Ac/Ds transposon mutagenesis. Having established the essentiality
 of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, and ET5036 functions in
 plants and having identified the genes encoding these essential activities, the inventors
 thereby provide an important and sought after tool for new herbicide development.

25 *Arabidopsis* insertional mutant lines segregating for seedling lethal mutations are identified as
 a first step in the identification of essential proteins. Ds transposon insertion lines were
 produced as described in Sundareson et al. (1995) *Genes and Dev.*, 9:1797-1810),
 incorporated herein by reference. Starting with F3 or F4 seeds collected from single F2 or F3
 kanamycin-resistant plants containing Ds insertions in their genomes (see Figure 3 of
 30 Sundareson et al. (1995) *Genes and Dev.*, 9:1797-1810), those lines segregating homozygous
 seedling lethal seedlings are identified. These lines are found by placing seeds onto minimal
 plant growth media, which contains the fungicides benomyl and maxim, and screening for
 inviable seedlings after 7 and 14 days in the light at room temperature. Inviably phenotypes

include altered pigmentation or altered morphology. These phenotypes are observed either on plates directly or in soil following transplantation of seedlings.

When a line is identified as segregating a seedling lethal, it is determined if the resistance marker in the Ds transposon insertion co-segregates with the lethality (Errampalli et al. (1991) *The Plant Cell*, 3:149-157). Co-segregation analysis is done by placing the seeds on media containing the selective agent and scoring the seedlings for resistance or sensitivity to the agent. Examples of selective agents used are kanamycin, hygromycin, or phosphinothricin. About 35 resistant seedlings are transplanted to soil and their progeny are examined for the segregation of the seedling lethal. In the case in which the Ds transposon insertion disrupts an essential gene, there is co-segregation of the resistance phenotype and the seedling lethal phenotype in every plant. Therefore, in such a case, all resistant plants segregate seedling lethals in the next generation; this result indicates that each of the resistant plants is heterozygous for the DNA causing both phenotypes.

For the *Arabidopsis* lines showing co-segregation of the transposon-encoded resistance marker and the lethal phenotype, PCR-based molecular approaches such as, TAIL-PCR (Liu et al. (1995) *The Plant Journal*, 8:457-463; Liu and Whittier (1995), *Genomics*, 25: 674-681), vectorette PCR (Riley et al. (1990) *Nucleic Acids Research*, 18: 2887-2890), or the GenomeWalker™ kit (CLONTECH Laboratories, Inc., Palo Alto, CA), may be used to directly amplify the plant DNA fragments flanking the transposon. Each of these techniques utilizes the known sequence of the transposon, and can be used to recover small (less than 5 KB) fragments directly adjacent to the insertion. PCR products are isolated and their DNA sequence is determined. The resulting sequences are analyzed for the presence of non-Ds transposon vector sequences. When such sequences are found, they are used to search DNA and protein databases using the BLAST and BLAST2 programs (Altschul et al. (1990) *J Mol. Biol.* 215: 403-410; Altschul et al (1997) *Nucleic Acid Res.* 25:3389-3402, both incorporated herein by reference). Additional genomic and cDNA sequences for each gene are identified by standard molecular biology procedures.

II. Sequence of the *Arabidopsis* ET6497 Gene

The *Arabidopsis* ET6497 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line # ET6497. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence

(chromosome 3, clone T20O10, GenBank accession # AL163816). The inventors are the first to demonstrate that the ET6497 gene product is essential for normal growth and development in plants, as well as defining the function of the ET6497 gene through protein homology. The present invention discloses the cDNA coding nucleotide sequence of the *Arabidopsis* ET6497 gene as well as the amino acid sequence of the *Arabidopsis* ET6497 protein.

The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:1, wherein said amino acid sequence has ET6497 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the ET6497 gene shows similarity to translation release factor 1 (R-F1), from *Caenorhabditis elegans* (GenPept Accession # AAB94190), *Schizosaccharomyces pombe* (GenPept Accession # CAA90504), *Thermus thermophilus* (GenPept Accession # BAA13349.1), *Escherichia coli* (GenPept Accession # AAC43437), *Neisseria meningitidis* (GenPept Accession # AAF42034), *Synechocystis sps* (GenPept Accession # BAA18826).

III. Sequence of the *Arabidopsis* GT1773 Gene

The *Arabidopsis* GT1773 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line # GT1773. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence (chromosome 3, clone T21L8, GenBank accession number AL096860). The inventors are the first to demonstrate that the GT1773 gene product is essential for normal growth and development in plants, as well as defining the function of the GT1773 gene through protein homology. The present invention discloses cDNA coding nucleotide sequences of the *Arabidopsis* GT1773 gene as well as the amino acid sequence of the *Arabidopsis* GT1773 protein.

The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:3, wherein said amino acid sequence has GT1773 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the GT1773 gene shows similarity to diaminohydroxyphosphoribosylaminopyrimidine deaminases, 5-amino-6-(5-phosphoribosylamino) uracil reductases, and a hypothetical protein from *Arabidopsis thaliana* (GenPept Accession # CAB79096), *Methanococcus jannaschii* (GenPept Accession #

AAB98665), *Escherichia coli* (GenPept Accession # AAB40170 and SWISS PROT Accession # P30176), *Thermotoga maritima* (GenPept Accession # AAD36891), *Streptomyces coelicolor* (GenPept Accession # CAB92254), and *Synechocystis* sps (GenPept Accession # BAA10295).

5

IV. Sequence of the *Arabidopsis* GT0992 Gene

The *Arabidopsis* GT0992 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line #GT0992. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence (chromosome 5, clone MXC9, GenBank accession number AB007727). The inventors are the first to demonstrate that the GT0992 gene product is essential for normal growth and development in plants, as well as defining the function of the GT0992 gene through protein homology. The present invention discloses the cDNA coding nucleotide sequence of the *Arabidopsis* GT0992 gene as well as the amino acid sequence of the *Arabidopsis* GT0992 protein.

10

The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:5, wherein said amino acid sequence has GT0992 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the GT0992 gene shows similarity to proteins of unknown function, which confer tellurium resistance and may function as transmembrane transport proteins, from *Rickettsia prowazekii* (GenPept Accession # CAA15212), *Vibrio cholerae* (GenPept Accession # AAF96448), *Escherichia coli* (GenPept Accession # AAA57890), *Streptomyces coelicolor* (GenPept Accession # CAB38153), *Oryza sativa* (GenPept Accession # BAA90517), and *Neisseria meningitidis* (GenPept Accession # AAF40491).

15

20

25

V. Sequence of the *Arabidopsis* ET6233 Gene

The *Arabidopsis* ET6233 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line # ET6233. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence (chromosome 1, clone F14I3, GenBank accession number AC007980). The inventors are the first to demonstrate that the ET6233 gene product is essential for normal growth and development in plants, as well as defining the function of the ET6233 gene through protein

30

homology. The present invention discloses the cDNA coding nucleotide sequence of the *Arabidopsis* ET6233 gene as well as the amino acid sequence of the *Arabidopsis* ET6233 protein.

The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:7, wherein said amino acid sequence has ET6233 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the ET6233 gene shows similarity to the *Bacillus megaterium* cbiX protein, which is required for cohyric acid biosynthesis (GenPept Accession # CAA04308), and other related proteins from *Methanococcus jannaschii* (GenPept Accession # AAB98975) and *Bacillus halodurans* (GenPept Accession # BAB05215.1).

VI. Sequence of the *Arabidopsis* ET0763 Gene

The *Arabidopsis* ET0763 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line # ET0763. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence (chromosome 1, clone F14O23, GenBank accession number AC012654). The inventors are the first to demonstrate that the ET0763 gene product is essential for normal growth and development in plants, as well as defining the function of the ET0763 gene through protein homology. The present invention discloses the cDNA coding nucleotide sequence of the *Arabidopsis* ET0763 gene as well as the amino acid sequence of the *Arabidopsis* ET0763 protein.

The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:9, wherein said amino acid sequence has ET0763 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the ET0763 gene shows similarity to 30S ribosomal protein S1 from *Escherichia coli* (GenPept Accession # CAA23644), *Streptomyces coelicolor* (GenPept Accession # CAB52054), *Spinacia oleracea* (GenPept Accession # AAA34045), *Porphyra purpurea* (GenPept Accession # AAC08231), *Mycobacterium tuberculosis* (GenPept Accession # CAB08883), and *Synechococcus* sps (GenPept Accession # BAA05946).

VII. Sequence of the *Arabidopsis* ET5848 Gene

The *Arabidopsis* ET5848 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line # ET5848. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence (chromosome 1, clone F17O7, GenBank accession number AC003671). The inventors are the first to demonstrate that the ET5848 gene product is essential for normal growth and development in plants, as well as defining the function of the ET5848 gene through protein homology. The present invention discloses the cDNA coding nucleotide sequence of the *Arabidopsis* ET5848 gene as well as the amino acid sequence of the *Arabidopsis* ET5848 protein.

10 The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:11, wherein said amino acid sequence has ET5848 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the ET5848 gene shows similarity to proteins of unknown function from *Arabidopsis thaliana* (GenPept Accession numbers AAF27677, CAB61960, AAF25966, AAC18801, AAD55510; GenBank accession number AC016162.3).

VIII. Sequence of the *Arabidopsis* GT5062 Gene

The *Arabidopsis* GT5062 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line # GT5062. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence (chromosome 3, clone F5K20, GenBank accession number AL132960). The inventors are the first to demonstrate that the GT5062 gene product is essential for normal growth and development in plants, as well as defining the function of the GT5062 gene through protein homology. The present invention discloses the cDNA coding nucleotide sequence of the *Arabidopsis* GT5062 gene as well as the amino acid sequence of the *Arabidopsis* GT5062 protein.

The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:13, wherein said amino acid sequence has GT5062 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the GT5062 gene shows similarity to uracil phosphoribosyltransferase (also known as UMP pyrophosphorylase) from

Schizosaccharomyces pombe (GenPept Accession # CAB11230), *Methanobacterium thermoautotrophicum* (GenPept Accession # AAB85603), *Thermotoga maritima* (GenPept Accession # AAD35803), *Escherichia coli* (GenPept Accession # BAA16386), *Synechocystis* sps (GenPept Accession # BAA16768), and *Nicotiana tabacum* (SWISS PROT Accession # P93394).

IX. Sequence of the *Arabidopsis* ET5036 Gene

The *Arabidopsis* ET5036 gene is identified by isolating DNA flanking the Ds transposon border from the tagged seedling-lethal line # GT5062. A region of the *Arabidopsis* DNA flanking the Ds transposon border corresponds to *Arabidopsis* genomic sequence (chromosome 1, clone T10024, GenBank accession number AC007067). The inventors are the first to demonstrate that the GT5062 gene product is essential for normal growth and development in plants, as well as defining the function of the GT5062 gene through protein homology. The present invention discloses the cDNA coding nucleotide sequence of the *Arabidopsis* GT5062 gene as well as the amino acid sequence of the *Arabidopsis* GT5062 protein.

The present invention also encompasses an isolated amino acid sequence derived from a plant, wherein said amino acid sequence is identical or substantially similar to the amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO:15, wherein said amino acid sequence has ET5036 activity. Using BLAST and BLAST2 programs with the default settings, the sequence of the ET5036 gene does not show similarity to any genes in public sequence databases.

X. Recombinant Production of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 Activities and Uses Thereof

For recombinant production of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activities in a host organism, a nucleotide sequence encoding a protein having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, is inserted into an expression cassette designed for the chosen host and introduced into the host where it is recombinantly produced. For example, SEQ ID NO:1, nucleotide sequences substantially similar to SEQ ID NO:1, or homologs of the ET6497 gene are used for the recombinant production of a protein having ET6497 activity. For example, SEQ ID NO:3, nucleotide sequences substantially similar to SEQ ID NO:3, or homologs of

the GT1773 gene are used for the recombinant production of a protein having GT1773 activity. For example, SEQ ID NO:5, nucleotide sequences substantially similar to SEQ ID NO:5, or homologs of the GT0992 gene are used for the recombinant production of a protein having GT0992 activity. For example, SEQ ID NO:7, nucleotide sequences substantially similar to SEQ ID NO:7, or homologs of the ET6233 gene are used for the recombinant production of a protein having ET6233 activity. For example, SEQ ID NO:9, nucleotide sequences substantially similar to SEQ ID NO:9, or homologs of the ET0763 gene are used for the recombinant production of a protein having ET0763 activity. For example, SEQ ID NO:11, nucleotide sequences substantially similar to SEQ ID NO:11, or homologs of the ET5848 gene are used for the recombinant production of a protein having ET5848 activity. For example, SEQ ID NO:13, nucleotide sequences substantially similar to SEQ ID NO:13, or homologs of the GT5062 gene are used for the recombinant production of a protein having GT5062 activity. For example, SEQ ID NO:15, nucleotide sequences substantially similar to SEQ ID NO:15, or homologs of the ET5036 gene are used for the recombinant production of a protein having ET5036 activity. The choice of specific regulatory sequences such as promoter, signal sequence, 5' and 3' untranslated sequences, and enhancer appropriate for the chosen host is within the level of skill of the routineer in the art. The resultant molecule, containing the individual elements operably linked in proper reading frame, may be inserted into a vector capable of being transformed into the host cell. Suitable expression vectors and methods for recombinant production of proteins are well known for host organisms such as *E. coli*, yeast, and insect cells (see, e.g., Luckow and Summers, *Bio/Technol.* 6: 47 (1988), and baculovirus expression vectors, e.g., those derived from the genome of *Autographica californica* nuclear polyhedrosis virus (AcMNPV). A preferred baculovirus/insect system is pAcHLT (Pharmlngen, San Diego, CA) used to transfect *Spodoptera frugiperda* Sf9 cells (ATCC) in the presence of linear *Autographa californica* baculovirus DNA (Pharmlngen, San Diego, CA). The resulting virus is used to infect HighFive *Tricoplusia ni* cells (Invitrogen, La Jolla, CA).

In a preferred embodiment, the nucleotide sequence encoding a protein having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity is derived from an eukaryote, such as a mammal, a fly or a yeast, but is preferably derived from a plant, preferably a monocotyledonous or a dicotyledonous plant. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:1, or encodes a protein having ET6497 activity, whose

amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:2. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:3, or encodes a protein having GT1773 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:4. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:5, or encodes a protein having GT0992 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:6. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:7, or encodes a protein having ET6233 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:8. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:9, or encodes a protein having ET0763 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:10. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:11, or encodes a protein having ET5848 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:12. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:13, or encodes a protein having GT5062 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:14. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:15, or encodes a protein having ET5036 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:16. In another preferred embodiment, the nucleotide sequence encoding a protein having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, is derived from a prokaryote. Recombinantly produced protein having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity is isolated and purified using a variety of standard techniques. The actual techniques that may be used will vary depending upon the host organism used, whether the protein is designed for secretion, and other such factors

familiar to the skilled artisan (*see, e.g.* chapter 16 of Ausubel, F. *et al.*, "Current Protocols in Molecular Biology", pub. by John Wiley & Sons, Inc. (1994).

Assays Utilizing the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or
5 ET5036 Proteins

Recombinantly produced proteins having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity are useful for a variety of purposes. For example, they can be used in *in vitro* assays to screen known herbicidal chemicals whose target has not been identified to determine if they inhibit ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, 10 GT5062, or ET5036, respectively. Such *in vitro* assays may also be used as more general screens to identify chemicals that inhibit such enzymatic activity and that are therefore novel herbicide candidates. Alternatively, recombinantly produced proteins having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity may be used to elucidate the complex structure of these molecules and to further characterize their association 15 with known inhibitors in order to rationally design new inhibitory herbicides as well as herbicide tolerant forms of the enzymes.

In Vitro Inhibitor Assays: Discovery of Small Molecule Ligand that Interacts with the Gene Product of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ
20 ID NO:11, SEQ ID NO:13, or SEQ ID NO:15

Once a protein has been identified as a potential herbicide target, the next step is to develop an assay that allows screening large number of chemicals to determine which ones interact with the protein. Although it is straightforward to develop assays for proteins of known function, developing assays with proteins of unknown functions is more difficult. This 25 difficulty can be overcome by using technologies that can detect interactions between a protein and a compound without knowing the biological function of the protein. A short description of three methods is presented, including fluorescence correlation spectroscopy, surface-enhanced laser desorption/ionization, and biacore technologies.

Fluorescence Correlation Spectroscopy (FCS) theory was developed in 1972 but it is only in 30 recent years that the technology to perform FCS became available (Madge et al. (1972) Phys. Rev. Lett., 29: 705-708; Maiti et al. (1997) Proc. Natl. Acad. Sci. USA, 94: 11753-11757). FCS measures the average diffusion rate of a fluorescent molecule within a small sample

volume. The sample size can be as low as 10^3 fluorescent molecules and the sample volume as low as the cytoplasm of a single bacterium. The diffusion rate is a function of the mass of the molecule and decreases as the mass increases. FCS can therefore be applied to protein-ligand interaction analysis by measuring the change in mass and therefore in diffusion rate of a molecule upon binding. In a typical experiment, the target to be analyzed is expressed as a recombinant protein with a sequence tag, such as a poly-histidine sequence, inserted at the N or C-terminus. The expression takes place in *E. coli*, yeast or insect cells. The protein is purified by chromatography. For example, the poly-histidine tag can be used to bind the expressed protein to a metal chelate column such as Ni^{2+} chelated on iminodiacetic acid agarose. The protein is then labeled with a fluorescent tag such as carboxytetramethylrhodamine or BODIPY® (Molecular Probes, Eugene, OR). The protein is then exposed in solution to the potential ligand, and its diffusion rate is determined by FCS using instrumentation available from Carl Zeiss, Inc. (Thornwood, NY). Ligand binding is determined by changes in the diffusion rate of the protein.

Surface-Enhanced Laser Desorption/Ionization (SELDI) was invented by Hutchens and Yip during the late 1980's (Hutchens and Yip (1993) Rapid Commun. Mass Spectrom. 7: 576-580). When coupled to a time-of-flight mass spectrometer (TOF), SELDI provides a mean to rapidly analyze molecules retained on a chip. It can be applied to ligand-protein interaction analysis by covalently binding the target protein on the chip and analyze by MS the small molecules that bind to this protein (Worrall et al. (1998) Anal. Biochem. 70: 750-756). In a typical experiment, the target to be analyzed is expressed as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the SELDI chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via, for example, a delivery system capable to pipet the ligands in a sequential manner (autosampler). The chip is then submitted to washes of increasing stringency, for example a series of washes with buffer solutions containing an increasing ionic strength. After each wash, the bound material is analyzed by submitting the chip to SELDI-TOF. Ligands that specifically bind the target will be identified by the stringency of the wash needed to elute them.

Biacore relies on changes in the refractive index at the surface layer upon binding of a ligand to a protein immobilized on the layer. In this system, a collection of small ligands is injected

sequentially in a 2-5 microlitre cell with the immobilized protein. Binding is detected by surface plasmon resonance (SPR) by recording laser light refracting from the surface. In general, the refractive index change for a given change of mass concentration at the surface layer, is practically the same for all proteins and peptides, allowing a single method to be applicable for any protein (Liedberg et al. (1983) *Sensors Actuators* 4: 299-304; Malmquist (1993) *Nature*, 361: 186-187). In a typical experiment, the target to be analyzed is expressed as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the Biacore chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via the delivery system incorporated in the instruments sold by Biacore (Uppsala, Sweden) to pipet the ligands in a sequential manner (autosampler). The SPR signal on the chip is recorded and changes in the refractive index indicate an interaction between the immobilized target and the ligand. Analysis of the signal kinetics on rate and off rate allows the discrimination between non-specific and specific interaction.

Also, an assay for small molecule ligands that interact with a polypeptide is an inhibitor assay. For example, such an inhibitor assay useful for identifying inhibitors of the products essential plant genes, such as ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes, comprises the steps of:

a) reacting an ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein, respectively, and a substrate thereof in the presence of a suspected inhibitor of the protein's respective function; b) comparing the rate of enzymatic activity of the protein in the presence of the suspected inhibitor to the rate of enzymatic activity under the same conditions in the absence of the suspected inhibitor; and c) determining whether the suspected inhibitor inhibits the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein, respectively.

For example, the inhibitory effect on ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, may be determined by a reduction or complete inhibition of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, in the assay. Such a determination may be made by comparing, in the presence and absence of the candidate inhibitor, the amount of substrate used or intermediate or product made during the reaction.

XI. *In Vivo* Inhibitor Assay

In one embodiment, a suspected herbicide, for example identified by *in vitro* screening, is applied to plants at various concentrations. The suspected herbicide is preferably sprayed on the plants. After application of the suspected herbicide, its effect on the plants, for example death or suppression of growth is recorded.

In another embodiment, an *in vivo* screening assay for inhibitors of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity uses transgenic plants, plant tissue, plant seeds or plant cells capable of overexpressing a nucleotide sequence having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, wherein the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene product is enzymatically active in the transgenic plants, plant tissue, plant seeds or plant cells. The nucleotide sequence is preferably derived from an eukaryote, such as a yeast, but is preferably derived from a plant. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:1, or encodes an enzyme having ET6497 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:2. In another preferred embodiment, the nucleotide sequence is derived from a prokaryote. In a further embodiment, the nucleotide sequence is derived from an eukaryote, such as a yeast, but is preferably derived from a plant. In a preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:3, or encodes an enzyme having GT1773 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:4. In another preferred embodiment, the nucleotide sequence is derived from a prokaryote. In a further preferred embodiment, the nucleotide sequence is derived from an eukaryote, such as a yeast, but is preferably derived from a plant. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:5, or encodes an enzyme having GT0992 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:6. In another preferred embodiment, the nucleotide sequence is derived from a prokaryote. In a further preferred embodiment, the nucleotide sequence is derived from an eukaryote, such as a yeast, but is preferably derived from a plant. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:7, or encodes an enzyme having ET6233 activity, whose amino acid sequence is identical

or substantially similar to the amino acid sequence set forth in SEQ ID NO:8. In another preferred embodiment, the nucleotide sequence is derived from a prokaryote. In a further preferred embodiment, the nucleotide sequence is derived from an eukaryote, such as a yeast, but is preferably derived from a plant. In a further preferred embodiment, the nucleotide
5 sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:9, or encodes an enzyme having ET0763 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:10. In another preferred embodiment, the nucleotide sequence is derived from a prokaryote. In a further preferred embodiment, the nucleotide sequence is derived from an eukaryote, such as a yeast,
10 but is preferably derived from a plant. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:11, or encodes an enzyme having ET5848 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:12. In another preferred embodiment, the nucleotide sequence is derived from a prokaryote. In a
15 further preferred embodiment, the nucleotide sequence is derived from an eukaryote, such as a yeast, but is preferably derived from a plant. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:13, or encodes an enzyme having GT5062 activity, whose amino acid sequence is identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:14. In
20 another preferred embodiment, the nucleotide sequence is derived from a prokaryote. In a further preferred embodiment, the nucleotide sequence is derived from an eukaryote, such as a yeast, but is preferably derived from a plant. In a further preferred embodiment, the nucleotide sequence is identical or substantially similar to the nucleotide sequence set forth in SEQ ID NO:15, or encodes an enzyme having ET5036 activity, whose amino acid sequence is
25 identical or substantially similar to the amino acid sequence set forth in SEQ ID NO:16. In another preferred embodiment, the nucleotide sequence is derived from a prokaryote.

A chemical is then applied to the transgenic plants, plant tissue, plant seeds or plant cells and to the isogenic non-transgenic plants, plant tissue, plant seeds or plant cells, and the growth or
30 viability of the transgenic and non-transformed plants, plant tissue, plant seeds or plant cells are determined after application of the chemical and compared. Compounds capable of inhibiting the growth of the non-transgenic plants, but not affecting the growth of the

transgenic plants are selected as specific inhibitors of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity.

XII. Herbicide Tolerant Plants

5 The present invention is further directed to plants, plant tissue, plant seeds, and plant cells tolerant to herbicides that inhibit the naturally occurring ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity in these plants, wherein the tolerance is conferred by an altered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively. Altered ET6497, GT1773, GT0992, ET6233, ET0763, 10 ET5848, GT5062, or ET5036 activity may be conferred upon a plant according to the invention by increasing expression of wild-type herbicide-sensitive ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene, respectively, for example by providing additional wild-type ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes and/or by overexpressing the endogenous ET6497, GT1773, 15 GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene, for example by driving expression with a strong promoter. Altered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity also may be accomplished by expressing nucleotide sequences that are substantially similar to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, or 20 homologs in a plant. Still further altered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity is conferred on a plant by expressing modified herbicide-tolerant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes, respectively, in the plant. Combinations of these techniques may also be used. Representative plants include any plants to which these herbicides are applied for their 25 normally intended purpose. Preferred are agronomically important crops such as cotton, soybean, oilseed rape, sugar beet, maize, rice, wheat, barley, oats, rye, sorghum, millet, turf, forage, turf grasses, and the like.

A. Increased Expression of Wild-Type ET6497, GT1773, GT0992, ET6233, 30 ET0763, ET5848, GT5062, or ET5036

Achieving altered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity through increased expression results in a level of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, in the plant cell at least

sufficient to overcome growth inhibition caused by the herbicide when applied in amounts sufficient to inhibit normal growth of control plants. The level of expressed enzyme generally is at least two times, preferably at least five times, and more preferably at least ten times the natively expressed amount. Increased expression may be due to multiple copies of a wild-type ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene; multiple occurrences of the coding sequence within the gene (*i.e.* gene amplification) or a mutation in the non-coding, regulatory sequence of the endogenous gene in the plant cell. Plants having such altered gene activity can be obtained by direct selection in plants by methods known in the art (see, *e.g.* U.S. Patent No. 5,162,602, and U.S. Patent No. 4,761,373, and references cited therein). These plants also may be obtained by genetic engineering techniques known in the art. Increased expression of a herbicide-sensitive ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene can also be accomplished by transforming a plant cell with a recombinant or chimeric DNA molecule comprising a promoter capable of driving expression of an associated structural gene in a plant cell operatively linked to a homologous or heterologous structural gene encoding the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein, respectively, or a homolog thereof. Preferably, the transformation is stable, thereby providing a heritable transgenic trait.

B. Expression of Modified Herbicide-Tolerant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 Proteins

According to this embodiment, plants, plant tissue, plant seeds, or plant cells are stably transformed with a recombinant DNA molecule comprising a suitable promoter functional in plants operatively linked to a coding sequence encoding a herbicide tolerant form of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein. A herbicide tolerant form of the enzyme has at least one amino acid substitution, addition or deletion that confers tolerance to a herbicide that inhibits the unmodified, naturally occurring form of the enzyme. The transgenic plants, plant tissue, plant seeds, or plant cells thus created are then selected by conventional selection techniques, whereby herbicide tolerant lines are isolated, characterized, and developed. Below are described methods for obtaining genes that encode herbicide tolerant forms of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein.

One general strategy involves direct or indirect mutagenesis procedures on microbes. For instance, a genetically manipulatable microbe such as *E. coli* or *S. cerevisiae* may be subjected to random mutagenesis *in vivo* with mutagens such as UV light or ethyl or methyl methane sulfonate. Mutagenesis procedures are described, for example, in Miller, 5 *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1972); Davis *et al.*, *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1980); Sherman *et al.*, *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1983); and U.S. Patent No. 4,975,374. For example, the microbe selected for mutagenesis contains a normal, inhibitor-sensitive ET6497, GT1773, 10 GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene, or nucleotide sequence substantially similar thereto, which encodes a protein having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene product activity, and is dependent upon the activity conferred by this gene for growth. The mutagenized cells are grown in the presence of the inhibitor at concentrations that inhibit the unmodified gene. Colonies of the 15 mutagenized microbe that grow better than the unmutagenized microbe in the presence of the inhibitor (i.e. exhibit resistance to the inhibitor) are selected for further analysis. ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes conferring tolerance to the inhibitor are isolated from these colonies, either by cloning or by PCR amplification, and their sequences are elucidated. Sequences encoding altered gene products 20 are then cloned back into the microbe to confirm their ability to confer inhibitor tolerance.

A method of obtaining mutant herbicide-tolerant alleles of a plant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene involves direct selection in plants. For example, the effect of a mutagenized ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene on the growth inhibition of plants such as *Arabidopsis*, soybean, or 25 maize is determined by plating seeds sterilized by art-recognized methods on plates on a simple minimal salts medium containing increasing concentrations of the inhibitor. Such concentrations are in the range of 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 110, 300, 1000 and 3000 parts per million (ppm). The lowest dose at which significant growth inhibition can be reproducibly detected is used for subsequent experiments. Determination of 30 the lowest dose is routine in the art.

Mutagenesis of plant material is utilized to increase the frequency at which resistant alleles occur in the selected population. Mutagenized seed material is derived from a variety of sources, including chemical or physical mutagenesis or seeds, or chemical or physical

mutagenesis or pollen (Neuffer, In *Maize for Biological Research* Sheridan, ed. Univ. Press, Grand Forks, ND., pp. 61-64 (1982)), which is then used to fertilize plants and the resulting M₁ mutant seeds collected. Typically for *Arabidopsis*, M₂ seeds (Lehle Seeds, Tucson, AZ), which are progeny seeds of plants grown from seeds mutagenized with chemicals, such as ethyl methane sulfonate, or with physical agents, such as gamma rays or fast neutrons, are plated at densities of up to 10,000 seeds/plate (10 cm diameter) on minimal salts medium containing an appropriate concentration of inhibitor to select for tolerance. Seedlings that continue to grow and remain green 7-21 days after plating are transplanted to soil and grown to maturity and seed set. Progeny of these seeds are tested for tolerance to the herbicide. If the tolerance trait is dominant, plants whose seed segregate 3:1 / resistant:sensitive are presumed to have been heterozygous for the resistance at the M₂ generation. Plants that give rise to all resistant seed are presumed to have been homozygous for the resistance at the M₂ generation. Such mutagenesis on intact seeds and screening of their M₂ progeny seed can also be carried out on other species, for instance soybean (*see, e.g.* U.S. Pat. No. 5,084,082). Alternatively, mutant seeds to be screened for herbicide tolerance are obtained as a result of fertilization with pollen mutagenized by chemical or physical means.

Confirmation that the genetic basis of the herbicide tolerance is a ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene is ascertained as exemplified below. First, alleles of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene from plants exhibiting resistance to the inhibitor are isolated using PCR with primers based either upon the *Arabidopsis* cDNA coding sequences shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15, respectively, or, more preferably, based upon the unaltered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene sequence from the plant used to generate tolerant alleles. After sequencing the alleles to determine the presence of mutations in the coding sequence, the alleles are tested for their ability to confer tolerance to the inhibitor on plants into which the putative tolerance-conferring alleles have been transformed. These plants can be either *Arabidopsis* plants or any other plant whose growth is susceptible to the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 inhibitors. Second, the inserted ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes are mapped relative to known restriction fragment length polymorphisms (RFLPs) (*See, for example, Chang et al. Proc. Natl. Acad. Sci. USA* 85:

6856-6860 (1988); Nam *et al.*, *Plant Cell* 1: 699-705 (1989), cleaved amplified polymorphic sequences (CAPS) (Konieczny and Ausubel (1993) *The Plant Journal*, 4(2): 403-410), or SSLPs (Bell and Ecker (1994) *Genomics*, 19: 137-144). The ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 inhibitor tolerance trait is independently
5 mapped using the same markers. When tolerance is due to a mutation in that ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene, the tolerance trait maps to a position indistinguishable from the position of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene.

Another method of obtaining herbicide-tolerant alleles of a ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene is by selection in plant cell cultures.
10 Explants of plant tissue, *e.g.* embryos, leaf disks, etc. or actively growing callus or suspension cultures of a plant of interest are grown on medium in the presence of increasing concentrations of the inhibitory herbicide or an analogous inhibitor suitable for use in a laboratory environment. Varying degrees of growth are recorded in different cultures. In
15 certain cultures, fast-growing variant colonies arise that continue to grow even in the presence of normally inhibitory concentrations of inhibitor. The frequency with which such faster-growing variants occur can be increased by treatment with a chemical or physical mutagen before exposing the tissues or cells to the inhibitor. Putative tolerance-conferring alleles of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene are
20 isolated and tested as described in the foregoing paragraphs. Those alleles identified as conferring herbicide tolerance may then be engineered for optimal expression and transformed into the plant. Alternatively, plants can be regenerated from the tissue or cell cultures containing these alleles.

Still another method involves mutagenesis of wild-type, herbicide sensitive plant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes in genetically
25 manipulatable microbes, followed by culturing the microbe on medium that contains inhibitory concentrations (i.e. sufficient to cause abnormal growth, inhibit growth or cause cell death) of the inhibitor, and then selecting those colonies that grow normally in the presence of the inhibitor. More specifically, a plant cDNA, such as the *Arabidopsis* cDNA
30 encoding the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein, is cloned into a microbe that is dependent on ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene product activity, respectively, for growth, or that otherwise lacks the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or

ET5036 activity. The transformed microbe is then subjected to *in vivo* mutagenesis or to *in vitro* mutagenesis by any of several chemical or enzymatic methods known in the art, e.g. sodium bisulfite (Shortle *et al.*, *Methods Enzymol.* 100:457-468 (1983); methoxylamine (Kadonaga *et al.*, *Nucleic Acids Res.* 13:1733-1745 (1985); oligonucleotide-directed
5 saturation mutagenesis (Hutchinson *et al.*, *Proc. Natl. Acad. Sci. USA*, 83:710-714 (1986); or various polymerase misincorporation strategies (see, e.g. Shortle *et al.*, *Proc. Natl. Acad. Sci. USA*, 79:1588-1592 (1982); Shiraishi *et al.*, *Gene* 64:313-319 (1988); and Leung *et al.*, *Technique* 1:11-15 (1989). Colonies that grow normally in the presence of normally inhibitory concentrations of inhibitor are picked and purified by repeated restreaking. Their
10 plasmids are purified and tested for the ability to confer tolerance to the inhibitor by retransforming them into the microbe lacking ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively. The DNA sequences of cDNA inserts from plasmids that pass this test are then determined.

Herbicide resistant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or
15 ET5036 proteins are also obtained using methods involving *in vitro* recombination, also called DNA shuffling. By DNA shuffling, mutations, preferably random mutations, are introduced into nucleotide sequences encoding ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity. DNA shuffling also leads to the recombination and rearrangement of sequences within a ET6497, GT1773, GT0992, ET6233, ET0763, ET5848,
20 GT5062, or ET5036 gene or to recombination and exchange of sequences between two or more different of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes. These methods allow for the production of millions of mutated ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 coding sequences. The mutated genes, or shuffled genes, are screened for desirable properties, e.g. improved tolerance to
25 herbicides and for mutations that provide broad spectrum tolerance to the different classes of inhibitor chemistry. Such screens are well within the skills of a routineer in the art.

In a preferred embodiment, a mutagenized ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene is formed from at least one template ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene, wherein the template
30 ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene has been cleaved into double-stranded random fragments of a desired size, and comprising the steps of adding to the resultant population of double-stranded random fragments one or more single or double-stranded oligonucleotides, wherein said oligonucleotides comprise an area of identity

and an area of heterology to the double-stranded random fragments; denaturing the resultant mixture of double-stranded random fragments and oligonucleotides into single-stranded fragments; incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of said single-stranded fragments at said areas
5 of identity to form pairs of annealed fragments, said areas of identity being sufficient for one member of a pair to prime replication of the other, thereby forming a mutagenized double-stranded polynucleotide; and repeating the second and third steps for at least two further cycles, wherein the resultant mixture in the second step of a further cycle includes the mutagenized double-stranded polynucleotide from the third step of the previous cycle, and the
10 further cycle forms a further mutagenized double-stranded polynucleotide, wherein the mutagenized polynucleotide is a mutated ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene having enhanced tolerance to a herbicide which inhibits naturally occurring ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively. In a preferred embodiment, the concentration of a single
15 species of double-stranded random fragment in the population of double-stranded random fragments is less than 1% by weight of the total DNA. In a further preferred embodiment, the template double-stranded polynucleotide comprises at least about 100 species of polynucleotides. In another preferred embodiment, the size of the double-stranded random fragments is from about 5 bp to 5 kb. In a further preferred embodiment, the fourth step of
20 the method comprises repeating the second and the third steps for at least 10 cycles. Such method is described e.g. in Stemmer et al. (1994) Nature 370: 389-391, in US Patent 5,605,793, US Patent 5,811,238 and in Crameri et al. (1998) Nature 391: 288-291, as well as in WO 97/20078, and these references are incorporated herein by reference.

In another preferred embodiment, any combination of two or more different ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes are mutagenized *in*
25 *vitro* by a staggered extension process (StEP), as described e.g. in Zhao et al. (1998) Nature Biotechnology 16: 258-261. The two or more ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes are used as template for PCR amplification with the extension cycles of the PCR reaction preferably carried out at a lower temperature than the
30 optimal polymerization temperature of the polymerase. For example, when a thermostable polymerase with an optimal temperature of approximately 72°C is used, the temperature for the extension reaction is desirably below 72°C, more desirably below 65°C, preferably below 60°C, more preferably the temperature for the extension reaction is 55°C. Additionally, the

duration of the extension reaction of the PCR cycles is desirably shorter than usually carried out in the art, more desirably it is less than 30 seconds, preferably it is less than 15 seconds, more preferably the duration of the extension reaction is 5 seconds. Only a short DNA fragment is polymerized in each extension reaction, allowing template switch of the extension products between the starting DNA molecules after each cycle of denaturation and annealing, thereby generating diversity among the extension products. The optimal number of cycles in the PCR reaction depends on the length of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes to be mutagenized but desirably over 40 cycles, more desirably over 60 cycles, preferably over 80 cycles are used. Optimal extension conditions and the optimal number of PCR cycles for every combination of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes are determined as described in using procedures well-known in the art. The other parameters for the PCR reaction are essentially the same as commonly used in the art. The primers for the amplification reaction are preferably designed to anneal to DNA sequences located outside of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes, e.g. to DNA sequences of a vector comprising the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes, whereby the different ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes used in the PCR reaction are preferably comprised in separate vectors. The primers desirably anneal to sequences located less than 500 bp away from ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 sequences, preferably less than 200 bp away from the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 sequences, more preferably less than 120 bp away from the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 sequences. Preferably, the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 sequences are surrounded by restriction sites, which are included in the DNA sequence amplified during the PCR reaction, thereby facilitating the cloning of the amplified products into a suitable vector. In another preferred embodiment, fragments of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes having cohesive ends are produced as described in WO 98/05765. The cohesive ends are produced by ligating a first oligonucleotide corresponding to a part of a ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene to a second oligonucleotide not present in the gene or corresponding to a part of the gene not adjoining to the part of the gene corresponding to the first oligonucleotide, wherein the second oligonucleotide contains at least one

ribonucleotide. A double-stranded DNA is produced using the first oligonucleotide as template and the second oligonucleotide as primer. The ribonucleotide is cleaved and removed. The nucleotide(s) located 5' to the ribonucleotide is also removed, resulting in double-stranded fragments having cohesive ends. Such fragments are randomly reassembled
5 by ligation to obtain novel combinations of gene sequences.

Any ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene or any combination of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes, or homologs thereof, is used for *in vitro* recombination in the context of the present invention, for example, a ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or
10 ET5036 gene derived from a plant, such as, e.g. *Arabidopsis thaliana*, e.g. a ET6497 gene set forth in SEQ ID NO:1, a GT1773 gene set forth in SEQ ID NO:3, a GT0992 gene set forth in SEQ ID NO:5, a ET6233 gene set forth in SEQ ID NO:7, a ET0763 gene set forth in SEQ ID NO:9, a ET5848 gene set forth in SEQ ID NO:11, a GT5062 gene set forth in SEQ ID NO:13, and a ET5036 gene set forth in SEQ ID NO:15. Whole ET6497, GT1773, GT0992, ET6233,
15 ET0763, ET5848, GT5062, or ET5036 genes or portions thereof are used in the context of the present invention. The library of mutated ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes obtained by the methods described above are cloned into appropriate expression vectors and the resulting vectors are transformed into an appropriate host, for example a plant cell, an algae like *Chlamydomonas*, a yeast or a bacteria. An
20 appropriate host requires ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene product activity for growth. Host cells transformed with the vectors comprising the library of mutated ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 genes are cultured on medium that contains inhibitory concentrations of the inhibitor and those colonies that grow in the presence of the inhibitor are selected. Colonies that grow
25 in the presence of normally inhibitory concentrations of inhibitor are picked and purified by repeated restreaking. Their plasmids are purified and the DNA sequences of cDNA inserts from plasmids that pass this test are then determined.

An assay for identifying a modified ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene that is tolerant to an inhibitor may be performed in the same manner
30 as the assay to identify inhibitors of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity (Inhibitor Assay, above) with the following modifications: First, a mutant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein is substituted in one of the reaction mixtures for the wild-type

ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein of the inhibitor assay. Second, an inhibitor of wild-type enzyme is present in both reaction mixtures. Third, mutated activity (activity in the presence of inhibitor and mutated enzyme) and unmutated activity (activity in the presence of inhibitor and wild-type enzyme) are compared to determine whether a significant increase in enzymatic activity is observed in the mutated activity when compared to the unmutated activity. Mutated activity is any measure of activity of the mutated enzyme while in the presence of a suitable substrate and the inhibitor. Unmutated activity is any measure of activity of the wild-type enzyme while in the presence of a suitable substrate and the inhibitor.

In addition to being used to create herbicide-tolerant plants, genes encoding herbicide-tolerant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 protein can also be used as selectable markers in plant cell transformation methods. For example, plants, plant tissue, plant seeds, or plant cells transformed with a heterologous DNA sequence can also be transformed with a sequence encoding an altered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity capable of being expressed by the plant. The transformed cells are transferred to medium containing an inhibitor of the enzyme in an amount sufficient to inhibit the growth or survivability of plant cells not expressing the modified coding sequence, wherein only the transformed cells will grow. The method is applicable to any plant cell capable of being transformed with a modified ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene, and can be used with any heterologous DNA sequence of interest. Expression of the heterologous DNA sequence and the modified gene can be driven by the same promoter functional in plant cells, or by separate promoters.

XIII. Plant Transformation Technology

A wild type or herbicide-tolerant form of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene, or homologs thereof, can be incorporated in plant or bacterial cells using conventional recombinant DNA technology. Generally, this involves inserting a DNA molecule encoding the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene into an expression system to which the DNA molecule is heterologous (i.e., not normally present) using standard cloning procedures known in the art. The vector contains the necessary elements for the transcription and translation of the inserted protein-coding sequences in a host cell containing the vector. A large number of vector

systems known in the art can be used, such as plasmids, bacteriophage viruses and other modified viruses. The components of the expression system may also be modified to increase expression. For example, truncated sequences, nucleotide substitutions, nucleotide optimization or other modifications may be employed. Expression systems known in the art can be used to transform virtually any crop plant cell under suitable conditions. A heterologous DNA sequence comprising a wild-type or herbicide-tolerant form of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene is preferably stably transformed and integrated into the genome of the host cells. In another preferred embodiment, the heterologous DNA sequence comprising a wild-type or herbicide-tolerant form of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene located on a self-replicating vector. Examples of self-replicating vectors are viruses, in particular gemini viruses. Transformed cells can be regenerated into whole plants such that the chosen form of the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene confers herbicide tolerance in the transgenic plants.

A. Requirements for Construction of Plant Expression Cassettes

Gene sequences intended for expression in transgenic plants are first assembled in expression cassettes behind a suitable promoter expressible in plants. The expression cassettes may also comprise any further sequences required or selected for the expression of the heterologous DNA sequence. Such sequences include, but are not restricted to, transcription terminators, extraneous sequences to enhance expression such as introns, vital sequences, and sequences intended for the targeting of the gene product to specific organelles and cell compartments. These expression cassettes can then be easily transferred to the plant transformation vectors described *infra*. The following is a description of various components of typical expression cassettes.

1. Promoters

The selection of the promoter used in expression cassettes will determine the spatial and temporal expression pattern of the heterologous DNA sequence in the plant transformed with this DNA sequence. Selected promoters will express heterologous DNA sequences in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific tissues or organs (roots, leaves or flowers, for example) and the selection will reflect the desired location of accumulation of the gene product. Alternatively, the selected promoter may drive expression of the gene under various inducing conditions. Promoters vary in their strength,

i.e., ability to promote transcription. Depending upon the host cell system utilized, any one of a number of suitable promoters known in the art can be used. For example, for constitutive expression, the CaMV 35S promoter, the rice actin promoter, or the ubiquitin promoter may be used. For regulatable expression, the chemically inducible PR-1 promoter from tobacco or
5 *Arabidopsis* may be used (*see, e.g.*, U.S. Patent No. 5,689,044).

2. Transcriptional Terminators

A variety of transcriptional terminators are available for use in expression cassettes. These are responsible for the termination of transcription beyond the heterologous DNA sequence and its correct polyadenylation. Appropriate transcriptional terminators are those that are
10 known to function in plants and include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator and the pea *rbcS* E9 terminator. These can be used in both monocotyledonous and dicotyledonous plants.

3. Sequences for the Enhancement or Regulation of Expression

Numerous sequences have been found to enhance gene expression from within the
15 transcriptional unit and these sequences can be used in conjunction with the genes of this invention to increase their expression in transgenic plants. For example, various intron sequences such as introns of the maize *Adhl* gene have been shown to enhance expression, particularly in monocotyledonous cells. In addition, a number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are
20 particularly effective in dicotyledonous cells.

4. Coding Sequence Optimization

The coding sequence of the selected gene optionally is genetically engineered by altering the coding sequence for optimal expression in the crop species of interest. Methods for modifying coding sequences to achieve optimal expression in a particular crop species are
25 well known (*see, e.g.* Perlak *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 3324 (1991); and Koziel *et al.*, *Bio/technol.* 11: 194 (1993); Fennoy and Bailey-Serres. *Nucl. Acids Res.* 21: 5294-5300 (1993). Methods for modifying coding sequences by taking into account codon usage in plant genes and in higher plants, green algae, and cyanobacteria are well known (*see table 4 in:* Murray *et al.* *Nucl. Acids Res.* 17: 477-498 (1989); Campbell and Gowri *Plant Physiol.* 92: 1-
30 11(1990).

5. Targeting of the Gene Product Within the Cell

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

detail. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various proteins which is cleaved during chloroplast import to yield the mature protein (*e.g.* Comai *et al.* J. Biol. Chem. 263: 15104-15109 (1988)). Other gene products are localized to other organelles such as the
5 mitochondrion and the peroxisome (*e.g.* Unger *et al.* Plant Molec. Biol. 13: 411-418 (1989)). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous products encoded by DNA sequences to these organelles. In addition, sequences have been characterized which cause the targeting of products encoded by DNA sequences to other cell compartments. Amino terminal sequences are responsible for targeting to the ER,
10 the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, Plant Cell 2: 769-783 (1990)). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al.* Plant Molec. Biol. 14: 357-368 (1990)). By the fusion of the appropriate targeting sequences described above to heterologous DNA sequences of interest it is possible to direct this product to any
15 organelle or cell compartment.

B. Construction of Plant Transformation Vectors

Numerous transformation vectors available for plant transformation are known to those of ordinary skill in the plant transformation arts, and the genes pertinent to this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the
20 preferred transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the *nptII* gene, which confers resistance to kanamycin and related antibiotics (Messing & Vierra. Gene 19: 259-268 (1982); Bevan *et al.*, Nature 304:184-187 (1983)), the *bar* gene, which confers resistance to the herbicide phosphinothricin (White *et al.*, Nucl. Acids Res 18: 1062 (1990), Spencer *et al.* Theor. Appl. Genet 79: 625-631 (1990)), the *hph* gene, which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, Mol Cell Biol 4: 2929-2931), the *manA* gene, which
25 allows for positive selection in the presence of mannose (Miles and Guest (1984) Gene, 32:41-48; U.S. Patent No. 5,767,378), the *dhfr* gene, which confers resistance to methotrexate (Bourouis *et al.*, EMBO J. 2(7): 1099-1104 (1983)), and the EPSPS gene, which confers
30 resistance to glyphosate (U.S. Patent Nos. 4,940,935 and 5,188,642).

1. Vectors Suitable for *Agrobacterium* Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, Nucl. Acids Res. (1984)). Typical vectors suitable for *Agrobacterium* transformation include the binary vectors pCIB200 and pCIB2001, as well as the binary vector pCIB10 and hygromycin selection derivatives thereof. (See, for example, U.S. Patent No. 5,639,949).

2. Vectors Suitable for non-*Agrobacterium* Transformation

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques that do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake (e.g. PEG and electroporation) and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Typical vectors suitable for non-*Agrobacterium* transformation include pCIB3064, pSOG19, and pSOG35. (See, for example, U.S. Patent No. 5,639,949).

C. Transformation Techniques

Once the coding sequence of interest has been cloned into an expression system, it is transformed into a plant cell. Methods for transformation and regeneration of plants are well known in the art. For example, Ti plasmid vectors have been utilized for the delivery of foreign DNA, as well as direct DNA uptake, liposomes, electroporation, micro-injection, and microprojectiles. In addition, bacteria from the genus *Agrobacterium* can be utilized to transform plant cells.

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques that do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, particle bombardment into callus tissue, as well as *Agrobacterium*-mediated transformation.

D. Plastid Transformation

In another preferred embodiment, a nucleotide sequence encoding a polypeptide having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity is directly transformed into the plastid genome. Plastid expression, in which genes are inserted
5 by homologous recombination into the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclear-expressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein. In a preferred embodiment, the nucleotide sequence is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants
10 homoplasmic for plastid genomes containing the nucleotide sequence are obtained, and are preferentially capable of high expression of the nucleotide sequence.

Plastid transformation technology is for example extensively described in U.S. Patent Nos. 5,451,513, 5,545,817, 5,545,818, and 5,877,462 in PCT application no. WO 95/16783 and WO 97/32977, and in McBride *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91, 7301-7305, all
15 incorporated herein by reference in their entirety. The basic technique for plastid transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the nucleotide sequence into a suitable target tissue, e.g., using biolistics or protoplast transformation (e.g., calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate homologous recombination
20 with the plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab, Z., Hajdukiewicz, P., and Maliga, P. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8526-8530; Staub, J. M., and Maliga, P. (1992) *Plant Cell* 4, 39-45). The presence of
25 cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub, J.M., and Maliga, P. (1993) *EMBO J.* 12, 601-606). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial *aadA* gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3'-adenyltransferase (Svab, Z., and Maliga, P. (1993) *Proc. Natl. Acad. Sci. USA* 90, 913-917).
30 Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention.

XV. Breeding

The wild-type or altered form of a ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene of the present invention can be utilized to confer herbicide tolerance to a wide variety of plant cells, including those of gymnosperms, monocots, and dicots.

5 Although the gene can be inserted into any plant cell falling within these broad classes, it is particularly useful in crop plant cells, such as rice, wheat, barley, rye, corn, potato, carrot, sweet potato, sugar beet, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot,
10 strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tobacco, tomato, sorghum and sugarcane.

The high-level expression of a wild-type ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene and/or the expression of herbicide-tolerant forms of a ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene conferring
15 herbicide tolerance in plants, in combination with other characteristics important for production and quality, can be incorporated into plant lines through breeding approaches and techniques known in the art.

Where a herbicide tolerant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene allele is obtained by direct selection in a crop plant or plant cell culture from
20 which a crop plant can be regenerated, it is moved into commercial varieties using traditional breeding techniques to develop a herbicide tolerant crop without the need for genetically engineering the allele and transforming it into the plant.

The invention will be further described by reference to the following detailed examples.
25 These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

EXAMPLES

Standard recombinant DNA and molecular cloning techniques used here are well known in
30 the art and are described by Sambrook, *et al.*, Molecular Cloning, eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989) and by T.J. Silhavy, M.L. Berman, and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and by Ausubel, F.M. *et al.*, Current Protocols in Molecular Biology, pub.

by Greene Publishing Assoc. and Wiley-Interscience (1987), Reiter, et al., Methods in Arabidopsis Research, World Scientific Press (1992), and Schultz et al., Plant Molecular Biology Manual, Kluwer Academic Publishers (1998). These references describe the standard techniques used for all steps in tagging and cloning genes from Ac/Ds transposon mutagenized populations of *Arabidopsis*: plant infection and transformation; screening for the identification of seedling mutants; and cosegregation analysis. Ds transposon insertion lines produced as described in Sundareson et al. (1995) *Genes and Dev.*, 9:1797-1810) were used in these experiments.

Example 1: Transposon Border Isolation

Arabidopsis genomic DNA is isolated from line ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 using the Nucleon PhytoPure™ Plant DNA Isolation Kit (Amersham International plc, Buckinghamshire, England). Fragments of genomic DNA flanking the borders of the transposon are isolated using the TAIL-PCR technique (Liu et al. (1995) *The Plant Journal*, 8:457-463; Liu and Whittier (1995), *Genomics*, 25: 674-681). Three sets of 12 TAIL-PCR reactions, referred to as the primary, secondary and tertiary reactions, are performed. In each reaction, one arbitrary degenerate primer and one transposon-specific primer are used. The arbitrary degenerate primer is chosen from among six primers, LWAD1, CA51, CA52, CA53, CA54, and CA55 (Table 1), which are used to prime the genomic DNA flanking the insertion. These degenerate primers are used in combination with two sets of three, nested, transposon-specific primers (Table 2). These primers are homologous to regions of the Ds elements which lie at the outermost ends of the transposons, DS5 at the 5' end (primers 5A, 5B, and 5C) and DS3 at the 3' end (primers 3A, 3B, and 3C). When the degenerate and nested primer pairs are used in a series of low and high-stringency PCR amplifications, as described in the TAIL-PCR protocol (Liu and Whittier (1995), *Genomics*, 25: 674-681), DNA fragments are produced which correspond to the genomic DNA that is directly adjacent to the transposon insertion. The nucleic acid sequence of the PCR products from the tertiary TAIL-PCR reactions are then determined by standard molecular biology techniques. The resulting sequences are analyzed for the presence of non-Ds transposon vector sequence. To confirm the integrity of the resultant products, PCR primers specific to the flanking genomic region are designed and used in conjunction with the tertiary nested primer in a PCR reaction, to confirm the transposon

insertion point within the genomic DNA. Finding a PCR product of the appropriate size, based on the sequence of the TAIL-PCR clone confirms a valid rescue.

TABLE 1: DEGENERATE PRIMERS

	<u>ID NO</u>	<u>PRIMER</u>	<u>DEGEN.</u>	<u>PRIMER SEQUENCE</u>	<u>NOTES AND REFS.</u>
5	25	LWAD1	1026	NGT TGW GNA TWT SGW GNT	designed by L. Wegrich
	26	CA51	128	TGW GNA GSA NCA SAG	derivative of primer AD1 ⁽²⁾
	27	CA52	128	AGW GNA GWA NCA WAG G	identical to primer AD2 ⁽²⁾
	28	CA53	256	STT GNT AST NCT NTG C	identical to primer AD5 ⁽³⁾
10	29	CA54	64	NTC GAS TWT SGW GTT	identical to primer AD1 ⁽¹⁾
	30	CA55	256	WGT GNA GWA NCA NAG A	identical to primer AD3 ⁽¹⁾

TABLE 2: NESTED PRIMERS

	<u>ID NO</u>	<u>PRIMER</u>	<u>PRIMER SEQUENCE</u>	<u>NOTES</u>
15	31	5A	ACTAGCTCTACCGTTTCCGTTTCCGTTTAC	DS5 PRIMARY
	32	5B	TTACCTCGGGTTCGAAATCGATCGGGATAA	DS5 SECONDARY
	33	5C	AAAATCGGTTATACGATAACGGTCGGTACGGGA	DS5 TERTIARY
	34	3A	GGGTCTTGCGGATCTGAATATATGTTTTTCATGTGTG	DS3 PRIMARY
	35	3B	TACCGAAGAAAAATACCGGTTCCCGTCCGATTTCGAC	DS3 SECONDARY
20	36	3C	GGATCGTATCGGTTTTTCGATTACCGTATTTATCC	DS3 TERTIARY

REFERENCES: (1) Liu et al. (1995) *The Plant Journal*, 8:457-463; (2) Liu and Whittier (1995) *Genomics*, 25: 674-681; (3) Tsugeki et al. (1996) *The Plant Journal*, 10: 479-489.

Example 2: Sequence Analysis of Tagged Seedling Lethal Line ET6497

25 For transposant line ET6497, PCR products are obtained from the Ds5 border. The preliminary sequences, obtained from the TAIL-PCR, are used in BLASTn searches against nucleotide databases (Altschul et al. (1990) *J Mol. Biol.* 215:403-410; Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402). The initial sequence of the region bordering the Ds5 end indicates that the transposon has inserted with the Ds5 end adjacent to *Arabidopsis* genomic

30 DNA (base numbers 3162 and lower of *Arabidopsis* chromosome 3, BAC T20O10, GenBank accession number AL163816). The transposon insertion region of BAC T20O10 is annotated as encoding a putative Prokaryotic-type class I peptide chain release factor (GenPept accession number CAB87736).

To identify the complete open reading frame (ORF) for this gene, 5' RACE primers are designed in the 3' portion of the predicted ORF, which is present in three *Arabidopsis* ESTs (GenBank accession numbers AV558327, AV558301, T41756). PCR is performed using template DNA from a Gene Racer (Invitrogen) cDNA library prepared from seedling tissue.

5 A resulting PCR product is TA-cloned (Original TA-Cloning kit, Invitrogen) and sequenced. The cDNA sequence (SEQ ID NO:1) differs from the sequence predicted in the GenBank annotation, thus identifying for the first time the actual ORF. Analysis of the cDNA sequence from this gene reveals a high degree of similarity to the Prokaryotic-type class I peptide chain release factor (see homolog table ET6497).

10	<u>ET6497 HOMOLOGS</u>			
	Description	Accession #	Database	%ID
	Prokaryotic peptide	AAB94190	GenPept	32.3
	Release factors (<i>Caenorhabditis elegans</i>)			
15	Mitochondrial peptide	CAA90504	GenPept	34.8
	Chain release factor			
	1 precursor (<i>Schizosaccharomyces pombe</i>)			
	Peptide chain release	BAA13349.1	GenPept	40.8
20	Factor 1			
	(Thermus thermophilus)			
	Peptide chain release	AAC43437	GenPept	41.1
	Factor 1			
	(Escherichia coli)			
25	Protein	AAF42034	GenPept	44.1
	(Neisseria meningitidis)			
	Peptide chain release factor	BAA18826	GenPept	58.9
	(Synechocystis sps)			
	<i>Hordeum vulgare</i> DNA	BE231015	GenBank	70.5
30	(pre-anthesis spike EST BE231015)			

% ID: percent identity relative to the protein encoded by the *Arabidopsis thaliana* ET6497 gene

Example 3: Sequence Analysis of Tagged Seedling Lethal Line GT1773

5 For transposant line GT1773, PCR products are obtained from the Ds3 border and from the Ds5 border. The preliminary sequences obtained from the TAIL-PCR are used in BLASTn searches against nucleotide databases (Altschul et al. (1990) J Mol. Biol. 215:403-410; Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The initial sequence of the region bordering the Ds3 end indicates that the transposon has inserted with the Ds3 end adjacent to
10 *Arabidopsis* genomic DNA (base numbers 63475 and lower of BAC T21L8 *Arabidopsis* chromosome 3, GenBank accession number AL096860). The initial sequence of the region bordering the Ds5 end indicates that the transposon has inserted with the Ds5 end adjacent to *Arabidopsis* genomic DNA (base numbers 63474 and higher of *Arabidopsis* BAC T21L8 on chromosome 3). This region of BAC T21L8 is annotated as having similarity to the
15 riboflavin biosynthesis protein ribG of *Synechocystis* sp., PIR2:S74377 (GenPept accession number CAB51211).

To resolve the ORF for this gene, 5' and 3' RACE primers are designed to the predicted ORF. PCR is performed using template from a Gene Racer (Invitrogen) cDNA library prepared from seedling tissue. The resulting PCR products are TA-cloned (Original TA-Cloning kit,
20 Invitrogen) and sequenced. The cDNA sequence (SEQ ID NO:3) differs from the cDNA sequence predicted in the GenBank annotation, thus identifying for the first time the actual ORF. Analysis of this cDNA sequence from this gene reveals a high degree of similarity to Diaminohydroxyphosphoribosylaminopyrimidine deaminases, 5-amino-6-(5-phosphoribosylamino) uracil reductases, and an *E. coli* protein of unknown function (see
25 homolog table GT1773).

A second cDNA PCR product is identified, TA cloned, and sequenced (SEQ ID NO:37). The resulting sequence is analyzed and may represent an incompletely spliced mRNA transcript of the above mentioned cDNA. The higher molecular weight clone (1796 nucleotides) contains 5 extra bases corresponding to nucleotides 61969-61973 of BAC T21L8, when
30 compared to the lower molecular weight cDNA clone (1791 nucleotides). These base differences noted, result in an ORF of only 1053 nucleotides in the higher molecular weight clone (the corresponding translation is in SEQ ID NO:38), while the ORF of the lower molecular weight cDNA is 1791 nucleotides in length.

<u>GT1773 HOMOLOGS</u>			
	Description	Accession #	Database %ID
5	RibG-like riboflavin biosynthetic protein (<i>Arabidopsis thaliana</i>)	CAB79096	GenPept 26.6
	5-amino-6-(5-phospho-ribosylamino) uracil reductase (<i>Methanococcus jannaschii</i>)	AAB98665	GenPept 29.4
10	Riboflavin biosynthetic Protein ribG (<i>Escherichia coli</i>)	AAB40170	GenPept 30.7
	Riboflavin-specific Deaminase (<i>Thermotoga maritima</i>)	AAD36891	GenPept 32.1
15	Riboflavin-specific Deaminase (<i>Streptomyces coelicolor</i>)	CAB92254	GenPept 38.4
20	Riboflavin bifunctional Biosynthetic protein ribG (<i>Synechocystis</i> sps)	BAA10295	GenPept 42.5
	Hypothetical protein ybiA (<i>Escherichia coli</i>)	P30176	SWISSPROT 45.0
25	<i>Glycine max</i> DNA (EST AW100925)	AW100925	GenBank 83.1

% ID: percent identity relative to the protein encoded by the *Arabidopsis thaliana* GT1773 gene

30 Example 4: Sequence Analysis of Tagged Seedling Lethal Line GT0992

For transposant line GT0992, PCR products are obtained from the Ds3 border and from the Ds5 border. The preliminary sequences obtained from the TAIL-PCR are used in BLASTn searches against nucleotide databases (Altschul et al. (1990) J Mol. Biol. 215:403-410; Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The initial sequence of the region

bordering the Ds3 end indicates that the transposon has inserted with the Ds3 end adjacent to *Arabidopsis* genomic DNA (base numbers 28738 and higher of *Arabidopsis* P1 clone MXC9 on chromosome 5, GenBank accession number AB007727). The initial sequence of the region bordering the Ds5 end indicates that the transposon has inserted with the Ds5 end adjacent to *Arabidopsis* genomic DNA (base numbers 28735 and lower of *Arabidopsis* P1 clone MXC9 on chromosome 5). This region of MXC9 is annotated as encoding a putative protein with similarity to transmembrane transport proteins. (GenPept accession number BAB10031).

To identify the ORF for this gene, 5' RACE primers are designed to in the 3' portion of the predicted ORF, which is present in two *Arabidopsis* ESTs (GenBank accession numbers AV556256, AV557070). PCR is performed using template DNA from a Gene Racer (Invitrogen) cDNA library prepared from seedling tissue.. The resulting PCR product is TA-cloned (Original TA-Cloning kit, Invitrogen) and sequenced. The identified cDNA (SEQ ID NO:5) differs from the cDNA sequence predicted in the GenBank annotation, thus identifying for the first time the actual ORF. Analysis of the cDNA sequence from this gene reveals a high degree of sequence similarity to tellurium resistance proteins from several species of bacteria (see homolog table GT0992).

GT0992 HOMOLOGS

Description	Accession #	Database	% ID
Tellurium resistance Protein terC (RP786) (<i>Rickettsia prowazekii</i>)	CAA15212	GenPept	30.0
Hypothetical protein (VCA0546) (<i>Vibrio cholerae</i>)	AAF96448	GenPept	34.3
Hypothetical protein ygjT (<i>Escherichia coli</i>)	AAA57890	GenPept	37.9
Putative integral membrane Export protein (<i>Streptomyces coelicolor</i>)	CAB38153	GenPept	38.9

	Unknown protein (<i>Oryza sativa</i>)	BAA90517	GenPept	40.5
	Hypothetical protein (<i>Neisseria meningitidis</i>)	AAF40491	GenPept	42.1
5	<i>Oryza sativa</i> DNA	AP001111	GenBank	51.5

% ID: percent identity relative to the protein encoded by the *Arabidopsis thaliana* GT0992 gene

10 Example 5: Sequence Analysis of Tagged Seedling Lethal Line ET6233

For transposant line ET6233, PCR products are obtained from the Ds3 border and from the Ds5 border. The preliminary sequences obtained from TAIL-PCR are used in BLASTn searches against nucleotide databases (Altschul et al. (1990) J Mol. Biol. 215:403-410; Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The initial sequence of the region
15 bordering the Ds3 end indicates that the transposon has inserted with the Ds3 end adjacent to *Arabidopsis* genomic DNA (base numbers 94094 and lower of BAC clone F14I3 *Arabidopsis* chromosome 1, GenBank accession number AC007980). The initial sequence of the region bordering the Ds5 end indicates that the transposon has inserted with the Ds5 end adjacent to
20 *Arabidopsis* genomic DNA (base numbers 94085 and higher of BAC clone F14I3 *Arabidopsis* chromosome 1). Analysis of the border sequences reveals a nine base pair duplication that occurred during the transposon insertion, corresponding to bases 94085 through 94083 of BAC F14I3. This region of BAC clone F14I3 on chromosome 1 is annotated as encoding a hypothetical protein (GenPept accession number AAD50047).

To identify the ORF for this gene, primers are designed to the 5' and 3' ends of the predicted
25 ORF. PCR is performed using template DNA from a cDNA library prepared from seedling tissue. The resulting PCR product is TA-cloned (Original TA-Cloning kit, Invitrogen) and sequenced. The cDNA sequence (SEQ ID NO: 7) differs from the sequence predicted in the GenBank annotation, thus identifying for the first time the actual ORF. Analysis of the cDNA sequence from this gene reveals that it is similar to the *Bacillus megaterium* cbix
30 gene, which is necessary for cobyric acid biosynthesis (see homolog table ET6233).

ET6233 HOMOLOGS

Description	Accession #	Database	%ID
-------------	-------------	----------	-----

	CbiX protein (<i>Bacillus megaterium</i>)	CAA04308	GenPept	27.7
	Hypothetical protein (<i>Methanococcus jannaschii</i>)	AAB98975	GenPept	34.5
5	Unknown protein (<i>Bacillus halodurans</i>)	BAB05215.1	GenPept	36.1
	<i>Beta vulgaris</i> DNA (EST BE590398)	BE590398	GenBank	57.2
10	% ID: percent identity relative to the protein encoded by the <i>Arabidopsis thaliana</i> ET6233 gene			

Example 6: Sequence Analysis of Tagged Seedling Lethal Line ET0763

15 For transposant line ET0763, PCR products are obtained from the Ds3 border and from the Ds5 border. The preliminary sequences obtained from TAIL-PCR are used in BLASTn searches against nucleotide databases (Altschul et al. (1990) J Mol. Biol. 215:403-410; Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The initial sequence of the region bordering the Ds3 end indicates that the transposon has inserted with the Ds3 end adjacent to

20 *Arabidopsis* genomic DNA (base numbers 43150 and higher of BAC clone F14O23 *Arabidopsis* chromosome 1, GenBank accession number AC012654). The initial sequence of the region bordering the Ds5 end indicates that the transposon has inserted with the Ds5 end adjacent to *Arabidopsis* genomic DNA (base numbers 43160 and lower of BAC clone F14O23 *Arabidopsis* chromosome 1, GenBank accession number AC012654). Analysis of

25 the border sequences reveals a nine base pair duplication that occurred during the transposon insertion, corresponding to bases 43151 through 43159 of BAC F14O23. This region of BAC clone F14O23 on chromosome 1 is annotated as encoding a hypothetical protein that contains a putative S1 RNA binding domain. (GenPept accession number AAF43225).

To identify the ORF for this gene, primers are designed to the 5' and 3' ends of the predicted

30 ORF. PCR is performed using template DNA from a cDNA library prepared from seedling tissue. The resulting PCR product is TA-cloned (Original TA-Cloning kit, Invitrogen) and sequenced. The cDNA sequence (SEQ ID NO:9) is the same as the sequence predicted in the GenBank annotation, thus confirming for the first time the actual ORF. Analysis of the

cDNA sequence from this gene reveals that it is similar to 30S Ribosomal Protein S1 in several other species (see homolog table ET0763).

ET0763 HOMOLOGS

5	Description	Accession #	Database	%ID
	S1 ribosomal protein (<i>Escherichia coli</i>)	CAA23644	GenPept	24.7
	30S ribosomal protein S1 (<i>Streptomyces coelicolor</i>)	CAB52054	GenPept	25.5
10	Chloroplast ribosomal protein S1 (<i>Spinacia oleracea</i>)	AAA34045	GenPept	26.0
	30S ribosomal protein S1 (<i>Porphyra purpurea</i>)	AAC08231	GenPept	26.8
15	Ribosomal protein S1 (<i>Mycobacterium tuberculosis</i>)	CAB08883	GenPept	27.2
	Ribosomal protein S1 (<i>Synechococcus</i> sps)	BAA05946	GenPept	32.8
	Tomato DNA (EST AW030202)	AW030202	GenBank	73.9
20				

% ID: percent identity relative to the protein encoded by the *Arabidopsis thaliana* ET0763 gene

25 Example 7: Sequence Analysis of Tagged Seedling Lethal Line ET5848

For transposant line ET5848, PCR products are obtained from the Ds3 border. The preliminary sequences obtained from TAIL-PCR are used in BLASTn searches against nucleotide databases (Altschul et al. (1990) J Mol. Biol. 215:403-410; Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The initial sequence of the region bordering the Ds3 end indicates that the transposon has inserted with the Ds3 end adjacent to *Arabidopsis* genomic DNA (base numbers 27530 and higher of BAC clone F17O7 *Arabidopsis* chromosome 1, GenBank accession number AC003671). This region of BAC clone F17O7, on chromosome 1, is annotated as encoding a hypothetical protein. (GenPept accession number AAC18802).

To identify the ORF for this gene, primers are designed to the 5' and 3' ends of the predicted ORF. PCR is performed using template DNA from a cDNA library prepared from seedling tissue. The resulting PCR product is TA-cloned (Original TA-Cloning kit, Invitrogen) and sequenced. The cDNA sequence (SEQ ID NO:11) is the same as the sequence predicted in the GenBank annotation, thus confirming for the first time the actual ORF. Analysis of the cDNA sequence from this gene reveals that it is similar to a large family of *Arabidopsis* genes, including some putative myb transcription factors and some with unknown function (see homolog table ET5848).

10

ET5848 HOMOLOGS

Description	Accession #	Database	%ID
Hypothetical protein (<i>Arabidopsis thaliana</i>)	AAF27677	GenPept	25.4
Hypothetical protein (<i>Arabidopsis thaliana</i>)	CAB61960	GenPept	27.5
MYB transcription factor-like protein (<i>Arabidopsis thaliana</i>)	AAF25966	GenPept	28.5
MYB transcription factor-like protein (<i>Arabidopsis thaliana</i>)	AAC18801	GenPept	49.4
Hypothetical protein (<i>Arabidopsis thaliana</i>)	AAD55510	GenPept	51.4
Hypothetical gene (F3I17.6) DNA (<i>Arabidopsis thaliana</i>)	AC016162.3	GenPept	75.7

% ID: percent identity relative to the protein encoded by the *Arabidopsis thaliana* ET5848 gene

30

Example 8: Sequence Analysis of Tagged Seedling Lethal Line GT5062

For transposant line GT5062, PCR products are obtained from the Ds3 border and from the Ds5 border. The preliminary sequences obtained from TAIL-PCR are used in BLASTn

searches against nucleotide databases (Altschul et al. (1990) J Mol. Biol. 215:403-410; Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The initial sequence of the region bordering the Ds3 end indicates that the transposon has inserted with the Ds3 end adjacent to *Arabidopsis* genomic DNA (base numbers 60591 and lower of BAC clone F5K20 *Arabidopsis* chromosome 3, GenBank accession number AL132960). The initial sequence of the region bordering the Ds5 end indicates that the transposon has inserted with the Ds5 end adjacent to *Arabidopsis* genomic DNA (base numbers 60592 and higher of BAC clone F5K20 *Arabidopsis* chromosome 3). This region of BAC clone F5K20 on chromosome 3 is annotated as encoding a hypothetical protein that shows similarity to a uracil phosphoribosyltransferase (GenPept accession number CAB88352).

To identify the ORF for this gene, primers are designed to the 5' and 3' ends of the predicted ORF. PCR is performed using template DNA from a cDNA library prepared from seedling tissue. The resulting PCR product is TA-cloned (Original TA-Cloning kit, Invitrogen) and sequenced. The cDNA sequence (SEQ ID NO:13) is the same as the sequence predicted in the GenBank annotation, thus confirming for the first time the actual ORF. Analysis of the cDNA sequence from this gene reveals that it is similar to uracil phosphoribosyltransferases from other species (see homolog table GT5062).

GT5062 HOMOLOGS

Description	Accession #	Database	%ID
Uracil phospho- ribosyltransferase (<i>Schizosaccharomyces pombe</i>)	CAB11230	GenPept	29.2
Uracil phospho- ribosyltransferase (<i>Methanobacterium thermoautotrophicum</i>)	AAB85603	GenPept	30.8
Uracil phospho- ribosyltransferase (<i>Thermotoga maritima</i>)	AAD35803	GenPept	40.7

	Uracil phospho- ribosyltransferase (<i>Escherichia coli</i>)	BAA16386	GenPept	41.3
5	Uracil phospho- ribosyltransferase (<i>Synechocystis</i> sps)	BAA16768	GenPept	53.7
	Uracil phospho- ribosyltransferase (<i>Nicotiana tabacum</i>)	P93394	SWISSPROT	79.0
10	DNA Uracil phosphoribosyltransferase (<i>Nicotiana tabacum</i>)	Y11210	GenBank	67.1

15 % ID: percent identity relative to the protein encoded by the *Arabidopsis thaliana* GT5062 gene

Example 9: Sequence Analysis of Tagged Seedling Lethal Line ET5036

For transposant line ET5036, PCR products are obtained from the Ds3 border. The preliminary sequences obtained from TAIL-PCR are used in BLASTn searches against
 20 nucleotide databases (Altschul et al. (1990) J Mol. Biol. 215:403-410; Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The initial sequence of the region bordering the Ds3 end indicates that the transposon has inserted with the Ds3 end adjacent to *Arabidopsis* genomic DNA (base numbers 51094 and lower of BAC clone T10O24 *Arabidopsis* chromosome 1, GenBank accession number AC007067). This region of BAC clone T10O24, on
 25 chromosome 1, is annotated as encoding a hypothetical protein. (GenPept accession number AAD39574).

To identify the ORF for this gene, 5' and 3' RACE primers are designed to the predicted ORF. PCR is performed using template DNA from a Gene Racer cDNA library (Gene Racer kit, Invitrogen) prepared from seedling tissue. The resulting PCR products are TA-cloned
 30 (Original TA-Cloning kit, Invitrogen) and sequenced. The cDNA sequence (SEQ ID NO:15) is the same as the sequence predicted in the GenBank annotation, thus confirming for the first time the actual ORF. BLAST analysis with the nucleotide and amino acid sequence does not identify similar sequences.

Example 10: Expression of Recombinant ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 Protein in *E. coli*

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:1 is subcloned into an appropriate expression vector, and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the ET6497 activity is confirmed. Protein conferring ET6497 activity is isolated using standard techniques.

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:3 is subcloned into an appropriate expression vector, and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the GT1773 activity is confirmed. Protein conferring GT1773 activity is isolated using standard techniques.

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:5, is subcloned into an appropriate expression vector, and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the GT0992 activity is confirmed. Protein conferring GT0992 activity is isolated using standard techniques.

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:7, is subcloned into an appropriate expression vector, and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the ET6233 activity is confirmed. Protein conferring ET6233 activity is isolated using standard techniques.

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:9, is subcloned into an appropriate expression vector, and transformed into *E. coli* using the

manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the ET0763 activity is confirmed. Protein conferring ET0763 activity is isolated using standard techniques.

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:11, is subcloned into an appropriate expression vector, and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the ET5848 activity is confirmed. Protein conferring ET5848 activity is isolated using standard techniques.

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:13, is subcloned into an appropriate expression vector, and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the GT5062 activity is confirmed. Protein conferring GT5062 activity is isolated using standard techniques.

The coding region of the protein, corresponding to the cDNA clone SEQ ID NO:15, is subcloned into an appropriate expression vector, and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the ET5036 activity is confirmed. Protein conferring ET5036 activity is isolated using standard techniques.

Example 11: *In vitro* Recombination of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 Genes by DNA Shuffling

The nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15 is amplified by PCR. The resulting DNA fragment is digested by DNaseI treatment essentially as described (Stemmer et al. (1994) PNAS 91: 10747-10751) and the PCR primers are removed from the reaction

mixture. A PCR reaction is carried out without primers and is followed by a PCR reaction with the primers, both as described (Stemmer et al. (1994) PNAS 91: 10747-10751). The resulting DNA fragments are cloned into pTRC99a (Pharmacia, Cat no: 27-5007-01) for use in bacteria, and transformed into a bacterial strain deficient in ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity by electroporation using the Biorad Gene Pulser and the manufacturer's conditions. The transformed bacteria are grown on medium that contains inhibitory concentrations of an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity, respectively, and those colonies that grow in the presence of the inhibitor are selected. Colonies that grow in the presence of normally inhibitory concentrations of inhibitor are picked and purified by repeated restreaking. Their plasmids are purified and the DNA sequences of cDNA inserts from plasmids that pass this test are then determined. Alternatively, the DNA fragments are cloned into expression vectors for transient or stable transformation into plant cells, which are screened for differential survival and/or growth in the presence of an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity. In a similar reaction, PCR-amplified DNA fragments comprising the *Arabidopsis* ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene encoding the protein and PCR-amplified DNA fragments derived from or comprising another ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 gene are recombined *in vitro* and resulting variants with improved tolerance to the inhibitor are recovered as described above.

Example 12: *In vitro* Recombination of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 Genes by Staggered Extension Process

The *Arabidopsis* ET6497 gene encoding the protein and another ET6497 gene, or homolog thereof, or fragment thereof, are each cloned into the polylinker of a pBluescript vector. A PCR reaction is carried out essentially as described (Zhao et al. (1998) Nature Biotechnology 16: 258-261) using the "reverse primer" and the "M13 -20 primer" (Stratagene Catalog). Amplified PCR fragments are digested with appropriate restriction enzymes and cloned into pTRC99a and mutated ET6497 genes are screened as described in Example 10. The same procedure is carried out with genes encoding GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 proteins, respectively.

Example 13: *In Vitro* Binding Assays

Recombinant ET6497 protein is obtained, for example, according to Example 10. The protein is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample compound in solution according to methods well known in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Recombinant GT1773 protein is obtained, for example, according to Example 10. The protein is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample compound in solution according to methods well known in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Recombinant GT0992 protein is obtained, for example, according to Example 10. The protein is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample compound in solution according to methods well known in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Recombinant ET6233 protein is obtained, for example, according to Example 10. The protein is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample compound in solution according to methods well known in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Recombinant ET0763 protein is obtained, for example, according to Example 10. The protein is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample compound in solution according to methods well know in the art. While the sample compound is in contact
5 with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Recombinant ET5848 protein is obtained, for example, according to Example 10. The protein
10 is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample compound in solution according to methods well know in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described
15 above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Recombinant GT5062 protein is obtained, for example, according to Example 10. The protein is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample
20 compound in solution according to methods well know in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Recombinant ET5036 protein is obtained, for example, according to Example 10. The protein is immobilized on chips appropriate for ligand binding assays using techniques which are well known in the art. The protein immobilized on the chip is exposed to sample compound in solution according to methods well know in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting
25 protein-ligand interactions are conducted. Examples of such measurements are SELDI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

Example 14: Plastid Transformation

Transformation vectors

For expression of a nucleotide sequence encoding a polypeptide having ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity encoding in plant plastids, plastid transformation vector pPH143 or pPH145 (WO 97/32011) is used; and this reference is incorporated herein by reference. The nucleotide sequence is inserted into pPH143 thereby replacing the PROTOX coding sequence. This vector is then used for plastid transformation and selection of transformants for spectinomycin resistance. Alternatively, the nucleotide sequence is inserted in pPH143 so that it replaces the *aadH* gene. In this case, transformants are selected for resistance to PROTOX inhibitors.

Plastid Transformation

Seeds of *Nicotiana tabacum* c.v. 'Xanthi nc' are germinated seven per plate in a 1" circular array on T agar medium and bombarded 12-14 days after sowing with 1 μ m tungsten particles (M10, Biorad, Hercules, CA) coated with DNA from plasmids pPH143 and pPH145 essentially as described (Svab, Z. and Maliga, P. (1993) *Proc. Natl. Acad. Sci. USA* 90, 913–917). Bombarded seedlings are incubated on T medium for two days after which leaves are excised and placed abaxial side up in bright light (350-500 μ mol photons/m²/s) on plates of RMOP medium (Svab, Z., Hajdukiewicz, P. and Maliga, P. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8526–8530) containing 500 μ g/ml spectinomycin dihydrochloride (Sigma, St. Louis, MO). Resistant shoots appearing underneath the bleached leaves three to eight weeks after bombardment are subcloned onto the same selective medium, allowed to form callus, and secondary shoots isolated and subcloned. Complete segregation of transformed plastid genome copies (homoplasmy) in independent subclones is assessed by standard techniques of Southern blotting (Sambrook et al., (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor). Homoplasmic shoots are rooted aseptically on spectinomycin-containing MS/IBA medium (McBride, K. E. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91, 7301-7305) and transferred to the greenhouse.

The above disclosed embodiments are illustrative. This disclosure of the invention will place one skilled in the art in possession of many variations of the invention. All such obvious and foreseeable variations are intended to be encompassed by the appended claims.

What is Claimed Is:

1. An isolated DNA molecule comprising a nucleotide sequence encoding SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID
5 NO:14, or SEQ ID NO:16.
2. The DNA molecule of claim 1, comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or SEQ ID NO:15.
3. The DNA molecule according to claim 1, wherein said nucleotide sequence is a plant nucleotide sequence.
- 10 4. The DNA molecule of claim 1, wherein the amino acid sequence has ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity.
5. A polypeptide comprising an amino acid sequence selected from the group consisting of: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, and SEQ ID NO:16.
- 15 6. The polypeptide of claim 5, wherein said amino acid sequence has ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity.
7. A polypeptide comprising an amino acid sequence comprising at least 20 consecutive amino acid residues of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16,
20 wherein the amino acid sequence has ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity.
8. An expression cassette comprising a promoter operatively linked to a DNA molecule according to claim 1.
9. A recombinant vector comprising an expression cassette according to claim 8.
- 25 10. A host cell comprising an expression cassette according to claim 8.
11. A host cell according to claim 10, wherein said host cell is selected from the group consisting of an insect cell, a yeast cell, a prokaryotic cell and a plant cell.
12. A host cell according to claim 10, which is a plant cell.
13. A plant or seed comprising a plant cell of claim 12.

14. A plant of claim 13, wherein said plant is tolerant to an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity.

15. A method for selecting a compound for its ability to interact with a polypeptide according to claim 5, comprising:

- 5 a) combining said polypeptide and a compound to be tested for the ability to interact with said polypeptide, under conditions conducive to interaction; and
 b) selecting a compound identified in step (a) that is capable of interacting with said polypeptide.

16. The method according to claim 15, further comprising:

- 10 c) applying a compound selected in step (b) to a plant to test for herbicidal activity; and
 d) selecting compounds having herbicidal activity.

17. A process of identifying an inhibitor of ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity comprising:

- 15 a) introducing a DNA molecule according to claim 4 into a plant cell, such that said sequence is functionally expressible at levels that are higher than wild-type expression levels;
 b) combining said plant cell with a compound to be tested for the ability to inhibit the ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036
20 activity under conditions conducive to such inhibition;
 c) measuring plant cell growth under the conditions of step (b);
 d) comparing the growth of said plant cell with the growth of a plant cell having unaltered ET6497, GT1773, GT0992, ET6233, ET0763, ET5848, GT5062, or ET5036 activity under identical conditions; and
25 e) selecting said compound that inhibits plant cell growth in step (d).

SEQUENCE LISTING

<110> Syngenta Participations AG

<120> Herbicide Target Genes and Methods

<130> PB/5-50025A

<140>

<141>

<150> US 60/260,967

<151> 2001-01-11

<160> 38

<170> PatentIn Ver. 2.1

<210> 1

<211> 1269

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (1)..(1269)

<400> 1

atg aat agc tcg atg acg acg agt cca acg gct cga gcc ttc ctt cat	48
Met Asn Ser Ser Met Thr Thr Ser Pro Thr Ala Arg Ala Phe Leu His	
1 5 10 15	

ctt cac tcg tca cca ttc act tta cgt tcg aga aag tca ccg tct gga	96
Leu His Ser Ser Pro Phe Thr Leu Arg Ser Arg Lys Ser Pro Ser Gly	
20 25 30	

gtt tgt cta gtc tcg tgg ccg cct tct cgg tct cgc agg act cga cga	144
Val Cys Leu Val Ser Trp Pro Pro Ser Arg Ser Arg Arg Thr Arg Arg	
35 40 45	

ctc gtt tgt atg gcc gaa cct tac ctc att agg aaa atg gag tct gtt	192
Leu Val Cys Met Ala Glu Pro Tyr Leu Ile Arg Lys Met Glu Ser Val	
50 55 60	

gag aag aca tgg aag gag tta tcg gtg aaa cta gct gac cca gat gtt	240
Glu Lys Thr Trp Lys Glu Leu Ser Val Lys Leu Ala Asp Pro Asp Val	
65 70 75 80	

gtt tcg aat caa tct gaa tac caa aaa ctg gca caa tcc atg tcg gag	288
Val Ser Asn Gln Ser Glu Tyr Gln Lys Leu Ala Gln Ser Met Ser Glu	
85 90 95	

ctt gac gag gtt gta aca gtc ttc aga aga ttc aag gac tgt gag aag	336
Leu Asp Glu Val Val Thr Val Phe Arg Arg Phe Lys Asp Cys Glu Lys	
100 105 110	

cag tta cta gaa agc aaa gtt ttg gcg aaa gag gct gga gat gat gaa	384
---	-----

Gln	Leu	Leu	Glu	Ser	Lys	Val	Leu	Ala	Lys	Glu	Ala	Gly	Asp	Asp	Glu		
	115						120					125					
gat	atg	gct	gaa	atg	ata	gga	agt	gaa	atc	aac	tcg	ctc	act	aaa	gaa	432	
Asp	Met	Ala	Glu	Met	Ile	Gly	Ser	Glu	Ile	Asn	Ser	Leu	Thr	Lys	Glu		
	130					135				140							
att	gag	gag	ctt	gaa	aaa	caa	ctt	aag	gtg	tta	ctt	ctt	cct	tcc	gat	480	
Ile	Glu	Glu	Leu	Glu	Lys	Gln	Leu	Lys	Val	Leu	Leu	Leu	Pro	Ser	Asp		
145					150				155						160		
cct	ctt	gat	gcc	aga	aac	att	ttg	ctt	gaa	ggt	aga	gca	gga	act	ggt	528	
Pro	Leu	Asp	Ala	Arg	Asn	Ile	Leu	Leu	Glu	Val	Arg	Ala	Gly	Thr	Gly		
				165					170					175			
ggc	gat	gaa	gct	gca	att	tgg	act	ggt	gac	ctt	gtg	cgg	atg	tat	cag	576	
Gly	Asp	Glu	Ala	Ala	Ile	Trp	Thr	Gly	Asp	Leu	Val	Arg	Met	Tyr	Gln		
		180						185					190				
agg	tac	agc	gag	cgc	agc	tct	tgg	aag	ttt	tca	atg	ggt	tca	tgc	tca	624	
Arg	Tyr	Ser	Glu	Arg	Ser	Ser	Trp	Lys	Phe	Ser	Met	Val	Ser	Cys	Ser		
		195					200					205					
gag	gct	gaa	cat	gga	ggt	tac	aag	acg	tgt	gtg	atg	gag	att	aaa	gga	672	
Glu	Ala	Glu	His	Gly	Gly	Tyr	Lys	Thr	Cys	Val	Met	Glu	Ile	Lys	Gly		
	210					215				220							
aat	cgt	gtc	tat	agc	aag	ctc	aag	tat	gaa	tct	ggt	ggt	cat	cga	gtc	720	
Asn	Arg	Val	Tyr	Ser	Lys	Leu	Lys	Tyr	Glu	Ser	Gly	Val	His	Arg	Val		
225					230				235						240		
cag	cgt	ggt	cct	caa	aca	gaa	aca	cag	ggt	cgt	gta	cac	act	tct	aca	768	
Gln	Arg	Val	Pro	Gln	Thr	Glu	Thr	Gln	Gly	Arg	Val	His	Thr	Ser	Thr		
				245					250					255			
gca	act	gta	gcc	atc	atg	cca	gag	gct	gat	gaa	gtg	gaa	gtg	gtg	att	816	
Ala	Thr	Val	Ala	Ile	Met	Pro	Glu	Ala	Asp	Glu	Val	Glu	Val	Val	Ile		
		260					265					270					
gac	cca	aag	gac	att	gaa	ctt	aca	tct	gca	aga	tct	ggc	ggc	gct	gga	864	
Asp	Pro	Lys	Asp	Ile	Glu	Leu	Thr	Ser	Ala	Arg	Ser	Gly	Gly	Ala	Gly		
		275					280					285					
ggc	caa	aat	gtg	aac	aag	gtg	gag	acg	gct	ata	gat	cta	ttc	cac	aaa	912	
Gly	Gln	Asn	Val	Asn	Lys	Val	Glu	Thr	Ala	Ile	Asp	Leu	Phe	His	Lys		
	290				295				300								
cag	tca	ggt	att	cgc	att	ttc	tgt	aca	gaa	gaa	agg	acc	cag	att	agg	960	
Pro	Ser	Gly	Ile	Arg	Ile	Phe	Cys	Thr	Glu	Glu	Arg	Thr	Gln	Ile	Arg		
305					310				315						320		
aat	aag	gcg	cgt	gct	ttc	cag	ctt	ctg	cga	gca	aaa	ctg	tat	gaa	ata	1008	
Asn	Lys	Ala	Arg	Ala	Phe	Gln	Leu	Leu	Arg	Ala	Lys	Leu	Tyr	Glu	Ile		
				325				330				335					
aaa	gta	agg	gaa	caa	caa	gag	aag	ata	agg	aat	gaa	agg	aaa	tct	cag	1056	
Lys	Val	Arg	Glu	Gln	Gln	Glu	Lys	Ile	Arg	Asn	Glu	Arg	Lys	Ser	Gln		

340	345	350	
gtt ggg act gga gct cgt tca gaa aag ata aga aca tac aat tat aag			1104
Val Gly Thr Gly Ala Arg Ser Glu Lys Ile Arg Thr Tyr Asn Tyr Lys			
355	360	365	
gac agt aga gtg act gat cat agg cta aag atg aac ttt gcg ctt aca			1152
Asp Ser Arg Val Thr Asp His Arg Leu Lys Met Asn Phe Ala Leu Thr			
370	375	380	
aca ttt ctt gac ggt gct tta gag gac gca gtg cag gct tgt gct gct			1200
Thr Phe Leu Asp Gly Ala Leu Glu Asp Ala Val Gln Ala Cys Ala Ala			
385	390	395	400
ttg gag caa aag gaa ctc atg gaa gaa ctc tcc gag tct gta gct gct			1248
Leu Glu Gln Lys Glu Leu Met Glu Glu Leu Ser Glu Ser Val Ala Ala			
405	410	415	
tct tct gct act tct gga taa			1269
Ser Ser Ala Thr Ser Gly			
420			

<210> 2

<211> 422

<212> PRT

<213> Arabidopsis thaliana

<400> 2

Met Asn Ser Ser Met Thr Thr Ser Pro Thr Ala Arg Ala Phe Leu His	
1 5 10 15	
Leu His Ser Ser Pro Phe Thr Leu Arg Ser Arg Lys Ser Pro Ser Gly	
20 25 30	
Val Cys Leu Val Ser Trp Pro Pro Ser Arg Ser Arg Arg Thr Arg Arg	
35 40 45	
Leu Val Cys Met Ala Glu Pro Tyr Leu Ile Arg Lys Met Glu Ser Val	
50 55 60	
Glu Lys Thr Trp Lys Glu Leu Ser Val Lys Leu Ala Asp Pro Asp Val	
65 70 75 80	
Val Ser Asn Gln Ser Glu Tyr Gln Lys Leu Ala Gln Ser Met Ser Glu	
85 90 95	
Leu Asp Glu Val Val Thr Val Phe Arg Arg Phe Lys Asp Cys Glu Lys	
100 105 110	
Gln Leu Leu Glu Ser Lys Val Leu Ala Lys Glu Ala Gly Asp Asp Glu	
115 120 125	
Asp Met Ala Glu Met Ile Gly Ser Glu Ile Asn Ser Leu Thr Lys Glu	
130 135 140	
Ile Glu Glu Leu Glu Lys Gln Leu Lys Val Leu Leu Leu Pro Ser Asp	
145 150 155 160	
Pro Leu Asp Ala Arg Asn Ile Leu Leu Glu Val Arg Ala Gly Thr Gly	
165 170 175	
Gly Asp Glu Ala Ala Ile Trp Thr Gly Asp Leu Val Arg Met Tyr Gln	
180 185 190	
Arg Tyr Ser Glu Arg Ser Ser Trp Lys Phe Ser Met Val Ser Cys Ser	
195 200 205	
Glu Ala Glu His Gly Gly Tyr Lys Thr Cys Val Met Glu Ile Lys Gly	
210 215 220	

```

Asn Arg Val Tyr Ser Lys Leu Lys Tyr Glu Ser Gly Val His Arg Val
225          230          235          240
Gln Arg Val Pro Gln Thr Glu Thr Gln Gly Arg Val His Thr Ser Thr
          245          250          255
Ala Thr Val Ala Ile Met Pro Glu Ala Asp Glu Val Glu Val Val Ile
          260          265          270
Asp Pro Lys Asp Ile Glu Leu Thr Ser Ala Arg Ser Gly Gly Ala Gly
          275          280          285
Gly Gln Asn Val Asn Lys Val Glu Thr Ala Ile Asp Leu Phe His Lys
          290          295          300
Pro Ser Gly Ile Arg Ile Phe Cys Thr Glu Glu Arg Thr Gln Ile Arg
305          310          315          320
Asn Lys Ala Arg Ala Phe Gln Leu Leu Arg Ala Lys Leu Tyr Glu Ile
          325          330          335
Lys Val Arg Glu Gln Gln Glu Lys Ile Arg Asn Glu Arg Lys Ser Gln
          340          345          350
Val Gly Thr Gly Ala Arg Ser Glu Lys Ile Arg Thr Tyr Asn Tyr Lys
          355          360          365
Asp Ser Arg Val Thr Asp His Arg Leu Lys Met Asn Phe Ala Leu Thr
          370          375          380
Thr Phe Leu Asp Gly Ala Leu Glu Asp Ala Val Gln Ala Cys Ala Ala
385          390          395          400
Leu Glu Gln Lys Glu Leu Met Glu Glu Leu Ser Glu Ser Val Ala Ala
          405          410          415
Ser Ser Ala Thr Ser Gly
          420

```

<210> 3
 <211> 1791
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> CDS
 <222> (1)..(1791)

```

<400> 3
atg gct ctt tca ttt cga att tct tcg tcg tct ccg tta atc tgc aga      48
Met Ala Leu Ser Phe Arg Ile Ser Ser Ser Ser Pro Leu Ile Cys Arg
   1              5              10              15

gcc act ctc tca aac ggc gac aac agt cgt aac tac cac acg act gat      96
Ala Thr Leu Ser Asn Gly Asp Asn Ser Arg Asn Tyr His Thr Thr Asp
          20              25              30

gcc gct ttt atc cga cgc gcc gcc gat ttg tcc gaa atg tcc gcc gga      144
Ala Ala Phe Ile Arg Arg Ala Ala Asp Leu Ser Glu Met Ser Ala Gly
          35              40              45

ctt act tca cct cac cct aat ttc ggc tgt gta atc gcc act tct tcc      192
Leu Thr Ser Pro His Pro Asn Phe Gly Cys Val Ile Ala Thr Ser Ser
          50              55              60

ggg aaa gtc gcc gga gaa gga tat cta tac gct caa ggt act aaa ccg      240
Gly Lys Val Ala Gly Glu Gly Tyr Leu Tyr Ala Gln Gly Thr Lys Pro

```

65	70					75					80					
gcg gaa gcg ttg gct gtt gaa gct gcc gga gag ttt tct aga ggc gcc	288															
Ala Glu Ala Leu Val Glu Ala Ala Gly Glu Phe Ser Arg Gly Ala	85					90					95					
act gct tat tta aac atg gag cct ggt gat tgt cat ggt gac cac act	336															
Thr Ala Tyr Leu Asn Met Glu Pro Gly Asp Cys His Gly Asp His Thr	100					105					110					
gct gta tct gct ctt gtt cag gca gga att gaa cga gtt gtt gtt gga	384															
Ala Val Ser Ala Leu Val Gln Ala Gly Ile Glu Arg Val Val Val Gly	115					120					125					
ata agg cat cct ttg caa cat ctt aga ggc tct gcg att cgt gaa tta	432															
Ile Arg His Pro Leu Gln His Leu Arg Gly Ser Ala Ile Arg Glu Leu	130					135					140					
aga agt cat ggt atc gaa gtt aat gtt ctt gga gaa gac ttt gag agt	480															
Arg Ser His Gly Ile Glu Val Asn Val Leu Gly Glu Asp Phe Glu Ser	145					150					155					160
aaa gtt ctt gag gat gct cgg aaa tcg tgt ctt ctt gtg aat gct cct	528															
Lys Val Leu Glu Asp Ala Arg Lys Ser Cys Leu Leu Val Asn Ala Pro	165					170					175					
ttg att cat aga gct tgt tct cga gtt ccg ttc tct gtt ctt aag tat	576															
Leu Ile His Arg Ala Cys Ser Arg Val Pro Phe Ser Val Leu Lys Tyr	180					185					190					
gcc atg act ctt gat ggt aag att gca gcg agt agt gga cat gca gca	624															
Ala Met Thr Leu Asp Gly Lys Ile Ala Ala Ser Ser Gly His Ala Ala	195					200					205					
tgg ata agt agt aaa ctc tct aga acc cga gtg ttc gag ttg cga gga	672															
Trp Ile Ser Ser Lys Leu Ser Arg Thr Arg Val Phe Glu Leu Arg Gly	210					215					220					
gga agt gat gct gtg ata gtt gga gga aac aca gta cgc caa gat gat	720															
Gly Ser Asp Ala Val Ile Val Gly Gly Asn Thr Val Arg Gln Asp Asp	225					230					235					240
cct cga ttg act gca aga cac gga cag ggt cac acc cct aca cga att	768															
Pro Arg Leu Thr Ala Arg His Gly Gln His Thr Pro Thr Arg Ile	245					250					255					
gtg atg act cag agt ctt gat ctt cct gag aaa gct aat cta tgg gat	816															
Val Met Thr Gln Ser Leu Asp Leu Pro Glu Lys Ala Asn Leu Trp Asp	260					265					270					
gtc tcg gaa gta tct acc ata gtt gtt acc caa agg ggt gca aga aag	864															
Val Ser Glu Val Ser Thr Ile Val Val Thr Gln Arg Gly Ala Arg Lys	275					280					285					
agt ttc cag aaa ttg ctt gcg tcc aag ggt gtt gaa gtt gtg gag ttt	912															
Ser Phe Gln Lys Leu Leu Ala Ser Lys Gly Val Glu Val Val Glu Phe	290					295					300					

gac atg ttg aat cca aga gaa gtt atg gaa tat ttc cat ctc cgt gga	960
Asp Met Leu Asn Pro Arg Glu Val Met Glu Tyr Phe His Leu Arg Gly	
305 310 315 320	
tat ctt tct ata cta tgg gag tgt ggg gga aca tta gct gca tct gct	1008
Tyr Leu Ser Ile Leu Trp Glu Cys Gly Gly Thr Leu Ala Ala Ser Ala	
325 330 335	
ata tca tcc agt gtc atc cac aag gtt gtt gct ttc gtt gct cca aaa	1056
Ile Ser Ser Ser Val Ile His Lys Val Val Ala Phe Val Ala Pro Lys	
340 345 350	
att atc ggt ggg agt aaa gca cct tcc cct gtg ggt gac ctt ggg atg	1104
Ile Ile Gly Gly Ser Lys Ala Pro Ser Pro Val Gly Asp Leu Gly Met	
355 360 365	
gta gag atg aca caa gca ttg aat tta att gat gtt tgt tat gag cag	1152
Val Glu Met Thr Gln Ala Leu Asn Leu Ile Asp Val Cys Tyr Glu Gln	
370 375 380	
gtt ggt ccc gac atg ctt gtc agc gga ttt ctt cag ccc ata cca gac	1200
Val Gly Pro Asp Met Leu Val Ser Gly Phe Leu Gln Pro Ile Pro Asp	
385 390 395 400	
ctc tta cct gtt atc cca tca gag gat gca act gtc gag att gac ccc	1248
Leu Leu Pro Val Ile Pro Ser Glu Asp Ala Thr Val Glu Ile Asp Pro	
405 410 415	
agt gtt gac ccc ttt gaa cct agt atc ata ttc ttc tac aaa aca tgg	1296
Ser Val Asp Pro Phe Glu Pro Ser Ile Ile Phe Phe Tyr Lys Thr Trp	
420 425 430	
gac ctc tat gga agt ttc tca aat ttt tct cca cat cca att cga atg	1344
Asp Leu Tyr Gly Ser Phe Ser Asn Phe Ser Pro His Pro Ile Arg Met	
435 440 445	
ccc tat ggc gag gag aat gat gat tac atg acc tgg tcc agt gtg gaa	1392
Pro Tyr Gly Glu Glu Asn Asp Asp Tyr Met Thr Trp Ser Ser Val Glu	
450 455 460	
cat tac tat cag gca aac aag ttt gtt ggg gtc gaa aat cca aag gca	1440
His Tyr Tyr Gln Ala Asn Lys Phe Val Gly Val Glu Asn Pro Lys Ala	
465 470 475 480	
tat gat tgc gtt gag aaa att aga aca gct aga agt cca gag gag gca	1488
Tyr Asp Cys Val Glu Lys Ile Arg Thr Ala Arg Ser Pro Glu Glu Ala	
485 490 495	
gct ctg ata ggc aga tca aca ctg aga cag aaa cct gaa ctg gtt cgg	1536
Ala Leu Ile Gly Arg Ser Thr Leu Arg Gln Lys Pro Glu Leu Val Arg	
500 505 510	
aat gat tgg gaa gat gtg aag ata gaa gta atg tac aag gct tta aaa	1584
Asn Asp Trp Glu Asp Val Lys Ile Glu Val Met Tyr Lys Ala Leu Lys	
515 520 525	

tgc aag ttc tca act tat cct cac ttg aaa tca atg ttg ctc tcg aca 1632
 Cys Lys Phe Ser Thr Tyr Pro His Leu Lys Ser Met Leu Leu Ser Thr
 530 535 540

 att gga aca gtt ctc gtg gag gct tca ccg cac gat cta ttc tgg gga 1680
 Ile Gly Thr Val Leu Val Glu Ala Ser Pro His Asp Leu Phe Trp Gly
 545 550 555 560

 ggt ggc cgt gaa gga gaa ggg ctt aac tat ctt ggt cga ctg ctt atg 1728
 Gly Gly Arg Glu Gly Glu Gly Leu Asn Tyr Leu Gly Arg Leu Leu Met
 565 570 575

 cag ctt cgt tct gag tat ctg ggc gaa tcc tct gtt tca gct gaa aag 1776
 Gln Leu Arg Ser Glu Tyr Leu Gly Glu Ser Ser Val Ser Ala Glu Lys
 580 585 590

 act tct tcg gcc tga 1791
 Thr Ser Ser Ala
 595

<210> 4
 <211> 596
 <212> PRT
 <213> Arabidopsis thaliana

<400> 4
 Met Ala Leu Ser Phe Arg Ile Ser Ser Ser Ser Pro Leu Ile Cys Arg
 1 5 10 15
 Ala Thr Leu Ser Asn Gly Asp Asn Ser Arg Asn Tyr His Thr Thr Asp
 20 25 30
 Ala Ala Phe Ile Arg Arg Ala Ala Asp Leu Ser Glu Met Ser Ala Gly
 35 40 45
 Leu Thr Ser Pro His Pro Asn Phe Gly Cys Val Ile Ala Thr Ser Ser
 50 55 60
 Gly Lys Val Ala Gly Glu Gly Tyr Leu Tyr Ala Gln Gly Thr Lys Pro
 65 70 75 80
 Ala Glu Ala Leu Ala Val Glu Ala Ala Gly Glu Phe Ser Arg Gly Ala
 85 90 95
 Thr Ala Tyr Leu Asn Met Glu Pro Gly Asp Cys His Gly Asp His Thr
 100 105 110
 Ala Val Ser Ala Leu Val Gln Ala Gly Ile Glu Arg Val Val Val Gly
 115 120 125
 Ile Arg His Pro Leu Gln His Leu Arg Gly Ser Ala Ile Arg Glu Leu
 130 135 140
 Arg Ser His Gly Ile Glu Val Asn Val Leu Gly Glu Asp Phe Glu Ser
 145 150 155 160
 Lys Val Leu Glu Asp Ala Arg Lys Ser Cys Leu Leu Val Asn Ala Pro
 165 170 175
 Leu Ile His Arg Ala Cys Ser Arg Val Pro Phe Ser Val Leu Lys Tyr
 180 185 190
 Ala Met Thr Leu Asp Gly Lys Ile Ala Ala Ser Ser Gly His Ala Ala
 195 200 205
 Trp Ile Ser Ser Lys Leu Ser Arg Thr Arg Val Phe Glu Leu Arg Gly
 210 215 220
 Gly Ser Asp Ala Val Ile Val Gly Gly Asn Thr Val Arg Gln Asp Asp
 225 230 235 240

```

Pro Arg Leu Thr Ala Arg His Gly Gln Gly His Thr Pro Thr Arg Ile
                245                250                255
Val Met Thr Gln Ser Leu Asp Leu Pro Glu Lys Ala Asn Leu Trp Asp
                260                265                270
Val Ser Glu Val Ser Thr Ile Val Val Thr Gln Arg Gly Ala Arg Lys
                275                280                285
Ser Phe Gln Lys Leu Leu Ala Ser Lys Gly Val Glu Val Val Glu Phe
                290                295                300
Asp Met Leu Asn Pro Arg Glu Val Met Glu Tyr Phe His Leu Arg Gly
305                310                315                320
Tyr Leu Ser Ile Leu Trp Glu Cys Gly Gly Thr Leu Ala Ala Ser Ala
                325                330                335
Ile Ser Ser Ser Val Ile His Lys Val Val Ala Phe Val Ala Pro Lys
                340                345                350
Ile Ile Gly Gly Ser Lys Ala Pro Ser Pro Val Gly Asp Leu Gly Met
                355                360                365
Val Glu Met Thr Gln Ala Leu Asn Leu Ile Asp Val Cys Tyr Glu Gln
                370                375                380
Val Gly Pro Asp Met Leu Val Ser Gly Phe Leu Gln Pro Ile Pro Asp
385                390                395                400
Leu Leu Pro Val Ile Pro Ser Glu Asp Ala Thr Val Glu Ile Asp Pro
                405                410                415
Ser Val Asp Pro Phe Glu Pro Ser Ile Ile Phe Phe Tyr Lys Thr Trp
                420                425                430
Asp Leu Tyr Gly Ser Phe Ser Asn Phe Ser Pro His Pro Ile Arg Met
                435                440                445
Pro Tyr Gly Glu Glu Asn Asp Asp Tyr Met Thr Trp Ser Ser Val Glu
                450                455                460
His Tyr Tyr Gln Ala Asn Lys Phe Val Gly Val Glu Asn Pro Lys Ala
465                470                475                480
Tyr Asp Cys Val Glu Lys Ile Arg Thr Ala Arg Ser Pro Glu Glu Ala
                485                490                495
Ala Leu Ile Gly Arg Ser Thr Leu Arg Gln Lys Pro Glu Leu Val Arg
                500                505                510
Asn Asp Trp Glu Asp Val Lys Ile Glu Val Met Tyr Lys Ala Leu Lys
                515                520                525
Cys Lys Phe Ser Thr Tyr Pro His Leu Lys Ser Met Leu Leu Ser Thr
                530                535                540
Ile Gly Thr Val Leu Val Glu Ala Ser Pro His Asp Leu Phe Trp Gly
545                550                555                560
Gly Gly Arg Glu Gly Glu Gly Leu Asn Tyr Leu Gly Arg Leu Leu Met
                565                570                575
Gln Leu Arg Ser Glu Tyr Leu Gly Glu Ser Ser Val Ser Ala Glu Lys
                580                585                590
Thr Ser Ser Ala
                595

```

<210> 5
 <211> 1155
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> CDS
 <222> (1) .. (1155)

```

<400> 5
atg agc tta gct tca gtt atc cac cac gga att ctc cct ccg gcc aaa 48
Met Ser Leu Ala Ser Val Ile His His Gly Ile Leu Pro Pro Ala Lys
  1             5             10             15

tcg gat cga atc ttt ctt acc att ccg gtg ttc ccg ccg gat ttc aga 96
Ser Asp Arg Ile Phe Leu Thr Ile Pro Val Phe Pro Pro Asp Phe Arg
             20             25             30

gct cgt gga tgg act aag tct cct ttc tct ctc ctt ata aat ccc tct 144
Ala Arg Gly Trp Thr Lys Ser Pro Phe Ser Leu Leu Ile Asn Pro Ser
             35             40             45

ctt gct tca gct gcc aat cgt cgt ctc tct cat ctt cct cct atc gcg 192
Leu Ala Ser Ala Ala Asn Arg Arg Leu Ser His Leu Pro Pro Ile Ala
             50             55             60

tgt tcc aga ggt att gat cag gaa gat gaa gag aag gag agc aga gaa 240
Cys Ser Arg Gly Ile Asp Gln Glu Asp Glu Glu Lys Glu Ser Arg Glu
             65             70             75             80

ctt cta cca cac aaa aac gat gag aat gca acc act tca cgc tct tct 288
Leu Leu Pro His Lys Asn Asp Glu Asn Ala Thr Thr Ser Arg Ser Ser
             85             90             95

tca tca gta gat agt gga ggc ctg aaa gat tat caa caa gaa gaa act 336
Ser Ser Val Asp Ser Gly Gly Leu Lys Asp Tyr Gln Gln Glu Glu Thr
             100             105             110

tac aaa act tcc ttc aag act gtt gcc ttg tgt gtt ggc act gcg gtg 384
Tyr Lys Thr Ser Phe Lys Thr Val Ala Leu Cys Val Gly Thr Ala Val
             115             120             125

gca ttt gga att gga ata ggt ttg aag gaa ggt gtt ggc aag gct tca 432
Ala Phe Gly Ile Gly Ile Gly Leu Lys Glu Gly Val Gly Lys Ala Ser
             130             135             140

gag ttt ttc gca ggc tac ata ttg gag caa agc ttg tca gtg gac aac 480
Glu Phe Phe Ala Gly Tyr Ile Leu Glu Gln Ser Leu Ser Val Asp Asn
             145             150             155             160

ttg ttt gtt ttt gtt ctt gtt ttc aag tat ttc aaa gtg cca ctc atg 528
Leu Phe Val Phe Val Leu Val Phe Lys Tyr Phe Lys Val Pro Leu Met
             165             170             175

tat cag aat aag gta ctt acc tac ggt ata gct ggt gca att gtt ttc 576
Tyr Gln Asn Lys Val Leu Thr Tyr Gly Ile Ala Gly Ala Ile Val Phe
             180             185             190

cgt ttc acg ctc att tta cta gga aca gct act ctt cag aaa ttt gaa 624
Arg Phe Thr Leu Ile Leu Leu Gly Thr Ala Thr Leu Gln Lys Phe Glu
             195             200             205

gca gtc aac cta ctg ctt gct gct gta ctc tta tat tcg tca ttc aag 672
Ala Val Asn Leu Leu Leu Ala Ala Val Leu Leu Tyr Ser Ser Phe Lys
             210             215             220

```

```

tta ttt gca agt gaa gaa gat gat aca gat cta tcg gac aac ttc att 720
Leu Phe Ala Ser Glu Glu Asp Asp Thr Asp Leu Ser Asp Asn Phe Ile
225                230                235                240

gta aag aca tgc cag aga ttt att cct gtc aca tca agt tac gat gga 768
Val Lys Thr Cys Gln Arg Phe Ile Pro Val Thr Ser Ser Tyr Asp Gly
                245                250                255

aat cga ttt ttc aca aag cat gat ggc att ttg aaa gcc aca cca ttg 816
Asn Arg Phe Phe Thr Lys His Asp Gly Ile Leu Lys Ala Thr Pro Leu
                260                265                270

cta ctt aca gta gca gtg atc gag ctc agt gac ata gca ttt gcg gtt 864
Leu Leu Thr Val Ala Val Ile Glu Leu Ser Asp Ile Ala Phe Ala Val
                275                280                285

gac tca ata cca gct gtt ttt ggc gtc aca agg gat cct ttt att gtc 912
Asp Ser Ile Pro Ala Val Phe Gly Val Thr Arg Asp Pro Phe Ile Val
                290                295                300

ttg act tct aat ctc ttt gca atc cta gga cta agg tct ctc tat aca 960
Leu Thr Ser Asn Leu Phe Ala Ile Leu Gly Leu Arg Ser Leu Tyr Thr
305                310                315                320

ctg att tct gaa gga atg gac gag ctg gaa tac ttg cag cca tca ata 1008
Leu Ile Ser Glu Gly Met Asp Glu Leu Glu Tyr Leu Gln Pro Ser Ile
                325                330                335

gca gtt gtt ctt ggc ttc ata gga gtc aag atg atc ctt gat ttc ttt 1056
Ala Val Val Leu Gly Phe Ile Gly Val Lys Met Ile Leu Asp Phe Phe
                340                345                350

ggc ttc cac ata tca aca gag gca tca ctt ggt gtc gtt gca ctt tct 1104
Gly Phe His Ile Ser Thr Glu Ala Ser Leu Gly Val Val Ala Leu Ser
                355                360                365

ctc agt acc gga gta ctg ttg agc cta aca aac aaa tcc agc gac agc 1152
Leu Ser Thr Gly Val Leu Leu Ser Leu Thr Asn Lys Ser Ser Asp Ser
                370                375                380

taa 1155

385

```

<210> 6

<211> 384

<212> PRT

<213> Arabidopsis thaliana

<400> 6

```

Met Ser Leu Ala Ser Val Ile His His Gly Ile Leu Pro Pro Ala Lys
 1          5          10          15
Ser Asp Arg Ile Phe Leu Thr Ile Pro Val Phe Pro Pro Asp Phe Arg
                20                25                30
Ala Arg Gly Trp Thr Lys Ser Pro Phe Ser Leu Leu Ile Asn Pro Ser

```

```
<210> 7
<211> 678
<212> DNA
<213> Arabidopsis thaliana
```

<400> 7
atg acg act cag tcg cag ttc tta gtg aac ctc tct tac gga ggc ctc 48

Met	Thr	Thr	Gln	Ser	Gln	Phe	Leu	Val	Asn	Leu	Ser	Tyr	Gly	Gly	Leu	
1				5					10					15		
gcc	tcc	caa	agt	aat	ctc	aga	gcg	aat	aat	cgt	gtt	tca	cct	tct	tct	96
Ala	Ser	Gln	Ser	Asn	Leu	Arg	Ala	Asn	Asn	Arg	Val	Ser	Pro	Ser	Ser	
			20					25					30			
tgt	caa	att	aca	cga	aca	aat	cgc	tca	tggt	gca	tta	cca	gtt	tcc	ttt	144
Cys	Gln	Ile	Thr	Arg	Thr	Asn	Arg	Ser	Trp	Ala	Leu	Pro	Val	Ser	Phe	
		35					40					45				
aaa	gtt	gag	aaa	ttt	caa	ctc	caa	cga	gga	agg	agg	aga	aga	gga	tca	192
Lys	Val	Glu	Lys	Phe	Gln	Leu	Gln	Arg	Gly	Arg	Arg	Arg	Arg	Gly	Ser	
	50					55			60							
cca	tgt	ttt	ggg	gag	agc	gaa	ggg	ttg	gtc	aaa	aat	ggg	att	ggt	gat	240
Pro	Cys	Phe	Gly	Glu	Ser	Glu	Gly	Leu	Val	Lys	Asn	Gly	Ile	Gly	Asp	
	65				70				75						80	
gct	gat	ggg	att	ata	att	gtt	gat	cac	ggt	tct	cgt	cgc	cgt	gaa	tcc	288
Ala	Asp	Gly	Ile	Ile	Ile	Val	Asp	His	Gly	Ser	Arg	Arg	Arg	Glu	Ser	
				85				90						95		
aat	ctt	atg	ctt	gag	gag	ttt	gtg	aaa	atg	ttc	aaa	gaa	aaa	act	ggt	336
Asn	Leu	Met	Leu	Glu	Glu	Phe	Val	Lys	Met	Phe	Lys	Glu	Lys	Thr	Gly	
			100					105					110			
tac	cca	att	gta	gag	cct	gct	cat	atg	gaa	ttg	gct	gag	cca	tct	ata	384
Tyr	Pro	Ile	Val	Glu	Pro	Ala	His	Met	Glu	Leu	Ala	Glu	Pro	Ser	Ile	
		115					120					125				
aaa	gat	gca	ttc	agt	ttg	tgt	gta	caa	caa	ggg	gcg	aaa	cgt	gta	gtt	432
Lys	Asp	Ala	Phe	Ser	Leu	Cys	Val	Gln	Gln	Gly	Ala	Lys	Arg	Val	Val	
	130					135				140						
gtg	agc	ccg	ttt	ttt	cta	ttt	cct	gga	aga	cat	tggt	cat	acg	gac	att	480
Val	Ser	Pro	Phe	Phe	Leu	Phe	Pro	Gly	Arg	His	Trp	His	Thr	Asp	Ile	
	145				150				155					160		
ccc	tcg	ttg	acg	gct	gat	gct	gca	aag	gag	ttc	tct	ggc	atc	tca	tac	528
Pro	Ser	Leu	Thr	Ala	Asp	Ala	Ala	Lys	Glu	Phe	Ser	Gly	Ile	Ser	Tyr	
			165					170					175			
ctc	ata	act	gca	cct	ctc	ggc	ccc	cat	aat	ctc	ctc	ctg	gac	gtt	gta	576
Leu	Ile	Thr	Ala	Pro	Leu	Gly	Pro	His	Asn	Leu	Leu	Leu	Asp	Val	Val	
			180					185					190			
aat	gac	aga	atc	caa	cat	tgc	ttg	agc	cat	gtt	gaa	ggt	gat	gct	gac	624
Asn	Asp	Arg	Ile	Gln	His	Cys	Leu	Ser	His	Val	Glu	Gly	Asp	Ala	Asp	
		195					200					205				
gag	tgt	ctt	gtt	tgt	gct	gga	aca	aat	aag	tgc	aag	ctc	tac	aat	agt	672
Glu	Cys	Leu	Val	Cys	Ala	Gly	Thr	Asn	Lys	Cys	Lys	Leu	Tyr	Asn	Ser	
	210					215					220					
tct	tag															678
Ser																

225

<210> 8
 <211> 225
 <212> PRT
 <213> Arabidopsis thaliana

<400> 8
 Met Thr Thr Gln Ser Gln Phe Leu Val Asn Leu Ser Tyr Gly Gly Leu
 1 5 10 15
 Ala Ser Gln Ser Asn Leu Arg Ala Asn Asn Arg Val Ser Pro Ser Ser
 20 25 30
 Cys Gln Ile Thr Arg Thr Asn Arg Ser Trp Ala Leu Pro Val Ser Phe
 35 40 45
 Lys Val Glu Lys Phe Gln Leu Gln Arg Gly Arg Arg Arg Gly Ser
 50 55 60
 Pro Cys Phe Gly Glu Ser Glu Gly Leu Val Lys Asn Gly Ile Gly Asp
 65 70 75 80
 Ala Asp Gly Ile Ile Ile Val Asp His Gly Ser Arg Arg Arg Glu Ser
 85 90 95
 Asn Leu Met Leu Glu Glu Phe Val Lys Met Phe Lys Glu Lys Thr Gly
 100 105 110
 Tyr Pro Ile Val Glu Pro Ala His Met Glu Leu Ala Glu Pro Ser Ile
 115 120 125
 Lys Asp Ala Phe Ser Leu Cys Val Gln Gln Gly Ala Lys Arg Val Val
 130 135 140
 Val Ser Pro Phe Phe Leu Phe Pro Gly Arg His Trp His Thr Asp Ile
 145 150 155 160
 Pro Ser Leu Thr Ala Asp Ala Ala Lys Glu Phe Ser Gly Ile Ser Tyr
 165 170 175
 Leu Ile Thr Ala Pro Leu Gly Pro His Asn Leu Leu Leu Asp Val Val
 180 185 190
 Asn Asp Arg Ile Gln His Cys Leu Ser His Val Glu Gly Asp Ala Asp
 195 200 205
 Glu Cys Leu Val Cys Ala Gly Thr Asn Lys Cys Lys Leu Tyr Asn Ser
 210 215 220
 Ser
 225

<210> 9
 <211> 1503
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> CDS
 <222> (1)..(1503)

<400> 9
 atg caa acg ctt ctc tgt cag cca tgt aag tct ctt cct ata ctc act 48
 Met Gln Thr Leu Leu Cys Gln Pro Cys Lys Ser Leu Pro Ile Leu Thr
 1 5 10 15
 gct tcg tcg tcg tcg tcg ttg att cgt agc tcc ggt gat gtg cga gaa 96

Ala	Ser	Ser	Ser	Ser	Ser	Leu	Ile	Arg	Ser	Ser	Gly	Asp	Val	Arg	Glu		
			20					25					30				
tgt	atc	gat	ttt	aga	gcg	tcg	gag	aaa	gtc	tcg	aag	ttt	caa	ttt	cat	144	
Cys	Ile	Asp	Phe	Arg	Ala	Ser	Glu	Lys	Val	Ser	Lys	Phe	Gln	Phe	His		
		35					40					45					
gta	acc	cta	agc	ccc	ttt	gcc	ttt	cgt	ggg	ttc	tca	ata	tgt	cgc	gaa	192	
Val	Thr	Leu	Ser	Pro	Phe	Ala	Phe	Arg	Gly	Phe	Ser	Ile	Cys	Arg	Glu		
	50					55				60							
ttt	gct	gtt	cgt	gga	gct	tat	ggg	att	aga	ttt	tgc	tca	cga	gaa	gat	240	
Phe	Ala	Val	Arg	Gly	Ala	Tyr	Gly	Ile	Arg	Phe	Cys	Ser	Arg	Glu	Asp		
65					70				75						80		
gtt	tcc	ggg	gtt	gga	aat	gga	gga	atc	gtg	gcg	gag	gag	gag	att	gag	288	
Val	Ser	Gly	Val	Gly	Asn	Gly	Gly	Ile	Val	Ala	Glu	Glu	Glu	Ile	Glu		
				85				90						95			
cta	ttg	aac	aag	ccg	aat	cct	tta	cca	aaa	tca	gaa	aat	gag	gag	agt	336	
Leu	Leu	Asn	Lys	Pro	Asn	Pro	Leu	Pro	Lys	Ser	Glu	Asn	Glu	Glu	Ser		
			100					105					110				
ggg	aaa	gca	gac	gat	gac	gcg	att	ctg	gag	ccg	ttt	ttg	aaa	ttc	ttt	384	
Gly	Lys	Ala	Asp	Asp	Asp	Ala	Ile	Leu	Glu	Pro	Phe	Leu	Lys	Phe	Phe		
		115					120					125					
aaa	ccg	gaa	gaa	gaa	ggg	gaa	gga	att	gaa	tcg	gaa	gtt	tcg	gat	gag	432	
Lys	Pro	Glu	Glu	Glu	Gly	Glu	Gly	Ile	Glu	Ser	Glu	Val	Ser	Asp	Glu		
	130					135				140							
acg	gat	aga	gtc	agc	gta	gag	tat	tat	gat	cct	aaa	cca	ggg	gat	ttt	480	
Thr	Asp	Arg	Val	Ser	Val	Glu	Tyr	Tyr	Asp	Pro	Lys	Pro	Gly	Asp	Phe		
145					150					155					160		
gtg	gtt	ggg	gtt	gtg	gtt	tcg	ggg	aat	gag	aat	aag	ctt	gat	gtg	aat	528	
Val	Val	Gly	Val	Val	Val	Ser	Gly	Asn	Glu	Asn	Lys	Leu	Asp	Val	Asn		
				165					170					175			
att	ggg	gca	gat	atg	ttg	gga	act	atg	ttg	acg	aaa	gag	ata	ctt	cct	576	
Ile	Gly	Ala	Asp	Met	Leu	Gly	Thr	Met	Leu	Thr	Lys	Glu	Ile	Leu	Pro		
			180					185					190				
ttg	tat	gat	aaa	gag	ttg	gat	tat	ttg	ttg	tgt	gat	ttg	aag	tat	gat	624	
Leu	Tyr	Asp	Lys	Glu	Leu	Asp	Tyr	Leu	Leu	Cys	Asp	Leu	Lys	Tyr	Asp		
		195					200					205					
gct	gaa	gag	ttt	ttg	gtt	aat	ggg	aaa	atg	ggg	att	gtt	aag	gat	gat	672	
Ala	Glu	Glu	Phe	Leu	Val	Asn	Gly	Lys	Met	Gly	Ile	Val	Lys	Asp	Asp		
	210					215				220							
gat	gaa	ggg	gtt	gag	att	gcg	gag	ttt	gct	cgg	caa	ggg	agg	ccg	gtg	720	
Asp	Glu	Gly	Val	Glu	Ile	Ala	Glu	Phe	Ala	Arg	Gln	Gly	Arg	Pro	Val		
225					230					235					240		
gtg	gag	att	ggg	acg	gtt	gtt	ttc	gcg	gag	gtt	tta	ggg	agg	acg	ttg	768	
Val	Glu	Ile	Gly	Thr	Val	Val	Phe	Ala	Glu	Val	Leu	Gly	Arg	Thr	Leu		

245								250				255				
agt	gga	aga	cct	ttg	ctt	tct	tct	aga	cgg	tat	ttt	cgg	aga	atc	gct	816
Ser	Gly	Arg	Pro	Leu	Leu	Ser	Ser	Arg	Arg	Tyr	Phe	Arg	Arg	Ile	Ala	
			260					265					270			
tg	cac	cga	gtg	agg	cag	atc	aaa	caa	ctt	aat	gag	cct	att	gaa	gtt	864
Trp	His	Arg	Val	Arg	Gln	Ile	Lys	Gln	Leu	Asn	Glu	Pro	Ile	Glu	Val	
			275				280					285				
aag	ata	aca	gag	tg	aat	act	gga	ggg	ctt	ctc	acc	aga	att	gag	ggc	912
Lys	Ile	Thr	Glu	Trp	Asn	Thr	Gly	Gly	Leu	Leu	Thr	Arg	Ile	Glu	Gly	
			290			295					300					
ttg	aga	gct	ttt	att	ccg	aag	cag	gaa	ctg	gtg	aag	aaa	gtg	aac	act	960
Leu	Arg	Ala	Phe	Ile	Pro	Lys	Gln	Glu	Leu	Val	Lys	Lys	Val	Asn	Thr	
305					310					315					320	
ttc	acc	gag	ttg	aag	gaa	aac	gtg	ggt	cgt	aga	ttc	ctt	gtg	cag	ata	1008
Phe	Thr	Glu	Leu	Lys	Glu	Asn	Val	Gly	Arg	Arg	Phe	Leu	Val	Gln	Ile	
				325					330					335		
acg	cga	ttg	aac	gaa	gac	aaa	aac	gac	ttg	ata	ctg	agt	gaa	aaa	gta	1056
Thr	Arg	Leu	Asn	Glu	Asp	Lys	Asn	Asp	Leu	Ile	Leu	Ser	Glu	Lys	Val	
			340					345					350			
gcc	tg	gag	aag	ctg	tac	ctt	cga	gaa	gga	aca	ctc	tta	gaa	gga	aca	1104
Ala	Trp	Glu	Lys	Leu	Tyr	Leu	Arg	Glu	Gly	Thr	Leu	Leu	Glu	Gly	Thr	
			355				360					365				
gtt	gtc	aaa	atc	tta	cca	tat	gga	gct	caa	gtc	aaa	ctt	ggg	gat	agt	1152
Val	Val	Lys	Ile	Leu	Pro	Tyr	Gly	Ala	Gln	Val	Lys	Leu	Gly	Asp	Ser	
			370			375					380					
agc	aga	agt	gga	ctg	cta	cac	att	tca	aac	ata	act	cga	aga	aga	att	1200
Ser	Arg	Ser	Gly	Leu	Leu	His	Ile	Ser	Asn	Ile	Thr	Arg	Arg	Arg	Ile	
385					390					395					400	
gga	tcg	gtg	agc	gat	gtg	ctt	cag	gtg	gac	gaa	agt	gta	aag	gtt	ctg	1248
Gly	Ser	Val	Ser	Asp	Val	Leu	Gln	Val	Asp	Glu	Ser	Val	Lys	Val	Leu	
				405					410					415		
gtt	gtg	aaa	tct	cta	ttt	cca	gac	aag	atc	tct	ctt	agc	att	gct	gat	1296
Val	Val	Lys	Ser	Leu	Phe	Pro	Asp	Lys	Ile	Ser	Leu	Ser	Ile	Ala	Asp	
			420					425					430			
ctt	gaa	agc	gaa	cca	gga	ttg	ttc	atc	tca	gac	aga	gag	aaa	gtg	ttt	1344
Leu	Glu	Ser	Glu	Pro	Gly	Leu	Phe	Ile	Ser	Asp	Arg	Glu	Lys	Val	Phe	
			435				440					445				
acg	gaa	gct	gaa	gag	atg	gcg	aag	aag	tac	aga	gaa	aaa	atg	cct	ttg	1392
Thr	Glu	Ala	Glu	Glu	Met	Ala	Lys	Lys	Tyr	Arg	Glu	Lys	Met	Pro	Leu	
			450			455					460					
gtg	gct	aca	agt	cca	att	tct	gat	cgt	cct	ccg	atc	aca	agc	agt	ttc	1440
Val	Ala	Thr	Ser	Pro	Ile	Ser	Asp	Arg	Pro	Pro	Ile	Thr	Ser	Ser	Phe	
465					470					475					480	

cca cag ggc aag gat gaa gaa atc tat gct aac tgg gag tgg ttc aag 1488
 Pro Gln Gly Lys Asp Glu Glu Ile Tyr Ala Asn Trp Glu Trp Phe Lys
 485 490 495

ttc gag agt cag tga 1503
 Phe Glu Ser Gln
 500

<210> 10
 <211> 500
 <212> PRT
 <213> Arabidopsis thaliana

<400> 10
 Met Gln Thr Leu Leu Cys Gln Pro Cys Lys Ser Leu Pro Ile Leu Thr
 1 5 10 15
 Ala Ser Ser Ser Ser Ser Leu Ile Arg Ser Ser Gly Asp Val Arg Glu
 20 25 30
 Cys Ile Asp Phe Arg Ala Ser Glu Lys Val Ser Lys Phe Gln Phe His
 35 40 45
 Val Thr Leu Ser Pro Phe Ala Phe Arg Gly Phe Ser Ile Cys Arg Glu
 50 55 60
 Phe Ala Val Arg Gly Ala Tyr Gly Ile Arg Phe Cys Ser Arg Glu Asp
 65 70 75 80
 Val Ser Gly Val Gly Asn Gly Gly Ile Val Ala Glu Glu Glu Ile Glu
 85 90 95
 Leu Leu Asn Lys Pro Asn Pro Leu Pro Lys Ser Glu Asn Glu Glu Ser
 100 105 110
 Gly Lys Ala Asp Asp Asp Ala Ile Leu Glu Pro Phe Leu Lys Phe Phe
 115 120 125
 Lys Pro Glu Glu Glu Gly Glu Gly Ile Glu Ser Glu Val Ser Asp Glu
 130 135 140
 Thr Asp Arg Val Ser Val Glu Tyr Tyr Asp Pro Lys Pro Gly Asp Phe
 145 150 155 160
 Val Val Gly Val Val Val Ser Gly Asn Glu Asn Lys Leu Asp Val Asn
 165 170 175
 Ile Gly Ala Asp Met Leu Gly Thr Met Leu Thr Lys Glu Ile Leu Pro
 180 185 190
 Leu Tyr Asp Lys Glu Leu Asp Tyr Leu Leu Cys Asp Leu Lys Tyr Asp
 195 200 205
 Ala Glu Glu Phe Leu Val Asn Gly Lys Met Gly Ile Val Lys Asp Asp
 210 215 220
 Asp Glu Gly Val Glu Ile Ala Glu Phe Ala Arg Gln Gly Arg Pro Val
 225 230 235 240
 Val Glu Ile Gly Thr Val Val Phe Ala Glu Val Leu Gly Arg Thr Leu
 245 250 255
 Ser Gly Arg Pro Leu Leu Ser Ser Arg Arg Tyr Phe Arg Arg Ile Ala
 260 265 270
 Trp His Arg Val Arg Gln Ile Lys Gln Leu Asn Glu Pro Ile Glu Val
 275 280 285
 Lys Ile Thr Glu Trp Asn Thr Gly Gly Leu Leu Thr Arg Ile Glu Gly
 290 295 300
 Leu Arg Ala Phe Ile Pro Lys Gln Glu Leu Val Lys Lys Val Asn Thr
 305 310 315 320
 Phe Thr Glu Leu Lys Glu Asn Val Gly Arg Arg Phe Leu Val Gln Ile

Glu	Pro	Pro	Leu	Leu	Leu	Ser	Gly	Lys	Arg	Gln	Met	Leu	Thr	Asp	Asn		
				85					90					95			
cac	ata	gca	cca	gtt	ttc	acg	gtt	tct	aat	ccc	att	ctt	ggt	ttg	ctc	336	
His	Ile	Ala	Pro	Val	Phe	Thr	Val	Ser	Asn	Pro	Ile	Leu	Gly	Leu	Leu		
			100					105					110				
tgt	cat	caa	gga	tat	aat	aat	gag	att	gtg	att	tgc	aat	ccc	ggt	ctg	384	
Cys	His	Gln	Gly	Tyr	Asn	Asn	Glu	Ile	Val	Ile	Cys	Asn	Pro	Gly	Leu		
		115					120					125					
aaa	aag	ttt	cgg	aat	tta	ccg	caa	atc	caa	gtg	cct	gaa	gga	gca	tcc	432	
Lys	Lys	Phe	Arg	Asn	Leu		Gln	Ile	Gln	Val	Pro	Glu	Gly	Ala	Ser		
		130				135					140						
gtg	aga	agt	ttt	ttc	gga	tac	gat	aaa	gtt	gac	aaa	gta	ttc	aag	gtg	480	
Val	Arg	Ser	Phe	Phe	Gly	Tyr	Asp	Lys	Val	Asp	Lys	Val	Phe	Lys	Val		
145					150					155					160		
tta	tgt	gta	tgt	gta	cca	cag	aca	cag	ttc	gga	gct	ggt	tgg	aca	tca	528	
Leu	Cys	Val	Cys	Val	Pro	Gln	Thr	Gln	Phe	Gly	Ala	Gly	Trp	Thr	Ser		
				165					170					175			
aaa	gaa	cct	aag	aaa	tat	cag	gtc	tac	acg	gtg	aga	tcg	gat	cat	gaa	576	
Lys	Glu	Pro	Lys	Lys	Tyr	Gln	Val	Tyr	Thr	Val	Arg	Ser	Asp	His	Glu		
			180					185					190				
aaa	cct	tcc	ttt	tgg	aga	ccg	atc	att	tgt	aat	gat	gat	cat	tca	gta	624	
Lys	Pro	Ser	Phe	Trp	Arg	Pro	Ile	Ile	Cys	Asn	Asp	Asp	His	Ser	Val		
		195					200					205					
gtt	act	gat	gaa	gga	cta	ttt	aat	gga	ggt	ttc	ttg	tac	tat	gga	gct	672	
Val	Thr	Asp	Glu	Gly	Leu	Phe	Asn	Gly	Gly	Phe	Leu	Tyr	Tyr	Gly	Ala		
		210				215						220					
cgg	tcg	agt	tgc	aac	aaa	cct	atg	gta	atg	agg	ttc	aac	gtg	agt	tct	720	
Arg	Ser	Ser	Cys	Asn	Lys	Pro	Met	Val	Met	Arg	Phe	Asn	Val	Ser	Ser		
225					230					235					240		
gaa	gcg	ttt	gct	gtc	att	aaa	ttg	cct	gag	gaa	gtc	gac	att	gat	tat	768	
Glu	Ala	Phe	Ala	Val	Ile	Lys	Leu	Pro	Glu	Glu	Val	Asp	Ile	Asp	Tyr		
				245					250					255			
cat	tgg	aga	ttg	gtg	aac	tac	aaa	ggg	aaa	gtt	gcc	tta	gtg	aac	acc	816	
His	Trp	Arg	Leu	Val	Asn	Tyr	Lys	Gly	Lys	Val	Ala	Leu	Val	Asn	Thr		
			260					265					270				
ttt	gac	tct	aaa	ttt	aaa	ttc	aat	gga	gtt	ttt	gaa	ata	tgg	gtt	agg	864	
Phe	Asp	Ser	Lys	Phe	Lys	Phe	Asn	Gly	Val	Phe	Glu	Ile	Trp	Val	Arg		
		275					280					285					
aac	gag	gtt	gat	aga	aag	tgg	gac	aag	gac	atc	att	aag	att	cct	caa	912	
Asn	Glu	Val	Asp	Arg	Lys	Trp	Asp	Lys	Asp	Ile	Ile	Lys	Ile	Pro	Gln		
		290				295					300						
tgg	aca	gaa	tct	gtt	gat	aat	tat	aat	gtt	tat	ttc	aaa	ggt	acc	act	960	
Trp	Thr	Glu	Ser	Val	Asp	Asn	Tyr	Asn	Val	Tyr	Phe	Lys	Gly	Thr	Thr		

```

305                310                315                320
gga aca aaa gaa ctt gtt ttc gca cca ccc aat tgg ttc ggg gaa cca 1008
Gly Thr Lys Glu Leu Val Phe Ala Pro Pro Asn Trp Phe Gly Glu Pro
                325                330                335

ctc ttt gtc ttg tat tac gat aaa gcc aca act aac ctc aga aga ttt 1056
Leu Phe Val Leu Tyr Tyr Asp Lys Ala Thr Thr Asn Leu Arg Arg Phe
                340                345                350

gat att gaa gga atg gat gat caa cat tat tct ttt cat acc ttc ttg 1104
Asp Ile Glu Gly Met Asp Asp Gln His Tyr Ser Phe His Thr Phe Leu
                355                360                365

gat cat gtt gat agc act tgg ctt atg tga 1134
Asp His Val Asp Ser Thr Trp Leu Met
                370                375

```

<210> 12

<211> 377

<212> PRT

<213> Arabidopsis thaliana

<400> 12

```

Met Val Asn Thr Ser Phe Glu Thr Leu Ala Leu Asp Met Gln Ile Glu
 1          5          10          15
Ile Leu Ala Arg Leu Pro Leu Lys Tyr Leu Met Arg Cys Met Cys Val
          20          25          30
Ser Lys Lys Trp Ala Ser Leu Ile Arg Gly Glu Asp Phe Arg Ser Ala
          35          40          45
Tyr Leu Leu Arg Ser Met Ala Arg Pro Arg Leu Leu Phe Val Ala Ser
          50          55          60
Arg Ser Arg Lys Phe Thr Arg Glu Thr Cys Phe His Ser Val Tyr Gln
          65          70          75          80
Glu Pro Pro Leu Leu Leu Ser Gly Lys Arg Gln Met Leu Thr Asp Asn
          85          90          95
His Ile Ala Pro Val Phe Thr Val Ser Asn Pro Ile Leu Gly Leu Leu
          100          105          110
Cys His Gln Gly Tyr Asn Asn Glu Ile Val Ile Cys Asn Pro Gly Leu
          115          120          125
Lys Lys Phe Arg Asn Leu Pro Gln Ile Gln Val Pro Glu Gly Ala Ser
          130          135          140
Val Arg Ser Phe Phe Gly Tyr Asp Lys Val Asp Lys Val Phe Lys Val
          145          150          155          160
Leu Cys Val Cys Val Pro Gln Thr Gln Phe Gly Ala Gly Trp Thr Ser
          165          170          175
Lys Glu Pro Lys Lys Tyr Gln Val Tyr Thr Val Arg Ser Asp His Glu
          180          185          190
Lys Pro Ser Phe Trp Arg Pro Ile Ile Cys Asn Asp Asp His Ser Val
          195          200          205
Val Thr Asp Glu Gly Leu Phe Asn Gly Gly Phe Leu Tyr Tyr Gly Ala
          210          215          220
Arg Ser Ser Cys Asn Lys Pro Met Val Met Arg Phe Asn Val Ser Ser
          225          230          235          240
Glu Ala Phe Ala Val Ile Lys Leu Pro Glu Glu Val Asp Ile Asp Tyr
          245          250          255

```

```

His Trp Arg Leu Val Asn Tyr Lys Gly Lys Val Ala Leu Val Asn Thr
      260      265      270
Phe Asp Ser Lys Phe Lys Phe Asn Gly Val Phe Glu Ile Trp Val Arg
      275      280      285
Asn Glu Val Asp Arg Lys Trp Asp Lys Asp Ile Ile Lys Ile Pro Gln
      290      295      300
Trp Thr Glu Ser Val Asp Asn Tyr Asn Val Tyr Phe Lys Gly Thr Thr
      305      310      315      320
Gly Thr Lys Glu Leu Val Phe Ala Pro Pro Asn Trp Phe Gly Glu Pro
      325      330      335
Leu Phe Val Leu Tyr Tyr Asp Lys Ala Thr Thr Asn Leu Arg Arg Phe
      340      345      350
Asp Ile Glu Gly Met Asp Asp Gln His Tyr Ser Phe His Thr Phe Leu
      355      360      365
Asp His Val Asp Ser Thr Trp Leu Met
      370      375

```

<210> 13
 <211> 891
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> CDS
 <222> (1)..(891)

```

<400> 13
atg gcg tgc tca att ggt aac gct ttc agg tgt tct tct gat acg ctg 48
Met Ala Cys Ser Ile Gly Asn Ala Phe Arg Cys Ser Ser Asp Thr Leu
  1              5              10              15

cgt ttt gct ccc cga caa caa tgc agc agc cgc cta aac cct aat cct 96
Arg Phe Ala Pro Arg Gln Gln Cys Ser Ser Arg Leu Asn Pro Asn Pro
      20              25              30

tcc tct ttc ctc tcc ttt aat tcc tcc cca att ctc gca cag aat ttg 144
Ser Ser Phe Leu Ser Phe Asn Ser Ser Pro Ile Leu Ala Gln Asn Leu
      35              40              45

ggt gct tct tct tcg tca ctt tcc cga cgg act ata cga gct aga aca 192
Gly Ala Ser Ser Ser Ser Leu Ser Arg Arg Thr Ile Arg Ala Arg Thr
      50              55              60

aaa atg gct gcg tcg gag gct tcc att aac ggg agc aac agg atg ctg 240
Lys Met Ala Ala Ser Glu Ala Ser Ile Asn Gly Ser Asn Arg Met Leu
      65              70              75              80

gtc ttt gtg cca cca cat cct ttg atc aaa cac tgg atc tct gtt ctg 288
Val Phe Val Pro Pro His Pro Leu Ile Lys His Trp Ile Ser Val Leu
      85              90              95

aga aat gaa caa act cct tgt cct gtt ttc agg aat gca ata gct gaa 336
Arg Asn Glu Gln Thr Pro Cys Pro Val Phe Arg Asn Ala Ile Ala Glu
      100              105              110

```

```

cta ggg aga tta ctc atg tat gaa gca tca cga gag tgg ctg cca act 384
Leu Gly Arg Leu Leu Met Tyr Glu Ala Ser Arg Glu Trp Leu Pro Thr
      115                      120                      125

gtt gtg ggt gaa ata atg tct cca atg ggt cct gct tct gtt gaa ttc 432
Val Val Gly Glu Ile Met Ser Pro Met Gly Pro Ala Ser Val Glu Phe
      130                      135                      140

att gat cct aga gag cct att gcg gtt gtt ccg ata tta aga gct ggt 480
Ile Asp Pro Arg Glu Pro Ile Ala Val Val Pro Ile Leu Arg Ala Gly
145                      150                      155                      160

ctt gct ctt gcg gaa cac gca tcc tca gtt tta cct gct aat aaa att 528
Leu Ala Leu Ala Glu His Ala Ser Ser Val Leu Pro Ala Asn Lys Ile
      165                      170                      175

tat cat tta gga gta agt aga gat gag aaa aca ctt tta cca tcg gtg 576
Tyr His Leu Gly Val Ser Arg Asp Glu Lys Thr Leu Leu Pro Ser Val
      180                      185                      190

tat ctt aac aag ttg cct gat gaa ttt ccc aag aat tca aga gta ttt 624
Tyr Leu Asn Lys Leu Pro Asp Glu Phe Pro Lys Asn Ser Arg Val Phe
      195                      200                      205

ttg gtc gac cca gtg ctt gcc aca ggt ggt acc ata atg gcg gcg atg 672
Leu Val Asp Pro Val Leu Ala Thr Gly Gly Thr Ile Met Ala Ala Met
      210                      215                      220

gat cta ttg aag gaa cgt ggt ttg tct gtt cag caa atc aaa gtg att 720
Asp Leu Leu Lys Glu Arg Gly Leu Ser Val Gln Gln Ile Lys Val Ile
225                      230                      235                      240

tgt gcg att gct gca cct cct gca cta tcg aag ctt aac gag aaa ttc 768
Cys Ala Ile Ala Ala Pro Pro Ala Leu Ser Lys Leu Asn Glu Lys Phe
      245                      250                      255

ccc ggg ctt cat gta tat gct gga att att gat cct gaa gtc aat gaa 816
Pro Gly Leu His Val Tyr Ala Gly Ile Ile Asp Pro Glu Val Asn Glu
      260                      265                      270

aag ggg tac ata atc cct ggg ctt gga gac gct gga gac cgc agt ttc 864
Lys Gly Tyr Ile Ile Pro Gly Leu Gly Asp Ala Gly Asp Arg Ser Phe
      275                      280                      285

ggg aca gaa aca cat tgg gtg aag tga 891
Gly Thr Glu Thr His Trp Val Lys
      290                      295

```

<210> 14

<211> 296

<212> PRT

<213> Arabidopsis thaliana

<400> 14

```

Met Ala Cys Ser Ile Gly Asn Ala Phe Arg Cys Ser Ser Asp Thr Leu
  1           5           10           15

```

```

Arg Phe Ala Pro Arg Gln Gln Cys Ser Ser Arg Leu Asn Pro Asn Pro
      20      25      30
Ser Ser Phe Leu Ser Phe Asn Ser Ser Pro Ile Leu Ala Gln Asn Leu
      35      40      45
Gly Ala Ser Ser Ser Ser Leu Ser Arg Arg Thr Ile Arg Ala Arg Thr
      50      55      60
Lys Met Ala Ala Ser Glu Ala Ser Ile Asn Gly Ser Asn Arg Met Leu
      65      70      75      80
Val Phe Val Pro Pro His Pro Leu Ile Lys His Trp Ile Ser Val Leu
      85      90      95
Arg Asn Glu Gln Thr Pro Cys Pro Val Phe Arg Asn Ala Ile Ala Glu
      100      105      110
Leu Gly Arg Leu Leu Met Tyr Glu Ala Ser Arg Glu Trp Leu Pro Thr
      115      120      125
Val Val Gly Glu Ile Met Ser Pro Met Gly Pro Ala Ser Val Glu Phe
      130      135      140
Ile Asp Pro Arg Glu Pro Ile Ala Val Val Pro Ile Leu Arg Ala Gly
      145      150      155      160
Leu Ala Leu Ala Glu His Ala Ser Ser Val Leu Pro Ala Asn Lys Ile
      165      170      175
Tyr His Leu Gly Val Ser Arg Asp Glu Lys Thr Leu Leu Pro Ser Val
      180      185      190
Tyr Leu Asn Lys Leu Pro Asp Glu Phe Pro Lys Asn Ser Arg Val Phe
      195      200      205
Leu Val Asp Pro Val Leu Ala Thr Gly Gly Thr Ile Met Ala Ala Met
      210      215      220
Asp Leu Leu Lys Glu Arg Gly Leu Ser Val Gln Gln Ile Lys Val Ile
      225      230      235      240
Cys Ala Ile Ala Ala Pro Pro Ala Leu Ser Lys Leu Asn Glu Lys Phe
      245      250      255
Pro Gly Leu His Val Tyr Ala Gly Ile Ile Asp Pro Glu Val Asn Glu
      260      265      270
Lys Gly Tyr Ile Ile Pro Gly Leu Gly Asp Ala Gly Asp Arg Ser Phe
      275      280      285
Gly Thr Glu Thr His Trp Val Lys
      290      295

```

<210> 15

<211> 540

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (1)..(540)

<400> 15

```

atg gct tca atg cac gaa gct ctg ttc tct tct cgg ctt ctt caa gtc 48
Met Ala Ser Met His Glu Ala Leu Phe Ser Ser Arg Leu Leu Gln Val
  1          5          10          15

```

```

aat tct tct ttt tcc ttc cgt tgc gct ctt cca att atc tct tca cca 96
Asn Ser Ser Phe Ser Phe Arg Cys Ala Leu Pro Ile Ile Ser Ser Pro
      20          25          30

```

```

gct gca gta tct tgc gca atc aaa tcg acc cag ttt ttt aaa cag cgt 144
Ala Ala Val Ser Cys Ala Ile Lys Ser Thr Gln Phe Phe Lys Gln Arg
      35              40              45

tgt aga acc aaa gtg agg gac ttt tct cta tct tca ctg tcc cgg aga 192
Cys Arg Thr Lys Val Arg Asp Phe Ser Leu Ser Ser Leu Ser Arg Arg
      50              55              60

ggg ttc gtt tgc aga gct gct gag tac aag ttt ccg gat ccc ata cct 240
Gly Phe Val Cys Arg Ala Ala Glu Tyr Lys Phe Pro Asp Pro Ile Pro
      65              70              75              80

gaa ttc gcc gaa gct gag act gag aaa ttt agg gac cat atg tta aac 288
Glu Phe Ala Glu Ala Glu Thr Glu Lys Phe Arg Asp His Met Leu Asn
      85              90              95

aaa cta tca aaa aga gat ctt ttt gaa gac tct gtt gat gaa att gtt 336
Lys Leu Ser Lys Arg Asp Leu Phe Glu Asp Ser Val Asp Glu Ile Val
      100             105             110

gga gtc tgc act gag atc ttc gag acg ttc ttg cgc agt gag tat gga 384
Gly Val Cys Thr Glu Ile Phe Glu Thr Phe Leu Arg Ser Glu Tyr Gly
      115             120             125

gga cct ggt aca ctc ttg gtg ata cct ttc att gac atg gca gat act 432
Gly Pro Gly Thr Leu Leu Val Ile Pro Phe Ile Asp Met Ala Asp Thr
      130             135             140

ctc aac gaa agg gaa ttg cct gga ggt cca caa gcc gcc cga gca gct 480
Leu Asn Glu Arg Glu Leu Pro Gly Gly Pro Gln Ala Ala Arg Ala Ala
      145             150             155             160

att aaa tgg gct cag gat cat gta gac aag gac tgg aag gaa tgg acc 528
Ile Lys Trp Ala Gln Asp His Val Asp Lys Asp Trp Lys Glu Trp Thr
      165             170             175

ggc act gat tag 540
Gly Thr Asp
      180

```

<210> 16

<211> 179

<212> PRT

<213> Arabidopsis thaliana

<400> 16

```

Met Ala Ser Met His Glu Ala Leu Phe Ser Ser Arg Leu Leu Gln Val
  1          5          10          15
Asn Ser Ser Phe Ser Phe Arg Cys Ala Leu Pro Ile Ile Ser Ser Pro
      20          25          30
Ala Ala Val Ser Cys Ala Ile Lys Ser Thr Gln Phe Phe Lys Gln Arg
      35          40          45
Cys Arg Thr Lys Val Arg Asp Phe Ser Leu Ser Ser Leu Ser Arg Arg
      50          55          60
Gly Phe Val Cys Arg Ala Ala Glu Tyr Lys Phe Pro Asp Pro Ile Pro
      65          70          75          80

```

Glu	Phe	Ala	Glu	Ala	Glu	Thr	Glu	Lys	Phe	Arg	Asp	His	Met	Leu	Asn
			85					90						95	
Lys	Leu	Ser	Lys	Arg	Asp	Leu	Phe	Glu	Asp	Ser	Val	Asp	Glu	Ile	Val
			100					105					110		
Gly	Val	Cys	Thr	Glu	Ile	Phe	Glu	Thr	Phe	Leu	Arg	Ser	Glu	Tyr	Gly
			115				120					125			
Gly	Pro	Gly	Thr	Leu	Leu	Val	Ile	Pro	Phe	Ile	Asp	Met	Ala	Asp	Thr
			130				135				140				
Leu	Asn	Glu	Arg	Glu	Leu	Pro	Gly	Gly	Pro	Gln	Ala	Ala	Arg	Ala	Ala
145					150					155					160
Ile	Lys	Trp	Ala	Gln	Asp	His	Val	Asp	Lys	Asp	Trp	Lys	Glu	Trp	Thr
				165					170					175	
Gly	Thr	Asp													

<210> 17

<211> 2726

<212> DNA

<213> Arabidopsis thaliana

<400> 17

```

atgaatagct cgatgacgac gagtccaacg gctcgagcct tccttcacatc tcaactcgta 60
ccattcactt tacgttcgag aaagtcaccg tctggagttt gtctagtctc gtggccgcct 120
tctcgtcttc gcaggactcg acgactcggt tgtatggccg taagtgattc tagttctccc 180
aatgttggat ttttttttca aaagtttcag aattgttcga tgattttgtc tgttgaaaca 240
tttaggaacc ttacctcatt aggaaaatgg agtctgttga gaagacatgg aaggagttat 300
cggtttgttc ctacacttat tccattaatg gttaagggtt gtttcttaat ctcatcttta 360
agagttttga atgatttgta ggtgaaacta gctgaccagc atgttggttc gaatcaatct 420
gaataccaaa aactggcaca atccatgtcg gagcttgacg aggtatttgc acattctttt 480
agtgttcaaa atttgcagct ttttcggggg tttcttaaca ttgaattcaa ttttaggttg 540
taacagtctt cagaagattc aaggactgtg agaagcagtt actagaaagc aaaggtccct 600
tcaagtctcc tcttttcttt tactttgctt tgattctgtt cctcggaagc tggttatgat 660
ttgttcctat atttcattgc tcagttttgg cgaaagaggc tggagatgat gaagatatgg 720
ctgaaatgat aggaagtga atcaactcgc tcaactaaaga aattgaggag cttgaaaaac 780
aacttaaggt ttagagcctt gggtccatag ctgacttatt cggcagctga aaatcgaatg 840
ttgtctattt taaaaatttg attctgatac tcaactcgta acattgtgtt ggcttgtgtc 900
aggtgttact tcttccttcc gatcctcttg atgccagaaa cattttgctt gaaggtaact 960
gactaaacta tatgtcgaca agtagtatca tttgtctaca ctggcctttt tcttgaaaaa 1020
actgcctttt gctcggactt cccgccattg ttttacctat ctacctctgg acgagtctga 1080
aatcttagtt ccttcgcgag ttagagcagg aactggtggc gatgaagctg caatttggac 1140
tggtagcctt gtaggtacat gctttccttg atttcttctg atcctctatc tgattttaat 1200
gccttcgat agttcaccgt tttgtttcag tcttatttct taaaaaactt tcaaatttta 1260
ttgttcctat atttcattgc taagagaagc aagtgcata attaaaaact tcagatccta 1320
ccatgaatca gtatgctttg cattattctt ttgtctgcag gtgcggatgt atcagagcta 1380
cagcgagcgc agctcttggg agttttcaat gggttcatgc tcagaggtga gatgctctct 1440
gtgttttgtt ttggcattta aggagtgtgt aatgaaaatg gcagtagagc ataacaactt 1500
agagacatgt tctcaccagt atagcttttt aatcttctta cttttataag gctgaacatg 1560
gaggttacaa gacgtgtgtg atggagatta aaggaaatcg tgtctatagc aagctcaagt 1620
atgaatctgg tgttcacga gtccagcgtg ttctcaaac agaaacacag ggtcgtgtac 1680
acacttctac agcaactgta gccatcatgc cagaggtaaa ctagctttcg catgttccat 1740
ttttcgatat tctgattgat aattggtaac cagtccaaaa ctgtcatctc tgctttgttt 1800
tttttgtcat ttaggctga tgaagtggaa gtggtgattg acccaaagga cattgaactt 1860
acatctgcaa gatctggcgg cgctggaggt tatgctaaag aaattcgctt tctgtttctc 1920
tcacatttca cttgctaata atactgaaaa gaggtcttgg tgcaggccaa aatgtgaaca 1980
aggtggaagc ggctatagat ctattccaca aaccgtcagg tattcgatt ttctgtacag 2040
aagaaaggac ccagattagg aataaggcgc gtgctttcca gcttctgcga gcaaaactgt 2100

```

```

aagcaacatt caaggagcaa actctctact cttcttctcc tctatcaagt cccaccttag 2160
tctttttttg ttgaaacatt atatgtctat atataggtat gaaataaaag taagggaaca 2220
acaagagaag ataaggaatg aaaggaaaac tcagggtccac ttcattttat tatccccata 2280
gtcttttctaa actcaaggtc atcatcaaat attcatctta ttctgctcta tgtcgattct 2340
tgcggaacaa ggttgggact ggagctcggt cagaaaagat aagaacatac aattataagg 2400
taatacacgg atccagagcc ttgctttgaa aacttgatgg attctgtttt gcttatgatg 2460
tctgaaaatt acaggacagt agagtgactg atcataggct aaagatgaac ttgctgctta 2520
caacatttct tgacggtgct ttagaggacg cagtgcaggt cctaaacact aaaccttaaa 2580
ctttttctta cgtttgattt agttaatgtg aatcttgaag aatgtttgca ttattcttat 2640
tgcaggcttg tgctgctttg gagcaaaagg aactcatgga agaactctcc gagtctgtag 2700
ctgcttcttc tgctacttct ggataa 2726

```

<210> 18

<211> 2562

<212> DNA

<213> *Arabidopsis thaliana*

<400> 18

```

atggctcttt catttcgaat ttcttcgctg tctccgttaa tctgcagagc cactctctca 60
aacggcgaca acagtcgtaa ctaccacacg actgatgccg cttttatccg acgcgccgcc 120
gatttgtccg aaatgtccgc cggacttact tcacctcacc ctaatttcgg ctgtgtaatc 180
gccacttctt ccggttaaagt cgcggagaa ggatatctat acgctcaagg tactaaaccg 240
gcggaagcgt tggctgttga agctgccgga gagttttcta gaggcgccac tgcttattta 300
aacatggagc ctggtgattg tcatggtgac cacactgctg tatctgctct tgttcagggt 360
tttgcctctt tctcttagat ttttctgtga ataagccata agccaatgat tgtgtgatat 420
taggaacaaa ttgaccaa ttgagattgt atagattagg caggaattga acgagttgtt 480
gttgaataaa ggcacacctt gcaacatctt agaggctctg cgattcgtga attaagaagt 540
catggtatcg aagttaatgt tcttgagaa gactttgaga gtaaagttct tgaggatatc 600
atgtcttctt actcttcaat ttgaagttaa ttgtaggat tctgttaact aaatgttcaa 660
gttcttatag gatgctcgga aatcgtgtct tcttgtgaat gtccttttga ttcataagac 720
ttgttctcga gttccgttct ctgttcttaa gtatgccatg actcttgatg gtaactatct 780
caacctccat ttcatatgtt ttaggatag ctgttttagc taacgttttt tttgtttatt 840
gatagtaag attgcagcga ttagtgagaa tgcagcatgg ataagtagta aactctctag 900
aaccgcagtg ttcgagttgc gagggaggaag tgatgctgtg atagttggag gaaacacagt 960
acgccaaagt ggtaaatatt gtatcttaaa tctcgcactt ttacggatgc ttacaacata 1020
ttgaatagat atgtgaaagt tgtgcaattt tgtaactgca gatcctcgat tgactgcaag 1080
acacggacag ggtcacaccc ctacacgaat tgtgatgact cagagtcttg atcttctga 1140
gaaagcta atctatggatg tctcggaagt atctaccata gttgttacc aaaggggtgc 1200
aagaaagagt ttccagaaat tgcttgctgc caaggggtgt gaagttgtgg agtttgacat 1260
gttgaatcca agagaagtta tggaaatatt ccatctccgt ggatatcttt ctatactatg 1320
ggagtgtggg ggaacattag ctgcatctgc tatatcatcc agtgcaccc acaaggtacc 1380
cattatacta tatccatact tatttgtagt catcttatgt ttttgtctgg ctaaacctca 1440
tttctttaca cttctttctc tcaatgtaat tcatttttca ggttggtgct ttcgttgctc 1500
caaaaattat cgggtgggagt aaagcacctt cccctgtggg tgacctggg atggttaga 1560
tgacacaagc attgaattta attgatgttt gttatgagca ggtaatTTTT ttgctttatg 1620
aattaacatt cactgtgaca agttaaccat gttactgaat ctggcctctt agttccgtgg 1680
gtttaatcta aatccttgaa ctacagggtg gtcccacat gcttgtcagc ggatttcttc 1740
agccataacc agacctctta cctgttatcc catcagagga tgcaactgtc gagattgacc 1800
ccagtgttga cccctttgaa cctagtatca tattcttcta caaaacatgg gacctctatg 1860
gaagtttctc aaatttttct ccacatccaa ttcgaatgcc ctatggcgag gagaatgatg 1920
attacatgac ctggtccagt gtggaacatt actatcaggt accacacaac agtccatgta 1980
aaatggatc ttgctctttc aaaaaaacat aatctcttga tccttccacc aaaaacttta 2040
aaggcaaaca agtttggttg ggtcgaaaat ccaaaggcat atgattgcgt tgagaaaatt 2100
agaacagcta gaagtcagga ggaggcagct ctgtaggca gatcaacact gagacagaaa 2160
cctgaactgg tatgaataaa agttgaattg tccagtttcg agcatcaaag cagttttcac 2220
tcgtttgttg atctctcgag aatcgagatc gttattcagg ggatcaataa tactgcattc 2280

```

```

tcgctttaat gattgtttta ggttcggaat gattgggaag atgtgaagat agaagtaatg 2340
tacaaggctt taaaatgcaa gttctcaact taccctcact tgaaatcaat gttgctctcg 2400
acaattggaa cagttctcgt ggaggcttca ccgcacgac tattctgggg aggtggccgt 2460
gaaggagaag ggcttaacta tcttggtcga ctgcttatgc agcttcgttc tgagtatctg 2520
ggcgaatcct ctgtttcagc tgaaaagact tcttcggcct ga 2562

```

<210> 19

<211> 2542

<212> DNA

<213> *Arabidopsis thaliana*

<400> 19

```

atgagcttag cttcagttat ccaccacgga attctccctc cggccaaatc ggatcgaatc 60
tttcttacca ttccggtggt cccgcgggat ttcagagctc gtggatggac taagtctcct 120
ttctctctcc ttataaatcc ctctcttgct tcaggctctc cttctctctt tctttacatt 180
ttctatatcc cttatctgtg tgaatggaag aaggattcat ctatctgagt catgctaatt 240
tgcttctggg ttacgattct ttcagctgcc aatcgctcgc tctctcatct tctctctatc 300
gcgtgttcca gaggtattga tcaggaagat gaagggtgaat gcttcttctc caatattttc 360
ttctgttttt ttttttcctg aattgtctgt gtcagctcta cttttgctag caagacacaa 420
ttgtatataa aggaagaagc tttatcttgt gacaataatt tgtttatggc atgttagtac 480
agagaaggag agcagagaac ttctaccaca caaaaacgat gagaatgcaa ccacttcacg 540
ctcttcttca tcagtagata gtggaggcct gaaagattat caacaagaag aaacttacaa 600
aacttccttc aagactggtg ccttgtgtgt tagtattctc ttgcatctac taagtataaa 660
tgctgttttg atctatgtgc gagacctgat ttcatagtta tctcaatagg ttggcactgc 720
ggtggcattt ggaattggaa taggtttgaa ggaagggtgt ggcaaggctt cagagttttt 780
cgcagggtgt gtagagctta ctatatatga tatattgatt ttcacttaac tcttcaagga 840
gatatttctt atcttaactg ttttcaactc ttgcagctac atattggagc aaagcttgct 900
agtggacaac ttgtttggtt ttgttcttgt tttcaagtat ttcaaagtgc cactcatgta 960
tcaggtagca aaactatctt cacagagcca cagagattta gctcttgact acagagtatt 1020
ctgattctgt gtccctttct atagaataag gtacttacct acggtatagc tgggtgaatt 1080
gttttccgtt tcacgctcat tttactagga acagctactc ttcagggtgat ttctatgttt 1140
acatcaaaa atttgttagt gattatttgt ttcatttttc tgcctcattt acttttgaat 1200
tctctctact ttttttcaga aatttgaagc agtcaacctc ctgcttgctg ctgtactctt 1260
atattctgca ttcaagggtgc atactacctc atcttagcat ttataacct aatcacacac 1320
aattattgat atgggtattag tttttctgtc ctcaatgtgt taatacagtt atttgcaagt 1380
gaagaagatg atacagatct atcggacaac ttcattgtaa agacatgcc gagatttatt 1440
cctgtcacat gtaatatacc cggctttttt ctctttttta ctgtttggta tgcttccata 1500
gctgatttat tctctttttc ttttgtttat aactgcttgc agcaagttac gatggaaatc 1560
gattttttcac aaagcatgat ggcattttga aagtaagtat atgacaaaat tctctctata 1620
tattcaagggt ttatgtataa gtactgttta agttttattg catgtggact ccatgatcaa 1680
aacattcatg atagtacctt tgattacttc taattatttc aacaataatt tgggcaggcc 1740
acaccattgc tacttacagt agcagtgatc gagctcagtg acatagcatt tgcggtaact 1800
tccgtcacca ttttgtgaat ctgtttttgt ccttttagaa atgaagcaaa tcaactgata 1860
tggttgccca ttttcttagg ttgactcaat accagctgtt tttggcgtca caagggatcc 1920
ttttattgtc ttgacttcta atctctttgc aatcctaggt attttttccc taaacacaat 1980
ttttgttaac ttcttgcttg tctcttttga atactcttgg aactttgaat tgaagagttt 2040
atttatttca tctgggtgtt gaatttgccg caggactaag gtctctctat acactgattt 2100
ctgaaggaaat ggacgagctg gaatacttgc aggtaccttt accacttctt taaatcttca 2160
ttcatattct gtttcattac ttcatactcg agctaagcat cagatacata cttaagagtg 2220
ccagtaaatgc tcattatgta ccatggact agtatgattt gtatctaata ttgtctttgc 2280
ctttcccagc catcaatagc agttgttctt ggcttcatag gagtcaagat gatccttgat 2340
ttctttggta tagatattct ctcttcacgt gcaatttctg cttgaactaa accatatatc 2400
tacaagaaaa actaagctct ctcaacattt ctctctctt ctaggcttcc acatatcaac 2460
agaggcatca cttgggtgtc ttgcactttc tctcagtacc ggagtactgt tgagcctaac 2520
aaacaaatcc agcgacagct aa 2542

```

<210> 20
 <211> 1428
 <212> DNA
 <213> Arabidopsis thaliana

<400> 20
 atgacgactc agtcgcagtt cttagtgaac ctctcttacg gaggcctcgc ctcccaaagg 60
 ttcattcttc tttctcttat accccttcta tcgctttctt ctgttggtgt gctttcttct 120
 gaatgatttt tttttgtcg atttggtttt gtagtaatct cagagcgaat aatcgtgttt 180
 caccttcttc ttgtcaaatt acacgaacaa atcgctcatg ggcattacca gtttcttcta 240
 aagttgagaa atttcaactc caacgaggaa ggaggagaag aggatcacca tgttttgggg 300
 agagcgaagg gttggtcaaa aatgggattg gtgatgctga tgggattata attgttgatc 360
 acggttctcg tcgccgtgaa tccaatctta tgcttggtta tagctttgtt tctcttctct 420
 ctcttcttcc attttctttg ttgtccttac tttattgtat agtatacttc ctttacggct 480
 ttaccttggt cgagtttatg taattgtgtt gttgtttctg tatttcagag gagtttgtga 540
 aaatgttcaa agaaaaaact ggttacccaa ttgtagagcc tgctcatatg gtaactcttt 600
 ctaatgttct cacctttatg catagtgaaga accataccca tgagtttgat acagtctgat 660
 ttccgtttct tggatcatct ttgcatttag cttaatatgt tctatatcag gaaattagat 720
 gactttcaag ctctgcttgc tgtttgctga tttttacact ttacatcagg aattggctga 780
 gccatctata aaagatgcat tcagtttgtg tgtacaacaa ggggcgaaac gtgtagttgt 840
 gagcccgttt tttctatttc ctggaagaca ttggcatacg gtatggcttt gttcatctca 900
 atttgttttg ttttcttgaa gcttctagct tttcttattt atgggattat tggaggatgt 960
 acattaatgg cctagtgaaga tcatgacctc tatttgattc aggacattcc ctctgtgacg 1020
 gctgatgctg caaaggagtt ctctggcacc tcataacctc taactgcacc tctcgcccc 1080
 cataatctcc tcttggttaag ttataaagca tattttcccc ccaaaacgag cgagcacttt 1140
 aaacctaagg gatgtgagag catcctcatg cattagtga tggcactttt tttcttgaa 1200
 cttccaatga gagcatcttc atgcattagg gatcttctgc tgtttaatct tcacaagaga 1260
 tctgttcttc ttttggctct tgatctttca gtaatctact aatatttctt tctgcaggac 1320
 gttgtaaatt acagaatcca acattgcttg agccatgttg aaggtgatgc tgacgagtgt 1380
 cttgtttgtg ctggaacaaa taagtgaag ctctacaata gttcttag 1428

<210> 21
 <211> 2150
 <212> DNA
 <213> Arabidopsis thaliana

<400> 21
 atgcaaagc ttctctgtca gccatgtaag tctcttcta tactcactgc ttcgtcgtcg 60
 tcgtcgttga ttcgtagctc cggtagatgt cgagaatgta tcgatttttag agcgtcggag 120
 aaagtctcga agtttcaatt tcatgtaacc ctaagccctt ttgcctttcg tggtttctca 180
 atatgtcgcg aatttgctgt tcgtggagct tatgggatta gattttgctc acgagaagat 240
 gtttccggtg ttggaaatgg aggaatcgtg gcggaggagg agattgagct attgaacaag 300
 ccgaatcctt taccaaaatc agaaaatgag gagagtggta aagcagacga tgacgcgatt 360
 ctggagccgt ttttgaaatt ctttaaaccg gaagaagaag gggaaggaaat tgaatcgga 420
 gtttcggatg agacggatag agtcagcgta gagtattatg atcctaaacc aggtgatttt 480
 gtggttggtg ttgtgggttc gggtaatgag aataagcttg atgtgaatat tgggtcagat 540
 atgttgggaa ctatgttgac gaaagagata ctctcttctg atgataaaga gttggattat 600
 ttgttggtgt atttgaagta tgatgctgaa gagtttttgg ttaatgggaa aatggggatt 660
 gttaaggatg atgatgaagg tgttgagatt gcggagtttg ctccggcaagg taggccggtg 720
 gtggagattg ggacggttgt tttcgcggag gtttttagga ggacgttgag tgggaagacct 780
 ttgctttctt ctagacggta ttttcggaga atcgcttggc accgagttag gcaggttctt 840
 ttcttgtgcc ttgatttatt cattcttttg attcaagaac tgttcttggg ttgaagtatt 900
 ttgtgtcatt gagaagcatc taatgaagag tgaattgtta tgtggcagat caaacaactt 960
 aatgagccta ttgaagttaa gataacagag tggaatactg gagggcttct caccagaatt 1020
 gaggttgctg aatttttctg gttgattttg aggatttaca tcttgctcac aacatagctc 1080

```

attgttggct taatttatag ggcttgagag cttttattcc gaagcaggaa ctggtgaaga 1140
aagtgaacac tttcaccgag ttgaaggaaa acgtaaagca agagtttcat cttccactgt 1200
aacaattttg tttttctgct gaatcgagcaggcttttgg ctttactgtt ctgcgttttc 1260
tcaggcccttt tttcgtgttg caggtgggtc gtagattcct tgtgcagata acgcgattga 1320
acgaagacaa aaacgacttg atactgagtg aaaaagtagc ctgggttagt tccttcccct 1380
ttttcatgtc ttcataagaa aaaccaacaa tgaggctaag gctttatcac ataatgcctc 1440
aggagaagct gtaccttcga gaaggaacac tcttagaagg aacagttgtc aaaatcttac 1500
catatggagc tcaagtcaaa cttgggggata gtagcagaag gttagttcat tgcctattgt 1560
ttccataaat ttgtccttct cagtgcactac tagttttgaa ttcagtgcgt ttgccataat 1620
aatgtttcga cagtggactg ctacacattt caaacataac tcgaagaaga attggatcgg 1680
tgagcgatgt gcttcagggtg gacgaaagtg taaaggttct gggtgtgaaa tctctatttc 1740
cagacaagat ctctcttagg taaaaacaac taatattcca aaatcaacta ccacctcagt 1800
agtcagtaac aactattttg ttttctacag cattgtgat cttgaaagcg aaccaggatt 1860
gttcattctca gacagagagg tacaacttga gctatggacc taacttatct tttctaactc 1920
ggtacaaaaat tgccattcat gccctcttct tagaagttaa caccatgatg ttgtcgatgc 1980
agaaagtgtt tacggaagct gaagagatgg cgaagaagta cagagaaaaa atgcctttgg 2040
tggtacaaag tccaatttct gatcgctctc cgatcacaaag cagtttccca cagggaaggg 2100
atgaagaaat ctatgctaac tgggagtggt tcaagttcga gagtcagtga 2150

```

<210> 22

<211> 1134

<212> DNA

<213> *Arabidopsis thaliana*

<400> 22

```

atggtgaaca caagtttcga aacacttget ctagacatgc aaatagagat tctggcacga 60
ttgcctctga aatacttaat gagatgcatg tgcgtctcaa agaaatgggc atctttaata 120
cgtggggaag acttcagaag tgcttacttg ctccgatcaa tggcaaggcc tcgactgttg 180
ttcgtcgcta gtcgtagccg caaattcaca cgcgaaacct gctttcactc tgtttatcaa 240
gaaccacat tgttgttgtc tggtaaaccg cagatgctta ccgataatca catagacca 300
gttttcacgg tttctaattc cattcttggg ttgctctgtc atcaaggata taataatgag 360
attgtgattt gcaatcccgg tctgaaaaag tttcggaatt taccgcaaat ccaagtgcct 420
gaaggagcat ccgtgagaag ttttttcgga tacgataaag ttgacaaagt attcaagggt 480
ttatgtgtat gtgtaccaca gacacagttc ggagctggtt ggacatcaaa agaacctaa 540
aaatatcagg tctacacggg gagatcggat catgaaaaac cttccttttg gagaccgatc 600
atgtgtaatg atgatcattc agtagttact gatgaaggac tatttaatgg aggtttcttg 660
tactatggag ctcggtcgag ttgcaacaaa cctatggtaa tgaggttcaa cgtgagttct 720
gaagcgtttg ctgtcattaa attgcctgag gaagtcgaca ttgattatca ttggagattg 780
gtgaactaca aagggaagt tgccttagtg aacacctttg actctaaatt taaattcaat 840
ggagtttttg aaatatgggt taggaacgag gttgatagaa agtgggacaa ggacatcatt 900
aagattcctc aatggacaga atctgttgat aattataatg tttatttcaa aggtaccact 960
ggaacaaaag aacttgtttt cgcaccacc aattggttcg gggaaccact ctttgtcttg 1020
tattacgata aagccacaac taacctcaga agatttgata ttgaaggaaat ggatgatcaa 1080
cattattctt ttcatacctt cttggatcat gttgatagca cttggcttat gtga 1134

```

<210> 23

<211> 2093

<212> DNA

<213> *Arabidopsis thaliana*

<400> 23

```

tcacttcacc caatgtgttt ctgtcccgaa actgcggtct ccagcgtctc caagcccagg 60
gattatgtac ctagtctatg acaatacat aaggacatta aaatacaatc gccagttttc 120
atatggaaag ccaaagaacc taatgactta ccccttttca ttgacttcag gatcaataat 180
tccagcatat acatgaagcc tgaaatgttt caaatagtag ccaaacagtt tagtaaaaag 240

```

```

agagttaact aaatggtaag caaggtgaga gttgatcttt tgctattgat taccggggga 300
atttctcggt aagcttcgat agtgcaggag gtgcagcaat cgcacaaatc tgtaatcaga 360
aaaatcaaaa caattaacac ttcgatttct gacaaaagct gtttcctata tatatgctta 420
ccactttgat ttgctgaaca gacaaaccac gttccttcaa tagatccatc gccgccatta 480
tggtaccacc taaaatacaa gttactaagt tactaaacct acgcttctag ctatgaaaca 540
ttcgtattac catgtcatgc gttcacattc acatattatc gtgtacattt gaacatcaaa 600
ccctaactat acttctaata tggaggtggc taaaatgata agactttctt acctgtggca 660
agcactgggt cgacaaaaaa tactcttgaa ttcttgggaa attcatcagg caacctaaag 720
aagatgtatc aaaggtgtta ttataaacac ttcagtaact acaaacacaa tgctgcagag 780
aagttcttac ttgttaagat acaccgatgg taaaagtgtt ttctcatctc tacttactcc 840
tgtaaaagaa aagaagcacg gggacatcaa taagctattt gataaaagtt ttcattttttg 900
acatccgaag ttcaggcaag agatgatata tagctagatt actaacctaa atgataaatt 960
ttattagcag gtaaaactga ggtgcgtgt tccgcaagag caagaccagc tcttaatatc 1020
ggaacaacct acacagagac aaaaccaaga attagaaacg cttacatgta aggtcgaaaa 1080
gaaaatatag acaggaagag tttgagggat aaacaaaccg caataggctc tctaggatca 1140
atgaattcaa cagaagcagg acccattgga gacattattt caccacaaac agttggctta 1200
accgcataga gcaaaaaaca aaaaacaata agcaaaaaac gggtaggaag aattcaagaa 1260
taatagagat gtgaaattct acgcaccagc cactctcgtg atgcttcata catgagtaat 1320
ctccctagtt cagctattgc attccctaga taattaaaaa aactcaagct ttaaagatca 1380
acacagactt atactagtag caaacctaaa aaaactcata ttcaaacgca atgtagcaga 1440
taacaccaat ctaagagcta aagattgtcc aacgaatcaa aaaaaaaaaa aaaaaaaatg 1500
gatgtaaaat tcagttcttt aacaacaaac actgggttaa tgcgagaaat actcataaga 1560
atcttatcta ctaactgaaa acaggacaag gagtgtgttc atttctcaga acagagatcc 1620
agtgtttgat caaaggatgt ggtggcacia agacctgaaa caccagaatc aaattcacat 1680
ctttcacttt caaatcaaac tcaaaaaaaa tctgaaaatt gaaatcctag agttttatta 1740
ccagcatcct gttgctcccg ttaatggaag cctccgacgc agccattttt gttctagctc 1800
gtatagtccg tcgggaaagt gacgaagaag aagcacccaa attctttaa ataatccaga 1860
aattcgaaaa tcaaaattca gtttgaatat gcaaaaggcg ttactggaaa agtgtgcttt 1920
ggacattaga ataaatatat atatatactc gatacctgtg cgagaatttg ggaggaatta 1980
aaggagagga aagaggaagg attaggggtt aggcggctgc tgcattgttg tcggggagca 2040
aaacgcagcg tatcagaaga acacctgaaa gcgttaccaa ttgagcacgc cat 2093

```

<210> 24

<211> 1283

<212> DNA

<213> *Arabidopsis thaliana*

<400> 24

```

atggcttcaa tgcacgaagc tctgttctct tctcggttcc ttcaagtcaa ttcttctttt 60
tccttccggt gcgtcttccc aattatctct tcaccagctg cagtatcttg cgcaatcaaa 120
tcgaccaggt tttttaaaaca gcgttgtaga accaaagtga gggacttttc tctatcttca 180
ctgtcccggg gaggggttcgt ttgcagagct gctgagtaca agtttccgga tcccatacct 240
gaattcgccg aagctgtgag ctcatttcct ctcttcttcc aactcatttt tgatggatga 300
tcgaaacctt attgatgtgt gatataacca ggagactgag aaatttaggg accatatgtt 360
aaacaaacta tcaaaaagag atctttttga agactctgtt gatgaaattg ttggagtctg 420
cactgaggta atctcaaate ttctcacctt cactcatgta tatctgttcg tgtctatgtg 480
agatgtcttg aattgtggtg tttttagtaa ttgtgttcca tttcaaagaa ctcaaaagtg 540
gtgtaatttt tagtctcagt ttctgatatt ttcttaacat gcctcaaate tcttattctt 600
gtgatgaatg agaatcctga tctgaaattg tttaggtaat gtaatcatct tttactgtca 660
atatgctagt tttggaattg caatggtgat gtttgatcga agttttccta gtttgtatga 720
ttgtgattat gcaagcattt tccttaaatc ttgaggagtg attttgtggg ctgggactta 780
gtcgagctta aggcctagaa tggagagggt catgtgattc ggtttagaat gtgaaggagt 840
ggttttcaat ttgaacctat cttgaaagct cagtgaaaaa aagtttgbat tgtttgaaaa 900
tgagatgctg ttggtctagt aatgatgcct tggtaggag attaatctac gtagtcttta 960
caatgagtcg actgtaagtt tttgtttcat gagtatact ttaaacccca ttctcttaagt 1020
ctgttatatg tcttggcatg agactcaggg agttttaaca gttttttgtt ctgcaccatt 1080

```

```

gtgattattg acagatcttc gagacgttct tgcgcagtga gtatggagga cctggtacac 1140
tcttggatgat accttttcatt gacatggcag atactctcaa cgaaagggaa ttgcctggag 1200
gtccacaagc cgcccgagca gctattaaat gggctcagga tcatgtagac aaggactgga 1260
aggaatggac cggcactgat tag 1283

```

<210> 25
 <211> 18
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 oligonucleotide LWAD1

<400> 25
 ngttgwgnat wtsgwgnt 18

<210> 26
 <211> 15
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 oligonucleotide CA51

<400> 26
 tgwgnagsan casag 15

<210> 27
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 oligonucleotide CA52

<400> 27
 agwgnagwan cawagg 16

<210> 28
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 oligonucleotide CA53

<400> 28
 sttgntastn cnttgc 16

<210> 29
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide CA54

<400> 29
ntcgastwts gwggtt 15

<210> 30
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide CA55

<400> 30
wgtgnagwan canaga 16

<210> 31
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide 5A

<400> 31
actagctcta ccgtttccgt ttccgtttac 30

<210> 32
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide 5B

<400> 32
ttacctcggg ttcgaaatcg atcgggataa 30

<210> 33
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:
 oligonucleotide 5C

 <400> 33
 aaaatcgggtt atacgataac ggtcgggtacg gga 33

 <210> 34
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence:
 oligonucleotide 3A

 <400> 34
 gggctcttgcg gatctgaata tatgttttca tgtgtg 36

 <210> 35
 <211> 37
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence:
 oligonucleotide 3B

 <400> 35
 taccgaagaa aaataccggt tcccgtccga ttctgac 37

 <210> 36
 <211> 34
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence:
 oligonucleotide 3C

 <400> 36
 ggatcgtatc ggttttctgat taccgtatatt atcc 34

 <210> 37
 <211> 1796
 <212> DNA
 <213> Arabidopsis thaliana

 <220>
 <221> CDS
 <222> (585)..(1796)

<400> 37

```

atggctcttt catttcgaat ttcttcgctcg tctccgtaa tctgcagagc cactctctca 60
aacggcgaca acagtcgtaa ctaccacacg actgatgccg cttttatccg acgcgccgcc 120
gatttgctcg aaatgtccgc cggacttaact tcacctcacc ctaatttcgg ctgtgtaatc 180
gccacttctt ccggtaaagt cgccggagaa ggatatctat acgctcaagg tactaaaccg 240
gcggaagcgt tggctgttga agctgccgga gagttttcta gaggcgccac tgcttattta 300
aacatggagc ctggtgattg tcatggtgac cacactgctg tatctgctct tgttcagatt 360
aggcaggaat tgaacgagtt gttgttgga taaggcatcc ttgcaacat cttagaggct 420
ctgcgattcg tgaattaaga agtcatggta tcgaagttaa tgttcttgga gaagactttg 480
agagtaaagt tcttgaggat gctcggaaat cgtgtcttct tgtgaatgct cctttgattc 540
atagagcttg ttctcgagtt ccgttctctg ttcttaagta tgcc atg act ctt gat 596
                                     Met Thr Leu Asp
                                     1
ggt aag att gca gcg agt agt gga cat gca gca tgg ata agt agt aaa 644
Gly Lys Ile Ala Ala Ser Ser Gly His Ala Ala Trp Ile Ser Ser Lys
  5                10                15                20
ctc tct aga acc cga gtg ttc gag ttg cga gga gga agt gat gct gtg 692
Leu Ser Arg Thr Arg Val Phe Glu Leu Arg Gly Gly Ser Asp Ala Val
                25                30                35
ata gtt gga gga aac aca gta cgc caa gat gat cct cga ttg act gca 740
Ile Val Gly Gly Asn Thr Val Arg Gln Asp Asp Pro Arg Leu Thr Ala
                40                45                50
aga cac gga cag ggt cac acc cct aca cga att gtg atg act cag agt 788
Arg His Gly Gln Gly His Thr Pro Thr Arg Ile Val Met Thr Gln Ser
                55                60                65
ctt gat ctt cct gag aaa gct aat cta tgg gat gtc tcg gaa gta tct 836
Leu Asp Leu Pro Glu Lys Ala Asn Leu Trp Asp Val Ser Glu Val Ser
                70                75                80
acc ata gtt gtt acc caa agg ggt gca aga aag agt ttc cag aaa ttg 884
Thr Ile Val Val Thr Gln Arg Gly Ala Arg Lys Ser Phe Gln Lys Leu
                85                90                95                100
ctt gcg tcc aag ggt gtt gaa gtt gtg gag ttt gac atg ttg aat cca 932
Leu Ala Ser Lys Gly Val Glu Val Val Glu Phe Asp Met Leu Asn Pro
                105                110                115
aga gaa gtt atg gaa tat ttc cat ctc cgt gga tat ctt tct ata cta 980
Arg Glu Val Met Glu Tyr Phe His Leu Arg Gly Tyr Leu Ser Ile Leu
                120                125                130
tgg gag tgt ggg gga aca tta gct gca tct gct ata tca tcc agt gtc 1028
Trp Glu Cys Gly Gly Thr Leu Ala Ala Ser Ala Ile Ser Ser Ser Val

```

135	140	145	
atc cac aag gtt gtt gct ttc gtt gct cca aaa att atc ggt ggg agt Ile His Lys Val Val Ala Phe Val Ala Pro Lys Ile Ile Gly Gly Ser 150 155 160			1076
aaa gca cct tcc cct gtg ggt gac ctt ggg atg gta gag atg aca caa Lys Ala Pro Ser Pro Val Gly Asp Leu Gly Met Val Glu Met Thr Gln 165 170 175 180			1124
gca ttg aat tta att gat gtt tgt tat gag cag gtt ggt ccc gac atg Ala Leu Asn Leu Ile Asp Val Cys Tyr Glu Gln Val Gly Pro Asp Met 185 190 195			1172
ctt gtc agc gga ttt ctt cag ccc ata cca gac ctc tta cct gtt atc Leu Val Ser Gly Phe Leu Gln Pro Ile Pro Asp Leu Leu Pro Val Ile 200 205 210			1220
cca tca gag gat gca act gtc gag att gac ccc agt gtt gac ccc ttt Pro Ser Glu Asp Ala Thr Val Glu Ile Asp Pro Ser Val Asp Pro Phe 215 220 225			1268
gaa cct agt atc ata ttc ttc tac aaa aca tgg gac ctc tat gga agt Glu Pro Ser Ile Ile Phe Phe Tyr Lys Thr Trp Asp Leu Tyr Gly Ser 230 235 240			1316
ttc tca aat ttt tct cca cat cca att cga atg ccc tat ggc gag gag Phe Ser Asn Phe Ser Pro His Pro Ile Arg Met Pro Tyr Gly Glu Glu 245 250 255 260			1364
aat gat gat tac atg acc tgg tcc agt gtg gaa cat tac tat cag gca Asn Asp Asp Tyr Met Thr Trp Ser Ser Val Glu His Tyr Tyr Gln Ala 265 270 275			1412
aac aag ttt gtt ggg gtc gaa aat cca aag gca tat gat tgc gtt gag Asn Lys Phe Val Gly Val Glu Asn Pro Lys Ala Tyr Asp Cys Val Glu 280 285 290			1460
aaa att aga aca gct aga agt cca gag gag gca gct ctg ata ggc aga Lys Ile Arg Thr Ala Arg Ser Pro Glu Glu Ala Ala Leu Ile Gly Arg 295 300 305			1508
tca aca ctg aga cag aaa cct gaa ctg gtt cgg aat gat tgg gaa gat Ser Thr Leu Arg Gln Lys Pro Glu Leu Val Arg Asn Asp Trp Glu Asp 310 315 320			1556
gtg aag ata gaa gta atg tac aag gct tta aaa tgc aag ttc tca act Val Lys Ile Glu Val Met Tyr Lys Ala Leu Lys Cys Lys Phe Ser Thr 325 330 335 340			1604
tat cct cac ttg aaa tca atg ttg ctc tcg aca att gga aca gtt ctc Tyr Pro His Leu Lys Ser Met Leu Leu Ser Thr Ile Gly Thr Val Leu 345 350 355			1652
gtg gag gct tca ccg cac gat cta ttc tgg gga ggt ggc cgt gaa gga Val Glu Ala Ser Pro His Asp Leu Phe Trp Gly Gly Gly Arg Glu Gly 360 365 370			1700

gaa ggg ctt aac tat ctt ggt cga ctg ctt atg cag ctt cgt tct gag 1748
 Glu Gly Leu Asn Tyr Leu Gly Arg Leu Leu Met Gln Leu Arg Ser Glu
 375 380 385

tat ctg ggc gaa tcc tct gtt tca gct gaa aag act tct tcg gcc tga 1796
 Tyr Leu Gly Glu Ser Ser Val Ser Ala Glu Lys Thr Ser Ser Ala
 390 395 400

<210> 38

<211> 403

<212> PRT

<213> Arabidopsis thaliana

<400> 38

Met Thr Leu Asp Gly Lys Ile Ala Ala Ser Ser Gly His Ala Ala Trp
 1 5 10 15
 Ile Ser Ser Lys Leu Ser Arg Thr Arg Val Phe Glu Leu Arg Gly Gly
 20 25 30
 Ser Asp Ala Val Ile Val Gly Gly Asn Thr Val Arg Gln Asp Asp Pro
 35 40 45
 Arg Leu Thr Ala Arg His Gly Gly His Thr Pro Thr Arg Ile Val
 50 55 60
 Met Thr Gln Ser Leu Asp Leu Pro Glu Lys Ala Asn Leu Trp Asp Val
 65 70 75 80
 Ser Glu Val Ser Thr Ile Val Val Thr Gln Arg Gly Ala Arg Lys Ser
 85 90 95
 Phe Gln Lys Leu Leu Ala Ser Lys Gly Val Glu Val Val Glu Phe Asp
 100 105 110
 Met Leu Asn Pro Arg Glu Val Met Glu Tyr Phe His Leu Arg Gly Tyr
 115 120 125
 Leu Ser Ile Leu Trp Glu Cys Gly Gly Thr Leu Ala Ala Ser Ala Ile
 130 135 140
 Ser Ser Ser Val Ile His Lys Val Val Ala Phe Val Ala Pro Lys Ile
 145 150 155 160
 Ile Gly Gly Ser Lys Ala Pro Ser Pro Val Gly Asp Leu Gly Met Val
 165 170 175
 Glu Met Thr Gln Ala Leu Asn Leu Ile Asp Val Cys Tyr Glu Gln Val
 180 185 190
 Gly Pro Asp Met Leu Val Ser Gly Phe Leu Gln Pro Ile Pro Asp Leu
 195 200 205
 Leu Pro Val Ile Pro Ser Glu Asp Ala Thr Val Glu Ile Asp Pro Ser
 210 215 220
 Val Asp Pro Phe Glu Pro Ser Ile Ile Phe Phe Tyr Lys Thr Trp Asp
 225 230 235 240
 Leu Tyr Gly Ser Phe Ser Asn Phe Ser Pro His Pro Ile Arg Met Pro
 245 250 255
 Tyr Gly Glu Glu Asn Asp Asp Tyr Met Thr Trp Ser Ser Val Glu His
 260 265 270
 Tyr Tyr Gln Ala Asn Lys Phe Val Gly Val Glu Asn Pro Lys Ala Tyr
 275 280 285
 Asp Cys Val Glu Lys Ile Arg Thr Ala Arg Ser Pro Glu Glu Ala Ala
 290 295 300
 Leu Ile Gly Arg Ser Thr Leu Arg Gln Lys Pro Glu Leu Val Arg Asn
 305 310 315 320
 Asp Trp Glu Asp Val Lys Ile Glu Val Met Tyr Lys Ala Leu Lys Cys

				325					330					335			
Lys	Phe	Ser	Thr	Tyr	Pro	His	Leu	Lys	Ser	Met	Leu	Leu	Ser	Thr	Ile		
			340					345					350				
Gly	Thr	Val	Leu	Val	Glu	Ala	Ser	Pro	His	Asp	Leu	Phe	Trp	Gly	Gly		
		355					360				365						
Gly	Arg	Glu	Gly	Glu	Gly	Leu	Asn	Tyr	Leu	Gly	Arg	Leu	Leu	Met	Gln		
	370					375				380							
Leu	Arg	Ser	Glu	Tyr	Leu	Gly	Glu	Ser	Ser	Val	Ser	Ala	Glu	Lys	Thr		
385					390					395					400		
Ser	Ser	Ala															