



(86) **Date de dépôt PCT/PCT Filing Date:** 2012/09/06
(87) **Date publication PCT/PCT Publication Date:** 2013/03/21
(85) **Entrée phase nationale/National Entry:** 2014/02/26
(86) **N° demande PCT/PCT Application No.:** IB 2012/054595
(87) **N° publication PCT/PCT Publication No.:** 2013/038303
(30) **Priorités/Priorities:** 2011/09/16 (EP11181691.4);
2011/09/16 (US61/535,420); 2011/09/21 (EP11182171.6);
2011/09/21 (US61/537,081); 2011/10/21 (EP11186199.3);
2011/10/21 (US61/549,778); 2011/11/24 (EP11190546.9);
2011/11/24 (US61/563,563)

(51) **Cl.Int./Int.Cl.** **A01H 1/00** (2006.01),
A01H 5/00 (2006.01), **C07K 14/415** (2006.01),
C12N 15/29 (2006.01), **C12N 15/63** (2006.01)
(71) **Demandeur/Applicant:**
BASF PLANT SCIENCE COMPANY GMBH, DE
(72) **Inventeurs/Inventors:**
FRANKARD, VALERIE, BE;
REUZEAU, CHRISTOPHE, FR;
HATZFELD, YVES, FR;
SANZ MOLINERO, ANA ISABEL, ES;
VANDENABEELE, STEVEN, BE
(74) **Agent:** ROBIC

(54) **Titre : PLANTES PRESENTANT DES CARACTERES AMELIORES ASSOCIES AU RENDEMENT ET PROCEDES POUR LEUR FABRICATION**
(54) **Title: PLANTS HAVING ENHANCED YIELD-RELATED TRAITS AND METHODS FOR MAKING THE SAME**

(57) **Abrégé/Abstract:**

Provided are methods for enhancing yield-related traits in plants by modulating expression of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMGI-like polypeptide, or a GPx-related polypeptide in a plant. Also provided are plants having modulated expression of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMGI-like polypeptide, or a GPx-related polypeptide, which plants have enhanced yield-related traits relative to control plants. Further provided are constructs comprising a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMGI-like polypeptide, or a GPx-related polypeptide, useful in carrying out the methods of the invention.

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

(19) World Intellectual Property

Organization

International Bureau



WIPO | PCT



(10) International Publication Number

WO 2013/038303 A1

(43) International Publication Date

21 March 2013 (21.03.2013)

(51) International Patent Classification:

A01H 1/00 (2006.01) C12N 15/63 (2006.01)
A01H 5/00 (2006.01) C07K 14/415 (2006.01)
C12N 15/29 (2006.01)

(21) International Application Number:

PCT/IB2012/054595

(22) International Filing Date:

6 September 2012 (06.09.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/535,420	16 September 2011 (16.09.2011)	US
11181691.4	16 September 2011 (16.09.2011)	EP
61/537,081	21 September 2011 (21.09.2011)	US
11182171.6	21 September 2011 (21.09.2011)	EP
61/549,778	21 October 2011 (21.10.2011)	US
11186199.3	21 October 2011 (21.10.2011)	EP
61/563,563	24 November 2011 (24.11.2011)	US
11190546.9	24 November 2011 (24.11.2011)	EP

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

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

(72) Inventors; and

(75) **Inventors/Applicants** (for US only): **FRANKARD, Valerie** [BE/BE]; Avenue des Croix du Feu 84, 1410 Waterloo (BE). **REUZEAU, Christophe** [FR/FR]; Rue de Cimetière, F-24350 La Chapelle Gonaguet (FR). **HATZFELD, Yves** [FR/FR]; 6B rue du Chemin de Fer, F-59000 Lille (FR). **SANZ MOLINERO, Ana Isabel**

[ES/ES]; Calle Isla Malaita 3 9 D, 28035 Madrid (ES). **VANDENABEELE, Steven** [BE/BE]; Meersport 23, B-9700 Oudenaarde (BE).

(74) **Agent: POPP, Andreas**; BASF SE, Global Intellectual Property, GVX - C006, 67056 Ludwigshafen (DE).

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

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** PLANTS HAVING ENHANCED YIELD-RELATED TRAITS AND METHODS FOR MAKING THE SAME

(57) **Abstract:** Provided are methods for enhancing yield-related traits in plants by modulating expression of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMGI-like polypeptide, or a GPx-related polypeptide in a plant. Also provided are plants having modulated expression of a nucleic acid encoding an ELM2 -related polypeptide, or a WRKY-related polypeptide, or an EMGI-like polypeptide, or a GPx-related polypeptide, which plants have enhanced yield-related traits relative to control plants. Further provided are constructs comprising a nucleic acid encoding an ELM2 -related polypeptide, or a WRKY-related polypeptide, or an EMGI-like polypeptide, or a GPx-related polypeptide, useful in carrying out the methods of the invention.



WO 2013/038303 A1

PLANTS HAVING ENHANCED YIELD-RELATED TRAITS AND METHODS FOR MAKING THE SAME

Background

The present invention relates generally to the field of molecular biology and concerns a method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding an ELM2-related (Egl-27 and MTA1 homology 2 - related) polypeptide, or a WRKY-related polypeptide, or an EMG1-like (Essential for Mitotic Growth - like) polypeptide, or a GPx-related polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, which plants have enhanced yield-related traits relative to corresponding wild type plants or other control plants. The invention also provides constructs useful in the methods of the invention.

The ever-increasing world population and the dwindling supply of arable land available for agriculture fuels research towards increasing the efficiency of agriculture. Conventional means for crop and horticultural improvements utilise selective breeding techniques to identify plants having desirable characteristics. However, such selective breeding techniques have several drawbacks, namely that these techniques are typically labour intensive and result in plants that often contain heterogeneous genetic components that may not always result in the desirable trait being passed on from parent plants. Advances in molecular biology have allowed mankind to modify the germplasm of animals and plants. Genetic engineering of plants entails the isolation and manipulation of genetic material (typically in the form of DNA or RNA) and the subsequent introduction of that genetic material into a plant. Such technology has the capacity to deliver crops or plants having various improved economic, agronomic or horticultural traits.

A trait of particular economic interest is increased yield. Yield is normally defined as the measurable produce of economic value from a crop. This may be defined in terms of quantity and/or quality. Yield is directly dependent on several factors, for example, the number and size of the organs, plant architecture (for example, the number of branches), seed production, leaf senescence and more. Root development, nutrient uptake, stress tolerance and early vigour may also be important factors in determining yield. Optimizing the abovementioned factors may therefore contribute to increasing crop yield.

Seed yield is a particularly important trait, since the seeds of many plants are important for human and animal nutrition. Crops such as corn, rice, wheat, canola and soybean account for over half the total human caloric intake, whether through direct consumption of the seeds themselves or through consumption of meat products raised on processed seeds. They are also a source of sugars, oils and many kinds of metabolites used in industrial processes. Seeds contain an embryo (the source of new shoots and roots) and an endosperm (the source of nutrients for embryo growth during germination and during early

growth of seedlings). The development of a seed involves many genes, and requires the transfer of metabolites from the roots, leaves and stems into the growing seed. The endosperm, in particular, assimilates the metabolic precursors of carbohydrates, oils and proteins and synthesizes them into storage macromolecules to fill out the grain.

5

Another important trait for many crops is early vigour. Improving early vigour is an important objective of modern rice breeding programs in both temperate and tropical rice cultivars. Long roots are important for proper soil anchorage in water-seeded rice. Where rice is sown directly into flooded fields, and where plants must emerge rapidly through water, longer shoots are associated with vigour. Where drill-seeding is practiced, longer mesocotyls and coleoptiles are important for good seedling emergence. The ability to engineer early vigour into plants would be of great importance in agriculture. For example, poor early vigour has been a limitation to the introduction of maize (*Zea mays* L.) hybrids based on Corn Belt germplasm in the European Atlantic.

15

A further important trait is that of improved abiotic stress tolerance. Abiotic stress is a primary cause of crop loss worldwide, reducing average yields for most major crop plants by more than 50% (Wang et al., *Planta* 218, 1-14, 2003). Abiotic stresses may be caused by drought, salinity, extremes of temperature, chemical toxicity and oxidative stress. The ability to improve plant tolerance to abiotic stress would be of great economic advantage to farmers worldwide and would allow for the cultivation of crops during adverse conditions and in territories where cultivation of crops may not otherwise be possible.

25

Crop yield may therefore be increased by optimising one of the above-mentioned factors. Depending on the end use, the modification of certain yield traits may be favoured over others. For example for applications such as forage or wood production, or bio-fuel resource, an increase in the vegetative parts of a plant may be desirable, and for applications such as flour, starch or oil production, an increase in seed parameters may be particularly desirable. Even amongst the seed parameters, some may be favoured over others, depending on the application. Various mechanisms may contribute to increasing seed yield, whether that is in the form of increased seed size or increased seed number.

35

It has now been found that various yield-related traits may be improved in plants by modulating expression in a plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, in a plant.

40

With respect to ELM2-related polypeptides, the ELM2 (Egl-27 and MTA1 homology 2) domain is a small domain. It is found in the MTA1 protein that is part of the NuRD complex. The domain is usually found to the N-terminus of a myb-like DNA binding domain pfam00249. ELM2 is also found associated with an ARID (AT rich interactive domain) DNA

binding domain pfam01388 in a protein from *Arabidopsis thaliana*. This suggests that ELM2 may also be involved in DNA binding, or perhaps is a protein-protein interaction domain. According to Ding et al. (Mol. Cell Biol. 2003 Jan; 23(1):250-8), the ELM2 domain functions as a transcriptional repression domain through recruitment of a trichostatin A-sensitive histone deacetylase 1 (HDAC1).

With respect to WRKY-related polypeptides, however, nothing is reported on modification of yield-related traits in plants with regard to WRKY-related polypeptide until the present day. Rushton, 2010 reports on WRKY proteins playing roles in repression and de-repression of important plant processes. Furthermore, it is described that a single WRKY-related polypeptide might be involved in regulating several seemingly disparate processes. Mechanisms of signalling and transcriptional regulation are being dissected, uncovering WRKY protein functions via interactions with a diverse array of protein partners, including MAP kinases, MAP kinase kinases, 14-3-3 proteins, calmodulin, histone deacetylases, resistance proteins and other WRKY-related polypeptides. WRKY genes exhibit extensive autoregulation and cross-regulation that facilitates transcriptional reprogramming in a dynamic web with built-in redundancy.

Van der Ent, 2009 reports on *Pseudomonas fluorescens* WCS417r bacteria and β -aminobutyric acid which can induce disease resistance in *Arabidopsis* and is based on priming of defence. Particularly, the differences and similarities of WCS417r- and β -aminobutyric acid-induced priming has been examined, wherein both WCS417r and β -aminobutyric acid prime for enhanced deposition of callose-rich papillae after infection by the oomycete *Hyaloperonospora arabidopsis*. This priming is regulated by convergent pathways, which depend on phosphoinositide and ABA-dependent signalling components. Conversely, induced resistance by WCS417r and β -aminobutyric acid against the bacterial pathogen *Pseudomonas syringae* are controlled by distinct NPR1-dependent signalling pathways. As WCS417r and β -aminobutyric acid prime jasmonate- and salicylate-inducible genes, respectively, they subsequently investigated the role of transcription factors. A quantitative PCR-based genome-wide screen for putative WCS417r- and β -aminobutyric acid-responsive transcription factor genes revealed distinct sets of priming-responsive genes. Transcriptional analysis of a selection of these genes showed that they can serve as specific markers for priming. Promoter analysis of WRKY genes identified a putative cis-element that is strongly over-represented in promoters of 21 NPR1-dependent, β -aminobutyric acid-inducible WRKY genes. It is shown that priming of defence is regulated by different pathways, depending on the inducing agent and the challenging pathogen. Furthermore, it is demonstrated that priming is associated with the enhanced expression of transcription factors.

With respect to EMG1-like proteins, ribosome biogenesis is a complex stepwise process that begins with the transcription of ribosomal RNA (rRNA) and continues with a coordinated processing pathway by which the rRNA is processed and ribosomal proteins

are assembled. Ribosomal RNA processing and ribosome assembly require several non-ribosomal protein factors, as well as small nucleolar RNA (snoRNA) molecules. These factors not only guide cleavage steps, but also carry out site-specific rRNA modifications that include methylation and pseudouridylation.

5

Nucleolar essential protein 1 (Nep1) is a highly conserved protein required for ribosome biogenesis and found in organisms from archaea to humans. Nep1 has also been designated the name Emg1 for essential for mitotic growth 1. An Emg1 gene identified in *Arabidopsis thaliana*, the At3g57000 gene, is described to be a member of a family of genes which are essential for 40S ribosomal biogenesis. They play a role in the methylation reaction of pre-rRNA processing. The structure of EMG1 has revealed that it is a novel member of the superfamily of alpha/beta knot fold methyltransferases.

With respect to GPx-related polypeptides, however, nothing is reported on modification of yield-related traits in plants with regard to GPx-related polypeptide until the present day. GPx encodes glutathione peroxidase and belongs to the large family of Peroxidase (Px) enzymes. Green plants contain thousands of Px proteins classified in multiple classes. GPx belong to the class of Thiol peroxidase Selenium-dependent enzyme containing GPx and Peroxiredoxin proteins.

20

GPx are present across kingdom with hundreds of members identified in plants and animals so far. A role of GPx in alleviating oxidative stress generated from polyunsaturated fatty acid metabolism in breast cancer cells has been revealed recently (Margis, 2008). GPx could act as homo-multimers or homo-monomers and are known to have distinct subcellular localization (cytosolic, nuclear, mitochondrial or membrane-bound). Interestingly, plant GPx could be grouped into 5 distinct clusters clearly separated according to their predicted subcellular localization suggesting existence of multiple duplication events during the evolution of GPx in different plant cellular compartment.

Margis, 2008, reports on an evolutionary overview of Glutathione peroxidase family. Particularly, Glutathione peroxidases (EC 1.11.1.9 and EC 1.11.1.12) catalyze the reduction of H₂O₂ or organic hydroperoxides to water or corresponding alcohols using reduced glutathione. Some glutathione peroxidase isozymes have a selenium-dependent glutathione peroxidase activity and present a selenocysteine encoded by the opal TGA codon. In the study of Margis, insights into the evolution of the whole glutathione peroxidase gene family were obtained after a comprehensive phylogenetic analysis using the improved number of glutathione peroxidase sequences recorded in known PeroxiBase database. The identification of a common ancestral origin for the diverse glutathione peroxidase clusters was not possible. The complex relationships and evolutionary rates of this gene family suggest that basal glutathione peroxidase classes, present in all kingdoms, have originated from independent evolutionary events such as gene duplication, gene losses, lateral gene transfer among invertebrates and vertebrates or plants. In addition, the present study also

40

emphasizes the possibility of some members being submitted to strong selective forces that probably dictated functional convergences of taxonomically distant groups.

Definitions

5 The following definitions will be used throughout the present application. The section captions and headings in this application are for convenience and reference purpose only and should not affect in any way the meaning or interpretation of this application. The technical terms and expressions used within the scope of this application are generally to be given the meaning commonly applied to them in the pertinent art of plant biology,
10 molecular biology, bioinformatics and plant breeding. All of the following term definitions apply to the complete content of this application. The term “essentially”, “about”, “approximately” and the like in connection with an attribute or a value, particularly also define exactly the attribute or exactly the value, respectively. The term “about” in the context of a given numeric value or range relates in particular to a value or range that is within 20%,
15 within 10%, or within 5% of the value or range given. As used herein, the term “comprising” also encompasses the term “consisting of”.

Peptide(s)/Protein(s)

The terms “peptides”, “oligopeptides”, “polypeptide” and “protein” are used interchangeably
20 herein and refer to amino acids in a polymeric form of any length, linked together by peptide bonds, unless mentioned herein otherwise.

Polynucleotide(s)/Nucleic acid(s)/Nucleic acid sequence(s)/nucleotide sequence(s)

The terms “polynucleotide(s)”, “nucleic acid sequence(s)”, “nucleotide sequence(s)”,
25 “nucleic acid(s)”, “nucleic acid molecule” are used interchangeably herein and refer to nucleotides, either ribonucleotides or deoxyribonucleotides or a combination of both, in a polymeric unbranched form of any length.

Homologue(s)

30 “Homologues” of a protein encompass peptides, oligopeptides, polypeptides, proteins and enzymes having amino acid substitutions, deletions and/or insertions relative to the unmodified protein in question and having similar biological and functional activity as the unmodified protein from which they are derived.

35 Orthologues and paralogues are two different forms of homologues and encompass evolutionary concepts used to describe the ancestral relationships of genes. Paralogues are genes within the same species that have originated through duplication of an ancestral gene; orthologues are genes from different organisms that have originated through speciation, and are also derived from a common ancestral gene.

40

A “deletion” refers to removal of one or more amino acids from a protein.

An "insertion" refers to one or more amino acid residues being introduced into a predetermined site in a protein. Insertions may comprise N-terminal and/or C-terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Generally, insertions within the amino acid sequence will be smaller than N- or C-terminal fusions, of the order of about 1 to 10 residues. Examples of N- or C-terminal fusion proteins or peptides include the binding domain or activation domain of a transcriptional activator as used in the yeast two-hybrid system, phage coat proteins, (histidine)-6-tag, glutathione S-transferase-tag, protein A, maltose-binding protein, dihydrofolate reductase, Tag•100 epitope, c-myc epitope, FLAG[®]-epitope, lacZ, CMP (calmodulin-binding peptide), HA epitope, protein C epitope and VSV epitope.

A "substitution" refers to replacement of amino acids of the protein with other amino acids having similar properties (such as similar hydrophobicity, hydrophilicity, antigenicity, propensity to form or break α -helical structures or β -sheet structures). Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide and may range from 1 to 10 amino acids. The amino acid substitutions are preferably conservative amino acid substitutions. Conservative substitution tables are well known in the art (see for example Creighton (1984) Proteins. W.H. Freeman and Company (Eds) and Table 1 below).

Table 1: Examples of conserved amino acid substitutions

Residue	Conservative Substitutions	Residue	Conservative Substitutions
Ala	Ser	Leu	Ile; Val
Arg	Lys	Lys	Arg; Gln
Asn	Gln; His	Met	Leu; Ile
Asp	Glu	Phe	Met; Leu; Tyr
Gln	Asn	Ser	Thr; Gly
Cys	Ser	Thr	Ser; Val
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp; Phe
His	Asn; Gln	Val	Ile; Leu
Ile	Leu, Val		

Amino acid substitutions, deletions and/or insertions may readily be made using peptide synthetic techniques known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulation. Methods for the manipulation of DNA sequences to produce substitution, insertion or deletion variants of a protein are well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA are well known to those skilled in the art and include M13 mutagenesis, T7-Gen in vitro mutagenesis (USB, Cleveland, OH), QuickChange Site Directed mutagenesis (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols (see Current Protocols in Molecular Biology, John Wiley & Sons,

N.Y. (1989 and yearly updates)).

Derivatives

“Derivatives” include peptides, oligopeptides, polypeptides which may, compared to the amino acid sequence of the naturally-occurring form of the protein, such as the protein of interest, comprise substitutions of amino acids with non-naturally occurring amino acid residues, or additions of non-naturally occurring amino acid residues. “Derivatives” of a protein also encompass peptides, oligopeptides, polypeptides which comprise naturally occurring altered (glycosylated, acylated, prenylated, phosphorylated, myristoylated, sulphated etc.) or non-naturally altered amino acid residues compared to the amino acid sequence of a naturally-occurring form of the polypeptide. A derivative may also comprise one or more non-amino acid substituents or additions compared to the amino acid sequence from which it is derived, for example a reporter molecule or other ligand, covalently or non-covalently bound to the amino acid sequence, such as a reporter molecule which is bound to facilitate its detection, and non-naturally occurring amino acid residues relative to the amino acid sequence of a naturally-occurring protein. Furthermore, “derivatives” also include fusions of the naturally-occurring form of the protein with tagging peptides such as FLAG, HIS6 or thioredoxin (for a review of tagging peptides, see Terpe, Appl. Microbiol. Biotechnol. 60, 523-533, 2003).

20

Domain, Motif/Consensus sequence/Signature

The term “domain” refers to a set of amino acids conserved at specific positions along an alignment of sequences of evolutionarily related proteins. While amino acids at other positions can vary between homologues, amino acids that are highly conserved at specific positions indicate amino acids that are likely essential in the structure, stability or function of a protein. Identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers to determine if any polypeptide in question belongs to a previously identified polypeptide family.

30 The term “motif” or “consensus sequence” or “signature” refers to a short conserved region in the sequence of evolutionarily related proteins. Motifs are frequently highly conserved parts of domains, but may also include only part of the domain, or be located outside of conserved domain (if all of the amino acids of the motif fall outside of a defined domain).

35 Specialist databases exist for the identification of domains, for example, SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864; Letunic et al. (2002) Nucleic Acids Res 30, 242-244), InterPro (Mulder et al., (2003) Nucl. Acids. Res. 31, 315-318), Prosite (Bucher and Bairoch (1994), A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation. (In) ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., pp53-61, AAAI Press, Menlo Park; Hulo et al., Nucl. Acids. Res. 32:D134-D137, (2004)), or Pfam (Bateman et al., Nucleic Acids Research

40

30(1): 276-280 (2002)). The Pfam protein families database: R.D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, O.L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman Nucleic Acids Research (2010) Database Issue 38: 211-222). A set of tools for *in silico* analysis of protein sequences is available on the ExPASy proteomics server (Swiss Institute of Bioinformatics (Gasteiger et al., ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31:3784-3788(2003)). Domains or motifs may also be identified using routine techniques, such as by sequence alignment.

Methods for the alignment of sequences for comparison are well known in the art, such methods include GAP, BESTFIT, BLAST, FASTA and TFASTA. GAP uses the algorithm of Needleman and Wunsch ((1970) J Mol Biol 48: 443-453) to find the global (i.e. spanning the complete sequences) alignment of two sequences that maximizes the number of matches and minimizes the number of gaps. The BLAST algorithm (Altschul et al. (1990) J Mol Biol 215: 403-10) calculates percent sequence identity and performs a statistical analysis of the similarity between the two sequences. The software for performing BLAST analysis is publicly available through the National Centre for Biotechnology Information (NCBI). Homologues may readily be identified using, for example, the ClustalW multiple sequence alignment algorithm (version 1.83), with the default pairwise alignment parameters, and a scoring method in percentage. Global percentages of similarity and identity may also be determined using one of the methods available in the MatGAT software package (Campanella et al., BMC Bioinformatics. 2003 Jul 10;4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences.). Minor manual editing may be performed to optimise alignment between conserved motifs, as would be apparent to a person skilled in the art. Furthermore, instead of using full-length sequences for the identification of homologues, specific domains may also be used. The sequence identity values may be determined over the entire nucleic acid or amino acid sequence or over selected domains or conserved motif(s), using the programs mentioned above using the default parameters. For local alignments, the Smith-Waterman algorithm is particularly useful (Smith TF, Waterman MS (1981) J. Mol. Biol 147(1);195-7).

Reciprocal BLAST

Typically, this involves a first BLAST involving BLASTing a query sequence (for example using any of the sequences listed in Table A of the Examples section) against any sequence database, such as the publicly available NCBI database. BLASTN or TBLASTX (using standard default values) are generally used when starting from a nucleotide sequence, and BLASTP or TBLASTN (using standard default values) when starting from a protein sequence. The BLAST results may optionally be filtered. The full-length sequences of either the filtered results or non-filtered results are then BLASTed back (second BLAST) against sequences from the organism from which the query sequence is derived. The results of the first and second BLASTs are then compared. A paralogue is identified if a high-ranking hit from the first blast is from the same species as from which the query

sequence is derived, a BLAST back then ideally results in the query sequence amongst the highest hits; an orthologue is identified if a high-ranking hit in the first BLAST is not from the same species as from which the query sequence is derived, and preferably results upon BLAST back in the query sequence being among the highest hits.

5

High-ranking hits are those having a low E-value. The lower the E-value, the more significant the score (or in other words the lower the chance that the hit was found by chance). Computation of the E-value is well known in the art. In addition to E-values, comparisons are also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In the case of large families, ClustalW may be used, followed by a neighbour joining tree, to help visualize clustering of related genes and to identify orthologues and paralogues.

10

15 Hybridisation

The term "hybridisation" as defined herein is a process wherein substantially homologous complementary nucleotide sequences anneal to each other. The hybridisation process can occur entirely in solution, i.e. both complementary nucleic acids are in solution. The hybridisation process can also occur with one of the complementary nucleic acids immobilised to a matrix such as magnetic beads, Sepharose beads or any other resin. The hybridisation process can furthermore occur with one of the complementary nucleic acids immobilised to a solid support such as a nitro-cellulose or nylon membrane or immobilised by e.g. photolithography to, for example, a siliceous glass support (the latter known as nucleic acid arrays or microarrays or as nucleic acid chips). In order to allow hybridisation to occur, the nucleic acid molecules are generally thermally or chemically denatured to melt a double strand into two single strands and/or to remove hairpins or other secondary structures from single stranded nucleic acids.

20

25

The term "stringency" refers to the conditions under which a hybridisation takes place. The stringency of hybridisation is influenced by conditions such as temperature, salt concentration, ionic strength and hybridisation buffer composition. Generally, low stringency conditions are selected to be about 30°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Medium stringency conditions are when the temperature is 20°C below T_m , and high stringency conditions are when the temperature is 10°C below T_m . High stringency hybridisation conditions are typically used for isolating hybridising sequences that have high sequence similarity to the target nucleic acid sequence. However, nucleic acids may deviate in sequence and still encode a substantially identical polypeptide, due to the degeneracy of the genetic code. Therefore medium stringency hybridisation conditions may sometimes be needed to identify such nucleic acid molecules.

35

40

The T_m is the temperature under defined ionic strength and pH, at which 50% of the target

sequence hybridises to a perfectly matched probe. The T_m is dependent upon the solution conditions and the base composition and length of the probe. For example, longer sequences hybridise specifically at higher temperatures. The maximum rate of hybridisation is obtained from about 16°C up to 32°C below T_m . The presence of monovalent cations in the hybridisation solution reduce the electrostatic repulsion between the two nucleic acid strands thereby promoting hybrid formation; this effect is visible for sodium concentrations of up to 0.4M (for higher concentrations, this effect may be ignored). Formamide reduces the melting temperature of DNA-DNA and DNA-RNA duplexes with 0.6 to 0.7°C for each percent formamide, and addition of 50% formamide allows hybridisation to be performed at 30 to 45°C, though the rate of hybridisation will be lowered. Base pair mismatches reduce the hybridisation rate and the thermal stability of the duplexes. On average and for large probes, the T_m decreases about 1°C per % base mismatch. The T_m may be calculated using the following equations, depending on the types of hybrids:

1) DNA-DNA hybrids (Meinkoth and Wahl, Anal. Biochem., 138: 267-284, 1984):

$$T_m = 81.5^\circ\text{C} + 16.6 \times \log_{10}[\text{Na}^+]^a + 0.41 \times \%[\text{G/C}^b] - 500 \times [\text{L}^c]^{-1} - 0.61 \times \% \text{ formamide}$$

2) DNA-RNA or RNA-RNA hybrids:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]^a) + 0.58 (\% \text{G/C}^b) + 11.8 (\% \text{G/C}^b)^2 - 820/\text{L}^c$$

3) oligo-DNA or oligo-RNA^d hybrids:

$$\text{For } <20 \text{ nucleotides: } T_m = 2 (I_n)$$

$$\text{For } 20\text{--}35 \text{ nucleotides: } T_m = 22 + 1.46 (I_n)$$

^a or for other monovalent cation, but only accurate in the 0.01–0.4 M range.

^b only accurate for %GC in the 30% to 75% range.

^c L = length of duplex in base pairs.

^d oligo, oligonucleotide; I_n = effective length of primer = $2 \times (\text{no. of G/C}) + (\text{no. of A/T})$.

Non-specific binding may be controlled using any one of a number of known techniques such as, for example, blocking the membrane with protein containing solutions, additions of heterologous RNA, DNA, and SDS to the hybridisation buffer, and treatment with Rnase. For non-homologous probes, a series of hybridizations may be performed by varying one of (i) progressively lowering the annealing temperature (for example from 68°C to 42°C) or (ii) progressively lowering the formamide concentration (for example from 50% to 0%). The skilled artisan is aware of various parameters which may be altered during hybridisation and which will either maintain or change the stringency conditions.

Besides the hybridisation conditions, specificity of hybridisation typically also depends on the function of post-hybridisation washes. To remove background resulting from non-specific hybridisation, samples are washed with dilute salt solutions. Critical factors of such washes include the ionic strength and temperature of the final wash solution: the lower the salt concentration and the higher the wash temperature, the higher the stringency of the wash. Wash conditions are typically performed at or below hybridisation stringency. A positive hybridisation gives a signal that is at least twice of that of the background.

Generally, suitable stringent conditions for nucleic acid hybridisation assays or gene amplification detection procedures are as set forth above. More or less stringent conditions may also be selected. The skilled artisan is aware of various parameters which may be altered during washing and which will either maintain or change the stringency conditions.

5

For example, typical high stringency hybridisation conditions for DNA hybrids longer than 50 nucleotides encompass hybridisation at 65°C in 1x SSC or at 42°C in 1x SSC and 50% formamide, followed by washing at 65°C in 0.3x SSC. Examples of medium stringency hybridisation conditions for DNA hybrids longer than 50 nucleotides encompass
10 hybridisation at 50°C in 4x SSC or at 40°C in 6x SSC and 50% formamide, followed by washing at 50°C in 2x SSC. The length of the hybrid is the anticipated length for the hybridising nucleic acid. When nucleic acids of known sequence are hybridised, the hybrid length may be determined by aligning the sequences and identifying the conserved regions described herein. 1×SSC is 0.15M NaCl and 15mM sodium citrate; the hybridisation
15 solution and wash solutions may additionally include 5x Denhardt's reagent, 0.5-1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.5% sodium pyrophosphate.

For the purposes of defining the level of stringency, reference can be made to Sambrook et al. (2001) Molecular Cloning: a laboratory manual, 3rd Edition, Cold Spring Harbor
20 Laboratory Press, CSH, New York or to Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989 and yearly updates).

Splice variant

The term “splice variant” as used herein encompasses variants of a nucleic acid sequence
25 in which selected introns and/or exons have been excised, replaced, displaced or added, or in which introns have been shortened or lengthened. Such variants will be ones in which the biological activity of the protein is substantially retained; this may be achieved by selectively retaining functional segments of the protein. Such splice variants may be found in nature or may be manmade. Methods for predicting and isolating such splice variants are well known
30 in the art (see for example Foissac and Schiex (2005) BMC Bioinformatics 6: 25).

Allelic variant

“Alleles” or “allelic variants” are alternative forms of a given gene, located at the same chromosomal position. Allelic variants encompass Single Nucleotide Polymorphisms
35 (SNPs), as well as Small Insertion/Deletion Polymorphisms (INDELs). The size of INDELs is usually less than 100 bp. SNPs and INDELs form the largest set of sequence variants in naturally occurring polymorphic strains of most organisms.

Endogenous gene

Reference herein to an “endogenous” gene not only refers to the gene in question as found
40 in a plant in its natural form (i.e., without there being any human intervention), but also refers to that same gene (or a substantially homologous nucleic acid/gene) in an isolated

form subsequently (re)introduced into a plant (a transgene). For example, a transgenic plant containing such a transgene may encounter a substantial reduction of the transgene expression and/or substantial reduction of expression of the endogenous gene. The isolated gene may be isolated from an organism or may be manmade, for example by chemical synthesis.

Gene shuffling/Directed evolution

“Gene shuffling” or “directed evolution” consists of iterations of DNA shuffling followed by appropriate screening and/or selection to generate variants of nucleic acids or portions thereof encoding proteins having a modified biological activity (Castle et al., (2004) Science 304(5674): 1151-4; US patents 5,811,238 and 6,395,547).

Construct

Artificial DNA (such as but, not limited to plasmids or viral DNA) capable of replication in a host cell and used for introduction of a DNA sequence of interest into a host cell or host organism. Host cells of the invention may be any cell selected from bacterial cells, such as Escherichia coli or Agrobacterium species cells, yeast cells, fungal, algal or cyanobacterial cells or plant cells. The skilled artisan is well aware of the genetic elements that must be present on the genetic construct in order to successfully transform, select and propagate host cells containing the sequence of interest. The sequence of interest is operably linked to one or more control sequences (at least to a promoter) as described herein. Additional regulatory elements may include transcriptional as well as translational enhancers. Those skilled in the art will be aware of terminator and enhancer sequences that may be suitable for use in performing the invention. An intron sequence may also be added to the 5' untranslated region (UTR) or in the coding sequence to increase the amount of the mature message that accumulates in the cytosol, as described in the definitions section. Other control sequences (besides promoter, enhancer, silencer, intron sequences, 3'UTR and/or 5'UTR regions) may be protein and/or RNA stabilizing elements. Such sequences would be known or may readily be obtained by a person skilled in the art.

The genetic constructs of the invention may further include an origin of replication sequence that is required for maintenance and/or replication in a specific cell type. One example is when a genetic construct is required to be maintained in a bacterial cell as an episomal genetic element (e.g. plasmid or cosmid molecule). Preferred origins of replication include, but are not limited to, the f1-ori and colE1.

For the detection of the successful transfer of the nucleic acid sequences as used in the methods of the invention and/or selection of transgenic plants comprising these nucleic acids, it is advantageous to use marker genes (or reporter genes). Therefore, the genetic construct may optionally comprise a selectable marker gene. Selectable markers are described in more detail in the “definitions” section herein. The marker genes may be removed or excised from the transgenic cell once they are no longer needed. Techniques

for marker removal are known in the art, useful techniques are described above in the definitions section.

Regulatory element/Control sequence/Promoter

5 The terms “regulatory element”, “control sequence” and “promoter” are all used interchangeably herein and are to be taken in a broad context to refer to regulatory nucleic acid sequences capable of effecting expression of the sequences to which they are ligated. The term “promoter” typically refers to a nucleic acid control sequence located upstream from the transcriptional start of a gene and which is involved in recognising and binding of
10 RNA polymerase and other proteins, thereby directing transcription of an operably linked nucleic acid. Encompassed by the aforementioned terms are transcriptional regulatory sequences derived from a classical eukaryotic genomic gene (including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence) and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or
15 external stimuli, or in a tissue-specific manner. Also included within the term is a transcriptional regulatory sequence of a classical prokaryotic gene, in which case it may include a –35 box sequence and/or –10 box transcriptional regulatory sequences. The term “regulatory element” also encompasses a synthetic fusion molecule or derivative that
20 confers, activates or enhances expression of a nucleic acid molecule in a cell, tissue or organ.

A “plant promoter” comprises regulatory elements, which mediate the expression of a coding sequence segment in plant cells. Accordingly, a plant promoter need not be of plant
25 origin, but may originate from viruses or micro-organisms, for example from viruses which attack plant cells. The “plant promoter” can also originate from a plant cell, e.g. from the plant which is transformed with the nucleic acid sequence to be expressed in the inventive process and described herein. This also applies to other “plant” regulatory signals, such as “plant” terminators. The promoters upstream of the nucleotide sequences useful in the
30 methods of the present invention can be modified by one or more nucleotide substitution(s), insertion(s) and/or deletion(s) without interfering with the functionality or activity of either the promoters, the open reading frame (ORF) or the 3'-regulatory region such as terminators or other 3' regulatory regions which are located away from the ORF. It is furthermore possible that the activity of the promoters is increased by modification of their sequence, or that they
35 are replaced completely by more active promoters, even promoters from heterologous organisms. For expression in plants, the nucleic acid molecule must, as described above, be linked operably to or comprise a suitable promoter which expresses the gene at the right point in time and with the required spatial expression pattern.

40 For the identification of functionally equivalent promoters, the promoter strength and/or expression pattern of a candidate promoter may be analysed for example by operably linking the promoter to a reporter gene and assaying the expression level and pattern of the

reporter gene in various tissues of the plant. Suitable well-known reporter genes include for example beta-glucuronidase or beta-galactosidase. The promoter activity is assayed by measuring the enzymatic activity of the beta-glucuronidase or beta-galactosidase. The promoter strength and/or expression pattern may then be compared to that of a reference promoter (such as the one used in the methods of the present invention). Alternatively, promoter strength may be assayed by quantifying mRNA levels or by comparing mRNA levels of the nucleic acid used in the methods of the present invention, with mRNA levels of housekeeping genes such as 18S rRNA, using methods known in the art, such as Northern blotting with densitometric analysis of autoradiograms, quantitative real-time PCR or RT-PCR (Heid et al., 1996 Genome Methods 6: 986-994). Generally by “weak promoter” is intended a promoter that drives expression of a coding sequence at a low level. By “low level” is intended at levels of about 1/10,000 transcripts to about 1/100,000 transcripts, to about 1/500,000 transcripts per cell. Conversely, a “strong promoter” drives expression of a coding sequence at high level, or at about 1/10 transcripts to about 1/100 transcripts to about 1/1000 transcripts per cell. Generally, by “medium strength promoter” is intended a promoter that drives expression of a coding sequence at a lower level than a strong promoter, in particular at a level that is in all instances below that obtained when under the control of a 35S CaMV promoter.

20 Operably linked

The term “operably linked” as used herein refers to a functional linkage between the promoter sequence and the gene of interest, such that the promoter sequence is able to initiate transcription of the gene of interest.

25 Constitutive promoter

A “constitutive promoter” refers to a promoter that is transcriptionally active during most, but not necessarily all, phases of growth and development and under most environmental conditions, in at least one cell, tissue or organ. Table 2a below gives examples of constitutive promoters.

30

Table 2a: Examples of constitutive promoters

Gene Source	Reference
Actin	McElroy et al, Plant Cell, 2: 163-171, 1990
HMGP	WO 2004/070039
CAMV 35S	Odell et al, Nature, 313: 810-812, 1985
CaMV 19S	Nilsson et al., Physiol. Plant. 100:456-462, 1997
GOS2	de Pater et al, Plant J Nov;2(6):837-44, 1992, WO 2004/065596
Ubiquitin	Christensen et al, Plant Mol. Biol. 18: 675-689, 1992
Rice cyclophilin	Buchholz et al, Plant Mol Biol. 25(5): 837-43, 1994
Maize H3 histone	Lepetit et al, Mol. Gen. Genet. 231:276-285, 1992
Alfalfa H3 histone	Wu et al. Plant Mol. Biol. 11:641-649, 1988
Actin 2	An et al, Plant J. 10(1); 107-121, 1996

34S FMV	Sanger et al., Plant. Mol. Biol., 14, 1990: 433-443
Rubisco small subunit	US 4,962,028
OCS	Leisner (1988) Proc Natl Acad Sci USA 85(5): 2553
SAD1	Jain et al., Crop Science, 39 (6), 1999: 1696
SAD2	Jain et al., Crop Science, 39 (6), 1999: 1696
nos	Shaw et al. (1984) Nucleic Acids Res. 12(20):7831-7846
V-ATPase	WO 01/14572
Super promoter	WO 95/14098
G-box proteins	WO 94/12015

Ubiquitous promoter

A “ubiquitous promoter” is active in substantially all tissues or cells of an organism.

5 Developmentally-regulated promoter

A “developmentally-regulated promoter” is active during certain developmental stages or in parts of the plant that undergo developmental changes.

Inducible promoter

- 10 An “inducible promoter” has induced or increased transcription initiation in response to a chemical (for a review see Gatz 1997, Annu. Rev. Plant Physiol. Plant Mol. Biol., 48:89-108), environmental or physical stimulus, or may be “stress-inducible”, i.e. activated when a plant is exposed to various stress conditions, or a “pathogen-inducible” i.e. activated when a plant is exposed to exposure to various pathogens.

15

Organ-specific/Tissue-specific promoter

- An “organ-specific” or “tissue-specific promoter” is one that is capable of preferentially initiating transcription in certain organs or tissues, such as the leaves, roots, seed tissue etc. For example, a “root-specific promoter” is a promoter that is transcriptionally active
20 predominantly in plant roots, substantially to the exclusion of any other parts of a plant, whilst still allowing for any leaky expression in these other plant parts. Promoters able to initiate transcription in certain cells only are referred to herein as “cell-specific”.

Examples of root-specific promoters are listed in Table 2b below:

25

Table 2b: Examples of root-specific promoters

Gene Source	Reference
RCc3	Plant Mol Biol. 1995 Jan;27(2):237-48
Arabidopsis PHT1	Koyama et al. J Biosci Bioeng. 2005 Jan;99(1):38-42.; Mudge et al. (2002, Plant J. 31:341)
Medicago phosphate transporter	Xiao et al., 2006, Plant Biol (Stuttg). 2006 Jul;8(4):439-49
Arabidopsis Pyk10	Nitz et al. (2001) Plant Sci 161(2): 337-346

root-expressible genes	Tingey et al., EMBO J. 6: 1, 1987.
tobacco auxin-inducible gene	Van der Zaal et al., Plant Mol. Biol. 16, 983, 1991.
β -tubulin	Oppenheimer, et al., Gene 63: 87, 1988.
tobacco root-specific genes	Conkling, et al., Plant Physiol. 93: 1203, 1990.
B. napus G1-3b gene	United States Patent No. 5, 401, 836
SbPRP1	Suzuki et al., Plant Mol. Biol. 21: 109-119, 1993.
LRX1	Baumberger et al. 2001, Genes & Dev. 15:1128
BTG-26 Brassica napus	US 20050044585
LeAMT1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
The LeNRT1-1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
class I patatin gene (potato)	Liu et al., Plant Mol. Biol. 17 (6): 1139-1154
KDC1 (Daucus carota)	Downey et al. (2000, J. Biol. Chem. 275:39420)
TobRB7 gene	W Song (1997) PhD Thesis, North Carolina State University, Raleigh, NC USA
OsRAB5a (rice)	Wang et al. 2002, Plant Sci. 163:273
ALF5 (Arabidopsis)	Diener et al. (2001, Plant Cell 13:1625)
NRT2;1Np (N. plumbaginifolia)	Quesada et al. (1997, Plant Mol. Biol. 34:265)

A “seed-specific promoter” is transcriptionally active predominantly in seed tissue, but not necessarily exclusively in seed tissue (in cases of leaky expression). The seed-specific promoter may be active during seed development and/or during germination. The seed specific promoter may be endosperm/aleurone/embryo specific. Examples of seed-specific promoters (endosperm/aleurone/embryo specific) are shown in Table 2c to Table 2f below. Further examples of seed-specific promoters are given in Qing Qu and Takaiwa (Plant Biotechnol. J. 2, 113-125, 2004), which disclosure is incorporated by reference herein as if fully set forth.

Table 2c: Examples of seed-specific promoters

Gene source	Reference
seed-specific genes	Simon et al., Plant Mol. Biol. 5: 191, 1985;
	Scofield et al., J. Biol. Chem. 262: 12202, 1987.;
	Baszczynski et al., Plant Mol. Biol. 14: 633, 1990.
Brazil Nut albumin	Pearson et al., Plant Mol. Biol. 18: 235-245, 1992.
legumin	Ellis et al., Plant Mol. Biol. 10: 203-214, 1988.
glutelin (rice)	Takaiwa et al., Mol. Gen. Genet. 208: 15-22, 1986;
	Takaiwa et al., FEBS Letts. 221: 43-47, 1987.
zein	Matzke et al Plant Mol Biol, 14(3):323-32 1990

napA	Stalberg et al, Planta 199: 515-519, 1996.
wheat LMW and HMW glutenin-1	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
wheat SPA	Albani et al, Plant Cell, 9: 171-184, 1997
wheat α , β , γ -gliadins	EMBO J. 3:1409-15, 1984
barley ltr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	Mena et al, The Plant Journal, 116(1): 53-62, 1998
blz2	EP99106056.7
synthetic promoter	Vicente-Carbajosa et al., Plant J. 13: 629-640, 1998.
rice prolamin NRP33	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice a-globulin Glb-1	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
rice α -globulin REB/OHP-1	Nakase et al. Plant Mol. Biol. 33: 513-522, 1997
rice ADP-glucose pyrophosphorylase	Trans Res 6:157-68, 1997
maize ESR gene family	Plant J 12:235-46, 1997
sorghum α -kafirin	DeRose et al., Plant Mol. Biol 32:1029-35, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
rice oleosin	Wu et al, J. Biochem. 123:386, 1998
sunflower oleosin	Cummins et al., Plant Mol. Biol. 19: 873-876, 1992
PRO0117, putative rice 40S ribosomal protein	WO 2004/070039
PRO0136, rice alanine aminotransferase	unpublished
PRO0147, trypsin inhibitor ITR1 (barley)	unpublished
PRO0151, rice WSI18	WO 2004/070039
PRO0175, rice RAB21	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039
α -amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β -like gene	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Barley Ltp2	Kalla et al., Plant J. 6:849-60, 1994
Chi26	Leah et al., Plant J. 4:579-89, 1994
Maize B-Peru	Selinger et al., Genetics 149;1125-38,1998

Table 2d: examples of endosperm-specific promoters

Gene source	Reference
glutelin (rice)	Takaiwa et al. (1986) Mol Gen Genet 208:15-22; Takaiwa et al. (1987) FEBS Letts. 221:43-47
zein	Matzke et al., (1990) Plant Mol Biol 14(3): 323-32
wheat LMW and HMW glutenin-1	Colot et al. (1989) Mol Gen Genet 216:81-90, Anderson et al. (1989) NAR 17:461-2
wheat SPA	Albani et al. (1997) Plant Cell 9:171-184
wheat gliadins	Rafalski et al. (1984) EMBO 3:1409-15
barley ltr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Cho et al. (1999) Theor Appl Genet 98:1253-62; Muller et al. (1993) Plant J 4:343-55; Sorenson et al. (1996) Mol Gen Genet 250:750-60
barley DOF	Mena et al, (1998) Plant J 116(1): 53-62
blz2	Onate et al. (1999) J Biol Chem 274(14):9175-82
synthetic promoter	Vicente-Carbajosa et al. (1998) Plant J 13:629-640
rice prolamin NRP33	Wu et al, (1998) Plant Cell Physiol 39(8) 885-889
rice globulin Glb-1	Wu et al. (1998) Plant Cell Physiol 39(8) 885-889
rice globulin REB/OHP-1	Nakase et al. (1997) Plant Molec Biol 33: 513-522
rice ADP-glucose pyrophosphorylase	Russell et al. (1997) Trans Res 6:157-68
maize ESR gene family	Opsahl-Ferstad et al. (1997) Plant J 12:235-46
sorghum kafirin	DeRose et al. (1996) Plant Mol Biol 32:1029-35

Table 2e: Examples of embryo specific promoters:

Gene source	Reference
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
PRO0151	WO 2004/070039
PRO0175	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039

5 **Table 2f:** Examples of aleurone-specific promoters:

Gene source	Reference
α -amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β -like gene	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Barley Ltp2	Kalla et al., Plant J. 6:849-60, 1994
Chi26	Leah et al., Plant J. 4:579-89, 1994
Maize B-Peru	Selinger et al., Genetics 149;1125-38,1998

A “green tissue-specific promoter” as defined herein is a promoter that is transcriptionally active predominantly in green tissue, substantially to the exclusion of any other parts of a plant, whilst still allowing for any leaky expression in these other plant parts.

- 5 Examples of green tissue-specific promoters which may be used to perform the methods of the invention are shown in Table 2g below.

Table 2g: Examples of green tissue-specific promoters

Gene	Expression	Reference
Maize Orthophosphate dikinase	Leaf specific	Fukavama et al., Plant Physiol. 2001 Nov;127(3):1136-46
Maize Phosphoenolpyruvate carboxylase	Leaf specific	Kausch et al., Plant Mol Biol. 2001 Jan;45(1):1-15
Rice Phosphoenolpyruvate carboxylase	Leaf specific	Lin et al., 2004 DNA Seq. 2004 Aug;15(4):269-76
Rice small subunit Rubisco	Leaf specific	Nomura et al., Plant Mol Biol. 2000 Sep;44(1):99-106
rice beta expansin EXBP9	Shoot specific	WO 2004/070039
Pigeonpea small subunit Rubisco	Leaf specific	Panguluri et al., Indian J Exp Biol. 2005 Apr;43(4):369-72
Pea RBCS3A	Leaf specific	

- 10 Another example of a tissue-specific promoter is a meristem-specific promoter, which is transcriptionally active predominantly in meristematic tissue, substantially to the exclusion of any other parts of a plant, whilst still allowing for any leaky expression in these other plant parts. Examples of green meristem-specific promoters which may be used to perform the methods of the invention are shown in Table 2h below.

15

Table 2h: Examples of meristem-specific promoters

Gene source	Expression pattern	Reference
rice OSH1	Shoot apical meristem, from embryo globular stage to seedling stage	Sato <i>et al.</i> (1996) Proc. Natl. Acad. Sci. USA, 93: 8117-8122
Rice metallothionein	Meristem specific	BAD87835.1
WAK1 & WAK 2	Shoot and root apical meristems, and in expanding leaves and sepals	Wagner & Kohorn (2001) Plant Cell 13(2): 303–318

Terminator

- 20 The term “terminator” encompasses a control sequence which is a DNA sequence at the end of a transcriptional unit which signals 3' processing and polyadenylation of a primary transcript and termination of transcription. The terminator can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The terminator to be added may

be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

Selectable marker (gene)/Reporter gene

5 "Selectable marker", "selectable marker gene" or "reporter gene" includes any gene that confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells that are transfected or transformed with a nucleic acid construct of the invention. These marker genes enable the identification of a successful transfer of the nucleic acid molecules via a series of different principles. Suitable markers may be selected
10 from markers that confer antibiotic or herbicide resistance, that introduce a new metabolic trait or that allow visual selection. Examples of selectable marker genes include genes conferring resistance to antibiotics (such as nptII that phosphorylates neomycin and kanamycin, or hpt, phosphorylating hygromycin, or genes conferring resistance to, for example, bleomycin, streptomycin, tetracyclin, chloramphenicol, ampicillin, gentamycin, geneticin (G418), spectinomycin or blasticidin), to herbicides (for example bar which provides resistance to Basta®; aroA or gox providing resistance against glyphosate, or the genes conferring resistance to, for example, imidazolinone, phosphinothricin or sulfonylurea), or genes that provide a metabolic trait (such as manA that allows plants to use mannose as sole carbon source or xylose isomerase for the utilisation of xylose, or
20 antinutritive markers such as the resistance to 2-deoxyglucose). Expression of visual marker genes results in the formation of colour (for example β -glucuronidase, GUS or β -galactosidase with its coloured substrates, for example X-Gal), luminescence (such as the luciferin/luciferase system) or fluorescence (Green Fluorescent Protein, GFP, and derivatives thereof). This list represents only a small number of possible markers. The skilled worker is familiar with such markers. Different markers are preferred, depending on
25 the organism and the selection method.

It is known that upon stable or transient integration of nucleic acids into plant cells, only a minority of the cells takes up the foreign DNA and, if desired, integrates it into its genome,
30 depending on the expression vector used and the transfection technique used. To identify and select these integrants, a gene coding for a selectable marker (such as the ones described above) is usually introduced into the host cells together with the gene of interest. These markers can for example be used in mutants in which these genes are not functional by, for example, deletion by conventional methods. Furthermore, nucleic acid molecules encoding a selectable marker can be introduced into a host cell on the same vector that
35 comprises the sequence encoding the polypeptides of the invention or used in the methods of the invention, or else in a separate vector. Cells which have been stably transfected with the introduced nucleic acid can be identified for example by selection (for example, cells which have integrated the selectable marker survive whereas the other cells die).

40 Since the marker genes, particularly genes for resistance to antibiotics and herbicides, are no longer required or are undesired in the transgenic host cell once the nucleic acids have

been introduced successfully, the process according to the invention for introducing the nucleic acids advantageously employs techniques which enable the removal or excision of these marker genes. One such a method is what is known as co-transformation. The co-transformation method employs two vectors simultaneously for the transformation, one vector bearing the nucleic acid according to the invention and a second bearing the marker gene(s). A large proportion of transformants receives or, in the case of plants, comprises (up to 40% or more of the transformants), both vectors. In case of transformation with *Agrobacteria*, the transformants usually receive only a part of the vector, i.e. the sequence flanked by the T-DNA, which usually represents the expression cassette. The marker genes can subsequently be removed from the transformed plant by performing crosses. In another method, marker genes integrated into a transposon are used for the transformation together with desired nucleic acid (known as the Ac/Ds technology). The transformants can be crossed with a transposase source or the transformants are transformed with a nucleic acid construct conferring expression of a transposase, transiently or stable. In some cases (approx. 10%), the transposon jumps out of the genome of the host cell once transformation has taken place successfully and is lost. In a further number of cases, the transposon jumps to a different location. In these cases the marker gene must be eliminated by performing crosses. In microbiology, techniques were developed which make possible, or facilitate, the detection of such events. A further advantageous method relies on what is known as recombination systems; whose advantage is that elimination by crossing can be dispensed with. The best-known system of this type is what is known as the Cre/lox system. Cre1 is a recombinase that removes the sequences located between the loxP sequences. If the marker gene is integrated between the loxP sequences, it is removed once transformation has taken place successfully, by expression of the recombinase. Further recombination systems are the HIN/HIX, FLP/FRT and REP/STB system (Tribble et al., J. Biol. Chem., 275, 2000: 22255-22267; Velmurugan et al., J. Cell Biol., 149, 2000: 553-566). A site-specific integration into the plant genome of the nucleic acid sequences according to the invention is possible. Naturally, these methods can also be applied to microorganisms such as yeast, fungi or bacteria.

Transgenic/Transgene/Recombinant

For the purposes of the invention, "transgenic", "transgene" or "recombinant" means with regard to, for example, a nucleic acid sequence, an expression cassette, gene construct or a vector comprising the nucleic acid sequence or an organism transformed with the nucleic acid sequences, expression cassettes or vectors according to the invention, all those constructions brought about by recombinant methods in which either

- (a) the nucleic acid sequences encoding proteins useful in the methods of the invention, or
- (b) genetic control sequence(s) which is operably linked with the nucleic acid sequence according to the invention, for example a promoter, or
- (c) a) and b)

are not located in their natural genetic environment or have been modified by recombinant

methods, it being possible for the modification to take the form of, for example, a substitution, addition, deletion, inversion or insertion of one or more nucleotide residues. The natural genetic environment is understood as meaning the natural genomic or chromosomal locus in the original plant or the presence in a genomic library. In the case of a genomic library, the natural genetic environment of the nucleic acid sequence is preferably retained, at least in part. The environment flanks the nucleic acid sequence at least on one side and has a sequence length of at least 50 bp, preferably at least 500 bp, especially preferably at least 1000 bp, most preferably at least 5000 bp. A naturally occurring expression cassette – for example the naturally occurring combination of the natural promoter of the nucleic acid sequences with the corresponding nucleic acid sequence encoding a polypeptide useful in the methods of the present invention, as defined above – becomes a transgenic expression cassette when this expression cassette is modified by non-natural, synthetic ("artificial") methods such as, for example, mutagenic treatment. Suitable methods are described, for example, in US 5,565,350 or WO 00/15815.

A transgenic plant for the purposes of the invention is thus understood as meaning, as above, that the nucleic acids used in the method of the invention are not present in, or originating from, the genome of said plant, or are present in the genome of said plant but not at their natural locus in the genome of said plant, it being possible for the nucleic acids to be expressed homologously or heterologously. However, as mentioned, transgenic also means that, while the nucleic acids according to the invention or used in the inventive method are at their natural position in the genome of a plant, the sequence has been modified with regard to the natural sequence, and/or that the regulatory sequences of the natural sequences have been modified. Transgenic is preferably understood as meaning the expression of the nucleic acids according to the invention at an unnatural locus in the genome, i.e. homologous or, preferably, heterologous expression of the nucleic acids takes place. Preferred transgenic plants are mentioned herein.

It shall further be noted that in the context of the present invention, the term "isolated nucleic acid" or "isolated polypeptide" may in some instances be considered as a synonym for a "recombinant nucleic acid" or a "recombinant polypeptide", respectively and refers to a nucleic acid or polypeptide that is not located in its natural genetic environment and/or that has been modified by recombinant methods.

In one embodiment an isolated nucleic acid sequence or isolated nucleic acid molecule is one that is not in its native surrounding or its native nucleic acid neighbourhood, yet is physically and functionally connected to other nucleic acid sequences or nucleic acid molecules and is found as part of a nucleic acid construct, vector sequence or chromosome.

Modulation

The term "modulation" means in relation to expression or gene expression, a process in

which the expression level is changed by said gene expression in comparison to the control plant, the expression level may be increased or decreased. The original, unmodulated expression may be of any kind of expression of a structural RNA (rRNA, tRNA) or mRNA with subsequent translation. For the purposes of this invention, the original unmodulated expression may also be absence of any expression. The term “modulating the activity” or the term “modulating expression” shall mean any change of the expression of the inventive nucleic acid sequences or encoded proteins, which leads to increased yield and/or increased growth of the plants. The expression can increase from zero (absence of, or immeasurable expression) to a certain amount, or can decrease from a certain amount to immeasurable small amounts or zero.

Expression

The term “expression” or “gene expression” means the transcription of a specific gene or specific genes or specific genetic construct. The term “expression” or “gene expression” in particular means the transcription of a gene or genes or genetic construct into structural RNA (rRNA, tRNA) or mRNA with or without subsequent translation of the latter into a protein. The process includes transcription of DNA and processing of the resulting mRNA product.

Increased expression/overexpression

The term “increased expression” or “overexpression” as used herein means any form of expression that is additional to the original wild-type expression level. For the purposes of this invention, the original wild-type expression level might also be zero, i.e. absence of expression or immeasurable expression.

Methods for increasing expression of genes or gene products are well documented in the art and include, for example, overexpression driven by appropriate promoters, the use of transcription enhancers or translation enhancers. Isolated nucleic acids which serve as promoter or enhancer elements may be introduced in an appropriate position (typically upstream) of a non-heterologous form of a polynucleotide so as to upregulate expression of a nucleic acid encoding the polypeptide of interest. For example, endogenous promoters may be altered in vivo by mutation, deletion, and/or substitution (see, Kmiec, US 5,565,350; Zarling et al., WO9322443), or isolated promoters may be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence may also be added to the 5' untranslated region (UTR) or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg (1988) Mol. Cell biol. 8: 4395-4405; Callis et al. (1987) Genes Dev 1:1183-1200). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of the maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. For general information see: The Maize Handbook, Chapter 116, Freeling and Walbot, Eds., Springer, N.Y. (1994).

Decreased expression

Reference herein to "decreased expression" or "reduction or substantial elimination" of expression is taken to mean a decrease in endogenous gene expression and/or polypeptide levels and/or polypeptide activity relative to control plants. The reduction or substantial elimination is in increasing order of preference at least 10%, 20%, 30%, 40% or 50%, 60%, 70%, 80%, 85%, 90%, or 95%, 96%, 97%, 98%, 99% or more reduced compared to that of control plants.

For the reduction or substantial elimination of expression an endogenous gene in a plant, a sufficient length of substantially contiguous nucleotides of a nucleic acid sequence is required. In order to perform gene silencing, this may be as little as 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 or fewer nucleotides, alternatively this may be as much as the entire gene (including the 5' and/or 3' UTR, either in part or in whole). The stretch of substantially contiguous nucleotides may be derived from the nucleic acid encoding the protein of interest (target gene), or from any nucleic acid capable of encoding an orthologue, paralogue or homologue of the protein of interest. Preferably, the stretch of substantially contiguous nucleotides is capable of forming hydrogen bonds with the target gene (either sense or antisense strand), more preferably, the stretch of substantially contiguous nucleotides has, in increasing order of preference, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100% sequence identity to the target gene (either sense or antisense strand). A nucleic acid sequence encoding a (functional) polypeptide is not a requirement for the various methods discussed herein for the reduction or substantial elimination of expression of an endogenous gene.

This reduction or substantial elimination of expression may be achieved using routine tools and techniques. A preferred method for the reduction or substantial elimination of endogenous gene expression is by introducing and expressing in a plant a genetic construct into which the nucleic acid (in this case a stretch of substantially contiguous nucleotides derived from the gene of interest, or from any nucleic acid capable of encoding an orthologue, paralogue or homologue of any one of the protein of interest) is cloned as an

inverted repeat (in part or completely), separated by a spacer (non-coding DNA).

In such a preferred method, expression of the endogenous gene is reduced or substantially eliminated through RNA-mediated silencing using an inverted repeat of a nucleic acid or a part thereof (in this case a stretch of substantially contiguous nucleotides derived from the gene of interest, or from any nucleic acid capable of encoding an orthologue, paralogue or homologue of the protein of interest), preferably capable of forming a hairpin structure. The inverted repeat is cloned in an expression vector comprising control sequences. A non-coding DNA nucleic acid sequence (a spacer, for example a matrix attachment region fragment (MAR), an intron, a polylinker, etc.) is located between the two inverted nucleic acids forming the inverted repeat. After transcription of the inverted repeat, a chimeric RNA with a self-complementary structure is formed (partial or complete). This double-stranded RNA structure is referred to as the hairpin RNA (hpRNA). The hpRNA is processed by the plant into siRNAs that are incorporated into an RNA-induced silencing complex (RISC). The RISC further cleaves the mRNA transcripts, thereby substantially reducing the number of mRNA transcripts to be translated into polypeptides. For further general details see for example, Grierson et al. (1998) WO 98/53083; Waterhouse et al. (1999) WO 99/53050).

Performance of the methods of the invention does not rely on introducing and expressing in a plant a genetic construct into which the nucleic acid is cloned as an inverted repeat, but any one or more of several well-known "gene silencing" methods may be used to achieve the same effects.

One such method for the reduction of endogenous gene expression is RNA-mediated silencing of gene expression (downregulation). Silencing in this case is triggered in a plant by a double stranded RNA sequence (dsRNA) that is substantially similar to the target endogenous gene. This dsRNA is further processed by the plant into about 20 to about 26 nucleotides called short interfering RNAs (siRNAs). The siRNAs are incorporated into an RNA-induced silencing complex (RISC) that cleaves the mRNA transcript of the endogenous target gene, thereby substantially reducing the number of mRNA transcripts to be translated into a polypeptide. Preferably, the double stranded RNA sequence corresponds to a target gene.

Another example of an RNA silencing method involves the introduction of nucleic acid sequences or parts thereof (in this case a stretch of substantially contiguous nucleotides derived from the gene of interest, or from any nucleic acid capable of encoding an orthologue, paralogue or homologue of the protein of interest) in a sense orientation into a plant. "Sense orientation" refers to a DNA sequence that is homologous to an mRNA transcript thereof. Introduced into a plant would therefore be at least one copy of the nucleic acid sequence. The additional nucleic acid sequence will reduce expression of the endogenous gene, giving rise to a phenomenon known as co-suppression. The reduction of gene expression will be more pronounced if several additional copies of a nucleic acid

sequence are introduced into the plant, as there is a positive correlation between high transcript levels and the triggering of co-suppression.

Another example of an RNA silencing method involves the use of antisense nucleic acid sequences. An "antisense" nucleic acid sequence comprises a nucleotide sequence that is complementary to a "sense" nucleic acid sequence encoding a protein, i.e. complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA transcript sequence. The antisense nucleic acid sequence is preferably complementary to the endogenous gene to be silenced. The complementarity may be located in the "coding region" and/or in the "non-coding region" of a gene. The term "coding region" refers to a region of the nucleotide sequence comprising codons that are translated into amino acid residues. The term "non-coding region" refers to 5' and 3' sequences that flank the coding region that are transcribed but not translated into amino acids (also referred to as 5' and 3' untranslated regions).

Antisense nucleic acid sequences can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid sequence may be complementary to the entire nucleic acid sequence (in this case a stretch of substantially contiguous nucleotides derived from the gene of interest, or from any nucleic acid capable of encoding an orthologue, paralogue or homologue of the protein of interest), but may also be an oligonucleotide that is antisense to only a part of the nucleic acid sequence (including the mRNA 5' and 3' UTR). For example, the antisense oligonucleotide sequence may be complementary to the region surrounding the translation start site of an mRNA transcript encoding a polypeptide. The length of a suitable antisense oligonucleotide sequence is known in the art and may start from about 50, 45, 40, 35, 30, 25, 20, 15 or 10 nucleotides in length or less. An antisense nucleic acid sequence according to the invention may be constructed using chemical synthesis and enzymatic ligation reactions using methods known in the art. For example, an antisense nucleic acid sequence (e.g., an antisense oligonucleotide sequence) may be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acid sequences, e.g., phosphorothioate derivatives and acridine substituted nucleotides may be used. Examples of modified nucleotides that may be used to generate the antisense nucleic acid sequences are well known in the art. Known nucleotide modifications include methylation, cyclization and 'caps' and substitution of one or more of the naturally occurring nucleotides with an analogue such as inosine. Other modifications of nucleotides are well known in the art.

The antisense nucleic acid sequence can be produced biologically using an expression vector into which a nucleic acid sequence has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest). Preferably, production of antisense nucleic acid sequences

in plants occurs by means of a stably integrated nucleic acid construct comprising a promoter, an operably linked antisense oligonucleotide, and a terminator.

5 The nucleic acid molecules used for silencing in the methods of the invention (whether introduced into a plant or generated in situ) hybridize with or bind to mRNA transcripts and/or genomic DNA encoding a polypeptide to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid sequence which binds to DNA duplexes, through specific interactions
10 in the major groove of the double helix. Antisense nucleic acid sequences may be introduced into a plant by transformation or direct injection at a specific tissue site. Alternatively, antisense nucleic acid sequences can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense nucleic acid sequences can be modified such that they specifically bind to receptors or antigens
15 expressed on a selected cell surface, e.g., by linking the antisense nucleic acid sequence to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid sequences can also be delivered to cells using the vectors described herein.

20 According to a further aspect, the antisense nucleic acid sequence is an a-anomeric nucleic acid sequence. An a-anomeric nucleic acid sequence forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual b-units, the strands run parallel to each other (Gaultier et al. (1987) Nucl Ac Res 15: 6625-6641). The antisense nucleic acid sequence may also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) Nucl Ac Res 15, 6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) FEBS
25 Lett. 215, 327-330).

The reduction or substantial elimination of endogenous gene expression may also be performed using ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid sequence, such as an
30 mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) Nature 334, 585-591) can be used to catalytically cleave mRNA transcripts encoding a polypeptide, thereby substantially reducing the number of mRNA transcripts to be translated into a polypeptide. A ribozyme having specificity for a nucleic acid sequence can be designed (see for example: Cech et al.
35 U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742). Alternatively, mRNA transcripts corresponding to a nucleic acid sequence can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (Bartel and Szostak (1993) Science 261, 1411-1418). The use of ribozymes for gene silencing in plants is known in the art (e.g., Atkins et al. (1994) WO 94/00012; Lenne et al. (1995) WO
40 95/03404; Lutziger et al. (2000) WO 00/00619; Prinsen et al. (1997) WO 97/13865 and Scott et al. (1997) WO 97/38116).

Gene silencing may also be achieved by insertion mutagenesis (for example, T-DNA insertion or transposon insertion) or by strategies as described by, among others, Angell and Baulcombe ((1999) Plant J 20(3): 357-62), (Amplicon VIGS WO 98/36083), or Baulcombe (WO 99/15682).

5

Gene silencing may also occur if there is a mutation on an endogenous gene and/or a mutation on an isolated gene/nucleic acid subsequently introduced into a plant. The reduction or substantial elimination may be caused by a non-functional polypeptide. For example, the polypeptide may bind to various interacting proteins; one or more mutation(s) and/or truncation(s) may therefore provide for a polypeptide that is still able to bind interacting proteins (such as receptor proteins) but that cannot exhibit its normal function (such as signalling ligand).

A further approach to gene silencing is by targeting nucleic acid sequences complementary to the regulatory region of the gene (e.g., the promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See Helene, C., Anticancer Drug Res. 6, 569-84, 1991; Helene et al., Ann. N.Y. Acad. Sci. 660, 27-36 1992; and Maher, L.J. Bioassays 14, 807-15, 1992.

Other methods, such as the use of antibodies directed to an endogenous polypeptide for inhibiting its function in planta, or interference in the signalling pathway in which a polypeptide is involved, will be well known to the skilled man. In particular, it can be envisaged that manmade molecules may be useful for inhibiting the biological function of a target polypeptide, or for interfering with the signalling pathway in which the target polypeptide is involved.

Alternatively, a screening program may be set up to identify in a plant population natural variants of a gene, which variants encode polypeptides with reduced activity. Such natural variants may also be used for example, to perform homologous recombination.

30

Artificial and/or natural microRNAs (miRNAs) may be used to knock out gene expression and/or mRNA translation. Endogenous miRNAs are single stranded small RNAs of typically 19-24 nucleotides long. They function primarily to regulate gene expression and/ or mRNA translation. Most plant microRNAs (miRNAs) have perfect or near-perfect complementarity with their target sequences. However, there are natural targets with up to five mismatches. They are processed from longer non-coding RNAs with characteristic fold-back structures by double-strand specific RNases of the Dicer family. Upon processing, they are incorporated in the RNA-induced silencing complex (RISC) by binding to its main component, an Argonaute protein. MiRNAs serve as the specificity components of RISC, since they base-pair to target nucleic acids, mostly mRNAs, in the cytoplasm. Subsequent regulatory events include target mRNA cleavage and destruction and/or translational inhibition. Effects of miRNA overexpression are thus often reflected in decreased mRNA

40

levels of target genes.

Artificial microRNAs (amiRNAs), which are typically 21 nucleotides in length, can be genetically engineered specifically to negatively regulate gene expression of single or multiple genes of interest. Determinants of plant microRNA target selection are well known in the art. Empirical parameters for target recognition have been defined and can be used to aid in the design of specific amiRNAs, (Schwab et al., Dev. Cell 8, 517–527, 2005). Convenient tools for design and generation of amiRNAs and their precursors are also available to the public (Schwab et al., Plant Cell 18, 1121-1133, 2006).

For optimal performance, the gene silencing techniques used for reducing expression in a plant of an endogenous gene requires the use of nucleic acid sequences from monocotyledonous plants for transformation of monocotyledonous plants, and from dicotyledonous plants for transformation of dicotyledonous plants. Preferably, a nucleic acid sequence from any given plant species is introduced into that same species. For example, a nucleic acid sequence from rice is transformed into a rice plant. However, it is not an absolute requirement that the nucleic acid sequence to be introduced originates from the same plant species as the plant in which it will be introduced. It is sufficient that there is substantial homology between the endogenous target gene and the nucleic acid to be introduced.

Described above are examples of various methods for the reduction or substantial elimination of expression in a plant of an endogenous gene. A person skilled in the art would readily be able to adapt the aforementioned methods for silencing so as to achieve reduction of expression of an endogenous gene in a whole plant or in parts thereof through the use of an appropriate promoter, for example.

Transformation

The term “introduction” or “transformation” as referred to herein encompasses the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for transfer. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a genetic construct of the present invention and a whole plant regenerated there from. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem). The polynucleotide may be transiently or stably introduced into a host cell and may be maintained non-integrated, for example, as a plasmid. Alternatively, it may be integrated into the host genome. The resulting transformed plant cell may then be used to regenerate a transformed plant in a manner known to persons skilled in the art. Alternatively, a plant cell that cannot be regenerated into a plant

may be chosen as host cell, i.e. the resulting transformed plant cell does not have the capacity to regenerate into a (whole) plant.

The transfer of foreign genes into the genome of a plant is called transformation. Transformation of plant species is now a fairly routine technique. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable ancestor cell. The methods described for the transformation and regeneration of plants from plant tissues or plant cells may be utilized for transient or for stable transformation. Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts (Krens, F.A. et al., (1982) Nature 296, 72-74; Negrutiu I et al. (1987) Plant Mol Biol 8: 363-373); electroporation of protoplasts (Shillito R.D. et al. (1985) Bio/Technol 3, 1099-1102); microinjection into plant material (Crossway A et al., (1986) Mol. Gen Genet 202: 179-185); DNA or RNA-coated particle bombardment (Klein TM et al., (1987) Nature 327: 70) infection with (non-integrative) viruses and the like. Transgenic plants, including transgenic crop plants, are preferably produced via *Agrobacterium*-mediated transformation. An advantageous transformation method is the transformation *in planta*. To this end, it is possible, for example, to allow the agrobacteria to act on plant seeds or to inoculate the plant meristem with agrobacteria. It has proved particularly expedient in accordance with the invention to allow a suspension of transformed agrobacteria to act on the intact plant or at least on the flower primordia. The plant is subsequently grown on until the seeds of the treated plant are obtained (Clough and Bent, Plant J. (1998) 16, 735-743). Methods for *Agrobacterium*-mediated transformation of rice include well known methods for rice transformation, such as those described in any of the following: European patent application EP 1198985 A1, Aldemita and Hodges (Planta 199: 612-617, 1996); Chan et al. (Plant Mol Biol 22 (3): 491-506, 1993), Hiei et al. (Plant J 6 (2): 271-282, 1994), which disclosures are incorporated by reference herein as if fully set forth. In the case of corn transformation, the preferred method is as described in either Ishida et al. (Nat. Biotechnol 14(6): 745-50, 1996) or Frame et al. (Plant Physiol 129(1): 13-22, 2002), which disclosures are incorporated by reference herein as if fully set forth. Said methods are further described by way of example in B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press (1993) 128-143 and in Potrykus Annu. Rev. Plant Physiol. Plant Molec. Biol. 42 (1991) 205-225). The nucleic acids or the construct to be expressed is preferably cloned into a vector, which is suitable for transforming *Agrobacterium tumefaciens*, for example pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984) 8711). Agrobacteria transformed by such a vector can then be used in known manner for the transformation of plants, such as plants used as a model, like *Arabidopsis* (*Arabidopsis thaliana* is within the scope of the present invention not considered as a crop plant), or crop plants such as, by way of example, tobacco plants, for example by immersing bruised leaves or chopped leaves in an agrobacterial solution and

then culturing them in suitable media. The transformation of plants by means of *Agrobacterium tumefaciens* is described, for example, by Höfgen and Willmitzer in Nucl. Acid Res. (1988) 16, 9877 or is known inter alia from F.F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press, 1993, pp. 15-38.

In addition to the transformation of somatic cells, which then have to be regenerated into intact plants, it is also possible to transform the cells of plant meristems and in particular those cells which develop into gametes. In this case, the transformed gametes follow the natural plant development, giving rise to transgenic plants. Thus, for example, seeds of *Arabidopsis* are treated with agrobacteria and seeds are obtained from the developing plants of which a certain proportion is transformed and thus transgenic [Feldman, KA and Marks MD (1987). Mol Gen Genet 208:1-9; Feldmann K (1992). In: C Koncz, N-H Chua and J Shell, eds, Methods in Arabidopsis Research. Word Scientific, Singapore, pp. 274-289]. Alternative methods are based on the repeated removal of the inflorescences and incubation of the excision site in the center of the rosette with transformed agrobacteria, whereby transformed seeds can likewise be obtained at a later point in time (Chang (1994). Plant J. 5: 551-558; Katavic (1994). Mol Gen Genet, 245: 363-370). However, an especially effective method is the vacuum infiltration method with its modifications such as the "floral dip" method. In the case of vacuum infiltration of *Arabidopsis*, intact plants under reduced pressure are treated with an agrobacterial suspension [Bechthold, N (1993). C R Acad Sci Paris Life Sci, 316: 1194-1199], while in the case of the "floral dip" method the developing floral tissue is incubated briefly with a surfactant-treated agrobacterial suspension [Clough, SJ and Bent AF (1998) The Plant J. 16, 735-743]. A certain proportion of transgenic seeds are harvested in both cases, and these seeds can be distinguished from non-transgenic seeds by growing under the above-described selective conditions. In addition the stable transformation of plastids is of advantages because plastids are inherited maternally in most crops reducing or eliminating the risk of transgene flow through pollen. The transformation of the chloroplast genome is generally achieved by a process which has been schematically displayed in Klaus et al., 2004 [Nature Biotechnology 22 (2), 225-229]. Briefly the sequences to be transformed are cloned together with a selectable marker gene between flanking sequences homologous to the chloroplast genome. These homologous flanking sequences direct site specific integration into the plastome. Plastidal transformation has been described for many different plant species and an overview is given in Bock (2001) Transgenic plastids in basic research and plant biotechnology. J Mol Biol. 2001 Sep 21; 312 (3):425-38 or Maliga, P (2003) Progress towards commercialization of plastid transformation technology. Trends Biotechnol. 21, 20-28. Further biotechnological progress has recently been reported in form of marker free plastid transformants, which can be produced by a transient co-integrated marker gene (Klaus et al., 2004, Nature Biotechnology 22(2), 225-229).

The genetically modified plant cells can be regenerated via all methods with which the

skilled worker is familiar. Suitable methods can be found in the abovementioned publications by S.D. Kung and R. Wu, Potrykus or Höfgen and Willmitzer. Alternatively, the genetically modified plant cells are non-regenerable into a whole plant.

- 5 Generally after transformation, plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant-expressible genes co-transferred with the gene of interest, following which the transformed material is regenerated into a whole plant. To select transformed plants, the plant material obtained in the transformation is, as a rule, subjected to selective conditions so that transformed plants can be distinguished from
10 untransformed plants. For example, the seeds obtained in the above-described manner can be planted and, after an initial growing period, subjected to a suitable selection by spraying. A further possibility consists in growing the seeds, if appropriate after sterilization, on agar plates using a suitable selection agent so that only the transformed seeds can grow into plants. Alternatively, the transformed plants are screened for the presence of a selectable
15 marker such as the ones described above.

Following DNA transfer and regeneration, putatively transformed plants may also be evaluated, for instance using Southern analysis, for the presence of the gene of interest, copy number and/or genomic organisation. Alternatively or additionally, expression levels of
20 the newly introduced DNA may be monitored using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1)
25 transformed plant may be selfed and homozygous second-generation (or T2) transformants selected, and the T2 plants may then further be propagated through classical breeding techniques. The generated transformed organisms may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed cells; clonal transformants (e.g., all cells transformed to contain the expression cassette); grafts of
30 transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

Throughout this application a plant, plant part, seed or plant cell transformed with - or interchangeably transformed by - a construct or transformed with or by a nucleic acid is to
35 be understood as meaning a plant, plant part, seed or plant cell that carries said construct or said nucleic acid as a transgene due the result of an introduction of said construct or said nucleic acid by biotechnological means. The plant, plant part, seed or plant cell therefore comprises said recombinant construct or said recombinant nucleic acid. Any plant, plant part, seed or plant cell that no longer contains said recombinant construct or said
40 recombinant nucleic acid after introduction in the past, is termed null-segregant, nullizygote or null control, but is not considered a plant, plant part, seed or plant cell transformed with said construct or with said nucleic acid within the meaning of this application.

T-DNA activation tagging

“T-DNA activation” tagging (Hayashi et al. Science (1992) 1350-1353), involves insertion of T-DNA, usually containing a promoter (may also be a translation enhancer or an intron), in the genomic region of the gene of interest or 10 kb up- or downstream of the coding region of a gene in a configuration such that the promoter directs expression of the targeted gene. Typically, regulation of expression of the targeted gene by its natural promoter is disrupted and the gene falls under the control of the newly introduced promoter. The promoter is typically embedded in a T-DNA. This T-DNA is randomly inserted into the plant genome, for example, through *Agrobacterium* infection and leads to modified expression of genes near the inserted T-DNA. The resulting transgenic plants show dominant phenotypes due to modified expression of genes close to the introduced promoter.

TILLING

The term “TILLING” is an abbreviation of “Targeted Induced Local Lesions In Genomes” and refers to a mutagenesis technology useful to generate and/or identify nucleic acids encoding proteins with modified expression and/or activity. TILLING also allows selection of plants carrying such mutant variants. These mutant variants may exhibit modified expression, either in strength or in location or in timing (if the mutations affect the promoter for example). These mutant variants may exhibit higher activity than that exhibited by the gene in its natural form. TILLING combines high-density mutagenesis with high-throughput screening methods. The steps typically followed in TILLING are: (a) EMS mutagenesis (Redei GP and Koncz C (1992) In Methods in Arabidopsis Research, Koncz C, Chua NH, Schell J, eds. Singapore, World Scientific Publishing Co, pp. 16–82; Feldmann et al., (1994) In Meyerowitz EM, Somerville CR, eds, Arabidopsis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 137-172; Lightner J and Caspar T (1998) In J Martinez-Zapater, J Salinas, eds, Methods on Molecular Biology, Vol. 82. Humana Press, Totowa, NJ, pp 91-104); (b) DNA preparation and pooling of individuals; (c) PCR amplification of a region of interest; (d) denaturation and annealing to allow formation of heteroduplexes; (e) DHPLC, where the presence of a heteroduplex in a pool is detected as an extra peak in the chromatogram; (f) identification of the mutant individual; and (g) sequencing of the mutant PCR product. Methods for TILLING are well known in the art (McCallum et al., (2000) Nat Biotechnol 18: 455-457; reviewed by Stemple (2004) Nat Rev Genet 5(2): 145-50).

Homologous recombination

“Homologous recombination” allows introduction in a genome of a selected nucleic acid at a defined selected position. Homologous recombination is a standard technology used routinely in biological sciences for lower organisms such as yeast or the moss *Physcomitrella*. Methods for performing homologous recombination in plants have been described not only for model plants (Offringa et al. (1990) EMBO J 9(10): 3077-84) but also for crop plants, for example rice (Terada et al. (2002) Nat Biotech 20(10): 1030-4; Iida and Terada (2004) Curr Opin Biotech 15(2): 132-8), and approaches exist that are generally

applicable regardless of the target organism (Miller et al, Nature Biotechnol. 25, 778-785, 2007).

Yield related Trait(s)

5 A "Yield related trait" is a trait or feature which is related to plant yield. Yield-related traits may comprise one or more of the following non-limitative list of features: early flowering time, yield, biomass, seed yield, early vigour, greenness index, growth rate, agronomic traits, such as e.g. tolerance to submergence (which leads to yield in rice), Water Use Efficiency (WUE), Nitrogen Use Efficiency (NUE), etc.

10 Reference herein to enhanced yield-related traits, relative to of control plants is taken to mean one or more of an increase in early vigour and/or in biomass (weight) of one or more parts of a plant, which may include (i) aboveground parts and preferably aboveground harvestable parts and/or (ii) parts below ground and preferably harvestable below ground.

15 In particular, such harvestable parts are roots such as taproots, stems, beets, leaves, flowers or seeds, and performance of the methods of the invention results in plants having increased seed yield relative to the seed yield of control plants, and/or increased stem biomass relative to the stem biomass of control plants, and/or increased root biomass relative to the root biomass of control plants and/or increased beet biomass relative to the

20 beet biomass of control plants. Moreover, it is particularly contemplated that the sugar content (in particular the sucrose content) in the above ground parts, particularly stem (in particular of sugar cane plants) and/or in the belowground parts, in particular in roots including taproots, tubers and/or beets (in particular in sugar beets) is increased relative to the sugar content (in particular the sucrose content) in corresponding part(s) of the control

25 plant.

In an alternative embodiment, such harvestable parts are seeds.

Yield

30 The term "yield" in general means a measurable produce of economic value, typically related to a specified crop, to an area, and to a period of time. Individual plant parts directly contribute to yield based on their number, size and/or weight, or the actual yield is the yield per square meter for a crop and year, which is determined by dividing total production (includes both harvested and appraised production) by planted square meters.

35 The terms "yield" of a plant and "plant yield" are used interchangeably herein and are meant to refer to vegetative biomass such as root and/or shoot biomass, to reproductive organs, and/or to propagules such as seeds of that plant.

40 Flowers in maize are unisexual; male inflorescences (tassels) originate from the apical stem and female inflorescences (ears) arise from axillary bud apices. The female inflorescence produces pairs of spikelets on the surface of a central axis (cob). Each of the female spikelets encloses two fertile florets, one of them will usually mature into a maize kernel

once fertilized. Hence a yield increase in maize may be manifested as one or more of the following: increase in the number of plants established per square meter, an increase in the number of ears per plant, an increase in the number of rows, number of kernels per row, kernel weight, thousand kernel weight, ear length/diameter, increase in the seed filling rate, which is the number of filled florets (i.e. florets containing seed) divided by the total number of florets and multiplied by 100), among others.

Inflorescences in rice plants are named panicles. The panicle bears spikelets, which are the basic units of the panicles, and which consist of a pedicel and a floret. The floret is borne on the pedicel and includes a flower that is covered by two protective glumes: a larger glume (the lemma) and a shorter glume (the palea). Hence, taking rice as an example, a yield increase may manifest itself as an increase in one or more of the following: number of plants per square meter, number of panicles per plant, panicle length, number of spikelets per panicle, number of flowers (or florets) per panicle; an increase in the seed filling rate which is the number of filled florets (i.e. florets containing seeds) divided by the total number of florets and multiplied by 100; an increase in thousand kernel weight, among others.

Early flowering time

Plants having an "early flowering time" as used herein are plants which start to flower earlier than control plants. Hence this term refers to plants that show an earlier start of flowering. Flowering time of plants can be assessed by counting the number of days ("time to flower") between sowing and the emergence of a first inflorescence. The "flowering time" of a plant can for instance be determined using the method as described in WO 2007/093444.

Early vigour

"Early vigour" refers to active healthy well-balanced growth especially during early stages of plant growth, and may result from increased plant fitness due to, for example, the plants being better adapted to their environment (i.e. optimizing the use of energy resources and partitioning between shoot and root). Plants having early vigour also show increased seedling survival and a better establishment of the crop, which often results in highly uniform fields (with the crop growing in uniform manner, i.e. with the majority of plants reaching the various stages of development at substantially the same time), and often better and higher yield. Therefore, early vigour may be determined by measuring various factors, such as thousand kernel weight, percentage germination, percentage emergence, seedling growth, seedling height, root length, root and shoot biomass and many more.

Increased growth rate

The increased growth rate may be specific to one or more parts of a plant (including seeds), or may be throughout substantially the whole plant. Plants having an increased growth rate may have a shorter life cycle. The life cycle of a plant may be taken to mean the time needed to grow from a mature seed up to the stage where the plant has produced mature

seeds, similar to the starting material. This life cycle may be influenced by factors such as speed of germination, early vigour, growth rate, greenness index, flowering time and speed of seed maturation. The increase in growth rate may take place at one or more stages in the life cycle of a plant or during substantially the whole plant life cycle. Increased growth rate during the early stages in the life cycle of a plant may reflect enhanced vigour. The increase in growth rate may alter the harvest cycle of a plant allowing plants to be sown later and/or harvested sooner than would otherwise be possible (a similar effect may be obtained with earlier flowering time). If the growth rate is sufficiently increased, it may allow for the further sowing of seeds of the same plant species (for example sowing and harvesting of rice plants followed by sowing and harvesting of further rice plants all within one conventional growing period). Similarly, if the growth rate is sufficiently increased, it may allow for the further sowing of seeds of different plants species (for example the sowing and harvesting of corn plants followed by, for example, the sowing and optional harvesting of soybean, potato or any other suitable plant). Harvesting additional times from the same rootstock in the case of some crop plants may also be possible. Altering the harvest cycle of a plant may lead to an increase in annual biomass production per square meter (due to an increase in the number of times (say in a year) that any particular plant may be grown and harvested). An increase in growth rate may also allow for the cultivation of transgenic plants in a wider geographical area than their wild-type counterparts, since the territorial limitations for growing a crop are often determined by adverse environmental conditions either at the time of planting (early season) or at the time of harvesting (late season). Such adverse conditions may be avoided if the harvest cycle is shortened. The growth rate may be determined by deriving various parameters from growth curves, such parameters may be: T-Mid (the time taken for plants to reach 50% of their maximal size) and T-90 (time taken for plants to reach 90% of their maximal size), amongst others.

Stress resistance

An increase in yield and/or growth rate occurs whether the plant is under non-stress conditions or whether the plant is exposed to various stresses compared to control plants. Plants typically respond to exposure to stress by growing more slowly. In conditions of severe stress, the plant may even stop growing altogether. Mild stress on the other hand is defined herein as being any stress to which a plant is exposed which does not result in the plant ceasing to grow altogether without the capacity to resume growth. Mild stress in the sense of the invention leads to a reduction in the growth of the stressed plants of less than 40%, 35%, 30% or 25%, more preferably less than 20% or 15% in comparison to the control plant under non-stress conditions. Due to advances in agricultural practices (irrigation, fertilization, pesticide treatments) severe stresses are not often encountered in cultivated crop plants. As a consequence, the compromised growth induced by mild stress is often an undesirable feature for agriculture. Abiotic stresses may be due to drought or excess water, anaerobic stress, salt stress, chemical toxicity, oxidative stress and hot, cold or freezing temperatures.

“Biotic stresses” are typically those stresses caused by pathogens, such as bacteria, viruses, fungi, nematodes and insects.

The “abiotic stress” may be an osmotic stress caused by a water stress, e.g. due to drought, salt stress, or freezing stress. Abiotic stress may also be an oxidative stress or a cold stress. “Freezing stress” is intended to refer to stress due to freezing temperatures, i.e. temperatures at which available water molecules freeze and turn into ice. “Cold stress”, also called “chilling stress”, is intended to refer to cold temperatures, e.g. temperatures below 10°, or preferably below 5°C, but at which water molecules do not freeze. As reported in Wang et al. (Planta (2003) 218: 1-14), abiotic stress leads to a series of morphological, physiological, biochemical and molecular changes that adversely affect plant growth and productivity. Drought, salinity, extreme temperatures and oxidative stress are known to be interconnected and may induce growth and cellular damage through similar mechanisms. Rabbani et al. (Plant Physiol (2003) 133: 1755-1767) describes a particularly high degree of “cross talk” between drought stress and high-salinity stress. For example, drought and/or salinisation are manifested primarily as osmotic stress, resulting in the disruption of homeostasis and ion distribution in the cell. Oxidative stress, which frequently accompanies high or low temperature, salinity or drought stress, may cause denaturing of functional and structural proteins. As a consequence, these diverse environmental stresses often activate similar cell signalling pathways and cellular responses, such as the production of stress proteins, up-regulation of anti-oxidants, accumulation of compatible solutes and growth arrest. The term “non-stress” conditions as used herein are those environmental conditions that allow optimal growth of plants. Persons skilled in the art are aware of normal soil conditions and climatic conditions for a given location. Plants with optimal growth conditions, (grown under non-stress conditions) typically yield in increasing order of preference at least 97%, 95%, 92%, 90%, 87%, 85%, 83%, 80%, 77% or 75% of the average production of such plant in a given environment. Average production may be calculated on harvest and/or season basis. Persons skilled in the art are aware of average yield productions of a crop.

In particular, the methods of the present invention may be performed under non-stress conditions. In an example, the methods of the present invention may be performed under non-stress conditions such as mild drought to give plants having increased yield relative to control plants.

In another embodiment, the methods of the present invention may be performed under stress conditions.

In an example, the methods of the present invention may be performed under stress conditions such as drought to give plants having increased yield relative to control plants.

In another example, the methods of the present invention may be performed under stress conditions such as nutrient deficiency to give plants having increased yield relative to control plants.

Nutrient deficiency may result from a lack of nutrients such as nitrogen, phosphates and other phosphorous-containing compounds, potassium, calcium, magnesium, manganese, iron and boron, amongst others.

In yet another example, the methods of the present invention may be performed under stress conditions such as salt stress to give plants having increased yield relative to control plants. The term salt stress is not restricted to common salt (NaCl), but may be any one or more of: NaCl, KCl, LiCl, MgCl₂, CaCl₂, amongst others.

In yet another example, the methods of the present invention may be performed under stress conditions such as cold stress or freezing stress to give plants having increased yield relative to control plants.

Increase/Improve/Enhance

The terms “increase”, “improve” or “enhance” are interchangeable and shall mean in the sense of the application at least a 3%, 4%, 5%, 6%, 7%, 8%, 9% or 10%, preferably at least 15% or 20%, more preferably 25%, 30%, 35% or 40% more yield and/or growth in comparison to control plants as defined herein.

Seed yield

Increased seed yield may manifest itself as one or more of the following:

- (a) an increase in seed biomass (total seed weight) which may be on an individual seed basis and/or per plant and/or per square meter;
- (b) increased number of flowers per plant;
- (c) increased number of seeds;
- (d) increased seed filling rate (which is expressed as the ratio between the number of filled florets divided by the total number of florets);
- (e) increased harvest index, which is expressed as a ratio of the yield of harvestable parts, such as seeds, divided by the biomass of aboveground plant parts; and
- (f) increased thousand kernel weight (TKW), which is extrapolated from the number of seeds counted and their total weight. An increased TKW may result from an increased seed size and/or seed weight, and may also result from an increase in embryo and/or endosperm size.

The terms “filled florets” and “filled seeds” may be considered synonyms.

An increase in seed yield may also be manifested as an increase in seed size and/or seed volume. Furthermore, an increase in seed yield may also manifest itself as an increase in seed area and/or seed length and/or seed width and/or seed perimeter.

Greenness Index

The “greenness index” as used herein is calculated from digital images of plants. For each pixel belonging to the plant object on the image, the ratio of the green value versus the red value (in the RGB model for encoding color) is calculated. The greenness index is expressed as the percentage of pixels for which the green-to-red ratio exceeds a given

threshold. Under normal growth conditions, under salt stress growth conditions, and under reduced nutrient availability growth conditions, the greenness index of plants is measured in the last imaging before flowering. In contrast, under drought stress growth conditions, the greenness index of plants is measured in the first imaging after drought.

5

Biomass

The term “biomass” as used herein is intended to refer to the total weight of a plant. Within the definition of biomass, a distinction may be made between the biomass of one or more parts of a plant, which may include any one or more of the following:

- 10 - aboveground parts such as but not limited to shoot biomass, seed biomass, leaf biomass, etc.;
- aboveground harvestable parts such as but not limited to shoot biomass, seed biomass, leaf biomass, etc.;
- parts below ground, such as but not limited to root biomass, tubers, bulbs, etc.;
- 15 - harvestable parts below ground, such as but not limited to root biomass, tubers, bulbs, etc.;
- harvestable parts partially below ground such as but not limited to beets and other hypocotyl areas of a plant, rhizomes, stolons or creeping rootstalks;
- vegetative biomass such as root biomass, shoot biomass, etc.;
- 20 - reproductive organs; and
- propagules such as seed.

Marker assisted breeding

Such breeding programmes sometimes require introduction of allelic variation by mutagenic treatment of the plants, using for example EMS mutagenesis; alternatively, the programme may start with a collection of allelic variants of so called “natural” origin caused unintentionally. Identification of allelic variants then takes place, for example, by PCR. This is followed by a step for selection of superior allelic variants of the sequence in question and which give increased yield. Selection is typically carried out by monitoring growth performance of plants containing different allelic variants of the sequence in question. Growth performance may be monitored in a greenhouse or in the field. Further optional steps include crossing plants in which the superior allelic variant was identified with another plant. This could be used, for example, to make a combination of interesting phenotypic features.

35

Use as probes in (gene mapping)

Use of nucleic acids encoding the protein of interest for genetically and physically mapping the genes requires only a nucleic acid sequence of at least 15 nucleotides in length. These nucleic acids may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning, A Laboratory Manual) of restriction-digested plant genomic DNA may be probed with the nucleic acids encoding the protein of interest. The resulting banding patterns may then be

40

subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1: 174-181) in order to construct a genetic map. In addition, the nucleic acids may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the nucleic acid encoding the protein of interest in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4: 37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

The nucleic acid probes may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al. In: *Non-mammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, the nucleic acid probes may be used in direct fluorescence in situ hybridisation (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favour use of large clones (several kb to several hundred kb; see Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods for genetic and physical mapping may be carried out using the nucleic acids. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Plant

The term "plant" as used herein encompasses whole plants, ancestors and progeny of the plants and plant parts, including seeds, shoots, stems, leaves, roots (including tubers),

flowers, and tissues and organs, wherein each of the aforementioned comprise the gene/nucleic acid of interest. The term "plant" also encompasses plant cells, suspension cultures, callus tissue, embryos, meristematic regions, gametophytes, sporophytes, pollen and microspores, again wherein each of the aforementioned comprises the gene/nucleic acid of interest.

Plants that are particularly useful in the methods of the invention include all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants including fodder or forage legumes, ornamental plants, food crops, trees or shrubs selected from the list comprising *Acer* spp., *Actinidia* spp., *Abelmoschus* spp., *Agave sisalana*, *Agropyron* spp., *Agrostis stolonifera*, *Allium* spp., *Amaranthus* spp., *Ammophila arenaria*, *Ananas comosus*, *Annona* spp., *Apium graveolens*, *Arachis* spp., *Artocarpus* spp., *Asparagus officinalis*, *Avena* spp. (e.g. *Avena sativa*, *Avena fatua*, *Avena byzantina*, *Avena fatua* var. *sativa*, *Avena hybrida*), *Averrhoa carambola*, *Bambusa* sp., *Benincasa hispida*, *Bertholletia excelsea*, *Beta vulgaris*, *Brassica* spp. (e.g. *Brassica napus*, *Brassica rapa* ssp. [canola, oilseed rape, turnip rape]), *Cadaba farinosa*, *Camellia sinensis*, *Canna indica*, *Cannabis sativa*, *Capsicum* spp., *Carex elata*, *Carica papaya*, *Carissa macrocarpa*, *Carya* spp., *Carthamus tinctorius*, *Castanea* spp., *Ceiba pentandra*, *Cichorium endivia*, *Cinnamomum* spp., *Citrullus lanatus*, *Citrus* spp., *Cocos* spp., *Coffea* spp., *Colocasia esculenta*, *Cola* spp., *Corchorus* sp., *Coriandrum sativum*, *Corylus* spp., *Crataegus* spp., *Crocus sativus*, *Cucurbita* spp., *Cucumis* spp., *Cynara* spp., *Daucus carota*, *Desmodium* spp., *Dimocarpus longan*, *Dioscorea* spp., *Diospyros* spp., *Echinochloa* spp., *Elaeis* (e.g. *Elaeis guineensis*, *Elaeis oleifera*), *Eleusine coracana*, *Eragrostis tef*, *Erianthus* sp., *Eriobotrya japonica*, *Eucalyptus* sp., *Eugenia uniflora*, *Fagopyrum* spp., *Fagus* spp., *Festuca arundinacea*, *Ficus carica*, *Fortunella* spp., *Fragaria* spp., *Ginkgo biloba*, *Glycine* spp. (e.g. *Glycine max*, *Soja hispida* or *Soja max*), *Gossypium hirsutum*, *Helianthus* spp. (e.g. *Helianthus annuus*), *Hemerocallis fulva*, *Hibiscus* spp., *Hordeum* spp. (e.g. *Hordeum vulgare*), *Ipomoea batatas*, *Juglans* spp., *Lactuca sativa*, *Lathyrus* spp., *Lens culinaris*, *Linum usitatissimum*, *Litchi chinensis*, *Lotus* spp., *Luffa acutangula*, *Lupinus* spp., *Luzula sylvatica*, *Lycopersicon* spp. (e.g. *Lycopersicon esculentum*, *Lycopersicon lycopersicum*, *Lycopersicon pyriforme*), *Macrotyloma* spp., *Malus* spp., *Malpighia emarginata*, *Mammea americana*, *Mangifera indica*, *Manihot* spp., *Manilkara zapota*, *Medicago sativa*, *Melilotus* spp., *Mentha* spp., *Miscanthus sinensis*, *Momordica* spp., *Morus nigra*, *Musa* spp., *Nicotiana* spp., *Olea* spp., *Opuntia* spp., *Ornithopus* spp., *Oryza* spp. (e.g. *Oryza sativa*, *Oryza latifolia*), *Panicum miliaceum*, *Panicum virgatum*, *Passiflora edulis*, *Pastinaca sativa*, *Pennisetum* sp., *Persea* spp., *Petroselinum crispum*, *Phalaris arundinacea*, *Phaseolus* spp., *Phleum pratense*, *Phoenix* spp., *Phragmites australis*, *Physalis* spp., *Pinus* spp., *Pistacia vera*, *Pisum* spp., *Poa* spp., *Populus* spp., *Prosopis* spp., *Prunus* spp., *Psidium* spp., *Punica granatum*, *Pyrus communis*, *Quercus* spp., *Raphanus sativus*, *Rheum rhabarbarum*, *Ribes* spp., *Ricinus communis*, *Rubus* spp., *Saccharum* spp., *Salix* sp., *Sambucus* spp., *Secale cereale*, *Sesamum* spp., *Sinapis* sp., *Solanum* spp. (e.g. *Solanum tuberosum*, *Solanum integrifolium* or *Solanum lycopersicum*), *Sorghum bicolor*, *Spinacia* spp.,

Syzygium spp., *Tagetes* spp., *Tamarindus indica*, *Theobroma cacao*, *Trifolium* spp., *Tripsacum dactyloides*, *Triticosecale rimpai*, *Triticum* spp. (e.g. *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum*, *Triticum monococcum* or *Triticum vulgare*), *Tropaeolum minus*, *Tropaeolum majus*, *Vaccinium* spp.,
5 *Vicia* spp., *Vigna* spp., *Viola odorata*, *Vitis* spp., *Zea mays*, *Zizania palustris*, *Ziziphus* spp., amongst others.

With respect to the sequences of the invention, a nucleic acid or a polypeptide sequence of plant origin has the characteristic of a codon usage optimised for expression in plants, and
10 of the use of amino acids and regulatory sites common in plants, respectively. The plant of origin may be any plant, but preferably those plants as described in the previous paragraph.

Control plant(s)

The choice of suitable control plants is a routine part of an experimental setup and may
15 include corresponding wild type plants or corresponding plants without the gene of interest. The control plant is typically of the same plant species or even of the same variety as the plant to be assessed. The control plant may also be a nullizygote of the plant to be assessed. Nullizygotes (or null control plants) are individuals missing the transgene by segregation. Further, control plants are grown under equal growing conditions to the
20 growing conditions of the plants of the invention, i.e. in the vicinity of, and simultaneously with, the plants of the invention. A “control plant” as used herein refers not only to whole plants, but also to plant parts, including seeds and seed parts.

Detailed description of the invention

25 The present invention shows that modulating expression in a plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, gives plants having enhanced yield-related traits relative or compared to control plants. The terms “relative to” and “compared to” can be used interchangeably throughout the text of this patent application.

30 According to a first embodiment, the present invention provides a method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, and optionally
35 selecting for plants having enhanced yield-related traits. According to another embodiment, the present invention provides a method for producing plants having enhanced yield-related traits relative to control plants, wherein said method comprises the steps of modulating expression in said plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as
40 described herein and optionally selecting for plants having enhanced yield-related traits.

A preferred method for modulating, preferably increasing, expression of a nucleic acid

encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, is by introducing and expressing in a plant a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, respectively. With respect to GPx-related polypeptides, a preferred method for modulating, preferably increasing, expression of a nucleic acid encoding a GPx-related polypeptide is by introducing and expressing in a plant a nucleic acid encoding a GPx-related polypeptide.

Any reference hereinafter to a “protein useful in the methods of the invention” is taken to mean an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined herein. Any reference hereinafter to a “nucleic acid useful in the methods of the invention” is taken to mean a nucleic acid capable of encoding such an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide. In one embodiment any reference to a protein or nucleic acid “useful in the methods of the invention” is to be understood to mean proteins or nucleic acids “useful in the methods, constructs, plants, harvestable parts and products of the invention”. The nucleic acid to be introduced into a plant (and therefore useful in performing the methods of the invention) is any nucleic acid encoding the type of protein which will now be described, hereafter also named “*ELM2-related* nucleic acid”, or “*WRKY-related* nucleic acid”, or “*EMG1-like* nucleic acid”, or “*GPx-related* nucleic acid”, or “*ELM2-related* gene”, or “*WRKY-related* gene”, or “*EMG1-like* gene”, or “*GPx-related* gene”.

The term “ELM2-related” or “ELM2-related polypeptide” as used herein also intends to include homologues as defined hereunder of “ELM2-related polypeptide”.

An “ELM2-related polypeptide” as defined herein refers to any polypeptide comprising one or more of the following motifs:

(i) Motif 1: [I/V]GKGR[S/Q]DSC[G/R]CQV[Q/P][K/G]S[I/V][K/E]CVRFH[I/V][T/A]E[R/K][S/R][S/A/L][R/K][V/L][M/K][R/L]E[L/I]G[K/V/S]AF[N/Y][Q/H/A]W[R/G/N][F/L]D[K/R][M/A]GEE (SEQ ID NO: 93),

(ii) Motif 2: [R/T/K]xFP[S/K][R/K][R/S/G]R[E/K][D/S/E]LVSYY[Y/F]NVFLL[Q/R]RR[A/G][N/Y]QNR[S/H/V]TP[D/N/K][S/N]I (SEQ ID NO: 94),

(iii) Motif 3: [P/S][I/P][T/R]xxIP[V/L]GP[V/N][F/H]QAE[V/I]PEWT (SEQ ID NO: 95),

wherein x can be any amino acid.

Motifs 1 to 3 were derived using the MEME algorithm (Bailey and Elkan, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994). At each position within a MEME motif, the residues are shown that are present in the query set of sequences with a frequency higher than 0.2. Residues within square brackets represent alternatives.

Preferably, the ELM2-related polypeptide comprises in increasing order of preference, at least 2, or all 3 motifs.

Additionally or alternatively, the homologue of an ELM2-related protein has in increasing order of preference at least 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% overall sequence identity to the polypeptide sequence represented by SEQ ID NO: 2 or SEQ ID NO: 4, provided that the homologous protein comprises any one or more of the motifs as outlined above. The overall sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG Wisconsin Package, Accelrys), preferably with default parameters and preferably with sequences of mature proteins (i.e. without taking into account secretion signals or transit peptides). In one embodiment the sequence identity level is determined by comparison of the polypeptide sequences over the entire length of the sequence of SEQ ID NO: 2 or SEQ ID NO: 4.

Compared to overall sequence identity, the sequence identity will generally be higher when only motifs are considered. Preferably the motifs in an ELM2-related polypeptide have, in increasing order of preference, at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one or more of the motifs represented by SEQ ID NO: 93 to SEQ ID NO: 95 (Motifs 1 to 3), provided that at least one of the motifs as represented in SEQ ID NO: 93 to SEQ ID NO: 95 (Motifs 1 to 3) is present in the ELM2-related polypeptide.

In other words, in another embodiment a method is provided wherein said ELM2-related polypeptide comprises a motif with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the motifs starting with amino acid 271 up to amino acid 320 in SEQ ID NO: 2, starting with amino acid 354 up to amino acid 388 in SEQ ID NO: 2 or starting with amino acid 220 up to amino acid 239 in SEQ ID NO: 2, provided the ELM2-related polypeptide comprises any one or more of the motifs as represented in SEQ ID NO: 93 to SEQ ID NO: 95 (Motifs 1 to 3).

The WRKY-related polypeptide as used herein comprises a 'WRKYGQK' motif that is invariant in most WRKY domains and the cysteine and histidine residues of the zinc-finger motif that is conserved in WRKY domains. Particularly, the WRKY-related polypeptide comprises the zinc-finger motif being a C2H2 zinc finger and comprising one or more WRKY motifs, preferably 'WRKYGQK' motif. Further, the WRKY-related polypeptide comprises a basic motif resembling a large T-antigen-type nuclear localization signal (NLS).

Particularly, the basic motif is NLS comprising 'KKAR' motif. The WRKY-related polypeptide also comprises a coiled-coil motif, preferably a WRKY-specific coiled-coil structure, more preferably a putative type II WRKY-specific coiled-coil structure, identified by 'coils' at the N-termini, and bulky hydrophobic residues that form the characteristic heptad repeat pattern of coiled coils. The WRKY-related polypeptide as used herein may also comprise three motifs as described herein above and exemplified in Figure 8.

Concerning the search of the motifs as used and defined herein, the information given in InterPro database as InterPro entry IPR003657 provides further information thereon: The WRKY domain is a 60 amino acid region that is defined by the conserved amino acid sequence WRKYGQK at its N-terminal end, together with a zinc-finger-like motif. The WRKY domain is found in one or two copies in a superfamily of plant transcription factors involved in the regulation of various physiological programs that are unique to plants, including pathogen defence, senescence, trichome development and the biosynthesis of secondary metabolites. The WRKY domain binds specifically to the DNA sequence motif (T)(T)TGAC(C/T), which is known as the W box. The invariant TGAC core of the W box is essential for function and WRKY binding.

Structural studies indicate that this domain is a four-stranded beta-sheet with a zinc binding pocket, forming a novel zinc and DNA binding structure. The WRKYGQK residues correspond to the most N-terminal beta-strand, which enables extensive hydrophobic interactions, contributing to the structural stability of the beta-sheet.

As used herein, the term "WRKY-related polypeptide" also intends to include homologues as defined hereunder of "WRKY-related polypeptide".

Additionally or alternatively, the homologue of a WRKY-related polypeptide has in increasing order of preference at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% overall sequence identity to the polypeptide sequence represented by SEQ ID NO: 103, provided that the homologous protein comprises any one or more of the conserved motifs as outlined above. The overall sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG Wisconsin Package, Accelrys), preferably with default parameters and preferably with sequences of mature proteins (i.e. without taking into account secretion signals or transit peptides). In one embodiment the sequence identity level is determined by comparison of the polypeptide sequences over the entire length of the sequence of SEQ ID NO: 103.

WRKY-related polypeptide as used herein has DNA-binding activity. Hybridising sequences useful in the methods of the invention encode a WRKY-related polypeptide as defined herein, having substantially the same biological activity as the polypeptide sequences given in Table A2 of the Examples section. Preferably, the hybridising sequence is capable of hybridising to the complement of any one of the nucleic acids given in Table A2 of the Examples section, or to a portion of any of these sequences, a portion being as defined above, or the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A2 of the Examples section. Most preferably, the hybridising sequence is capable of hybridising to the complement of a nucleic acid as represented by SEQ ID NO: 102 or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined above.

A "EMG1-like polypeptide" as defined herein refers to any polypeptide comprising an InterPro accession IPR005304 EMG1 domain corresponding to PFAM accession number PF03587. In a preferred embodiment, the EMG1-like polypeptide comprises an EMG1 domain having at least 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the EMG1 domain from amino acid 81 to 276 in SEQ ID NO: 179.

The term "EMG1-like" or "EMG1-like polypeptide" as used herein also intends to include homologues as defined hereunder of "EMG1-like polypeptide".

Most preferably, the EMG1-like polypeptide further comprises one of the signature sequences or motifs represented by SEQ ID NO: 286, SEQ ID NO: 287 or SEQ ID NO: 288.

Motifs 4 to 6 were derived using the MEME algorithm (Bailey and Elkan, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994). At each position within a MEME motif, the residues are shown that are present in the query set of sequences with a frequency higher than 0.2. Residues within square brackets represent alternatives.

More preferably, the EMG1-like polypeptide comprises in increasing order of preference, at least 2, or all 3 motifs.

Additionally or alternatively, the homologue of a EMG1-like protein has in increasing order of preference at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%

overall sequence identity to the polypeptide sequence represented by SEQ ID NO: 179, provided that the homologous protein comprises any one or more of the conserved domains and/or motifs as outlined above. The overall sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG
 5 Wisconsin Package, Accelrys), preferably with default parameters and preferably with sequences of mature proteins (i.e. without taking into account secretion signals or transit peptides). In one embodiment the sequence identity level is determined by comparison of the polypeptide sequences over the entire length of the sequence of SEQ ID NO: 179.

10 Compared to overall sequence identity, the sequence identity will generally be higher when only conserved domains or motifs are considered. Preferably the motifs in a EMG1-like polypeptide have, in increasing order of preference, at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one or more
 15 of the motifs represented by SEQ ID NO: 286 to SEQ ID NO: 288 (Motifs 4 to 6).

In other words, in another embodiment a method is provided wherein said EMG1-like polypeptide comprises a conserved motif with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%,
 20 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the conserved domain starting with amino acid 92 up to amino acid 141, amino acid 152 up to amino acid 201 and/or amino acid 232 up to amino acid 281 in SEQ ID NO: 179.

A "GPx-related polypeptide" as used and defined herein comprises peroxidase (Px)
 25 enzymes, particularly Glutathione peroxidase as described herein under Example 4. The GPx-related polypeptide" further comprises thioredoxin-like fold and/or thioredoxin fold protein as described herein under Example 4. Moreover, "GPx-related polypeptide" as used herein comprises twelve motifs and seven signatures as used and defined herein and further exemplified in Figure 17.

30 According one embodiment, there is provided a method for improving yield-related traits as provided herein in plants relative to or compared to control plants, comprising modulating expression in a plant of a nucleic acid encoding a GPx-related polypeptide as defined herein.

35 In one embodiment, the GPx-related polypeptide as used herein comprises one or more of the following signatures 1 to 6:

Signature 1 (SEQ ID NO: 373): G[K/R][L/V][I/L]LI[V/E/T]NVA[S/T/A][E/Q/L/Y][C/U]G
 [L/T]T

40 Signature 2 (SEQ ID NO: 374): VNVAS[R/K/Q]CG

Signature 3 (SEQ ID NO: 375): VN[T/V]A[S/T/A][R/K/E/Q/A]CG

Signature 4 (SEQ ID NO: 376): VN[T/V]AS[K/R/H/L/Q]C[G/S/A]

Signature 5 (SEQ ID NO: 377): LAFPCNQF

Signature 6 (SEQ ID NO: 378): WNF[S/T]KF

In another embodiment, the GPx-related polypeptide comprises in increasing order of preference, at least 2, at least 3, at least 4, at least 5, or all 6 signatures as defined herein.

Motifs 7 to 18 were derived using the MEME algorithm (Bailey and Elkan, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994). At each position within a MEME motif, the residues are shown that are present in the query set of sequences with a frequency higher than 0.2. Residues within square brackets represent alternatives.

In one embodiment, the GPx-related polypeptide as used herein comprises at least one of the motifs 7, 8 or 9.

Motif 7 (SEQ ID NO: 361): VN[V/A]AS[R/K/Q]CG

Motif 8 (SEQ ID NO: 362): L[A/G]FP[S/C]NQF

Motif 9 (SEQ ID NO: 363): WNF[S/T]KF

In a further embodiment, the GPx-related polypeptide as used herein comprises at least one of the motifs represented by Group A comprising motifs 10 to 12:

Group A

Motif 10 (SEQ ID NO: 364): EILAFPCNQFGGQEPG[S/T]NE[E/Q]I[V/Q][Q/E]

Motif 11 (SEQ ID NO: 365): CTRFKAE[Y/F]PIFDKVDVNG[D/N]N[A/T]AP[L/I]YKFLKSSKGG

Motif 12 (SEQ ID NO: 366): IKWNF[S/T]KFLVDK[E/D]G[N/H/R]VV[D/E]RYAPTTSPLS[I/M]

In a further embodiment, the GPx-related polypeptide as used herein comprises at least one of the motifs represented by Group B comprising motifs 13 to 15:

Group B

Motif 13 (SEQ ID NO: 367): GKVL[L/I]IVNVASQCGLTNSNYT[E/D][L/M][T/S][Q/E]LYE KYKDQG[L/F]EILAFPCNQFGGQEP

Motif 14 (SEQ ID NO: 368): CTRFKAE[Y/F]PIFDKVDVNGDN[A/T]AP[L/I]YKFLKSSKGG

Motif 15 (SEQ ID NO: 369): FGD[S/G/N]IKWNF[S/T]KFLVDK[E/D]G[N/K/R]VV[D/E]RYAPTTSP[L/A]S[I/M]EKD[I/V]KKLL

In a further embodiment, the GPX-related polypeptide as used herein comprises at least one of the motifs represented by Group C comprising motifs 16 to 18:

Group C

Motif 16 (SEQ ID NO: 370): FEILAFPCNQFGGQEPGTNEEIVQFACTRFKAEYPIFDKVDVNG

Motif 17 (SEQ ID NO: 371):

P[I/L]YKFLKSSKG[G/S]LFG[D/E][S/N]IKWNFSKFLVDKEG[H/R]VV[D/E]RYAPT
TSPLS[I/M]EKDI

Motif 18 (SEQ ID NO: 372): VHDFTVKDASGKDVDLS[T/V][Y/F]KGKVLLIVNVASQ

5

In still another embodiment, the GPx-related polypeptide comprises in increasing order of preference, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, or all 12 motifs as defined herein.

10 Additionally or alternatively, the GPx-related polypeptide or protein has in increasing order of preference at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%,
15 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% overall sequence identity to the polypeptide sequence represented by SEQ ID NO: 293, 295 or 297, provided that the homologous protein comprises any one or more of the conserved motifs as outlined above. The overall sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG
20 Wisconsin Package, Accelrys), preferably with default parameters and preferably with sequences of mature proteins (i.e. without taking into account secretion signals or transit peptides). In one embodiment the sequence identity level is determined by comparison of the polypeptide sequences over the entire length of the sequence of SEQ ID NO: 293, 295 or 297.

25

In another embodiment, the sequence identity level is determined by comparison of one or more conserved domains or motifs in SEQ ID NO: 293, 295 or 297 with corresponding conserved domains or motifs in other GPx-related polypeptides. Compared to overall sequence identity, the sequence identity will generally be higher when only conserved
30 domains or motifs are considered. Preferably the motifs in a GPx-related polypeptide have, in increasing order of preference, at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one or more of the motifs represented by SEQ ID NO: 361 to SEQ ID NO: 378 (Motifs 7 to 18, Signatures 1 to 6).

35

The terms “domain”, “signature” and “motif” are defined in the “definitions” section herein.

With respect to ELM2-related polypeptides, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 4, preferably
40 clusters with the group of ELM2-related polypeptides (in bold) comprising the amino acid sequence represented by SEQ ID NO: 2, indicated as Poptr_ELM2-related, rather than with any other group.

Furthermore, ELM2-related polypeptides, at least in their native form, typically have transcriptional repression activity, involved in inhibiting chromatin remodeling. Tools and techniques for measuring such activity are well known in the art and e.g. described in Ding et al., Mol. Cell Biol. 2003 Jan; 23(1):250-8.

In addition, ELM2-related polypeptides, when expressed in rice according to the methods of the present invention as outlined in Examples 7 and 8, give plants having increased yield related traits, in particular increased biomass and increased seed yield with amongst others increased total weight of seeds, increased number of (filled) seeds, increased fill rate, increased harvest index, increased number of florets.

In one embodiment of the present invention the function of the nucleic acid sequences of the invention is to confer information for synthesis of the ELM2-related polypeptide that increases yield or yield related traits, when such a nucleic acid sequence of the invention is transcribed and translated in a living plant cell.

With respect to WRKY-related polypeptides, in addition, WRKY-related polypeptides, when expressed in rice according to the methods of the present invention as outlined in Examples 7 and 8, give plants having increased yield related traits, in particular one or more of emergence vigour (early vigour), total seed yield (Totalwgseeds), fillrate, TKW, number of filled seeds, taller more erect plants (HeightMax), amount of thick roots (RootThickMax). Moreover, one or more lines showed in one or more parameters an increase of number of florets per panicle on a plant (flowerperpan), increased harvestindex (HI), increased greenness of a plant before flowering (GNbfFlow), increased height of the plant (GravityYMax), increased quick early development (AreaEmer), increased Cycle Time (AreaCycl).

In one embodiment of the present invention the function of the nucleic acid sequences of the invention is to confer information for synthesis of the WRKY-related polypeptide that increases yield or yield related traits, when such a nucleic acid sequence of the invention is transcribed and translated in a living plant cell.

With respect to EMG1-like polypeptides, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 13, preferably clusters with the group of EMG1-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 179, indicated as *P. trichocarpa* EMG1-like, rather than with any other group.

Furthermore, EMG1-like polypeptides, at least in their native form, typically have RNA binding activity as e.g. described by Leulliot et al., Nucleic acid Res. 2008, 36(2), 629-39.

In addition, EMG1-like polypeptides, when expressed in rice according to the methods of the present invention as outlined in Examples 7 and 8, give plants having increased yield related traits, in particular seed yield, such as increased total seed weight, increased harvest index and increased fill rate.

5

In one embodiment of the present invention the function of the nucleic acid sequences of the invention is to confer information for synthesis of the EMG1-like polypeptide that increases yield or yield related traits, when such a nucleic acid sequence of the invention is transcribed and translated in a living plant cell.

10

With respect to GPx-related polypeptides, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 21, preferably clusters with the group of GPx-related polypeptides comprising the amino acid sequence represented by SEQ ID NO: 293, 295 or 297 rather than with any other group.

15

Furthermore, GPx-related polypeptides (at least in their native form) typically have enzymatic activity, particularly peroxidase activity. Tools and techniques for measuring enzymatic activity, particularly peroxidase activity are known in the art. In addition, nucleic acids encoding GPx-related polypeptides, when expressed in rice according to the methods of the present invention as outlined in Examples 7 and 8, give plants having increased yield related traits, in particular seed yield, seed size, number total seeds, fillrate, number of filled seeds as well as flowering time. Another function of the nucleic acid sequences encoding GPx-related polypeptides is to confer information for synthesis of the GPx-related protein that increases yield or yield related traits as described herein, when such a nucleic acid sequence of the invention is transcribed and translated in a living plant cell.

20

25

With respect to ELM2-related polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 1, encoding the polypeptide sequence of SEQ ID NO: 2, and by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 3, encoding the polypeptide sequence of SEQ ID NO: 4. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any ELM2-related-encoding nucleic acid or ELM2-related polypeptide as defined herein.

30

35

Further examples of nucleic acids encoding ELM2-related polypeptides are given in Table A1 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The polypeptide sequences given in Table A1 of the Examples section are example sequences of orthologues and paralogues of the ELM2-related polypeptide represented by SEQ ID NO: 2, the terms "orthologues" and "paralogues" being as defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search as described in the definitions section; where the query sequence is SEQ ID NO: 1 or SEQ ID NO: 2, the second BLAST (back-BLAST) would be

40

against *Populus trichocarpa* sequences or where the query sequence is SEQ ID NO: 3 or SEQ ID NO: 4, the second BLAST (back-BLAST) would be against *Medicago truncatula*.

In a further preferred embodiment, a method is provided wherein said ELM2-related polypeptide comprises an Interpro accession IPR 000949, according to PFAM accession number PF01448 ELM2 domain.

In another embodiment a method is provided wherein said ELM2-related polypeptide comprises an ELM2 domain with at least 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the ELM2 domain starting with amino acid 225 up to amino acid 262 in SEQ ID NO: 2.

In another preferred embodiment, a method is provided wherein said nucleic acid encodes the polypeptide represented by any one of SEQ ID NO: 2 or SEQ ID NO: 4 or a homologue thereof which has at least 90% overall sequence identity to SEQ ID NO : 2 or SEQ ID NO: 4.

According to a further embodiment of the present invention, there is therefore provided an isolated nucleic acid molecule selected from:

- (i) a nucleic acid encoding an ELM2-related polypeptide having in increasing order of preference at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the polypeptide sequence represented by SEQ ID NO: 2, provided the ELM2-related polypeptide comprises any one or more of the motifs as represented in SEQ ID NO: 93 to SEQ ID NO: 95 (Motifs 1 to 3), and additionally or alternatively comprising one or more motifs having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one or more of the motifs given in SEQ ID NO: 93 to SEQ ID NO: 95, and further preferably conferring enhanced yield-related traits relative to control plants.
- (ii) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iii) under high stringency hybridization conditions and preferably confers enhanced yield-related traits relative to control plants.

According to a further embodiment of the present invention, there is also provided an isolated polypeptide selected from:

- (i) an amino acid sequence having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the polypeptide sequence represented by SEQ ID NO: 2, provided the ELM2-related polypeptide comprises any one or more of the motifs as represented in SEQ ID NO: 93 to SEQ ID NO: 95 (Motifs 1 to 3), and additionally or alternatively comprising one or more motifs having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one or more of the motifs given in SEQ ID NO: 93 to SEQ ID NO: 95, and further preferably conferring enhanced yield-related traits relative to control plants;
- (ii) derivatives of any of the amino acid sequences given in (i) above.

With respect to WRKY-related polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 102, encoding the polypeptide sequence of SEQ ID NO: 103. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any WRKY-related polypeptide-encoding nucleic acid or WRKY-related polypeptide as defined herein.

Examples of nucleic acids encoding WRKY-related polypeptides are given in Table A2 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The polypeptide sequences given in Table A2 of the Examples section are example sequences of orthologues and paralogues of the WRKY-related polypeptide represented by SEQ ID NO: 103, the terms "orthologues" and "paralogues" being as defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search as described in the definitions section.

With respect to EMG1-like polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 178, encoding the polypeptide sequence of SEQ ID NO: 179. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any EMG1-like-encoding nucleic acid or EMG1-like polypeptide as defined herein.

Examples of nucleic acids encoding EMG1-like polypeptides are given in Table A3 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The polypeptide acid sequences given in Table A3 of the Examples section are example sequences of orthologues and paralogues of the EMG1-like polypeptide represented by SEQ ID NO: 179, the terms "orthologues" and "paralogues" being as

defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search as described in the definitions section; where the query sequence is SEQ ID NO: 178 or SEQ ID NO: 179, the second BLAST (back-BLAST) would be against *Populus trichocarpa* sequences.

5

With respect to GPx-related polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NOs 292, 294 and 296 encoding the polypeptide sequence of SEQ ID NO: 293, 295 and 297, respectively. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any GPx-related -encoding nucleic acid or GPx-related polypeptide as defined herein. The term "GPx-related" or "GPx-related polypeptide" as used herein also intends to include homologues as defined hereunder of SEQ ID NO: 293, 295 or 297.

10

15

Examples of nucleic acids encoding GPx-related polypeptides are given in Table A4 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The polypeptide sequences given in Table A4 of the Examples section are example sequences of orthologues and paralogues of the GPx-related polypeptide represented by SEQ ID NO: 293, 295 or 297, the terms "orthologues" and "paralogues" being as defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search as described in the definitions section.

20

25

The invention also provides hitherto unknown GPx-related-encoding nucleic acids and GPx-related polypeptides useful for conferring enhanced yield-related traits in plants relative to or compared to control plants.

30

35

Nucleic acid variants may also be useful in practising the methods of the invention. Examples of such variants include nucleic acids encoding homologues and derivatives of any one of the amino acid sequences given in Table A1 to A4 of the Examples section, the terms "homologue" and "derivative" being as defined herein. Also useful in the methods, constructs, plants, harvestable parts and products of the invention are nucleic acids encoding homologues and derivatives of orthologues or paralogues of any one of the amino acid sequences given in Table A1 to A4 of the Examples section. Homologues and derivatives useful in the methods of the present invention have substantially the same biological and functional activity as the unmodified protein from which they are derived. Further variants useful in practising the methods of the invention are variants in which codon usage is optimised or in which miRNA target sites are removed.

40

Further nucleic acid variants useful in practising the methods of the invention include portions of nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, nucleic acids hybridising to nucleic acids encoding ELM2-related polypeptides, or WRKY-related

polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, splice variants of nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, allelic variants of nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides and variants of nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, obtained by gene shuffling. The terms hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

Nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, need not be full-length nucleic acids, since performance of the methods of the invention does not rely on the use of full-length nucleic acid sequences. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a portion of any one of the nucleic acid sequences given in Table A1 to A4 of the Examples section, or a portion of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 to A4 of the Examples section.

A portion of a nucleic acid may be prepared, for example, by making one or more deletions to the nucleic acid. The portions may be used in isolated form or they may be fused to other coding (or non-coding) sequences in order to, for example, produce a protein that combines several activities. When fused to other coding sequences, the resultant polypeptide produced upon translation may be bigger than that predicted for the protein portion.

With respect to the ELM2-related polypeptides, portions useful in the methods, constructs, plants, harvestable parts and products of the invention, encode an ELM2-related polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A1 of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A1 of the Examples section, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A1 of the Examples section. Preferably the portion is at least 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A1 of the Examples section, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A1 of the Examples section. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 1 or a portion of the nucleic acid of SEQ ID NO: 3. Preferably, the portion encodes a fragment of an amino acid sequence which, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 4, clusters with the group of ELM2-related polypeptides as indicated in bold,

comprising the amino acid sequences represented by SEQ ID NO: 2 and 4, rather than with any other group, and comprises any one or more of the motifs as represented by SEQ ID NO: 93 to SEQ ID NO: 95, and/or has transcriptional repression biological activity, and/or has at least 9% sequence identity to SEQ ID NO: 2.

5

With respect to WRKY-related polypeptides, portions useful in the methods, constructs, plants, harvestable parts and products of the invention, encode a WRKY-related polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A2 of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A2 of the Examples section, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A2 of the Examples section. Preferably the portion is at least 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A2 of the Examples section, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A2 of the Examples section. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 102.

20

With respect to EMG1-like polypeptides, portions useful in the methods, constructs, plants, harvestable parts and products of the invention, encode a EMG1-like polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A3 of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A3 of the Examples section, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A3 of the Examples section. Preferably the portion is at least 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A3 of the Examples section, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A3 of the Examples section. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 178. Preferably, the portion encodes a fragment of an amino acid sequence which, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 13, clusters with the group of EMG1-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 179, indicated as *P. trichocarpa* EMG1-like, rather than with any other group, and/or comprises any of the motifs 4 to 6, and/or has RNA binding activity, and/or has at least 80% sequence identity to SEQ ID NO: 179.

40

With respect to GPx-related polypeptides, portions useful in the methods, constructs, plants, harvestable parts and products of the invention, encode a GPx-related polypeptide

as defined herein or at least part thereof, and have substantially the same biological activity as the amino acid sequences given in Table A4 of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A4 of the Examples section, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A4 of the Examples section. Preferably the portion is at least 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A4 of the Examples section, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A4 of the Examples section. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 292, 294 or 296. Preferably, the portion encodes a fragment of an amino acid sequence which comprises one or more motifs selected from a group consisting of motifs 7 to 18 as defined herein and/or one or more signatures 1 to 6 as defined herein, and/or has biological activity particularly enzymatic activity, more particularly peroxidase activity, and/or has at least 50% sequence identity to SEQ ID NO: 293, 295 or 297.

Another nucleic acid variant useful in the methods, constructs, plants, harvestable parts and products of the invention is a nucleic acid capable of hybridising, under reduced stringency conditions, preferably under stringent conditions, with a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined herein, or with a portion as defined herein. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a nucleic acid capable of hybridizing to the complement of a nucleic acid encoding any one of the proteins given in Table A1 to A4 of the Examples section, or to the complement of a nucleic acid encoding an orthologue, paralogue or homologue of any one of the proteins given in Table A1 to A4 of the Examples section.

Hybridising sequences useful in the methods, constructs, plants, harvestable parts and products of the invention encode an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined herein, having substantially the same biological activity as the amino acid sequences given in Table A1 to A4 of the Examples section. Preferably, the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding any one of the proteins given in Table A1 to A4 of the Examples section, or to a portion of any of these sequences, a portion being as defined herein, or the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A1 to A4 of the Examples section.

With respect to ELM2-related polypeptides, the hybridising sequence is most preferably capable of hybridising to the complement of a nucleic acid as represented by SEQ ID NO: 1 or to a portion thereof. In one embodiment, the hybridization conditions are of medium

stringency, preferably of high stringency, as defined above.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence which, when full-length and used in the construction of a phylogenetic tree, such as the one depicted in Figure 4, clusters with the group of ELM2-related polypeptides as indicated in bold, comprising the amino acid sequences represented by SEQ ID NO: 2 and 4, rather than with any other group, and comprises any one or more of the motifs as represented by SEQ ID NO: 93 to SEQ ID NO: 95, and/or has transcriptional repression biological activity, and/or has at least 9% sequence identity to SEQ ID NO: 2.

With respect to WRKY-related polypeptides, the hybridising sequence most preferably is capable of hybridising to the complement of a nucleic acid as represented by SEQ ID NO: 102 or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined above.

With respect to EMG1-like polypeptides, the hybridising sequence most preferably is capable of hybridising to the complement of a nucleic acid as represented by SEQ ID NO: 178 or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined above.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence which, when full-length and used in the construction of a phylogenetic tree, such as the one depicted in Figure 13, clusters with the group of EMG1-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 179, indicated as *P. trichocarpa* EMG1-like, rather than with any other group, and/or comprises any of the motifs 4 to 6, and/or has RNA binding activity, and/or has at least 80% sequence identity to SEQ ID NO: 179.

With respect to GPx-related polypeptides, the hybridising sequence is most preferably capable of hybridising to the complement of a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 293, 295 or 297 or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined herein.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence which comprises one or more motifs selected from a group consisting of motifs 7 to 18 as defined herein and/or one or more signatures 1 to 6 as defined herein, and/or has biological activity particularly enzymatic activity, more particularly peroxidase activity, and/or has at least 50% sequence identity to SEQ ID NO: 293, 295 or 297.

Another nucleic acid variant useful in the methods of the invention is a splice variant encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined hereinabove, a splice variant being as

defined herein.

In another embodiment, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a splice variant of a nucleic acid encoding any one of the proteins given in Table A1 to A4 of the Examples section, or a splice variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 to A4 of the Examples section.

With respect to ELM2-related polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 1, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2. Preferably, the amino acid sequence encoded by the splice variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 4, clusters with the group of ELM2-related polypeptides as indicated in bold, comprising the amino acid sequences represented by SEQ ID NO: 2 and 4, rather than with any other group, and comprises any one or more of the motifs as represented by SEQ ID NO: 93 to SEQ ID NO: 95, and/or has transcriptional repression biological activity, and/or has at least 9% sequence identity to SEQ ID NO: 2.

With respect to WRKY-related polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 102, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 103.

With respect to EMG1-like polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 178, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 179. Preferably, the amino acid sequence encoded by the splice variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 13, clusters with the group of EMG1-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 179 rather than with any other group, and/or comprises any of the motifs 4 to 6, and/or has RNA binding activity, and/or has at least 80% sequence identity to SEQ ID NO: 179.

With respect to GPx-related polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 292, 294 or 296, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 293. Preferably, the amino acid sequence encoded by the splice variant comprises one or more motifs selected from a group consisting of motifs 7 to 18 as defined herein and/or one or more signatures 1 to 6 as defined herein, and/or has biological activity particularly enzymatic activity, more particularly peroxidase activity, and/or has at least 50% sequence identity to SEQ ID NO: 293, 295 or 297.

Another nucleic acid variant useful in performing the methods of the invention is an allelic variant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related

polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined hereinabove, an allelic variant being as defined herein.

In yet another embodiment, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant an allelic variant of a nucleic acid encoding any one of the proteins given in Table A1 to A4 of the Examples section, or comprising introducing and expressing in a plant an allelic variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 to A4 of the Examples section.

With respect to ELM2-related polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the ELM2-related polypeptide of SEQ ID NO: 2 and any of the amino acids depicted in Table A1 of the Examples section. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 1 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2. Preferably, the amino acid sequence encoded by the allelic variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 4, clusters with the group of ELM2-related polypeptides as indicated in bold, comprising the amino acid sequences represented by SEQ ID NO: 2 and 4, rather than with any other group, and comprises any one or more of the motifs as represented by SEQ ID NO: 93 to SEQ ID NO: 95, and/or has transcriptional repression biological activity, and/or has at least 9% sequence identity to SEQ ID NO: 2.

With respect to WRKY-related polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the WRKY-related polypeptide of SEQ ID NO: 103 and any of the amino acids depicted in Table A2 of the Examples section. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 102 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 103.

With respect to EMG1-like polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the EMG1-like polypeptide of SEQ ID NO: 179 and any of the amino acids depicted in Table A3 of the Examples section. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 178 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 179. Preferably, the amino acid sequence encoded by the allelic variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 13, clusters with the group of EMG1-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 179 rather than with any other group,

and/or comprises any of the motifs 4 to 6, and/or has RNA binding activity, and/or has at least 80% sequence identity to SEQ ID NO: 179.

5 With respect to GPx-related polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the GPx-related polypeptide of SEQ ID NO: 293, 295, or 297 and any of the amino acid sequences depicted in Table A4 of the Examples section. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 292, 294 or 296 or
10 an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 293, 295, or 297. Preferably, the amino acid sequence encoded by the allelic variant comprises one or more motifs selected from a group consisting of motifs 7 to 18 as defined herein and/or one or more signatures 1 to 6 as defined herein, and/or has biological activity particularly enzymatic activity, more particularly peroxidase activity, and/or has at least 50%
15 sequence identity to SEQ ID NO: 293, 295, or 297.

Gene shuffling or directed evolution may also be used to generate variants of nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, as defined above; the term "gene shuffling"
20 being as defined herein.

In yet another embodiment, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a variant of a nucleic acid encoding any one of the proteins given in Table A1 to A4 of the Examples section, or comprising
25 introducing and expressing in a plant a variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 to A4 of the Examples section, which variant nucleic acid is obtained by gene shuffling.

With respect to ELM2-related polypeptides, the amino acid sequence encoded by the
30 variant nucleic acid obtained by gene shuffling, when used in the construction of a phylogenetic tree such as the one depicted in Figure 4, preferably clusters with the group of ELM2-related polypeptides as indicated in bold, comprising the amino acid sequences represented by SEQ ID NO: 2 and 4, rather than with any other group, and comprises any one or more of the motifs as represented by SEQ ID NO: 93 to SEQ ID NO: 95, and/or has
35 transcriptional repression biological activity, and/or has at least 9% sequence identity to SEQ ID NO: 2.

With respect to EMG1-like polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling, when used in the construction of a phylogenetic
40 tree, such as the one depicted in Figure 13, preferably clusters with the group of EMG1-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 179 rather than with any other group, and/or comprises any of the motifs 4 to 6, and/or has RNA

binding activity, and/or has at least 80% sequence identity to SEQ ID NO: 179.

With respect to GPx-related polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling preferably comprises one or more motifs selected from a group consisting of motifs 7 to 18 as defined herein and/or one or more signatures 1 to 6 as defined herein, and/or has biological activity particularly enzymatic activity, more particularly peroxidase activity, and/or has at least 50% sequence identity to SEQ ID NO: 293, 295 or 297.

Furthermore, nucleic acid variants may also be obtained by site-directed mutagenesis. Several methods are available to achieve site-directed mutagenesis, the most common being PCR based methods (Current Protocols in Molecular Biology. Wiley Eds.).

ELM2-related polypeptides differing from the sequence of SEQ ID NO: 2 by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined above) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention.

Nucleic acids encoding ELM2-related polypeptides may be derived from any natural or artificial source. The nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. Preferably the ELM2-related polypeptide-encoding nucleic acid is from a plant, further preferably from a dicotyledonous plant, more preferably from the family Salicaceae, most preferably the nucleic acid is from *Populus trichocarpa*. In an alternative more preferred embodiment, the ELM2-related polypeptide-encoding nucleic acid is from the family Fabaceae, most preferably the nucleic acid is from *Medicago truncatula*. Preferably, the nucleic acid from *Medicago truncatula* is as represented by SEQ ID NO: 1.

WRKY-related polypeptides differing from the sequence of SEQ ID NO: 103 by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined above) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention.

Nucleic acids encoding WRKY-related polypeptides may be derived from any natural or artificial source. The nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. Preferably the WRKY-related polypeptide-encoding nucleic acid is from a plant, preferably from a dicotyledonous plant, further preferably from a leguminous plant, more preferably from the genus *Medicago*, most preferably from *Medicago truncatula*. Preferably, the nucleic acid from *Medicago truncatula* is as represented by SEQ ID NO: 102.

EMG1-like polypeptides differing from the sequence of SEQ ID NO: 179 by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined above) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention.

Nucleic acids encoding EMG1-like polypeptides may be derived from any natural or artificial source. The nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. Preferably the EMG1-like polypeptide-encoding nucleic acid is from a plant, further preferably from a dicotyledonous plant, more preferably from the family Salicaceae, most preferably the nucleic acid is from *Populus trichocarpa*. Preferably, the nucleic acid from *Populus trichocarpa* is as represented by SEQ ID NO: 178.

GPx-related polypeptides differing from the sequence of SEQ ID NO: 293, 295 or 297 by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined herein) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention.

Nucleic acids encoding GPx-related polypeptides may be derived from any natural or artificial source. The nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. Preferably the GPx-related polypeptide-encoding nucleic acid is from a plant, further preferably from a monocotyledonous plant, more preferably from the family Poaceae, most preferably the nucleic acid is from *Oryza sativa*. Preferably, the nucleic acid from *Oryza sativa* is as represented by SEQ ID NO: 292.

In another embodiment, the GPx-related polypeptide-encoding nucleic acid is from a plant, further preferably from a dicotyledonous plant, more preferably from the family Salicaceae, more preferably from the genus *Populus*, most preferably the nucleic acid is from *Populus trichocarpa*. Preferably, the nucleic acid from *Populus trichocarpa* is as represented by SEQ ID NO: 294 or 296.

In another embodiment the present invention extends to recombinant chromosomal DNA comprising a nucleic acid sequence useful in the methods of the invention, wherein said nucleic acid is present in the chromosomal DNA as a result of recombinant methods, i.e. said nucleic acid is not in the chromosomal DNA in its natural genetic environment. In a further embodiment the recombinant chromosomal DNA of the invention is comprised in a plant cell. DNA comprised within a cell, particularly a cell with cell walls like a plant cell, is better protected from degradation than a bare nucleic acid sequence. The same holds true for a DNA construct comprised in a host cell, for example a plant cell.

Performance of the methods of the invention gives plants having enhanced yield-related traits. In particular performance of the methods of the invention gives plants having increased early vigour and/or increased yield, especially increased biomass and/or increased seed yield relative to or compared to control plants. The terms “early vigour”
5 “yield” and “seed yield” are described in more detail in the “definitions” section herein.

Reference herein to enhanced yield-related traits is taken to mean an increase in early vigour and/or in biomass (weight) of one or more parts of a plant, which may include (i) aboveground parts and preferably aboveground harvestable parts and/or (ii) parts below
10 ground and preferably harvestable below ground. In particular, such harvestable parts are seeds, and performance of the methods of the invention results in plants having increased seed yield relative to or compared to the seed yield of control plants. In another particular embodiment, such harvestable parts are aboveground biomass, and performance of the methods of the invention results in plants having increased biomass relative to or compared
15 to the biomass of control plants.

The present invention provides a method for increasing yield-related traits, especially biomass and/or seed yield of plants, relative or compared to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding an ELM2-related
20 polypeptide as defined herein. The terms “relative to” and “compared to” can be used interchangeably.

The present invention also provides a method for increasing yield-related traits, especially seed yield of plants, relative to or compared to control plants, which method comprises
25 modulating expression in a plant of a nucleic acid encoding a WRKY-related polypeptide as defined herein. The terms “relative to” and “compared to” can be used interchangeably.

The present invention also provides a method for increasing yield-related traits, especially
30 seed yield of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding an EMG1-like polypeptide as defined herein. The terms “relative to” and “compared to” can be used interchangeably.

The present invention also provides a method for increasing yield-related traits, especially
35 biomass and/or seed yield of plants, relative to or compared to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding a GPx-related polypeptide as defined herein.

According to a preferred feature of the present invention, performance of the methods of the
40 invention gives plants having an increased growth rate relative to control plants. Therefore, according to the present invention, there is provided a method for increasing the growth rate of plants, which method comprises modulating expression in a plant of a nucleic acid

encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined herein.

Performance of the methods of the invention gives plants grown under non-stress conditions or under mild drought conditions increased yield-related traits relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield-related traits and/or yield in plants grown under non-stress conditions or under mild drought conditions, which method comprises modulating expression in a plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide.

With respect to WRKY-related polypeptides, in one embodiment, performance of the methods of the invention gives plants grown under abiotic conditions abiotic stress resistance.

Performance of the methods of the invention gives plants grown under conditions of drought, increased yield-related traits relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield-related traits and/or yield in plants grown under conditions of drought which method comprises modulating expression in a plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide.

Performance of the methods of the invention gives plants grown under conditions of nutrient deficiency, particularly under conditions of nitrogen deficiency, increased yield-related traits relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield-related traits and/or yield in plants grown under conditions of nutrient deficiency, which method comprises modulating expression in a plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide.

Performance of the methods of the invention gives plants grown under conditions of salt stress, increased yield-related traits relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield-related traits and/or yield in plants grown under conditions of salt stress, which method comprises modulating expression in a plant of a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide.

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides. The gene constructs

may be inserted into vectors, which may be commercially available, suitable for transforming into plants or host cells and suitable for expression of the gene of interest in the transformed cells. The invention also provides use of a gene construct as defined herein in the methods of the invention.

5

More specifically, the present invention provides a construct comprising:

- (a) a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined above;
- 10 (b) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
- (c) a transcription termination sequence.

Preferably, the nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, is as defined above. The term "control sequence" and "termination sequence" are as defined herein.

The genetic construct of the invention may be comprised in a host cell, plant cell, seed, agricultural product or plant. Plants or host cells are transformed with a genetic construct such as a vector or an expression cassette comprising any of the nucleic acids described above. Thus the invention furthermore provides plants or host cells transformed with a construct as described above. In particular, the invention provides plants transformed with a construct as described above, which plants have increased yield-related traits as described herein.

25

In one embodiment the genetic construct of the invention confers increased yield or yield related traits to a plant when it has been introduced into said plant, which plant expresses the nucleic acid encoding the ELM2-related polypeptide, or the WRKY-related polypeptide, or the EMG1-like polypeptide, or the GPx-related polypeptide, comprised in the genetic construct. In another embodiment the genetic construct of the invention confers increased yield or yield related traits(s) to a plant comprising plant cells in which the construct has been introduced, which plant cells express the nucleic acid encoding the ELM2-related polypeptide, or the WRKY-related polypeptide, or the EMG1-like polypeptide, or the GPx-related polypeptide, comprised in the genetic construct.

35

The promoter in such a genetic construct may be a non-native promoter to the nucleic acid described above, i.e. a promoter not regulating the expression of said nucleic acid in its native surrounding. The expression cassettes or the genetic construct of the invention may be comprised in a host cell, plant cell, seed, agricultural product or plant.

40

The skilled artisan is well aware of the genetic elements that must be present on the genetic construct in order to successfully transform, select and propagate host cells containing the

sequence of interest. The sequence of interest is operably linked to one or more control sequences (at least to a promoter).

Advantageously, any type of promoter, whether natural or synthetic, may be used to drive expression of the nucleic acid sequence, but preferably the promoter is of plant origin. A constitutive promoter is particularly useful in the methods. See the "Definitions" section herein for definitions of the various promoter types. With respect to WRKY-related polypeptides, and GPx-related polypeptides, also useful in the methods of the invention is a root-specific promoter.

The constitutive promoter is preferably a ubiquitous constitutive promoter of medium strength. More preferably, it is a plant derived promoter, e.g. a promoter of plant chromosomal origin, such as a GOS2 promoter or a promoter of substantially the same strength and having substantially the same expression pattern (a functionally equivalent promoter), more preferably the promoter is the GOS2 promoter from rice. Further preferably, the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 97, or to SEQ ID NO: 174, or to SEQ ID NO: 289, or to SEQ ID NO: 379, most preferably the constitutive promoter is as represented by SEQ ID NO: 97, or SEQ ID NO: 174, or SEQ ID NO: 289, or SEQ ID NO: 379. See the "Definitions" section herein for further examples of constitutive promoters.

With respect to WRKY-related polypeptides, and GPx-related polypeptides, according to another preferred embodiment of the invention, the nucleic acid encoding a WRKY-related polypeptide, or a GPx-related polypeptide, is operably linked to a root-specific promoter.

The root-specific promoter is preferably an RCc3 promoter (Plant Mol Biol. 1995 Jan; 27(2):237-48) or a promoter of substantially the same strength and having substantially the same expression pattern (a functionally equivalent promoter), more preferably the RCc3 promoter is from rice, further preferably the RCc3 promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 175, or to SEQ ID NO: 380, most preferably the promoter is as represented by SEQ ID NO: 175, or SEQ ID NO: 380. Examples of other root-specific promoters which may also be used to perform the methods of the invention are shown in Table 2b in the "Definitions" section.

With respect to ELM2-related polypeptides, it should be clear that the applicability of the present invention is not restricted to the ELM2-related polypeptide-encoding nucleic acid represented by SEQ ID NO: 1 or SEQ ID NO: 3, nor is the applicability of the invention restricted to expression of an ELM2-related polypeptide-encoding nucleic acid when driven by a constitutive promoter.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Preferably, the construct comprises an expression cassette comprising a GOS2 promoter, substantially similar to SEQ ID NO: 97, operably linked to the nucleic acid

encoding the ELM2-related polypeptide. More preferably, the construct comprises a zein terminator (t-zein) linked to the 3' end of the ELM2-related coding sequence. Most preferably, the expression cassette comprises a sequence having in increasing order of preference at least 95%, at least 96%, at least 97%, at least 98%, at least 99% identity to the sequence represented by SEQ ID NO: 96 (pPRO::ELM2-related::t-zein sequence). Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

With respect to WRKY-related polypeptides, it should be clear that the applicability of the present invention is not restricted to the WRKY-related polypeptide-encoding nucleic acid represented by SEQ ID NO: 102, nor is the applicability of the invention restricted to expression of a WRKY-related polypeptide-encoding nucleic acid when driven by a constitutive promoter, or when driven by a root-specific promoter.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Preferably, the construct comprises an expression cassette comprising a GOS2 promoter, substantially similar to SEQ ID NO: 174, operably linked to the nucleic acid encoding WRKY-related polypeptide. More preferably, the construct comprises a zein terminator (t-zein) linked to the 3' end of the WRKY-related polypeptide coding sequence. Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

With respect to EMG1-like polypeptides, it should be clear that the applicability of the present invention is not restricted to the EMG1-like polypeptide-encoding nucleic acid represented by SEQ ID NO: 178, nor is the applicability of the invention restricted to expression of a EMG1-like polypeptide-encoding nucleic acid when driven by a constitutive promoter.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Preferably, the construct comprises an expression cassette comprising a GOS2 promoter, substantially similar to SEQ ID NO: 289, operably linked to the nucleic acid encoding the EMG1-like polypeptide. More preferably, the construct comprises a zein terminator (t-zein) linked to the 3' end of the EMG1-like coding sequence. Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

With respect to GPx-related polypeptides, it should be clear that the applicability of the present invention is not restricted to the GPx-related polypeptide-encoding nucleic acid represented by SEQ ID NO: 292, 294 or 296, nor is the applicability of the invention restricted to the rice GOS2 promoter when expression of a GPx-related polypeptide-encoding nucleic acid is driven by a constitutive promoter, or when driven by a root-specific promoter.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Those skilled in the art will be aware of terminator sequences that may be suitable for use in performing the invention. Preferably, the construct comprises an expression cassette comprising a GOS2 promoter, substantially similar to SEQ ID NO: 379, operably
5 linked to the nucleic acid encoding the GPx-related polypeptide. More preferably, the construct furthermore comprises a zein terminator (t-zein) linked to the 3' end of the GPx-related coding sequence.

10 According to a preferred feature of the invention, the modulated expression is increased expression. Methods for increasing expression of nucleic acids or genes, or gene products, are well documented in the art and examples are provided in the definitions section.

As mentioned above, a preferred method for modulating expression of a nucleic acid
15 encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, is by introducing and expressing in a plant a nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, respectively; however the effects of performing the method, i.e. enhancing yield-related traits may also be achieved using other
20 well known techniques, including but not limited to T-DNA activation tagging, TILLING, homologous recombination. A description of these techniques is provided in the definitions section.

The invention also provides a method for the production of transgenic plants having
25 enhanced yield-related traits relative to control plants, comprising introduction and expression in a plant of any nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined herein.

30 With respect to ELM2-related polypeptides, the present invention more specifically provides a method for the production of transgenic plants having enhanced yield-related traits, particularly increased yield, more in particular increased biomass and increased seed yield, which method comprises:

- 35 (i) introducing and expressing in a plant or plant cell an ELM2-related polypeptide-encoding nucleic acid or a genetic construct comprising an ELM2-related polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development.

The nucleic acid of (i) may be any of the nucleic acids capable of encoding an ELM2-related
40 polypeptide as defined herein. Preferably the nucleic acid encoding the ELM2-related polypeptide and to be introduced into the plant is an isolated nucleic acid or is comprised in a genetic construct as described above.

With respect to WRKY-related polypeptides, the present invention more specifically provides a method for the production of transgenic plants having enhanced yield-related traits, particularly increased yield, which method comprises:

- 5 (i) introducing and expressing in a plant or plant cell a WRKY-related polypeptide-encoding nucleic acid or a genetic construct comprising a WRKY-related polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development.

10 The nucleic acid of (i) may be any of the nucleic acids capable of encoding a WRKY-related polypeptide as defined herein. Preferably the nucleic acid encoding the WRKY-related polypeptide and to be introduced into the plant is an isolated nucleic acid or is comprised in a genetic construct as described above.

15 With respect to EMG1-like polypeptides, the present invention more specifically provides a method for the production of transgenic plants having enhanced yield-related traits, particularly increased seed yield, which method comprises:

- (i) introducing and expressing in a plant or plant cell a EMG1-like polypeptide-encoding nucleic acid or a genetic construct comprising a EMG1-like
- 20 polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development.

 The nucleic acid of (i) may be any of the nucleic acids capable of encoding an EMG1-like polypeptide as defined herein. Preferably the nucleic acid encoding the EMG1-like

25 polypeptide and to be introduced into the plant is an isolated nucleic acid or is comprised in a genetic construct as described above.

With respect to GPx-related polypeptides, the present invention more specifically provides a method for the production of transgenic plants having enhanced yield-related traits,

30 particularly increased (seed) yield, which method comprises:

- (i) introducing and expressing in a plant or plant cell a GPx-related polypeptide-encoding nucleic acid or a genetic construct comprising a GPx-related polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and
- 35 development.

 The nucleic acid of (i) may be any of the nucleic acids capable of encoding a GPx-related polypeptide as defined herein. Preferably the nucleic acid encoding the GPx-related polypeptide and to be introduced into the plant is an isolated nucleic acid or is comprised in a genetic construct as described above.

40 Cultivating the plant cell under conditions promoting plant growth and development, may or may not include regeneration and/or growth to maturity. Accordingly, in a particular

embodiment of the invention, the plant cell transformed by the method according to the invention is regenerable into a transformed plant. In another particular embodiment, the plant cell transformed by the method according to the invention is not regenerable into a transformed plant, i.e. cells that are not capable to regenerate into a plant using cell culture techniques known in the art. While plants cells generally have the characteristic of totipotency, some plant cells can not be used to regenerate or propagate intact plants from said cells. In one embodiment of the invention the plant cells of the invention are such cells. In another embodiment the plant cells of the invention are plant cells that do not sustain themselves in an autotrophic way, such plant cells are not deemed to represent a plant variety. One example are plant cells that do not sustain themselves through photosynthesis by synthesizing carbohydrate and protein from such inorganic substances as water, carbon dioxide and mineral salt. In a further embodiment the plant cells of the invention are non-plant variety and non-propagative.

The nucleic acid may be introduced directly into a plant cell or into the plant itself (including introduction into a tissue, organ or any other part of a plant). According to a preferred feature of the present invention, the nucleic acid is preferably introduced into a plant or plant cell by transformation. The term "transformation" is described in more detail in the "definitions" section herein.

In one embodiment the present invention extends to any plant cell or plant produced by any of the methods described herein, and to all plant parts and propagules thereof.

The present invention encompasses plants or parts thereof (including seeds) obtainable by the methods according to the present invention. The plants or plant parts or plant cells comprise a nucleic acid transgene encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide as defined above, preferably in a genetic construct such as an expression cassette. The present invention extends further to encompass the progeny of a primary transformed or transfected cell, tissue, organ or whole plant that has been produced by any of the aforementioned methods, the only requirement being that progeny exhibit the same genotypic and/or phenotypic characteristic(s) as those produced by the parent in the methods according to the invention.

In a further embodiment the invention extends to seeds recombinantly comprising the expression cassettes of the invention, the genetic constructs of the invention, or the nucleic acids encoding the ELM2-related polypeptide, or WRKY-related polypeptide, or EMG1-like polypeptide, or GPx-related polypeptide, and/or the ELM2-related polypeptides, or the WRKY-related polypeptides, or the EMG1-like polypeptides, or the GPx-related polypeptides respectively, as described above.

The invention also includes host cells containing an isolated nucleic acid encoding an

ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, as defined above. In one embodiment host cells according to the invention are plant cells, yeasts, bacteria or fungi. Host plants for the nucleic acids, construct, expression cassette or the vector used in the method according to the invention are, in principle, advantageously all plants which are capable of synthesizing the polypeptides used in the inventive method. In a particular embodiment the plant cells of the invention overexpress the nucleic acid molecule of the invention.

The methods of the invention are advantageously applicable to any plant, in particular to any plant as defined herein. Plants that are particularly useful in the methods of the invention include all plants which belong to the superfamily Viridiplantae, in particular monocotyledonous and dicotyledonous plants including fodder or forage legumes, ornamental plants, food crops, trees or shrubs. According to an embodiment of the present invention, the plant is a crop plant. Examples of crop plants include but are not limited to chicory, carrot, cassava, trefoil, soybean, beet, sugar beet, sunflower, canola, alfalfa, rapeseed, linseed, cotton, tomato, potato and tobacco. According to another embodiment of the present invention, the plant is a monocotyledonous plant. Examples of monocotyledonous plants include sugarcane. According to another embodiment of the present invention, the plant is a cereal. Examples of cereals include rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo and oats. In a particular embodiment the plants of the invention or used in the methods of the invention are selected from the group consisting of maize, wheat, rice, soybean, cotton, oilseed rape including canola, sugarcane, sugar beet and alfalfa. Advantageously the methods of the invention are more efficient than the known methods, because the plants of the invention have increased yield and/or tolerance to an environmental stress compared to control plants used in comparable methods.

The invention also extends to harvestable parts of a plant such as, but not limited to seeds, leaves, fruits, flowers, stems, roots, rhizomes, tubers and bulbs, which harvestable parts comprise a recombinant nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide. The invention furthermore relates to products derived or produced, preferably directly derived or directly produced, from a harvestable part of such a plant, such as dry pellets, meal or powders, oil, fat and fatty acids, starch or proteins. In one embodiment the product comprises a recombinant nucleic acid encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, and/or a recombinant an ELM2-related polypeptide, or WRKY-related polypeptide, or EMG1-like polypeptide, or GPx-related polypeptide, for example as an indicator of the particular quality of the product.

The invention also includes methods for manufacturing a product comprising a) growing the plants of the invention and b) producing said product from or by the plants of the invention

or parts thereof, including seeds. In a further embodiment the methods comprise the steps of a) growing the plants of the invention, b) removing the harvestable parts as described herein from the plants and c) producing said product from, or with the harvestable parts of plants according to the invention.

5

In one embodiment the products produced by the methods of the invention are plant products such as, but not limited to, a foodstuff, feedstuff, a food supplement, feed supplement, fiber, cosmetic or pharmaceutical. In another embodiment the methods for production are used to make agricultural products such as, but not limited to, plant extracts, proteins, amino acids, carbohydrates, fats, oils, polymers, vitamins, and the like.

10

In yet another embodiment the polynucleotides or the polypeptides of the invention are comprised in an agricultural product. In a particular embodiment the nucleic acid sequences and protein sequences of the invention may be used as product markers, for example where an agricultural product was produced by the methods of the invention. Such a marker can be used to identify a product to have been produced by an advantageous process resulting not only in a greater efficiency of the process but also improved quality of the product due to increased quality of the plant material and harvestable parts used in the process. Such markers can be detected by a variety of methods known in the art, for example but not limited to PCR based methods for nucleic acid detection or antibody based methods for protein detection.

15

20

The present invention also encompasses use of nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, as described herein and use of these ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, in enhancing any of the aforementioned yield-related traits in plants. For example, nucleic acids encoding the ELM2-related polypeptides, or the WRKY-related polypeptides, or the EMG1-like polypeptides, or the GPx-related polypeptides, themselves, may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to above described polypeptide-encoding gene. The nucleic acids/genes, or the ELM2-related polypeptides, or the WRKY-related polypeptides, or the EMG1-like polypeptides, or the GPx-related polypeptides, themselves may be used to define a molecular marker. This DNA or protein marker may then be used in breeding programmes to select plants having enhanced yield-related traits as defined herein in the methods of the invention. Furthermore, allelic variants of a nucleic acid/gene encoding an ELM2-related polypeptide, or a WRKY-related polypeptide, or an EMG1-like polypeptide, or a GPx-related polypeptide, may find use in marker-assisted breeding programmes. Nucleic acids encoding ELM2-related polypeptides, or WRKY-related polypeptides, or EMG1-like polypeptides, or GPx-related polypeptides, may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired

25

30

35

40

phenotypes.

In the following, the expression “as defined in claim/embodiment X” is meant to direct the artisan to apply the definition as disclosed in claim/embodiment X. For example, “a nucleic acid as defined in embodiment 1” has to be understood so that the definition of the nucleic acid as in embodiment 1 is to be applied to the nucleic acid. In consequence the term “ as defined in item” or “ as defined in claim” may be replaced with the corresponding definition of that embodiment or claim, respectively.

10 With respect to ELM2-related polypeptides, the present invention moreover relates to the following specific embodiments:

1. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an ELM2-related poly-peptide, wherein said ELM2-related polypeptide comprises an Interpro
15 accession IPR 000949, according to PFAM accession number PF01448 ELM2 domain.

2. Method according to embodiment 1, wherein said ELM2-related polypeptide comprises one or more motifs having at least 80% sequence identity with one or more of the following motifs:

- 20 (i) Motif 1: [I/V]GKGR[S/Q]DSC[G/R]CQV[Q/P][K/G]S[I/V][K/E]CVRFH[I/V][T/A]E[R/K][SR][S/A/L][R/K][V/L][M/K][R/L]E[L/I]G[K/V/S]AF[N/Y][Q/H/A]W[R/G/N][F/L]D[K/R][M/A]GEE represented by SEQ ID NO: 93,
(ii) Motif 2: [R/T/K]xFP[S/K][R/K][R/S/G]R[E/K][D/S/E]LVSYY[Y/F]NVFLL[Q/R]RR[A/G][N/Y]QNR[S/H/V]TP[D/N/K][S/N]I represented by SEQ ID NO: 94,
25 (iii) Motif 3: [P/S][I/P][T/R]xxIP[V/L]GP[V/N][F/H]QAE[V/I]PEWT represented by SEQ ID NO: 95

wherein x can be any amino acid.

3. Method according to embodiment 1 or 2, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said
30 ELM2-related polypeptide.

4. Method according to any one of embodiments 1 to 3, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.

5. Method according to any one of embodiments 1 to 4, wherein said enhanced yield-related traits are obtained under non-stress conditions.
35

6. Method according to any one of embodiments 1 to 4, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

7. Method according to any one of embodiments 1 to 6, wherein said nucleic acid
40 encoding an ELM2-related polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Salicaceae, more preferably from the genus *Populus*, most preferably from *Populus trichocarpa*.

8. Method according to any one of embodiments 1 to 6, wherein said nucleic acid encoding an ELM2-related polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Fabaceae, more preferably from the genus *Medicago*, most preferably from *Medicago truncatula*.
- 5 9. Method according to any one of embodiments 1 to 8, wherein said nucleic acid encoding an ELM2-related encodes any one of the polypeptides listed in Table A1 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
- 10 10. Method according to any one of embodiments 1 to 9, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A1.
11. Method according to any one of embodiments 1 to 10, wherein said nucleic acid encodes the polypeptide represented by any one of SEQ ID NO: 2 or SEQ ID NO: 4 or a homologue thereof which has at least 90% overall sequence identity to SEQ ID NO :
15 2 or SEQ ID NO: 4.
12. Method according to any one of embodiments 1 to 11, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
- 20 13. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method according to any one of embodiments 1 to 12, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an ELM2-related polypeptide as defined in any of embodiments 1 and 6 to 11.
14. Construct comprising:
25 (i) nucleic acid encoding an ELM2-related polypeptide as defined in any of embodiments 1 and 6 to 11;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
(iii) a transcription termination sequence.
- 30 15. Construct according to embodiment 14, wherein one of said control sequences is a constitutive promoter, preferably a medium strength constitutive promoter, preferably to a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
16. Use of a construct according to embodiment 14 or 15 in a method for making plants
35 having enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
17. Plant, plant part or plant cell transformed with a construct according to embodiment 14 or 15.
- 40 18. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants,

comprising:

- (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an ELM2-related polypeptide as defined in any of embodiments 1 and 6 to 11; and
- (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.

19. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding an ELM2-related polypeptide as defined in any of embodiments 1 and 6 to 11 or a transgenic plant cell derived from said transgenic plant.

20. Transgenic plant according to embodiment 13, 17 or 19, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.

21. Harvestable parts of a plant according to embodiment 20, wherein said harvestable parts are preferably shoot biomass and/or seeds.

22. Products derived from a plant according to embodiment 20 and/or from harvestable parts of a plant according to embodiment 21.

23. Use of a nucleic acid encoding an ELM2-related polypeptide as defined in any of embodiments 1 and 6 to 11 for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.

24. A method for manufacturing a product comprising the steps of growing the plants according to embodiment 13, 17, 20 or 21 and producing said product from or by said plants; or parts thereof, including seeds.

25. Construct according to embodiment 14 or 15 comprised in a plant cell.

26. Recombinant chromosomal DNA comprising the construct according to embodiment 14 or 15.

27. A method of growing a transgenic plant comprising:

- a. planting a transformed seed comprising a nucleic acid sequence as defined in any of embodiments 1 and 6 to 11;
- b. growing a plant from said seed.

With respect to WRKY-related polypeptides, the present invention moreover relates to the following specific embodiments:

1. A method for enhancing yield in plants relative to or compared to control plants, comprising modulating expression in a plant of a nucleic acid molecule encoding a WRKY-related polypeptide, preferably comprising one or more motifs selected from a group consisting of WRKYGQ motif, basic motif, coiled coil motif as defined herein.

2. Method according to embodiment 1, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid molecule encoding a WRKY-

related polypeptide.

3. Method according to embodiment 1 or 2, wherein said polypeptide is encoded by a nucleic acid molecule comprising a nucleic acid molecule selected from the group consisting of:

- 5 (i) a nucleic acid represented by SEQ ID NO: 102;
 - (ii) the complement of a nucleic acid represented by SEQ ID NO: 102;
 - (iii) a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 103, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be deduced from a polypeptide sequence as represented by
10 SEQ ID NO: 103 and further preferably confers enhanced yield-related traits relative to or compared to control plants;
 - (iv) a nucleic acid having, in increasing order of preference at least 30 %, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%,
15 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the nucleic acid sequence of SEQ ID NO: 102, and further preferably conferring enhanced yield-related traits relative to or compared to
20 control plants, wherein the nucleic acid encodes a polypeptide that is not the polypeptide of the polypeptide sequence as represented by SEQ ID NO: 103;
 - (v) a first nucleic acid molecule which hybridizes with a second nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to or compared to control plants,
25 wherein the first nucleic acid encodes a polypeptide that is not the polypeptide of any of the polypeptide sequence as represented by SEQ ID NO: 103;
 - (vi) a nucleic acid encoding said polypeptide having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%,
30 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by SEQ ID NO: 103 and preferably conferring enhanced yield-related traits relative to or compared to control plants; or
 - 35 (vii) a nucleic acid comprising any combination(s) of features of (i) to (vi) above.
4. Method according to any embodiment 1 to 3, wherein said enhanced yield-related traits comprise one or more of increased yield, preferably increase of emergence vigour, total seed yield, fill rate, TKW, number of filled seeds, taller more erect plants, amount of thick roots, number of florets per panicle on a plant, increased harvest
40 index, increased greenness of a plant before flowering, increased height of the plant, increased quick early development, increased cycle time relative to or compared to control plants.

5. Method according to any one of embodiments 1 to 4, wherein said enhanced yield-related traits are obtained under non-stress conditions.
6. Method according to any one of embodiments 1 to 4, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.
7. Method according to any one of embodiments 1 to 5, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
8. Method according to any one of embodiments 1 to 7, wherein said nucleic acid molecule or said polypeptide, respectively, is of plant origin, preferably from a dicotyledonous plant, further preferably from a leguminous plant, more preferably from the genus *Medicago*, most preferably from *Medicago truncatula*.
9. Plant or part thereof, including seeds, obtainable by a method according to any one of embodiments 1 to 8, wherein said plant or part thereof comprises a recombinant nucleic acid encoding said polypeptide as defined in any one of embodiments 1, 3 or 8.
10. Construct comprising:
 - (i) nucleic acid encoding said polypeptide as defined in any one of embodiments 1, 3 or 8;
 - (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
 - (iii) a transcription termination sequence.
11. Construct according to embodiment 10, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
12. Use of a construct according to embodiment 10 or 11 in a method for making plants having increased yield, particularly seed yield and/or shoot biomass relative to or compared to control plants.
13. Plant, plant part or plant cell transformed with a construct according to embodiment 10 or 11 or obtainable by a method according to any one of embodiments 1 to 8, wherein said plant or part thereof comprises a recombinant nucleic acid encoding said polypeptide as defined in any one of embodiments 1, 3 or 8.
14. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to or compared to control plants, comprising:
 - (i) introducing and expressing in a plant a nucleic acid encoding said polypeptide as defined in any one of embodiments 1, 3 or 8; and
 - (ii) cultivating the plant cell under conditions promoting plant growth and development.
15. Plant having increased yield, particularly increased biomass and/or increased seed yield, relative to or compared to control plants, resulting from modulated expression of a nucleic acid encoding said polypeptide, or a transgenic plant cell originating from or

being part of said transgenic plant.

16. A method for the production of a product comprising the steps of growing the plants of the invention and producing said product from or by

(a) the plants of the invention; or

(b) parts, including seeds, of these plants.

17. Plant according to embodiment 9, 13 or 15 or a transgenic plant cell originating thereof, or a method according to embodiment 16, wherein said plant is a crop plant, preferably a dicot such as sugar beet, alfalfa, trefoil, chicory, carrot, cassava, cotton, soybean, canola or a monocot, such as sugarcane, or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum emmer, spelt, secale, einkorn, teff, milo and oats.

18. Harvestable parts of a plant according to embodiment 9, 13 or 15 wherein said harvestable parts are preferably shoot and/or root biomass and/or seeds.

19. Products produced from a plant according to embodiment 9, 13 or 15 and/or from harvestable parts of a plant according to embodiment 17.

20. Use of a nucleic acid encoding a polypeptide as defined in any one of embodiments 1, 3 or 8 in increasing yield, particularly seed yield and/or shoot biomass relative to or compared to control plants.

21. Construct according to embodiment 10 or 11 comprised in a plant cell.

22. Recombinant chromosomal DNA comprising the construct according to embodiment 10 or 11.

With respect to EMG1-like polypeptides, the present invention moreover relates to the following specific embodiments:

1. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an EMG1-like polypeptide, wherein said EMG1-like polypeptide comprises an InterPro accession IPR005304 EMG1 domain corresponding to PFAM accession number PF03587.

2. A method according to embodiment 1, wherein said EMG1-like polypeptide comprises an EMG1 domain having at least 60% sequence identity with the EMG1 domain from amino acid 81 to 276 in SEQ ID NO: 179.

3. A method according to any of the embodiments 1 or 2, wherein said EMG1-like polypeptide further comprises one of the motifs represented by SEQ ID NO: 286, SEQ ID NO: 287 or SEQ ID NO: 288.

4. Method according to any of the embodiments 1 to 3, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said EMG1-like polypeptide.

5. Method according to any of the embodiments 1 to 4, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.

6. Method according to any one of embodiments 1 to 5, wherein said enhanced yield-related traits are obtained under non-stress conditions.

7. Method according to any one of embodiments 1 to 5, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.
- 5 8. Method according to any one of embodiments 1 to 7, wherein said nucleic acid encoding an EMG1-like is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Salicaceae, more preferably from the genus *Populus*, most preferably from *Populus trichocarpa*.
- 10 9. Method according to any one of embodiments 1 to 8, wherein said nucleic acid encoding an EMG1-like polypeptide encodes any one of the polypeptides listed in Table A3 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
10. Method according to any one of embodiments 1 to 9, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A3.
- 15 11. Method according to any one of embodiments 1 to 10, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 179.
12. Method according to any one of embodiments 1 to 11, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
- 20 13. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method according to any one of embodiments 1 to 12, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an EMG1-like polypeptide as defined in any of embodiments 1 to 3 and 8 to 12.
- 25 14. Construct comprising:
 - (i) nucleic acid encoding an EMG1-like as defined in any of embodiments 1 to 3 and 8 to 12;
 - (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - 30 (iii) a transcription termination sequence.
15. Construct according to embodiment 14, wherein one of said control sequences is a constitutive promoter, preferably a medium strength constitutive promoter, preferably to a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
- 35 16. Use of a construct according to embodiment 14 or 15 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
- 40 17. Plant, plant part or plant cell transformed with a construct according to embodiment 14 or 15.
18. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more

preferably increased seed yield and/or increased biomass relative to control plants, comprising:

- (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an EMG1-like polypeptide as defined in any of embodiments 1 to 3 and 8 to 12; and
- (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.

19. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding an EMG1-like polypeptide as defined in any of embodiments 1 to 3 and 8 to 12 or a transgenic plant cell derived from said transgenic plant.

20. Transgenic plant according to embodiment 13, 17 or 19, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.

21. Harvestable parts of a plant according to embodiment 20, wherein said harvestable parts are preferably shoot biomass and/or seeds.

22. Products derived from a plant according to embodiment 20 and/or from harvestable parts of a plant according to embodiment 21.

23. Use of a nucleic acid encoding an EMG1-like polypeptide as defined in any of embodiments 1 to 3 and 8 to 12 for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.

24. A method for manufacturing a product comprising the steps of growing the plants according to embodiment 13, 17, 19 or 20 and producing said product from or by said plants; or parts thereof, including seeds.

With respect to GPx-related polypeptides, the present invention moreover relates to the following specific embodiments:

1. A method for enhancing yield-related traits in plants relative to or compared to control plants, comprising modulating expression in a plant of a nucleic acid encoding a GPx-related polypeptide, wherein said GPx-related polypeptide comprises one or more motifs and/or signatures having at least 80% sequence identity with a motif and/or signature selected from a group consisting of motifs 7 to 18, represented in SEQ ID NO: 361 to 372, respectively and/or one or more signatures 1 to 6 as represented in SEQ ID NO: 373 to 378, respectively.

2. Method according to embodiment 1, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said GPx-related polypeptide.

3. Method according to embodiment 1 or 2, wherein said polypeptide is encoded by a nucleic acid molecule comprising a nucleic acid molecule selected from the group

consisting of:

- (i) a nucleic acid represented by any one of SEQ ID NO: 292, 294, 296;
- (ii) the complement of a nucleic acid represented by any one of SEQ ID NO: 292, 294, 296;

5 (iii) a nucleic acid encoding the polypeptide as represented by any one of SEQ ID NO: 293, 295, 297, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be deduced from a polypeptide sequence as represented by any one of SEQ ID NO: 293, 295, 297 and further preferably confers enhanced yield-related traits relative to or compared to control plants;

10 (iv) a nucleic acid having, in increasing order of preference at least 30 %, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 15 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with any of the nucleic acid sequences of SEQ ID NO: 292, 294, 296 and further preferably conferring enhanced yield-related traits relative to or compared to control plants, wherein the nucleic acid encodes a polypeptide that is not the polypeptide of any of the polypeptide sequence as represented by any one of SEQ ID NO: 293, 295, 297;

20 (v) a first nucleic acid molecule which hybridizes with a second nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to or compared to control plants, wherein the first nucleic acid encodes a polypeptide that is not the polypeptide of any of the polypeptide sequence as represented by any one of SEQ ID NO: 293, 295, 297;

25 (vi) a nucleic acid encoding said polypeptide having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 30 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by any one of SEQ ID NO: 293, 295, 297 and preferably conferring enhanced yield-related traits relative to or compared to control plants; or

35 (vii) a nucleic acid comprising any combination(s) of features of (i) to (vi) above.

4. Method to any one of embodiments 1 to 3, wherein said enhanced seed yield traits relative or compared to control plants comprise one or more of increased seed yield, increased seed size, increased total weight of seeds, increased number of total seeds, increased fill rate, increased number of filled seeds, number of florets per panicle on a 40 plant, increased height of the plant, increased quick early development, harvest index, increased cycle time relative to or compared to control plants.

5. Method according to any one of embodiments 1 to 4, wherein said increase in seed

yield comprises an increase of at least 5 % in said plant when compared to control plants for at least one of said parameters.

6. Method according to any one of embodiments 1 to 5, wherein said enhanced yield-related traits are obtained under non-stress conditions.

5 7. Method according to any one of embodiments 1 to 6, wherein said enhanced yield-related traits are obtained under abiotic stress conditions, preferably drought stress, salt stress or nitrogen deficiency.

8. Method according to any one of embodiments 1 to 7, wherein said GPx-related polypeptide comprises one or more of the following motifs:

- 10 (i) Motif 7 represented by SEQ ID NO: 361,
- (ii) Motif 8 represented by SEQ ID NO: 362,
- (iii) Motif 9 represented by SEQ ID NO: 363,
- (iv) Motif 10 represented by SEQ ID NO: 364,
- (v) Motif 11 represented by SEQ ID NO: 365,
- 15 (vi) Motif 12 represented by SEQ ID NO: 366,
- (vii) Motif 13 represented by SEQ ID NO: 367,
- (viii) Motif 14 represented by SEQ ID NO: 368,
- (ix) Motif 15 represented by SEQ ID NO: 369,
- (x) Motif 16 represented by SEQ ID NO: 370,
- 20 (xi) Motif 17 represented by SEQ ID NO: 371,
- (xii) Motif 18 represented by SEQ ID NO: 372.

9. Method according to any one of embodiments 1 to 8, wherein said GPx-related polypeptide comprises one or more of the following signatures:

- (i) Signature 1 represented by SEQ ID NO: 373,
- 25 (ii) Signature 2 represented by SEQ ID NO: 374,
- (iii) Signature 3 represented by SEQ ID NO: 375,
- (iv) Signature 4 represented by SEQ ID NO: 376,
- (v) Signature 5 represented by SEQ ID NO: 377,
- (vi) Signature 6 represented by SEQ ID NO: 378.

30 10. Method according to any one of embodiments 1 to 9, wherein said nucleic acid encoding a GPx-related polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Poaceae, more preferably from the genus *Oryza*, most preferably from *Oryza sativa*.

35 11. Method according to any one of embodiments 1 to 10, wherein said nucleic acid encoding a GPx-related polypeptide encodes any one of the polypeptides listed in Table A4 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

40 12. Method according to any one of embodiments 1 to 11, wherein said nucleic acid is operably linked to a constitutive promoter of plant origin, preferably to a medium strength constitutive promoter of plant origin, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

13. Method according to any one of embodiments 1 to 12, wherein said plant is a

monocotyledonous plant.

14. Method according to embodiment 13, wherein said plant is a cereal.

15. Plant, or part thereof, or plant cell, obtainable by a method according to any one of embodiments 1 to 14, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding a GPx-related polypeptide as defined in any one of embodiments 1 to 14.

16. Construct comprising:

(i) nucleic acid encoding an GPx-related polypeptide as defined in any one of embodiments 1 to 14;

(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

17. Construct according to embodiment 16, wherein one of said control sequences is a constitutive promoter of plant origin, preferably to a medium strength constitutive promoter of plant origin, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

18. Use of a construct according to embodiment 16 or 17 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to or compared to control plants, and more preferably increased seed yield and/or increased biomass relative to or compared to control plants.

19. Plant, plant part or plant cell transformed with a construct according to embodiment 16 or 17.

20. Method for the production of a transgenic plant having enhanced yield-related traits relative to or compared to control plants, preferably increased yield relative to or compared to control plants, and more preferably increased seed yield and/or increased biomass relative to or compared to control plants, comprising:

(i) introducing and expressing in a plant cell or plant a nucleic acid encoding a GPx-related polypeptide as defined in any one of embodiments 1 to 14; and

(ii) cultivating said plant cell or plant under conditions promoting plant growth and development.

21. Transgenic plant having enhanced yield-related traits relative to or compared to control plants, preferably increased yield relative to or compared to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding a GPx-related polypeptide as defined in any one of embodiments 1 to 14 or a transgenic plant cell derived from said transgenic plant.

22. Transgenic plant according to embodiment 15, 19 or 21, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.

23. Harvestable parts of a plant according to embodiment 22, wherein said harvestable

parts are preferably biomass and/or seeds.

24. Products derived from a plant according to embodiment 22 and/or from harvestable parts of a plant according to embodiment 23.

25. Use of a nucleic acid encoding an GPx-related polypeptide as defined in any one of embodiments 1 to 14 for enhancing yield-related traits in plants relative to or compared to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to or compared to control plants.

26. A method for manufacturing a product comprising the steps of growing the plants according to embodiment 15, 19 or 21 and producing said product from or by said plants; or parts thereof, including seeds.

Description of figures

The present invention will now be described with reference to the following figures in which:

Figure 1 represents the domain structure of SEQ ID NO: 2 with conserved ELM2-related domain and the three MEME motifs.

Figure 2 represents the domain structure of SEQ ID NO: 4 with conserved ELM2-related domain and the three MEME motifs.

Figure 3 represents a multiple alignment of various ELM2-related polypeptides. These alignments can be used for defining further motifs or signature sequences, when using conserved amino acids.

The corresponding SEQ ID NO's for the aligned polypeptide sequences shown in Figure 3 are:

SEQ ID NO: 6 for A.thaliana_AT2G46040.1#1

SEQ ID NO: 8 for A.thaliana_AT4G11400.1#1

SEQ ID NO: 10 for G.hirsutum_ES795168 #1

SEQ ID NO: 12 for G.max_Glyma07g07530.1#1

SEQ ID NO: 24 for P.persica_TC15862#1

SEQ ID NO: 40 for V.vinifera_GSVIVT00026942001#1

SEQ ID NO: 16 for G.max_Glyma09g39360.1#1

SEQ ID NO: 18 for G.max_Glyma18g46950.1#1

SEQ ID NO: 4 for M.truncatula_AC149204_12.4#1

SEQ ID NO: 26 for P.trichocarpa_572279#1

SEQ ID NO: 2 for Poptr_ELM2-related

SEQ ID NO: 28 for P.trichocarpa_751302#1

SEQ ID NO: 30 for P.trichocarpa_757535#1

SEQ ID NO: 38 for V.vinifera_GSVIVT00002575001#1

SEQ ID NO: 20 for O.sativa_LOC_Os04g08410.1#1

SEQ ID NO: 32 for P.virgatum_TC11466#1

SEQ ID NO: 36 for S.bicolor_Sb06g001680.1#1

SEQ ID NO: 44 for Zea_mays_GRMZM2G424651_T01#1

SEQ ID NO: 14 for G.max_Glyma08g01950.1 #1

SEQ ID NO: 22 for O.sativa_LOC_Os10g30944.2#1

SEQ ID NO: 34 for S.bicolor_Sb01g020470.1#1

SEQ ID NO: 42 for Zea_mays_GRMZM2G365731_T01#1

Figure 4 shows a phylogenetic tree of ELM2-related polypeptides, as described in example 2.

Figure 5 shows the MATGAT table of Example 3 of the full length sequences of Table A1.

Figure 6 shows the MATGAT table of Example 3 of the conserved ELM2-domain in the sequences of Table A1.

Figure 7 represents the binary vector used for increased expression in *Oryza sativa* of an ELM2-related-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

Figure 8 shows the domain structure of WRKY-related polypeptide as represented by SEQ ID NO: 2. Coiled coil sequence is represented by amino acids LMNLYF (indicated in bold, underlined); basic motif as NLS is represented by KKAR (indicated in bold, italic, underlined); C2H2 zinc finger (represented by amino acids CCHH (bold, underlined)) containing the WRKY motif (WRKYGQK motif (bold, underlined)) represented by amino acid sequence indicated in bold.

Figure 9 represents the binary vector used for increased expression in *Oryza sativa* of WRKY-related polypeptide-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

Figure 10 shows the MATGAT table of Example 3.

Figure 11 represents the domain structure of SEQ ID NO: 179 with conserved motifs 4, 5 and 6 and the EMG1 domain.

Figure 12 represents a multiple alignment of various EMG1-like polypeptides. These alignments can be used for defining further motifs or signature sequences, when using conserved amino acids.

The corresponding SEQ ID NO's for the aligned polypeptide sequences shown in Figure 12 are:

SEQ ID NO: 267 for T.aestivum_TC329781

SEQ ID NO: 233 for T.aestivum_TC342076

SEQ ID NO: 221 for H.vulgare_TC168675

SEQ ID NO: 237 for O.sativa_LOC_Os02g18830.1

SEQ ID NO: 227 for P.virgatum_TC36436

SEQ ID NO: 231 for Z.mays_GRMZM2G130339_T02

SEQ ID NO: 229 for S.bicolor_Sb04g033060.1

SEQ ID NO: 225 for S.bicolor_Sb04g011280.1

SEQ ID NO: 219 for C.longa_TA2440_136217

SEQ ID NO: 251 for Y.filamentosa_DT581455

SEQ ID NO: 209 for A.sp_TC27214

SEQ ID NO: 253 for P.patens_46413

SEQ ID NO: 249 for P.patens_TC42118

SEQ ID NO: 247 for S.moellendorffii_407397

SEQ ID NO: 269 for A.trichopoda_CK754382

SEQ ID NO: 261 for *P.taeda*_TA10927_3352
 SEQ ID NO: 243 for *P.glauca*_TA18327_3330
 SEQ ID NO: 241 for *P.sitchensis*_TA12989_3332
 SEQ ID NO: 277 for *P.glauca*_CO481519
 5 SEQ ID NO: 239 for *C.japonica*_TA2268_3369
 SEQ ID NO: 211 for *M.truncatula*_CU012050_19.4
 SEQ ID NO: 203 for *L.japonicus*_TC43316
 SEQ ID NO: 217 for *G.max*_Glyma12g02820.2
 SEQ ID NO: 207 for *G.max*_TC278049
 10 SEQ ID NO: 235 for *G.max*_Glyma12g02820.5
 SEQ ID NO: 205 for *G.max*_Glyma11g10520.1
 SEQ ID NO: 201 for *G.max*_Glyma12g02820.1
 SEQ ID NO: 263 for *P.coccineus*_TA3737_3886
 SEQ ID NO: 259 for *P.coccineus*_TC15108
 15 SEQ ID NO: 193 for *P.vulgaris*_TC10901
 SEQ ID NO: 265 for *C.sinensis*_TC22219
 SEQ ID NO: 271 for *G.hirsutum*_ES833864
 SEQ ID NO: 185 for *G.raimondii*_TC1038
 SEQ ID NO: 183 for *P.trichocarpa*_667309
 20 SEQ ID NO: 181 for *P.trichocarpa*_831805
 SEQ ID NO: 179 for *P.trichocarpa* EMG1-like
 SEQ ID NO: 187 for *P.trifoliata*_TA8498_37690
 SEQ ID NO: 213 for *G.hirsutum*_TC158561
 SEQ ID NO: 245 for *E.esula*_TC3222
 25 SEQ ID NO: 223 for *E.esula*_DV156953
 SEQ ID NO: 189 for *V.vinifera*_GSVIVT00033904001
 SEQ ID NO: 199 for *B.napus*_TC70157
 SEQ ID NO: 195 for *B.napus*_TC76938
 SEQ ID NO: 197 for *A.thaliana*_AT3G57000.1
 30 SEQ ID NO: 191 for *A.lyrata*_486143
 SEQ ID NO: 275 for *P.americana*_CV001201
 SEQ ID NO: 255 for *M.crystallinum*_TC9042
 SEQ ID NO: 279 for *V.carteri*_106743
 SEQ ID NO: 273 for *C.reinhardtii*_178371
 35 SEQ ID NO: 285 for *M.sp*_77569
 SEQ ID NO: 283 for *O.lucimarinus*_39810
 SEQ ID NO: 281 for *O.sp*_18991
 SEQ ID NO: 277 for *O.tauri*_37777
 SEQ ID NO: 215 for *A.officinalis*_TA1655_4686

40 **Figure 13** shows a phylogenetic tree of EMG1-like polypeptides, as described in Example 2.
Figure 14 shows the MATGAT table of Example 3 over the full length of the sequences of Table A3.

Figure 5 shows the MATGAT table of Example 3 over the EMG1-like domain of the sequences of Table A3.

Figure 16 represents the binary vector used for increased expression in *Oryza sativa* of an EMG1-like-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

5 **Figure 17** represents the domain structure of SEQ ID NO: 293 with conserved motifs.

Figure 18 represents the binary vector used for increased expression in *Oryza sativa* of a GPx-related -encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

Figure 19 shows the MATGAT table of Example 3.

10 **Figure 20** represents a multiple alignment of various GPx-related polypeptides. The asterisks indicate identical amino acids among the various protein sequences, colons represent highly conserved amino acid substitutions, and the dots represent less conserved amino acid substitution; on other positions there is no sequence conservation. These alignments can be used for defining further motifs or signature sequences, when using conserved amino acids.

15 **Figure 21** shows phylogenetic tree of GPx-related polypeptides. SEQ ID NO: 293 as used and described herein is *Oryza sativa*_OsGPx03 gene from *Oryza sativa* and indicated by an arrow in the tree. SEQ ID NO: 295 as used and described herein is *Populus trichocarpa*_PtGPx06 gene from *Populus trichocarpa* and indicated by an arrow in the tree. SEQ ID NO: 297 as used and described herein is *Populus trichocarpa*_GPx gene
20 from *Populus trichocarpa* and indicated by an arrow in the tree.

Examples

The present invention will now be described with reference to the following examples, which are by way of illustration only. The following examples are not intended to limit the scope of
25 the invention. Unless otherwise indicated, the present invention employs conventional techniques and methods of plant biology, molecular biology, bioinformatics and plant breedings.

DNA manipulation: unless otherwise stated, recombinant DNA techniques are performed
30 according to standard protocols described in (Sambrook (2001) Molecular Cloning: a laboratory manual, 3rd Edition Cold Spring Harbor Laboratory Press, CSH, New York) or in Volumes 1 and 2 of Ausubel et al. (1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfax (1993) by R.D.D. Croy, published by BIOS Scientific Publications
35 Ltd (UK) and Blackwell Scientific Publications (UK).

Example 1: Identification of sequences related to the nucleic acid sequence used in the methods of intervention

1. ELM2-related polypeptides

40 Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 1 and SEQ ID NO: 2 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search

tools, such as the Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215:403-410; and Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The program is used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. For example, the polypeptide encoded by the nucleic acid of SEQ ID NO: 1 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflect the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search. For example the E-value may be increased to show less stringent matches. This way, short nearly exact matches may be identified.

Table A1 provides a list of nucleic acid and polypeptide sequences related to SEQ ID NO: 1 and SEQ ID NO: 2.

Table A1: Examples of ELM2-related nucleic acids and polypeptides:

Plant Source	Nucleic acid SEQ ID NO:	Protein SEQ ID NO:
Poptr_ELM2 related	1	2
M.truncatula_AC149204_12.4#1	3	4
A.thaliana_AT2G46040.1#1	5	6
A.thaliana_AT4G11400.1#1	7	8
G.hirsutum_ES795168#1	9	10
G.max_Glyma07g07530.1#1	11	12
G.max_Glyma08g01950.1#1	13	14
G.max_Glyma09g39360.1#1	15	16
G.max_Glyma18g46950.1#1	17	18
O.sativa_LOC_Os04g08410.1#1	19	20
O.sativa_LOC_Os10g30944.2#1	21	22
P.persica_TC15862#1	23	24
P.trichocarpa_572279#1	25	26
P.trichocarpa_751302#1	27	28
P.trichocarpa_757535#1	29	30
P.virgatum_TC11466#1	31	32
S.bicolor_Sb01g020470.1#1	33	34
S.bicolor_Sb06g001680.1#1	35	36
V.vinifera_GSVIVT00002575001#1	37	38
V.vinifera_GSVIVT00026942001#1	39	40

Zea_mays_GRMZM2G365731_T01#1	41	42
Zea_mays_GRMZM2G424651_T01#1	43	44

2. WRKY-related polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 102 and SEQ ID NO: 103 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215:403-410; and Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The program is used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. For example, the polypeptide encoded by the nucleic acid of SEQ ID NO: 102 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflect the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search. For example the E-value may be increased to show less stringent matches. This way, short nearly exact matches may be identified.

Table A2 provides a list of nucleic acid and polypeptide sequences related to SEQ ID NO: 102 and SEQ ID NO: 103.

Table A2: Examples of WRKY-related polypeptide nucleic acids and polypeptides:

Plant Source	Nucleic acid SEQ ID NO:	Protein SEQ ID NO:
<i>Medicago truncatula</i>	102	103
AL6G14950	104	105
AT5G15130	106	107
BD1G51030	108	109
CP00014G00310	110	111
CP00017G01970	112	113
GM04G34220	114	115
GM05G01280	116	117
GM06G20300	118	119
GM09G37470	120	121
GM17G10630	122	123

GM18G49140	124	125
GM19G02440	126	127
LJ1G009790	128	129
LJ1G030000	130	131
MD00G099380	132	133
MD00G142380	134	135
MD00G204530	136	137
MD00G279080	138	139
MD06G015440	140	141
MD06G015480	142	143
MD17G003750	144	145
ME02615G00410	146	147
ME04264G00010	148	149
ME10261G00730	150	151
MT7G02120	152	153
MT8G38070	154	155
OS06G05380	156	157
OSINDICA_06G04630	158	159
PT15G06580	160	161
PT17G11190	162	163
RC29822G01580	164	165
RC30147G07590	166	167
VV14G01690	168	169
VV17G04040	170	171
ZM03G02190	172	173

3. EMG1-like polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 178 and SEQ ID NO: 179 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215:403-410; and Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The program is used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. For example, the polypeptide encoded by the nucleic acid of SEQ ID NO: 178 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise

comparison, and ranked according to the probability score (E-value), where the score reflect the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search. For example the E-value may be increased to show less stringent matches. This way, short nearly exact matches may be identified.

- 10 Table A3 provides a list of nucleic acid and protein sequences related to SEQ ID NO: 178 and SEQ ID NO: 179.

Table A3: Examples of EMG1-like nucleic acids and polypeptides:

Plant Source	Nucleic acid SEQ ID NO:	Protein SEQ ID NO:
P.trichocarpa	178	179
P.trichocarpa	180	181
P.trichocarpa	182	183
G.raimondii	184	185
P.trifoliata	186	187
V.vinifera	188	189
A.lyrata	190	191
P.vulgaris	192	193
B.napus	194	195
A.thaliana	196	197
B.napus	198	199
G.max	200	201
L.japonicus	202	203
G.max	204	205
G.max	206	207
Aquilegia_sp_2.1	208	209
M.truncatula	210	211
G.hirsutum	212	213
A.officinalis	214	215
G.max	216	217
C.longa	218	219
H.vulgare	220	221

E.esula	222	223
S.bicolor	224	225
P.virgatum	226	227
S.bicolor	228	229
Z.mays	230	231
T.aestivum	232	233
G.max	234	235
O.sativa	236	237
C.japonica	238	239
P.sitchensis	240	241
P.glauca	242	243
E.esula	244	245
S.moellendorffii	246	247
P.patens	248	249
Y.filamentosa	250	251
P.patens	252	253
M.crystallinum	254	255
P.glauca	256	257
P.coccineus	258	259
P.taeda	260	261
P.coccineus	262	263
C.sinensis	264	265
T.aestivum	266	267
A.trichopoda	268	269
G.hirsutum	270	271
C.reinhardtii	272	273
P.americana	274	275
O.tauri	276	277
V.carteri	278	279
Ostreococcus_RCC809_1_JGI	280	281
O.lucimarinus	282	283
Micromonas_RCC299_3_JGI	284	285

4. GPx-related polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NOs: 292 to 297 were identified amongst those maintained in the Entrez Nucleotides database at the National

Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215:403-410; and Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The program is used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. For example, the polypeptide encoded by the nucleic acid of SEQ ID NO: 292 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflect the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search. For example the E-value may be increased to show less stringent matches. This way, short nearly exact matches may be identified.

Table A4 provides a list of nucleic acid and polypeptide sequences related to SEQ ID NOs: 293 to 297.

Table A4: Examples of GPx-related nucleic acids and polypeptides:

Plant source	Nucleic acid SEQ ID NO:	Protein SEQ ID NO:
Oryza_sativa_OsGPx03	292	293
Populus_trichocarpa_PtGPx06	294	295
Populus_trichocarpa_PtGPx	296	297
Zea_mays_ZmGPx03	298	299
Zea_mays_ZmGPx01-2	300	301
Zea_mays_ZmGPx01	302	303
Triticum_monococcum_TmGPx01		304
Triticum_aestivum-TaGPx03		305
Triticum_aestivum-TaGPx01	306	307
Setaria_italica_SiGPx01	308	309
Sorghum_bicolor_SbGPx06	310	311
Sorghum_bicolor_SbGPx03	312	313
Sorghum_bicolor_SbGPx01	314	315
Oryza_sativas_OsGPx06	316	317
Oryza_sativas_OsGPx01	318	319
Hordeum_vulgare_HvGPx06	320	321

Hordeum_vulgare_HvGPx03	322	323
Hordeum_vulgare_HvGPx01	324	325
Pinus_taeda_PtaGPx06-2		326
Pinus_taeda_PtaGPx06-1		327
Pinus_pinaster_PpGPx06-2		328
Gnetum_gnemon_GgGPx06		329
Zea_mays_ZmGPx03-2	330	331
Vitis_vinifera_VvGPx06		332
Solanum_tuberosum_StGPx06		333
Spinacia_oleracea_SoGPx06		334
Populus_trichocarpa_PtGPx06-2	335	336
Nicotiana_tabacum_NtGPx06		337
Nicotiana_sylvestris_NsGPx06		338
Medicago_truncatula_MtGPx06		339
Malus_domestica_MdGPx06		340
Mesembryanthemum_crystallinum_McGPx06		341
Lotus_japonicus_LjGPx06		342
Lycopersicon_esculentum_LeGPx06		343
Hevea_brasiliensis_HbGPx06		344
Helianthus_annuus_HaGPx06-3_ANN1312		345
Helianthus_annuus_HaGPx06-1		346
Gossypium_hirsutum_GhGPx06-3		347
Gossypium_hirsutum_GhGPx06-2		348
Gossypium_hirsutum_GhGPx06-1	349	350
Citrus_sinensis_CsGPx06		351
Cucumis_melo_CmGPx06		352
Capsicum_chinense_CcGPx06		353
Brassica_napus_BnGPx06-2		354
Brassica_napus_BnGPx06-1		355
Arabidopsis_thaliana_AtGPx06	356	357
Populus_trichocarpa_PtGPx01	358	359
Solanum_tuberosum_StGPx08-2		360

Sequences have been tentatively assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR; beginning with TA). For instance, the Eukaryotic Gene Orthologs (EGO) database may be used to identify such related

sequences, either by keyword search or by using the BLAST algorithm with the nucleic acid sequence or polypeptide sequence of interest. Special nucleic acid sequence databases have been created for particular organisms, e.g. for certain prokaryotic organisms, such as by the Joint Genome Institute. Furthermore, access to proprietary databases, has allowed
5 the identification of novel nucleic acid and polypeptide sequences.

Example 2: Alignment of sequences to the polypeptide sequences used in the methods of the invention

1. ELM2-related polypeptides

10 Alignment of the polypeptide sequences was performed using MAFFT (version 6.624, L-INS-I method). The ELM2-related polypeptides are aligned in Figure 3.

A phylogenetic tree of ELM2-related polypeptides (Figure 4) was constructed by aligning ELM2-related sequences using MAFFT (Kato and Toh (2008) - Briefings in Bioinformatics
15 9:286-298). A neighbour-joining tree was calculated using Quick-Tree (Howe et al. (2002), Bioinformatics 18(11): 1546-7), 100 bootstrap repetitions. The dendrogram was drawn using Dendroscope (Huson et al. (2007), BMC Bioinformatics 8(1):460). Confidence levels for 100 bootstrap repetitions are indicated for major branchings.

20 2. WRKY-related polypeptides

Alignment of the polypeptide sequences is performed using the ClustalW 2.0 algorithm of progressive alignment (Thompson et al. (1997) Nucleic Acids Res 25:4876-4882; Chenna et al. (2003). Nucleic Acids Res 31:3497-3500) with standard setting (slow alignment, similarity matrix: Gonnet, gap opening penalty 10, gap extension penalty: 0.2). Minor
25 manual editing was done to further optimise the alignment. In this context, the sequence as shown in Figure 8 can be used for alignment by the skilled person in order to identify homologues of WRKY-related polypeptide sequences.

3. EMG1-like polypeptides

30 Alignment of the polypeptide sequences was performed using the ClustalW 2.0 algorithm of progressive alignment (Thompson et al. (1997) Nucleic Acids Res 25:4876-4882; Chenna et al. (2003). Nucleic Acids Res 31:3497-3500) with standard setting (slow alignment, similarity matrix: Gonnet, gap opening penalty 10, gap extension penalty: 0.2). Minor manual editing was done to further optimise the alignment. The EMG1-like polypeptides are
35 aligned in Figure 12.

A phylogenetic tree of EMG1-like polypeptides (Figure 13) was constructed by aligning EMG1-like sequences using MAFFT (Kato and Toh (2008) - Briefings in Bioinformatics 9:286-298). A neighbour-joining tree was calculated using Quick-Tree (Howe et al. (2002),
40 Bioinformatics 18(11): 1546-7), 100 bootstrap repetitions. The dendrogram was drawn using Dendroscope (Huson et al. (2007), BMC Bioinformatics 8(1):460). Confidence levels for 100 bootstrap repetitions are indicated for major branchings.

4. GPx-related polypeptides

Alignment of the polypeptide sequences was performed using the ClustalW 2.0 algorithm of progressive alignment (Thompson et al. (1997) Nucleic Acids Res 25:4876-4882; Chenna et al. (2003). Nucleic Acids Res 31:3497-3500) with standard setting (slow alignment, similarity matrix: Gonnet, gap opening penalty 10, gap extension penalty: 0.2). Minor manual editing was done to further optimise the alignment. The GPx-related polypeptides are aligned in Figure 20.

A phylogenetic tree of GPx-related polypeptides (Figure 21) was constructed by aligning GPx-related sequences using MAFFT (Kato and Toh (2008) - Briefings in Bioinformatics 9:286-298) with default settings. A neighbour-joining tree was calculated using Quick-Tree (Howe et al. (2002), Bioinformatics 18(11): 1546-7), 100 bootstrap repetitions. The dendrogram was drawn using Dendroscope (Huson et al. (2007), BMC Bioinformatics 8(1):460). Confidence levels for 100 bootstrap repetitions are indicated for major branchings.

Example 3: Calculation of global percentage identity between polypeptide sequences

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention were determined using MatGAT (Matrix Global Alignment Tool) software (BMC Bioinformatics. 2003 4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. Campanella JJ, Bitincka L, Smalley J; software hosted by Ledion Bitincka). MatGAT generates similarity/identity matrices for DNA or protein sequences without needing pre-alignment of the data. The program performs a series of pair-wise alignments using the Myers and Miller global alignment algorithm, calculates similarity and identity, and then places the results in a distance matrix.

1. ELM2-related polypeptides

Results of the MatGAT analysis are shown in Figure 5 with global similarity and identity percentages over the full length of the polypeptide sequences. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring matrix: Blosom62, First Gap: 12, Extending Gap: 2. The sequence identity (in %) between the ELM2-related polypeptide sequences useful in performing the methods of the invention can be as low as 9 %, is generally higher than 9 %, compared to SEQ ID NO: 2.

The results of the MatGat analysis for the conserved ELM2 domain, according to an Interpro accession IPR 000949, according to PFAM accession number PF01448 ELM2 domain, of the ELM2-related polypeptide sequences are shown in Figure 6. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring

matrix: Blosum62, First Gap: 12, Extending Gap: 2. When percentage identity analysis is performed on the conserved domain instead of on the full length polypeptide sequences, an increase in percentage identity is observed as shown in Figure 6. Lowest values are now above 30% compared to SEQ ID NO: 2. Note that *P. virgatum* does not have an ELM2 domain, as can be seen in the alignment in Figure 3.

2. WRKY-related polypeptides

Results of the MatGAT analysis are shown in Figure 10 with global similarity and identity percentages over the full length of the polypeptide sequences. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2. The sequence identity (in %) between the WRKY-related polypeptide sequences useful in performing the methods of the invention can be as low as 34% compared to SEQ ID NO: 103.

A MATGAT table for local alignment of a specific domain, or data on % identity/similarity between specific domains may also be included.

3. EMG1-like polypeptides

Results of the MatGAT analysis are shown in Figure 14 with global similarity and identity percentages over the full length of the polypeptide sequences. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2. The sequence identity (in %) between the EMG1-like polypeptide sequences useful in performing the methods of the invention can be as low as 17.6 %, but is generally higher than 17.6% compared to SEQ ID NO: 179.

Results of a MATGAT analysis for the EMG1-like domain of the sequences of Table A3 are shown in Figure 15. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2. The sequence identity (in %) over the EMG1-like domain between the EMG1-like polypeptide sequences useful in performing the methods of the invention can be as low as 12.2 %, but is generally higher than 50% compared to SEQ ID NO: 179.

4. GPx-related polypeptides

Results of the MatGAT analysis are shown in Figure 19 with global similarity and identity percentages over the full length of the polypeptide sequences. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2. The sequence identity (in %) between the GPx-related polypeptide sequences useful in performing the methods of the invention can be as low as

50 % compared to SEQ ID NO: 293, 50 % compared to SEQ ID NO: 295, 46 % compared to SEQ ID NO: 297.

Like for full length sequences, a MATGAT table based on subsequences of a specific domain, may be generated. Based on a multiple alignment of GPx-related polypeptides, such as for example the one of Example 2, a skilled person may select conserved sequences and submit as input for a MatGAT analysis. This approach is useful where overall sequence conservation among GPx-related proteins is rather low.

10 Example 4: Identification of domains comprised in polypeptide sequences useful in performing the methods of the invention

The Integrated Resource of Protein Families, Domains and Sites (InterPro) database is an integrated interface for the commonly used signature databases for text- and sequence-based searches. The InterPro database combines these databases, which use different methodologies and varying degrees of biological information about well-characterized proteins to derive protein signatures. Collaborating databases include SWISS-PROT, PROSITE, TrEMBL, PRINTS, ProDom and Pfam, Smart and TIGRFAMs. Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. Pfam is hosted at the Sanger Institute server in the United Kingdom. Interpro is hosted at the European Bioinformatics Institute in the United Kingdom.

1. ELM2-related polypeptides

The results of the InterPro scan (see Zdobnov E.M. and Apweiler R.; "InterProScan - an integration platform for the signature-recognition methods in InterPro."; Bioinformatics, 2001, 17(9): 847-8 (InterPro database, version 4.7)) of the polypeptide sequence as represented by SEQ ID NO: 2 are presented in Table B1.

Table B1: InterPro scan results (major accession number IPR000949) of the polypeptide sequence as represented by SEQ ID NO: 2.

Database	Accession number	Accession name	Amino acid coordinates on SEQ ID NO: 2
PROFILE	PS51156	ELM2	223-267
PFAM	PF01448	ELM2	225-258

In an embodiment an ELM2-related polypeptide comprises a conserved domain or motif with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a conserved domain from amino acid 225 to 258 in SEQ ID NO: 2, more preferably to a conserved domain from amino acid 225 to 262 in SEQ ID NO: 2.

2. WRKY-related polypeptides

The results of the InterPro scan (see Zdobnov E.M. and Apweiler R.; "InterProScan - an integration platform for the signature-recognition methods in InterPro."; Bioinformatics, 2001, 17(9): 847-8 (InterPro database, release 31.0)) of the polypeptide sequence as represented by SEQ ID NO: 103 are presented in Table B2.

Table B2: InterPro scan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 103.

Method	Accession	Domain	start	stop	E-value
Gene3D	G3DSA:2.20.25.80	---	240	315	1.70E-28
PFAM	PF03106	WRKY	254	313	1.60E-27
SMART	SM00774	WRKY	254	314	1.10E-35
PROFILE	PS50811	WRKY	249	315	27.933
SUPERFAMILY	SSF118290	WRKY	245	316	3.2E-27

In an embodiment a WRKY-related polypeptide comprises a conserved domain (or motif) with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a conserved domain from amino acid 254 to 313 SEQ ID NO: 103).

3. EMG1-like polypeptides

The results of the InterPro scan (see Zdobnov E.M. and Apweiler R.; "InterProScan - an integration platform for the signature-recognition methods in InterPro."; Bioinformatics, 2001, 17(9): 847-8 (InterPro database, release version 4.8)) of the polypeptide sequence as represented by SEQ ID NO: 179 are presented in Table B3.

Table B3: InterPro scan results (major accession number IPR005304) of the polypeptide sequence as represented by SEQ ID NO: 179.

Database	Accession number	Accession name	Amino acid coordinates on SEQ ID NO: 179
PFAM	PF03587	EMG1/NEP1	81-276

In an embodiment an EMG1-like polypeptide comprises a conserved domain with at least 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a conserved domain from amino acid 81 to 276 in SEQ ID NO: 179.

4. GPx-related polypeptides

The results of the InterPro scan (see Zdobnov E.M. and Apweiler R.; "InterProScan - an integration platform for the signature-recognition methods in InterPro."; Bioinformatics,

2001, 17(9): 847-8 (InterPro database, release Ver 34.0)) of the polypeptide sequence as represented by SEQ ID NO: 293 are presented in Table B4.

Table B4

Method	Accession	Domain	start	stop	E-value
PRINTS	IPR000889	Glutathione peroxidase	98	115	8.40E-22
PRINTS	IPR000889	Glutathione peroxidase	134	150	8.40E-22
PRINTS	IPR000889	Glutathione peroxidase	199	208	8.40E-22
PIR	IPR000889	Glutathione peroxidase	51	238	1.10E-84
PANTHER	IPR000889	Glutathione peroxidase	40	237	1.20E-102
PFAM	IPR000889	Glutathione peroxidase	80	188	8.30E-43
PROSITE	IPR000889	Glutathione peroxidase	100	115	0.0
PROSITE	IPR000889	Glutathione peroxidase	137	144	0.0
PROSITE	IPR000889	Glutathione peroxidase	70	238	0.0
GENE3D	IPR012335	Thioredoxin fold	79	236	2.70E-77
SUPERF	IPR012336	Thioredoxin fold	78	237	9.30E-64

5

In one embodiment a GPx-related polypeptide comprises a conserved domain (or motif) with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to a conserved domain from amino acid 80 to 188 in SEQ ID NO: 293).

10

Example 5: Topology prediction of the polypeptide sequences useful in performing the methods of invention

TargetP 1.1 predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal pre-sequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP). Scores on which the final prediction is based are not really probabilities, and they do not necessarily add to one. However, the location with the highest score is the most likely according to TargetP, and the relationship between the scores (the reliability class) may be an indication of how certain the prediction is. The reliability class (RC) ranges from 1 to 5, where 1 indicates the strongest prediction. For the sequences predicted to contain an N-terminal presequence a potential cleavage site can also be predicted. TargetP is maintained at the server of the Technical University of Denmark.

15

20

25

For the sequences predicted to contain an N-terminal presequence a potential cleavage site can also be predicted.

A number of parameters must be selected before analysing a sequence, such as organism

group (non-plant or plant), cutoff sets (none, predefined set of cutoffs, or user-specified set of cutoffs), and the calculation of prediction of cleavage sites (yes or no).

Many other algorithms can be used to perform such analyses, including:

- 5 • ChloroP 1.1 hosted on the server of the Technical University of Denmark;
- Protein Prowler Subcellular Localisation Predictor version 1.2 hosted on the server of the Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia;
- PENCE Proteome Analyst PA-GOSUB 2.5 hosted on the server of the University
- 10 of Alberta, Edmonton, Alberta, Canada;
- TMHMM, hosted on the server of the Technical University of Denmark
- PSORT (URL: psort.org)
- PLOC (Park and Kanehisa, Bioinformatics, 19, 1656-1663, 2003).

15 1. WRKY-related polypeptides

The results of TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 103 are presented Table C. The “plant” organism group has been selected, no cutoffs defined, and the predicted length of the transit peptide requested. The subcellular localization of the polypeptide sequence as represented by SEQ ID NO: 103 may be the

20 cytoplasm or nucleus, no transit peptide is predicted.

Table C1: TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 103.

Length (AA)	631
Chloroplastic transit peptide	0.101
Mitochondrial transit peptide	0.119
Secretory pathway signal peptide	0.070
Other subcellular targeting	0.927
Predicted Location	/
Reliability class	1
Predicted transit peptide length	/

25 2. EMG1-like polypeptides

The results of TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 179 are presented Table C2. The “plant” organism group has been selected, no cutoffs defined, and the predicted length of the transit peptide requested. The subcellular localization of the polypeptide sequence as represented by SEQ ID NO: 179 may be the

30 cytoplasm or nucleus, no transit peptide is predicted.

Table C2: TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 179.

Length (AA)	284
Chloroplastic transit peptide	0.037
Mitochondrial transit peptide	0.329
Secretory pathway signal peptide	0.081
Other subcellular targeting	0.878
Predicted Location	/
Reliability class	3
Predicted transit peptide length	/

Example 6: Functional assay related to the polypeptide sequences useful in performing the methods of the invention

1. ELM2-related polypeptides

ELM2-related polypeptides, at least in their native form, typically have transcriptional repression activity, involved in inhibiting chromatin remodeling. Tools and techniques for measuring such activity are well known in the art and e.g. described in Ding et al., Mol. Cell Biol. 2003 Jan; 23(1): 250-8.

Example 7: Cloning of the nucleic acid sequence used in methods of the invention

1. ELM2-related polypeptides

In a first experiment, the nucleic acid sequence was amplified by PCR using as template a custom-made *Poplar trichocarpa* seedlings cDNA library. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm11394 (SEQ ID NO: 98; sense): 5'-ggggacaagtttgtaaaaaagcaggcttaacaatgatggacgggtcttttct-3' and prm11395 (SEQ ID NO: 99; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggtgcatcaacctgatcagacact-3', which include the AttB sites for Gateway recombination. The amplified PCR fragment was purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombined in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone", pELM2-related. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

In a second experiment, the nucleic acid sequence was amplified by PCR using as template a custom-made *Medicago truncatula* seedlings cDNA library. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm11280 (SEQ ID NO: 100; sense): 5'-ggggacaagtttgtaaaaaagcaggcttaacaatggtggtagtgagcaaact-3' and prm11281 (SEQ ID NO: 101; reverse, complementary): 5' ggggaccactttgtacaagaaagctgggtta cagctcatgatttgaggc 3', which include the AttB sites for Gateway recombination. The amplified PCR fragment was purified also using standard methods. The first step of the Gateway

procedure, the BP reaction, was then performed, during which the PCR fragment recombined in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an “entry clone”, pELM2-related. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

5

The entry clones comprising either SEQ ID NO: 1 or SEQ ID NO: 3 were then used in an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector contained as functional elements within the T-DNA borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the nucleic acid sequence of interest already cloned in the entry clone. A rice GOS2 promoter (SEQ ID NO: 97) for constitutive expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vectors pGOS2::ELM2-related (Figure 7) were transformed into *Agrobacterium* strain LBA4044 according to methods well known in the art.

2. WRKY-related polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made *Arabidopsis thaliana* seedlings cDNA library. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm10976 (SEQ ID NO: 176; sense): 5'-ggggacaagtttgtagcaaaaaagcaggcttaacaatggagtgcagttctgggagatc-3' and prm10977 (SEQ ID NO: 177; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggttcaccaattaattggtgtgtcactatt-3', which include the AttB sites for Gateway recombination. The amplified PCR fragment was purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombined in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an “entry clone”, pWRKY. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 102 was then used in an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector contained as functional elements within the T-DNA borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the nucleic acid sequence of interest already cloned in the entry clone. A rice GOS2 promoter (SEQ ID NO: 174) for constitutive expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::WRKY (Figure 9) was transformed into *Agrobacterium* strain LBA4044 according to methods well known in the art.

3. EMG1-like polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made *Populus trichocarpa* seedlings cDNA library. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm14800 (SEQ ID NO: 290; sense): 5'-ggggacaagtttgtacaaaaaagcaggcttaacaatggtgaggccttatggtaaa-3' and prm14801 (SEQ ID NO: 291; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggttgatcgacaaaatgcagagac-3', which include the AttB sites for Gateway recombination. The amplified PCR fragment was purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombined in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone", pEMG1. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 178 was then used in an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector contained as functional elements within the T-DNA borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the nucleic acid sequence of interest already cloned in the entry clone. A rice GOS2 promoter (SEQ ID NO: 289) for constitutive expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::EMG1-like (Figure 16) was transformed into *Agrobacterium* strain LBA4044 according to methods well known in the art.

4. GPx-related polypeptides

The nucleic acid sequence for SEQ ID NO: 292 was amplified by PCR using as template a custom-made *Oryza sativa* seedlings cDNA library. The nucleic acid sequence for SEQ ID NO: 294 and 296, respectively, was amplified by PCR using as template a custom-made *Populus trichocarpa* seedlings cDNA library. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used for amplification of:

- SEQ ID NO: 292 were prm15927: (SEQ ID NO: 381; sense): 5'-ggggacaagtttgtacaaaaaagcaggcttaacaatgccgtctcgacc-3' and prm15928 (SEQ ID NO: 382; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggtcggcggttaaggtttaag-3';
- SEQ ID NO: 294 were prm15929 (SEQ ID NO: 383; sense): 5'-ggggacaagtttgtacaaaaaagcaggcttaacaatggctagccaatccagtg-3' and prm15930 (SEQ ID NO: 384; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggtaattaacaacatgccttga-3';
- SEQ ID NO: 296 were prm15925 (SEQ ID NO: 385; sense): 5'-

ggggacaagtttgtacaaaaagcaggcttaacaatggcttccttacctttctcc-3' and prm15926
(SEQ ID NO: 386; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggt
ctcatacatgtcactgctgcc-3';

which include the AttB sites for Gateway recombination. The amplified PCR fragment was
5 purified also using standard methods. The first step of the Gateway procedure, the BP
reaction, was then performed, during which the PCR fragment recombined in vivo with the
pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone",
pGPx-related. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway®
technology.

10 The entry clone comprising SEQ ID NO: 292, 294, and 296, respectively, was then used in
an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector
contained as functional elements within the T-DNA borders: a plant selectable marker; a
screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo*
15 recombination with the nucleic acid sequence of interest already cloned in the entry clone.
A rice GOS2 promoter (SEQ ID NO: 379) for constitutive expression was located upstream
of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::GPx-related gene
20 (Figure 18) was transformed into *Agrobacterium* strain LBA4044 according to methods well
known in the art.

Example 8: Plant transformation

Rice transformation

25 The *Agrobacterium* containing the expression vector was used to transform *Oryza sativa*
plants. Mature dry seeds of the rice japonica cultivar Nipponbare were dehusked.
Sterilization was carried out by incubating for one minute in 70% ethanol, followed by 30 to
60 minutes, preferably 30 minutes in sodium hypochlorite solution (depending on the grade
of contamination), followed by a 3 to 6 times, preferably 4 time wash with sterile distilled
30 water. The sterile seeds were then germinated on a medium containing 2,4-D (callus
induction medium). After incubation in light for 6 days scutellum-derived calli is transformed
with *Agrobacterium* as described herein below.

Agrobacterium strain LBA4404 containing the expression vector was used for co-cultivation.
35 *Agrobacterium* was inoculated on AB medium with the appropriate antibiotics and cultured
for 3 days at 28°C. The bacteria were then collected and suspended in liquid co-cultivation
medium to a density (OD₆₀₀) of about 1. The calli were immersed in the suspension for 1 to
15 minutes. The callus tissues were then blotted dry on a filter paper and transferred to
solidified, co-cultivation medium and incubated for 3 days in the dark at 25°C. After washing
40 away the *Agrobacterium*, the calli were grown on 2,4-D-containing medium for 10 to 14
days (growth time for indica: 3 weeks) under light at 28°C - 32°C in the presence of a
selection agent. During this period, rapidly growing resistant callus developed. After transfer

of this material to regeneration media, the embryogenic potential was released and shoots developed in the next four to six weeks. Shoots were excised from the calli and incubated for 2 to 3 weeks on an auxin-containing medium from which they were transferred to soil. Hardened shoots were grown under high humidity and short days in a greenhouse.

5

Transformation of rice cultivar indica can also be done in a similar way as give above according to techniques well known to a skilled person.

35 to 90 independent T0 rice transformants were generated for one construct. The primary
10 transformants were transferred from a tissue culture chamber to a greenhouse. After a quantitative PCR analysis to verify copy number of the T-DNA insert, only single copy transgenic plants that exhibit tolerance to the selection agent were kept for harvest of T1 seed. Seeds were then harvested three to five months after transplanting. The method yielded single locus transformants at a rate of over 50 % (Aldemita and Hodges1996, Chan
15 et al. 1993, Hiei et al. 1994).

Example 9: Transformation of other crops

Corn transformation

Transformation of maize (*Zea mays*) is performed with a modification of the method
20 described by Ishida et al. (1996) Nature Biotech 14(6): 745-50. Transformation is genotype-dependent in corn and only specific genotypes are amenable to transformation and regeneration. The inbred line A188 (University of Minnesota) or hybrids with A188 as a parent are good sources of donor material for transformation, but other genotypes can be used successfully as well. Ears are harvested from corn plant approximately 11 days after
25 pollination (DAP) when the length of the immature embryo is about 1 to 1.2 mm. Immature embryos are cocultivated with *Agrobacterium tumefaciens* containing the expression vector, and transgenic plants are recovered through organogenesis. Excised embryos are grown on callus induction medium, then maize regeneration medium, containing the selection agent (for example imidazolinone but various selection markers can be used). The Petri
30 plates are incubated in the light at 25 °C for 2-3 weeks, or until shoots develop. The green shoots are transferred from each embryo to maize rooting medium and incubated at 25 °C for 2-3 weeks, until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

35

Wheat transformation

Transformation of wheat is performed with the method described by Ishida et al. (1996) Nature Biotech 14(6): 745-50. The cultivar Bobwhite (available from CIMMYT, Mexico) is commonly used in transformation. Immature embryos are co-cultivated with *Agrobacterium*
40 *tumefaciens* containing the expression vector, and transgenic plants are recovered through organogenesis. After incubation with *Agrobacterium*, the embryos are grown in vitro on callus induction medium, then regeneration medium, containing the selection agent (for

example imidazolinone but various selection markers can be used). The Petri plates are incubated in the light at 25 °C for 2-3 weeks, or until shoots develop. The green shoots are transferred from each embryo to rooting medium and incubated at 25 °C for 2-3 weeks, until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Soybean transformation

Soybean is transformed according to a modification of the method described in the Texas A&M patent US 5,164,310. Several commercial soybean varieties are amenable to transformation by this method. The cultivar Jack (available from the Illinois Seed foundation) is commonly used for transformation. Soybean seeds are sterilised for *in vitro* sowing. The hypocotyl, the radicle and one cotyledon are excised from seven-day old young seedlings. The epicotyl and the remaining cotyledon are further grown to develop axillary nodes. These axillary nodes are excised and incubated with *Agrobacterium tumefaciens* containing the expression vector. After the cocultivation treatment, the explants are washed and transferred to selection media. Regenerated shoots are excised and placed on a shoot elongation medium. Shoots no longer than 1 cm are placed on rooting medium until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Rapeseed/canola transformation

Cotyledonary petioles and hypocotyls of 5-6 day old young seedling are used as explants for tissue culture and transformed according to Babic et al. (1998, Plant Cell Rep 17: 183-188). The commercial cultivar Westar (Agriculture Canada) is the standard variety used for transformation, but other varieties can also be used. Canola seeds are surface-sterilized for *in vitro* sowing. The cotyledon petiole explants with the cotyledon attached are excised from the *in vitro* seedlings, and inoculated with *Agrobacterium* (containing the expression vector) by dipping the cut end of the petiole explant into the bacterial suspension. The explants are then cultured for 2 days on MSBAP-3 medium containing 3 mg/l BAP, 3 % sucrose, 0.7 % Phytagar at 23 °C, 16 hr light. After two days of co-cultivation with *Agrobacterium*, the petiole explants are transferred to MSBAP-3 medium containing 3 mg/l BAP, cefotaxime, carbenicillin, or timentin (300 mg/l) for 7 days, and then cultured on MSBAP-3 medium with cefotaxime, carbenicillin, or timentin and selection agent until shoot regeneration. When the shoots are 5 – 10 mm in length, they are cut and transferred to shoot elongation medium (MSBAP-0.5, containing 0.5 mg/l BAP). Shoots of about 2 cm in length are transferred to the rooting medium (MS0) for root induction. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Alfalfa transformation

A regenerating clone of alfalfa (*Medicago sativa*) is transformed using the method of (McKersie et al., 1999 Plant Physiol 119: 839–847). Regeneration and transformation of alfalfa is genotype dependent and therefore a regenerating plant is required. Methods to obtain regenerating plants have been described. For example, these can be selected from the cultivar Rangelander (Agriculture Canada) or any other commercial alfalfa variety as described by Brown DCW and A Atanassov (1985. Plant Cell Tissue Organ Culture 4: 111-112). Alternatively, the RA3 variety (University of Wisconsin) has been selected for use in tissue culture (Walker et al., 1978 Am J Bot 65:654-659). Petiole explants are cocultivated with an overnight culture of *Agrobacterium tumefaciens* C58C1 pMP90 (McKersie et al., 1999 Plant Physiol 119: 839–847) or LBA4404 containing the expression vector. The explants are cocultivated for 3 d in the dark on SH induction medium containing 288 mg/ L Pro, 53 mg/ L thioproline, 4.35 g/ L K₂SO₄, and 100 µm acetosyringinone. The explants are washed in half-strength Murashige-Skoog medium (Murashige and Skoog, 1962) and plated on the same SH induction medium without acetosyringinone but with a suitable selection agent and suitable antibiotic to inhibit *Agrobacterium* growth. After several weeks, somatic embryos are transferred to BOi2Y development medium containing no growth regulators, no antibiotics, and 50 g/ L sucrose. Somatic embryos are subsequently germinated on half-strength Murashige-Skoog medium. Rooted seedlings were transplanted into pots and grown in a greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Cotton transformation

Cotton is transformed using *Agrobacterium tumefaciens* according to the method described in US 5,159,135. Cotton seeds are surface sterilised in 3% sodium hypochlorite solution during 20 minutes and washed in distilled water with 500 µg/ml cefotaxime. The seeds are then transferred to SH-medium with 50µg/ml benomyl for germination. Hypocotyls of 4 to 6 days old seedlings are removed, cut into 0.5 cm pieces and are placed on 0.8% agar. An *Agrobacterium* suspension (approx. 10⁸ cells per ml, diluted from an overnight culture transformed with the gene of interest and suitable selection markers) is used for inoculation of the hypocotyl explants. After 3 days at room temperature and lighting, the tissues are transferred to a solid medium (1.6 g/l Gelrite) with Murashige and Skoog salts with B5 vitamins (Gamborg et al., Exp. Cell Res. 50:151-158 (1968)), 0.1 mg/l 2,4-D, 0.1 mg/l 6-furfurylaminopurine and 750 µg/ml MgCL₂, and with 50 to 100 µg/ml cefotaxime and 400-500 µg/ml carbenicillin to kill residual bacteria. Individual cell lines are isolated after two to three months (with subcultures every four to six weeks) and are further cultivated on selective medium for tissue amplification (30°C, 16 hr photoperiod). Transformed tissues are subsequently further cultivated on non-selective medium during 2 to 3 months to give rise to somatic embryos. Healthy looking embryos of at least 4 mm length are transferred to tubes with SH medium in fine vermiculite, supplemented with 0.1 mg/l indole acetic acid, 6-furfurylaminopurine and gibberellic acid. The embryos are cultivated at 30°C with a photoperiod of 16 hrs, and plantlets at the 2 to 3 leaf stage are transferred to pots with vermiculite and nutrients. The plants are hardened and subsequently moved to the

greenhouse for further cultivation.

Sugarbeet transformation

Seeds of sugarbeet (*Beta vulgaris* L.) are sterilized in 70% ethanol for one minute followed
5 by 20 min. shaking in 20% Hypochlorite bleach e.g. Clorox® regular bleach (commercially
available from Clorox, 1221 Broadway, Oakland, CA 94612, USA). Seeds are rinsed with
sterile water and air dried followed by plating onto germinating medium (Murashige and
Skoog (MS) based medium (Murashige, T., and Skoog, ., 1962. *Physiol. Plant*, vol. 15, 473-
497) including B5 vitamins (Gamborg et al.; *Exp. Cell Res.*, vol. 50, 151-8.) supplemented
10 with 10 g/l sucrose and 0,8% agar). Hypocotyl tissue is used essentially for the initiation of
shoot cultures according to Hussey and Hopher (Hussey, G., and Hopher, A., 1978. *Annals
of Botany*, 42, 477-9) and are maintained on MS based medium supplemented with 30g/l
sucrose plus 0,25mg/l benzylamino purine and 0,75% agar, pH 5,8 at 23-25°C with a 16-
hour photoperiod. *Agrobacterium tumefaciens* strain carrying a binary plasmid harbouring a
15 selectable marker gene, for example nptII, is used in transformation experiments. One day
before transformation, a liquid LB culture including antibiotics is grown on a shaker (28°C,
150rpm) until an optical density (O.D.) at 600 nm of ~1 is reached. Overnight-grown
bacterial cultures are centrifuged and resuspended in inoculation medium (O.D. ~1)
including Acetosyringone, pH 5,5. Shoot base tissue is cut into slices (1.0 cm x 1.0 cm x 2.0
20 mm approximately). Tissue is immersed for 30s in liquid bacterial inoculation medium.
Excess liquid is removed by filter paper blotting. Co-cultivation occurred for 24-72 hours on
MS based medium incl. 30g/l sucrose followed by a non-selective period including MS
based medium, 30g/l sucrose with 1 mg/l BAP to induce shoot development and cefotaxim
for eliminating the *Agrobacterium*. After 3-10 days explants are transferred to similar
25 selective medium harbouring for example kanamycin or G418 (50-100 mg/l genotype
dependent). Tissues are transferred to fresh medium every 2-3 weeks to maintain selection
pressure. The very rapid initiation of shoots (after 3-4 days) indicates regeneration of
existing meristems rather than organogenesis of newly developed transgenic meristems.
Small shoots are transferred after several rounds of subculture to root induction medium
30 containing 5 mg/l NAA and kanamycin or G418. Additional steps are taken to reduce the
potential of generating transformed plants that are chimeric (partially transgenic). Tissue
samples from regenerated shoots are used for DNA analysis. Other transformation methods
for sugarbeet are known in the art, for example those by Linsey & Gallois (Linsey, K., and
Gallois, P., 1990. *Journal of Experimental Botany*; vol. 41, No. 226; 529-36) or the methods
35 published in the international application published as WO 96/23891A.

Sugarcane transformation

Spindles are isolated from 6-month-old field grown sugarcane plants (Arencibia et al., 1998.
Transgenic Research, vol. 7, 213-22; Enriquez-Obregon et al., 1998. *Planta*, vol. 206, 20-
40 27). Material is sterilized by immersion in a 20% Hypochlorite bleach e.g. Clorox® regular
bleach (commercially available from Clorox, 1221 Broadway, Oakland, CA 94612, USA) for
20 minutes. Transverse sections around 0,5cm are placed on the medium in the top-up

direction. Plant material is cultivated for 4 weeks on MS (Murashige, T., and Skoog, ., 1962. *Physiol. Plant*, vol. 15, 473-497) based medium incl. B5 vitamins (Gamborg, O., et al., 1968. *Exp. Cell Res.*, vol. 50, 151-8) supplemented with 20g/l sucrose, 500 mg/l casein hydrolysate, 0,8% agar and 5mg/l 2,4-D at 23°C in the dark. Cultures are transferred after 4 weeks onto identical fresh medium. *Agrobacterium tumefaciens* strain carrying a binary plasmid harbouring a selectable marker gene, for example hpt, is used in transformation experiments. One day before transformation, a liquid LB culture including antibiotics is grown on a shaker (28°C, 150rpm) until an optical density (O.D.) at 600 nm of ~0,6 is reached. Overnight-grown bacterial cultures are centrifuged and resuspended in MS based inoculation medium (O.D. ~0,4) including acetosyringone, pH 5,5. Sugarcane embryogenic callus pieces (2-4 mm) are isolated based on morphological characteristics as compact structure and yellow colour and dried for 20 min. in the flow hood followed by immersion in a liquid bacterial inoculation medium for 10-20 minutes. Excess liquid is removed by filter paper blotting. Co-cultivation occurred for 3-5 days in the dark on filter paper which is placed on top of MS based medium incl. B5 vitamins containing 1 mg/l 2,4-D. After co-cultivation calli are washed with sterile water followed by a non-selective cultivation period on similar medium containing 500 mg/l cefotaxime for eliminating remaining *Agrobacterium* cells. After 3-10 days explants are transferred to MS based selective medium incl. B5 vitamins containing 1 mg/l 2,4-D for another 3 weeks harbouring 25 mg/l of hygromycin (genotype dependent). All treatments are made at 23°C under dark conditions. Resistant calli are further cultivated on medium lacking 2,4-D including 1 mg/l BA and 25 mg/l hygromycin under 16 h light photoperiod resulting in the development of shoot structures. Shoots are isolated and cultivated on selective rooting medium (MS based including, 20g/l sucrose, 20 mg/l hygromycin and 500 mg/l cefotaxime). Tissue samples from regenerated shoots are used for DNA analysis. Other transformation methods for sugarcane are known in the art, for example from the in-ternational application published as WO 2010/151634A and the granted European patent EP 1831378.

Example 10: Phenotypic evaluation procedure

10.1 Evaluation setup

35 to 90 independent T0 rice transformants were generated. The primary transformants were transferred from a tissue culture chamber to a greenhouse for growing and harvest of T1 seed. Six events, of which the T1 progeny segregated 3:1 for presence/absence of the transgene, were retained. For each of these events, approximately 10 T1 seedlings containing the transgene (hetero- and homo-zygotes) and approximately 10 T1 seedlings lacking the transgene (nullizygotes) were selected by monitoring visual marker expression. The transgenic plants and the corresponding nullizygotes were grown side-by-side at random positions. Greenhouse conditions were of shorts days (12 hours light), 28°C in the light and 22°C in the dark, and a relative humidity of 70%. Plants grown under non-stress conditions were watered at regular intervals to ensure that water and nutrients were not limiting and to satisfy plant needs to complete growth and development, unless they were used in a stress screen.

From the stage of sowing until the stage of maturity the plants were passed several times through a digital imaging cabinet. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles.

5

T1 events can be further evaluated in the T2 generation following the same evaluation procedure as for the T1 generation, e.g. with less events and/or with more individuals per event.

10 *Drought screen*

T1 or T2 plants are grown in potting soil under normal conditions until they approached the heading stage. They are then transferred to a "dry" section where irrigation is withheld. Soil moisture probes are inserted in randomly chosen pots to monitor the soil water content (SWC). When SWC goes below certain thresholds, the plants are automatically re-watered continuously until a normal level is reached again. The plants are then re-transferred again to normal conditions. The rest of the cultivation (plant maturation, seed harvest) is the same as for plants not grown under abiotic stress conditions. Growth and yield parameters are recorded as detailed for growth under normal conditions.

20 *Nitrogen use efficiency screen*

T1 or T2 plants are grown in potting soil under normal conditions except for the nutrient solution. The pots are watered from transplantation to maturation with a specific nutrient solution containing reduced N nitrogen (N) content, usually between 7 to 8 times less. The rest of the cultivation (plant maturation, seed harvest) is the same as for plants not grown under abiotic stress. Growth and yield parameters are recorded as detailed for growth under normal conditions.

Salt stress screen

T1 or T2 plants are grown on a substrate made of coco fibers and particles of baked clay (Argex) (3 to 1 ratio). A normal nutrient solution is used during the first two weeks after transplanting the plantlets in the greenhouse. After the first two weeks, 25 mM of salt (NaCl) is added to the nutrient solution, until the plants are harvested. Growth and yield parameters are recorded as detailed for growth under normal conditions.

35 10.2 Statistical analysis: F test

A two factor ANOVA (analysis of variants) was used as a statistical model for the overall evaluation of plant phenotypic characteristics. An F test was carried out on all the parameters measured of all the plants of all the events transformed with the gene of the present invention. The F test was carried out to check for an effect of the gene over all the transformation events and to verify for an overall effect of the gene, also known as a global gene effect. The threshold for significance for a true global gene effect was set at a 5% probability level for the F test. A significant F test value points to a gene effect, meaning that

it is not only the mere presence or position of the gene that is causing the differences in phenotype.

10.3 Parameters measured

5 From the stage of sowing until the stage of maturity the plants were passed several times through a digital imaging cabinet. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles as described in WO 2010/031780. These measurements were used to determine different parameters.

10 *Biomass-related parameter measurement*

The plant aboveground area (or leafy biomass) was determined by counting the total number of pixels on the digital images from aboveground plant parts discriminated from the background. This value was averaged for the pictures taken on the same time point from the different angles and was converted to a physical surface value expressed in square mm
15 by calibration. Experiments show that the aboveground plant area measured this way correlates with the biomass of plant parts above ground. The above ground area is the area measured at the time point at which the plant had reached its maximal leafy biomass.

Increase in root biomass is expressed as an increase in total root biomass (measured as
20 maximum biomass of roots observed during the lifespan of a plant); or as an increase in the root/shoot index, measured as the ratio between root mass and shoot mass in the period of active growth of root and shoot. In other words, the root/shoot index is defined as the ratio of the rapidity of root growth to the rapidity of shoot growth in the period of active growth of root and shoot. Root biomass can be determined using a method as described in WO
25 2006/029987.

A robust indication of the height of the plant is the measurement of the location of the centre of gravity, i.e. determining the height (in mm) of the gravity centre of the leafy biomass. This avoids influence by a single erect leaf, based on the asymptote of curve fitting or, if the fit is
30 not satisfactory, based on the absolute maximum.

Parameters related to development time

The early vigour is the plant aboveground area three weeks post-germination. Early vigour was determined by counting the total number of pixels from aboveground plant parts
35 discriminated from the background. This value was averaged for the pictures taken on the same time point from different angles and was converted to a physical surface value expressed in square mm by calibration.

AreaEmer is an indication of quick early development when this value is decreased
40 compared to control plants. It is the ratio (expressed in %) between the time a plant needs to make 30 % of the final biomass and the time needs to make 90 % of its final biomass.

The “time to flower” or “flowering time” of the plant can be determined using the method as described in WO 2007/093444.

Seed-related parameter measurements

- 5 The mature primary panicles were harvested, counted, bagged, barcode-labelled and then dried for three days in an oven at 37°C. The panicles were then threshed and all the seeds were collected and counted. The seeds are usually covered by a dry outer covering, the husk. The filled husks (herein also named filled florets) were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining
- 10 fraction was counted again. The filled husks were weighed on an analytical balance. The total number of seeds was determined by counting the number of filled husks that remained after the separation step. The total seed weight was measured by weighing all filled husks harvested from a plant.
- The total number of seeds (or florets) per plant was determined by counting the number of
- 15 husks (whether filled or not) harvested from a plant. Thousand Kernel Weight (TKW) is extrapolated from the number of seeds counted and their total weight.
- The Harvest Index (HI) in the present invention is defined as the ratio between the total seed weight and the above ground area (mm²), multiplied by a factor 10⁶.
- 20 The number of flowers per panicle as defined in the present invention is the ratio between the total number of seeds over the number of mature primary panicles.
- The “seed fill rate” or “seed filling rate” as defined in the present invention is the proportion (expressed as a %) of the number of filled seeds (i.e. florets containing seeds) over the total number of seeds (i.e. total number of florets). In other words, the seed filling rate is the
- 25 percentage of florets that are filled with seed.

Example 10: Results of the phenotypic evaluation of the transgenic plants

1. ELM2-related polypeptides

- 30 The results of the evaluation of transgenic rice plants in the T2 generation and expressing a nucleic acid encoding the ELM2-related polypeptide of SEQ ID NO: 2 under non-stress conditions are presented below in Table D1. When grown under non-stress conditions, an increase of at least 5 % was observed for seed yield, including total weight of seeds (totalwgseeds), number of seeds (nrfilledseed), fill rate and harvest index. In addition, 2
- 35 lines of plants expressing the *ELM2-related* nucleic acid of SEQ ID NO: 1 showed an increased root/shoot index and 3 lines showed an increase in Thousand Kernel Weight (TKW).

Table D1: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for T2 generation plants, for each parameter the p-value is <0.05.

Parameter	Overall increase
totalwgseeds	25.1
fillrate	22.9

harvestindex	23.1
nrfilledseed	21.6

The results of the evaluation of transgenic rice plants in the T2 generation and expressing a nucleic acid encoding the ELM2-related polypeptide of SEQ ID NO: 2 under nitrogen deficiency stress conditions are presented below in Table D2. When grown under nitrogen deficiency stress conditions, an increase of at least 5 % was observed for plant aboveground area or leafy biomass (AreaMax) and for seed yield, including total weight of seeds (totalwgseeds), number of florets (nrtotalseed), fill rate and number of seeds (nrfilledseed) In addition, 1 line of plants expressing the *ELM2-related* nucleic acid of SEQ ID NO: 1 under nitrogen deficient stress conditions showed an increased total root biomass and 3 lines showed an increase in the average amount of florets per panicle on a plant.

Table D2: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for T2 generation plants, for each parameter the p-value is <0.05.

Parameter	Overall increase
AreaMax	9.6
totalwgseeds	16.7
nrtotalseed	7.5
fillrate	9.4
nrfilledseed	17.7

The results of the evaluation of transgenic rice plants in the T2 generation and expressing a nucleic acid encoding the ELM2-related polypeptide of SEQ ID NO: 4 under non-stress conditions are presented below in Table D3. When grown under non-stress conditions, an increase of at least 5 % was observed for seed yield, including total weight of seeds (totalwgseeds), number of seeds (nrfilledseed), fill rate and harvest index. In addition, 1 line of plants expressing an *ELM2-related* nucleic acid of SEQ ID NO: 3 showed an increased total number of seeds.

Table D3: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for T2 generation plants, for each parameter the p-value is <0.05.

Parameter	Overall increase
totalwgseeds	70.6
fillrate	55.4
harvestindex	60.5
nrfilledseed	67.0

2. WRKY-related polypeptides

The results of the evaluation of transgenic rice plants expressing a WRKY-related polypeptidenucleic acid under non-stress conditions are presented hereunder. An increase was observed for emergence vigour (early vigour), total seed yield (Totalwgseeds), fillrate, TKW, number of filled seeds, taller more erect plants (HeightMax), amount of thick roots (RootThickMax) (Table D4).

In addition, one or more lines showed increased number of florets per panicle on a plant (flowerperpan), increased harvestindex (HI), increased greenness of a plant before flowering (GNbfFlow), increased height of the plant (GravityYMax), increased quick early
5 development (AreaEmer), increased Cycle Time (AreaCycl).

Table D4: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for the confirmation (T1 generation), for each parameter the p-value is <0.05.

Parameter	Overall increase
EmerVigor	13.1
totalwgseeds	16.3
nrfilledseed	12.2
fillrate	8.4
TKW	3.9
HeightMax	5.1
RootThickMax	7.1

10

3. EMG1-like polypeptides

The results of the evaluation of transgenic rice plants in the T1 generation and expressing a nucleic acid encoding the EMG1-like polypeptide of SEQ ID NO: 179 under non-stress conditions are presented below in Table D5. When grown under non-stress conditions,
15 increase of at least 5 % was observed for total seed yield (Totalwgseeds), seed fill rate (fillrate), harvest index (harvestindex). In addition 3 lines showed a positive trend for thousand kernel weight. See previous Examples for details on the generations of the transgenic plants.

20 **Table D5:** Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for T1 generation plants, for each parameter the p-value is <0.05.

Parameter	Overall increase
totalwgseeds	13.3
fillrate	14.5
harvestindex	15.5

4. GPx-related polypeptides

10.1 The results of the evaluation of transgenic rice plants in the T2 generation and
25 expressing a nucleic acid encoding the GPx-related polypeptide of SEQ ID NO: 293 under non-stress conditions are presented below in Table D6. When grown under non-stress conditions, an increase of at least 5% was observed for seed yield, particularly total weight of seeds, number of seeds, fill rate, harvest index. In addition, plants expressing a GPx-related nucleic acid of SEQ ID NO: 292 showed more number of flowers per panicle
30 (Flowerperpan), increased Thousand Kernel Weight (TKW), increased height of the plant

(HeightMax), increased robust indication of the height of the plant (GravityYMax).

Table D6: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for the confirmation (T1 generation), for each parameter the p-value is <0.05.

Parameter	Overall increase
totalwgseeds	26.6
nrfilledseed	23.2
Fillrate	21.4
harvestindex	25.8

10.2 The results of the evaluation of transgenic rice plants in the T2 generation and expressing a nucleic acid encoding the GPx-related polypeptide of SEQ ID NO: 295 under non-stress conditions are presented below in Table D7. When grown under non-stress conditions, an increase of at least 5% was observed for more number of flowers per panicle (Flowerperpan), increased height of the plant (HeightMax), increased robust indication of the height of the plant (GravityYMax), quick early development (AreaEmer), seed yield, particularly total weight of seeds, number of seeds, fill rate, harvest index. In addition, plants expressing a GPx-related nucleic acid of SEQ ID NO: 294 showed a faster growth rate (an earlier start of flowering (TimetoFlower: time (in days) between sowing and the emergence of the first panicle), increased greenness of a plant before flowering (GNbfFlow), a faster growth rate (a shorter time (in days) needed between sowing and the day the plant reaches 90 % of its final biomass (AreaCycl).

Table D7: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for T1 generation plants, for each parameter the p-value is <0.05.

Parameter	Overall increase
totalwgseeds	30.3
nrfilledseed	25.7
fillrate	36.9
harvestindex	31.0
flowerperpan	10.7
HeightMax	5.8
GravityYMax	6.1
AreaEmer	5.6

10.3 The results of the evaluation of transgenic rice plants in the T2 generation and expressing a nucleic acid encoding the GPx-related polypeptide of SEQ ID NO: 297 under non-stress conditions are presented in the following. When grown under non-stress conditions, an increase of at least 5% was observed for increased Thousand Kernel Weight (TKW), increased height of the plant (HeightMax), increased robust indication of the height of the plant (GravityYMax), a faster growth rate (a shorter time (in days) needed between sowing and the day the plant reaches 90 % of its final biomass (AreaCycl).

Claims

1. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an ELM2-related poly-peptide, wherein said ELM2-related polypeptide comprises an Interpro accession IPR 000949, according to PFAM accession number PF01448 ELM2 domain.
2. Method according to claim 1, wherein said ELM2-related polypeptide comprises one or more motifs having at least 80% sequence identity with one or more of the following motifs:
 - (i) Motif 1: [IV]GKGR[SQ]DSC[GR]CQV[QP][KG]S[IV][KE]CVRFH[IV][TA]E[RK][SR][SAL][RK][VL][MK][RL]E[LI]G[KVS]AF[NY][QHA]W[RGN][FL]D[KR][MA]GEE represented by SEQ ID NO: 93,
 - (ii) Motif 2: [RTK]xFP[SK][RK][RSG]R[EK][DSE]LVSY[YF]NVFLL[QR]RR[AG][NY]QNR[SHV]TP[DNK][SN]I represented by SEQ ID NO: 94,
 - (iii) Motif 3: [PS][IP][TR]xxIP[VL]GP[VN][FH]QAE[VI]PEWT represented by SEQ ID NO: 95
 wherein x can be any amino acid.
3. Method according to claim 1 or 2, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said ELM2-related polypeptide.
4. Method according to any one of claims 1 to 3, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.
5. Method according to any one of claims 1 to 4, wherein said enhanced yield-related traits are obtained under non-stress conditions.
6. Method according to any one of claims 1 to 4, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.
7. Method according to any one of claims 1 to 6, wherein said nucleic acid encoding an ELM2-related polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Salicaceae, more preferably from the genus *Populus*, most preferably from *Populus trichocarpa*.
8. Method according to any one of claims 1 to 6, wherein said nucleic acid encoding an ELM2-related polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Fabaceae, more preferably from the genus *Medicago*, most preferably from *Medicago truncatula*.
9. Method according to any one of claims 1 to 8, wherein said nucleic acid encoding an ELM2-related encodes any one of the polypeptides listed in Table A1 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
10. Method according to any one of claims 1 to 9, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A1.
11. Method according to any one of claims 1 to 10, wherein said nucleic acid encodes the

polypeptide represented by any one of SEQ ID NO: 2 or SEQ ID NO: 4 or a homologue thereof which has at least 90% overall sequence identity to SEQ ID NO : 2 or SEQ ID NO: 4.

12. Method according to any one of claims 1 to 11, wherein said nucleic acid is operably
5 linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
13. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method ac-
10 cording to any one of claims 1 to 12, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an ELM2-related polypeptide as defined in any of claims 1 and 6 to 11.
14. Construct comprising:
 - (i) nucleic acid encoding an ELM2-related polypeptide as defined in any of claims 1 and 6 to 11;
 - 15 (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - (iii) a transcription termination sequence.
15. Construct according to claim 14, wherein one of said control sequences is a
20 constitutive promoter, preferably a medium strength constitutive promoter, preferably to a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
16. Use of a construct according to claim 14 or 15 in a method for making plants having
25 enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
17. Plant, plant part or plant cell transformed with a construct according to claim 14 or 15.
18. Method for the production of a transgenic plant having enhanced yield-related traits
30 relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants, comprising:
 - (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an ELM2-related polypeptide as defined in any of claims 1 and 6 to 11; and
 - (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.
- 35 19. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding an ELM2-related polypeptide as defined in any of claims 1 and 6 to 11 or a transgenic plant cell derived from said transgenic plant.
- 40 20. Transgenic plant according to claim 13, 17 or 19, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat,

barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.

21. Harvestable parts of a plant according to claim 20, wherein said harvestable parts are preferably shoot biomass and/or seeds.
22. Products derived from a plant according to claim 20 and/or from harvestable parts of a plant according to claim 21.
23. Use of a nucleic acid encoding an ELM2-related polypeptide as defined in any of claims 1 and 6 to 11 for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.
24. A method for manufacturing a product comprising the steps of growing the plants according to claim 13, 17, 20 or 21 and producing said product from or by said plants; or parts thereof, including seeds.
25. Construct according to claim 14 or 15 comprised in a plant cell.
26. Recombinant chromosomal DNA comprising the construct according to claim 14 or 15.
27. A method of growing a transgenic plant comprising:
 - a) planting a transformed seed comprising a nucleic acid sequence as defined in any of claims 1 and 6 to 11;
 - b) growing a plant from said seed.
28. A method for enhancing yield-related traits in plants relative to or compared to control plants, comprising modulating expression in a plant of a nucleic acid encoding a GPx-related polypeptide, wherein said GPx-related polypeptide comprises one or more motifs and/or signatures having at least 80% sequence identity with a motif and/or signature selected from a group consisting of motifs 7 to 18,
 - (i) Motif 7 represented by SEQ ID NO: 361,
 - (ii) Motif 8 represented by SEQ ID NO: 362,
 - (iii) Motif 9 represented by SEQ ID NO: 363,
 - (iv) Motif 10 represented by SEQ ID NO: 364,
 - (v) Motif 11 represented by SEQ ID NO: 365,
 - (vi) Motif 12 represented by SEQ ID NO: 366,
 - (vii) Motif 13 represented by SEQ ID NO: 367,
 - (viii) Motif 14 represented by SEQ ID NO: 368,
 - (ix) Motif 15 represented by SEQ ID NO: 369,
 - (x) Motif 16 represented by SEQ ID NO: 370,
 - (xi) Motif 17 represented by SEQ ID NO: 371,
 - (xii) Motif 18 represented by SEQ ID NO: 372;and/or one or more signatures 1 to 6,
 - (i) Signature 1 represented by SEQ ID NO: 373,
 - (ii) Signature 2 represented by SEQ ID NO: 374,
 - (iii) Signature 3 represented by SEQ ID NO: 375,
 - (iv) Signature 4 represented by SEQ ID NO: 376,
 - (v) Signature 5 represented by SEQ ID NO: 377,

(vi) Signature 6 represented by SEQ ID NO: 378.

29. Method according to claim 28, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said GPx-related polypeptide.

5 30. Method according to claim 28 or 29, wherein said polypeptide is encoded by a nucleic acid molecule comprising a nucleic acid molecule selected from the group consisting of:

(i) a nucleic acid represented by any one of SEQ ID NO: 292, 294, 296;

10 (ii) the complement of a nucleic acid represented by any one of SEQ ID NO: 292, 294, 296;

15 (iii) a nucleic acid encoding the polypeptide as represented by any one of SEQ ID NO: 293, 295, 297, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be deduced from a polypeptide sequence as represented by any one of SEQ ID NO: 293, 295, 297 and further preferably confers enhanced yield-related traits relative to or compared to control plants;

20 (iv) a nucleic acid having, in increasing order of preference at least 30 %, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with any of the nucleic acid sequences of SEQ ID NO: 292, 294, 296 and further preferably conferring enhanced yield-related traits relative to or compared to control plants, wherein the nucleic acid encodes a polypeptide that is not the polypeptide of any of the polypeptide sequence as represented by any one of SEQ ID NO: 293, 295, 297;

25 (v) a first nucleic acid molecule which hybridizes with a second nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to or compared to control plants, wherein the first nucleic acid encodes a polypeptide that is not the polypeptide of any of the polypeptide sequence as represented by any one of SEQ ID NO: 293, 295, 297;

30 (vi) a nucleic acid encoding said polypeptide having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by any one of SEQ ID NO: 293, 295, 297 and preferably conferring enhanced yield-related traits relative to or compared to control plants; or

(vii) a nucleic acid comprising any combination(s) of features of (i) to (vi) above.

31. Method to any one of claims 28 to 30, wherein said enhanced seed yield traits relative

or compared to control plants comprise one or more of increased seed yield, increased seed size, increased total weight of seeds, increased number of total seeds, increased fill rate, increased number of filled seeds, number of florets per panicle on a plant, increased height of the plant, increased quick early development, harvest index, increased cycle time relative to or compared to control plants.

32. Method according to any one of claims 28 to 31, wherein said increase in seed yield comprises an increase of at least 5 % in said plant when compared to control plants for at least one of said parameters.

33. Method according to any one of claims 28 to 32, wherein said enhanced yield-related traits are obtained under non-stress conditions.

34. Method according to any one of claims 28 to 33, wherein said enhanced yield-related traits are obtained under abiotic stress conditions, preferably drought stress, salt stress or nitrogen deficiency.

35. Method according to any one of claims 28 to 34, wherein said nucleic acid encoding a GPx-related polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Poaceae, more preferably from the genus *Oryza*, most preferably from *Oryza sativa*.

36. Method according to any one of claims 28 to 35, wherein said nucleic acid encoding a GPx-related polypeptide encodes any one of the polypeptides listed in Table A4 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

37. Method according to any one of claims 28 to 36, wherein said nucleic acid is operably linked to a constitutive promoter of plant origin, preferably to a medium strength constitutive promoter of plant origin, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

38. Method according to any one of claims 28 to 37, wherein said plant is a monocotyledonous plant.

39. Method according to claim 38, wherein said plant is a cereal.

40. Plant, or part thereof, or plant cell, obtainable by a method according to any one of claims 28 to 39, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding a GPx-related polypeptide as defined in any one of claims 28 to 39.

41. Construct comprising:

- (i) nucleic acid encoding an GPx-related polypeptide as defined in any one of claims 28 to 39;
- (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
- (iii) a transcription termination sequence.

42. Construct according to claim 41, wherein one of said control sequences is a constitutive promoter of plant origin, preferably to a medium strength constitutive promoter of plant origin, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

43. Use of a construct according to claim 41 or 42 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to or compared to control plants, and more preferably increased seed yield and/or increased biomass relative to or compared to control plants.
- 5 44. Plant, plant part or plant cell transformed with a construct according to claim 41 or 42.
45. Method for the production of a transgenic plant having enhanced yield-related traits relative to or compared to control plants, preferably increased yield relative to or compared to control plants, and more preferably increased seed yield and/or increased biomass relative to or compared to control plants, comprising:
 - 10 (i) introducing and expressing in a plant cell or plant a nucleic acid encoding a GPx-related polypeptide as defined in any one of claims 28 to 39; and
 - (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.
- 15 46. Transgenic plant having enhanced yield-related traits relative to or compared to control plants, preferably increased yield relative to or compared to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding a GPx-related polypeptide as defined in any one of claims 28 to 39 or a transgenic plant cell derived from said transgenic plant.
- 20 47. Transgenic plant according to claim 40, 44 or 46, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.
- 25 48. Harvestable parts of a plant according to claim 47, wherein said harvestable parts are preferably biomass and/or seeds.
49. Products derived from a plant according to claim 47 and/or from harvestable parts of a plant according to claim 48.
- 30 50. Use of a nucleic acid encoding an GPx-related polypeptide as defined in any one of claims 28 to 39 for enhancing yield-related traits in plants relative to or compared to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to or compared to control plants.
51. A method for manufacturing a product comprising the steps of growing the plants according to claim 40, 44 or 46 and producing said product from or by said plants; or parts thereof, including seeds.
- 35 52. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an EMG1-like polypeptide, wherein said EMG1-like polypeptide comprises an InterPro accession IPR005304 EMG1 domain corresponding to PFAM accession number PF03587.
- 40 53. A method according to claim 52, wherein said EMG1-like polypeptide comprises an EMG1 domain having at least 60% sequence identity with the EMG1 domain from amino acid 81 to 276 in SEQ ID NO: 179.
54. A method according to any of the claims 52 or 53, wherein said EMG1-like

polypeptide further comprises one of the motifs represented by SEQ ID NO: 286, SEQ ID NO: 287 or SEQ ID NO: 288.

55. Method according to any of the claims 52 to 54, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said EMG1-like polypeptide.

56. Method according to any of the claims 52 to 55, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.

57. Method according to any one of claims 52 to 56, wherein said enhanced yield-related traits are obtained under non-stress conditions.

58. Method according to any one of claims 52 to 56, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

59. Method according to any one of claims 52 to 58, wherein said nucleic acid encoding an EMG1-like is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Salicaceae, more preferably from the genus *Populus*, most preferably from *Populus trichocarpa*.

60. Method according to any one of claims 52 to 59, wherein said nucleic acid encoding an EMG1-like polypeptide encodes any one of the polypeptides listed in Table A3 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

61. Method according to any one of claims 52 to 60, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A3.

62. Method according to any one of claims 52 to 61, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 179.

63. Method according to any one of claims 52 to 62, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

64. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method according to any one of claims 52 to 63, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an EMG1-like polypeptide as defined in any of claims 52 to 54 and 59 to 63.

65. Construct comprising:

(i) nucleic acid encoding an EMG1-like as defined in any of claims 52 to 54 and 59 to 63;

(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally

(iii) a transcription termination sequence.

66. Construct according to claim 65, wherein one of said control sequences is a constitutive promoter, preferably a medium strength constitutive promoter, preferably to a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2

promoter from rice.

67. Use of a construct according to claim 65 or 66 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
68. Plant, plant part or plant cell transformed with a construct according to claim 65 or 66.
69. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants, comprising:
 - (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an EMG1-like polypeptide as defined in any of claims 52 to 54 and 59 to 63; and
 - (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.
70. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding an EMG1-like polypeptide as defined in any of claims 52 to 54 and 59 to 63 or a transgenic plant cell derived from said transgenic plant.
71. Transgenic plant according to claim 64, 68 or 70, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.
72. Harvestable parts of a plant according to claim 71, wherein said harvestable parts are preferably shoot biomass and/or seeds.
73. Products derived from a plant according to claim 71 and/or from harvestable parts of a plant according to claim 72.
74. Use of a nucleic acid encoding an EMG1-like polypeptide as defined in any of claims 52 to 54 and 59 to 63 for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.
75. A method for manufacturing a product comprising the steps of growing the plants according to claim 64, 66, 68 or 71 and producing said product from or by said plants; or parts thereof, including seeds.
76. A method for enhancing yield in plants relative to or compared to control plants, comprising modulating expression in a plant of a nucleic acid molecule encoding a WRKY-related polypeptide, preferably comprising one or more motifs selected from a group consisting of WRKYGQ motif, basic motif, coiled coil motif as defined herein.
77. Method according to claim 76, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid molecule encoding a WRKY-related polypeptide.
78. Method according to claim 76 or 77, wherein said polypeptide is encoded by a nucleic

acid molecule comprising a nucleic acid molecule selected from the group consisting of:

(i) a nucleic acid represented by SEQ ID NO: 102;
 (ii) the complement of a nucleic acid represented by SEQ ID NO: 102;
 5 (iii) a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 103, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be deduced from a polypeptide sequence as represented by SEQ ID NO: 103 and further preferably confers enhanced yield-related traits relative to or compared to control plants;

10 (iv) a nucleic acid having, in increasing order of preference at least 30 %, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%,
 15 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the nucleic acid sequence of SEQ ID NO: 102, and further preferably conferring enhanced yield-related traits relative to or compared to control plants, wherein the nucleic acid encodes a polypeptide that is not the polypeptide of the polypeptide sequence as represented by SEQ ID NO: 103;

20 (v) a first nucleic acid molecule which hybridizes with a second nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to or compared to control plants, wherein the first nucleic acid encodes a polypeptide that is not the polypeptide of any of the polypeptide sequence as represented by SEQ ID NO: 103;

25 (vi) a nucleic acid encoding said polypeptide having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%
 30 sequence identity to the amino acid sequence represented by SEQ ID NO: 103 and preferably conferring enhanced yield-related traits relative to or compared to control plants; or

(vii) a nucleic acid comprising any combination(s) of features of (i) to (vi) above.

79. Method according to any claim 76 to 78, wherein said enhanced yield-related traits
 35 comprise one or more of increased yield, preferably increase of emergence vigour, total seed yield, fill rate, TKW, number of filled seeds, taller more erect plants, amount of thick roots, number of florets per panicle on a plant, increased harvest index, increased greenness of a plant before flowering, increased height of the plant, increased quick early development, increased cycle time relative to or compared to
 40 control plants.

80. Method according to any one of claims 76 to 79, wherein said enhanced yield-related traits are obtained under non-stress conditions.

81. Method according to any one of claims 76 to 79, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.
- 5 82. Method according to any one of claims 76 to 81, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
- 10 83. Method according to any one of claims 76 to 82, wherein said nucleic acid molecule or said polypeptide, respectively, is of plant origin, preferably from a dicotyledonous plant, further preferably from a leguminous plant, more preferably from the genus *Medicago*, most preferably from *Medicago truncatula*.
84. Plant or part thereof, including seeds, obtainable by a method according to any one of claims 76 to 83, wherein said plant or part thereof comprises a recombinant nucleic acid encoding said polypeptide as defined in any one of claims 76 to 83.
- 15 85. Construct comprising:
(i) nucleic acid encoding said polypeptide as defined in any one of claims 76 to 83;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
(iii) a transcription termination sequence.
- 20 86. Construct according to claim 85, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
87. Use of a construct according to claim 85 or 86 in a method for making plants having increased yield, particularly seed yield and/or shoot biomass relative to or compared to control plants.
- 25 88. Plant, plant part or plant cell transformed with a construct according to claim 85 or 86 or obtainable by a method according to any one of claims 76 to 83, wherein said plant or part thereof comprises a recombinant nucleic acid encoding said polypeptide as defined in any one of claims 76 to 83.
- 30 89. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to or compared to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding said polypeptide as defined in any one of claims 76 to 83; and
(ii) cultivating the plant cell under conditions promoting plant growth and
35 development.
90. Plant having increased yield, particularly increased biomass and/or increased seed yield, relative to or compared to control plants, resulting from modulated expression of a nucleic acid encoding said polypeptide, or a transgenic plant cell originating from or being part of said transgenic plant.
- 40 91. A method for the production of a product comprising the steps of growing the plants of the invention and producing said product from or by
(i) the plants of the invention; or

(ii) parts, including seeds, of these plants.

- 5 92. Plant according to claim 84, 88 or 90 or a transgenic plant cell originating thereof, or a method according to claim 91, wherein said plant is a crop plant, preferably a dicot such as sugar beet, alfalfa, trefoil, chicory, carrot, cassava, cotton, soybean, canola or a monocot, such as sugarcane, or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum emmer, spelt, secale, einkorn, teff, milo and oats.
93. Harvestable parts of a plant according to claim 84, 88 or 90 wherein said harvestable parts are preferably shoot and/or root biomass and/or seeds.
- 10 94. Products produced from a plant according to claim 84, 88 or 90 and/or from harvestable parts of a plant according to claim 92.
95. Use of a nucleic acid encoding a polypeptide as defined in any one of claims 76 to 83 in increasing yield, particularly seed yield and/or shoot biomass relative to or compared to control plants.
96. Construct according to claim 85 or 86 comprised in a plant cell.
- 15 97. Recombinant chromosomal DNA comprising the construct according to claim 85 or 86.

1/70

MMDGSFSNEDVKGKDV DENLISDSRKSEKVENED EVKSVVVVETDGDEEGNKGDNS
 EIEELDLATFEESVSSRKRKRDSIYRMLN WVTEIAKDPCDPVVGSLPEWSKWKSYG
 NEECWKQVLVTRESLFLKRNADSTSSAEGFVKQKYQKMHP CMYDDHAGSSYNLRER
 LKCRKLFPGETSSQALVCSQSSSAATQTDS DSCMEGVYDGDSSTECSVLDL **PITKR**
IPVGPVFQAEVPEWTGMASESDSKWLGT RVWPLETSIN **KFVIEREP** **IGKGRSDSCG**
Motif 3 ELM2 domain **Motif 1**
CQVPKSIKCVRFHITERRSRVMRELGKA FNQWRFDKMGEEVKLSWSLEDQKKFGAI
Motif 1
 VRSNPLSLDKCFWDEIF **RCFPSRRREDLVSYYYNVFLLQRRANQNRSTPDS**INSDD
Motif 2
 DESECELVNYGSRREAVNSPGSLISAKKQHKYVK*

FIGURE 1

MLGSEQTL DLYKLFMVVKDKGGYDVVCKNELWDLVGEEYGLGVKVGSSVKLVYSKY
 LSGLETPLKNVVDGEFPKRDLVGDRVKFGERLTELQAEVLDDYGE GDEVKSV
 YGCRKKLCDTNRVKVVP ELNASELEKVY EYIDGRKSCGTNKMKDTNLVSNMAKKV
 ESEGLVDVLMQDCKTNETSLRKLVNQNAAMEIMDDVSDVANSMPGLSDGSKSCAND
 DANDSGDDVLMLDPSSVNRESFGRKRKRDDL MSEMQSWVIRTAKNPCDPVLGSMPE
 KSKWKSYGNQEIWKKVLLFREAAFLKKDFGSDCEKLSWLAQRMHPSMYDVNLGVNY
 NLRQRIKRDNGVLVGKSASILQRTPDSEAKEKLLDSCVPESILD A**PPTVN** **IPLGPN**
Motif 3
HQAEVPEWTGTTHKSDSKWLGTQIWPLQIVKSKLLEGE P**VGKGRQDSCRCQVQGSV**
Motif 3 ELM2 domain **Motif 1**
ECVRFHIAEKSAKLEIGVAFYQWNLDKAGEEVRRCWTAEEEEKKFKDVVKSNPAS
Motif 1
 LDRCFWDDIF **TTFPKKSRESLVSYFFNVFLLQRRGYQNRHTPDNI**DSDDDEESEFTP
Motif 2
 LKGSFGHQTDKSNSFTLLSPKKPQPTPHTKGRQLVNGLKS*

FIGURE 2

2/70

	1	50
A.thaliana_AT2G46040.1#1	~~~~~	
A.thaliana_AT4G11400.1#1	~~~~~	
G.hirsutum_ES795168#1	~~~~~	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	~~~~~	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	~~~~~	
G.max_Glyma18g46950.1#1	~~~~~	
M.truncatula_AC149204_12.4#1	~~~~~	
P.trichocarpa_572279#1	~~~~~	
Poptr_ELM2-related	~~~~~	
P.trichocarpa_751302#1	~~~~~	
P.trichocarpa_757535#1	~~~~~	
V.vinifera_GSVIVT00002575001#1	~~~~~	
O.sativa_LOC_Os04g08410.1#1	~~~~~	
P.virgatum_TC11466#1	~~~~~	
S.bicolor_Sb06g001680.1#1	~~~~~	
Zea_mays_GRMZM2G424651_T01#1	~~~~~	
G.max_Glyma08g01950.1#1	~~~~~	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1	~~~~~	
Zea_mays_GRMZM2G365731_T01#1	MAGAPDPSMASALEAQSKSMEAQSRTLATQSQLLQQLCDRLESQDHRWSQ	
	51	100
A.thaliana_AT2G46040.1#1	~~~~~	
A.thaliana_AT4G11400.1#1	~~~~~	
G.hirsutum_ES795168#1	~~~~~	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	~~~~~	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	~~~~~	
G.max_Glyma18g46950.1#1	~~~~~	
M.truncatula_AC149204_12.4#1	~~~~~	
P.trichocarpa_572279#1	~~~~~	
Poptr_ELM2-related	~~~~~	
P.trichocarpa_751302#1	~~~~~	
P.trichocarpa_757535#1	~~~~~	
V.vinifera_GSVIVT00002575001#1	~~~~~	
O.sativa_LOC_Os04g08410.1#1	~~~~~	
P.virgatum_TC11466#1	~~~~~	
S.bicolor_Sb06g001680.1#1	~~~~~	
Zea_mays_GRMZM2G424651_T01#1	~~~~~	
G.max_Glyma08g01950.1#1	~~~~~	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1	~~~~~	
Zea_mays_GRMZM2G365731_T01#1	MEDAVKMNTTNISKLQEKLGDGEIEAFRTDLTRSLGDRVEAQIAQLEAAT	
	101	150
A.thaliana_AT2G46040.1#1	~~~~~	
A.thaliana_AT4G11400.1#1	~~~~~	
G.hirsutum_ES795168#1	~~~~~	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	~~~~~	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	~~~~~	
G.max_Glyma18g46950.1#1	~~~~~	
M.truncatula_AC149204_12.4#1	~~~~~	
P.trichocarpa_572279#1	~~~~~	
Poptr_ELM2-related	~~~~~	
P.trichocarpa_751302#1	~~~~~	
P.trichocarpa_757535#1	~~~~~	
V.vinifera_GSVIVT00002575001#1	~~~~~	
O.sativa_LOC_Os04g08410.1#1	~~~~~	
P.virgatum_TC11466#1	~~~~~	
S.bicolor_Sb06g001680.1#1	~~~~~	
Zea_mays_GRMZM2G424651_T01#1	~~~~~	
G.max_Glyma08g01950.1#1	~~~~~	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1	~~~~~	
Zea_mays_GRMZM2G365731_T01#1	LGRVDGLVTETSNRVAVLENATSAFESWRPRVEHSVDGIRNSVDTLRSDL	

FIGURE 3

3/70

A.thaliana_AT2G46040.1#1	151	200
A.thaliana_AT4G11400.1#1	~~~~~	~~~~~
G.hirsutum_ES795168#1	~~~~~	~~~~~
G.max_Glyma07g07530.1#1	~~~~~	~~~~~
P.persica_TC15862#1	~~~~~	~~~~~
V.vinifera_GSVIVT00026942001#1	~~~~~	~~~~~
G.max_Glyma09g39360.1#1	~~~~~	~~~~~
G.max_Glyma18g46950.1#1	~~~~~	~~~~~
M.truncatula_AC149204_12.4#1	~~~~~	~~~~~
P.trichocarpa_572279#1	~~~~~	~~~~~
Poptr_ELM2-related	~~~~~	~~~~~
P.trichocarpa_751302#1	~~~~~	~~~~~
P.trichocarpa_757535#1	~~~~~	~~~~~
V.vinifera_GSVIVT00002575001#1	~~~~~	~~~~~
O.sativa_LOC_Os04g08410.1#1	~~~~~	~~~~~
P.virgatum_TC11466#1	~~~~~	~~~~~
S.bicolor_Sb06g001680.1#1	~~~~~	~~~~~
Zea_mays_GRMZM2G424651_T01#1	~~~~~	~~~~~
G.max_Glyma08g01950.1#1	~~~~~	~~~~~
O.sativa_LOC_Os10g30944.2#1	~~~~~	~~~~~
S.bicolor_Sb01g020470.1#1	~~~~~	~~~~~
Zea_mays_GRMZM2G365731_T01#1	ARLTQFLERNNNTTGDHHGQDSILGPYVPAGEHTSASPTCTGRPDGHGTA	
A.thaliana_AT2G46040.1#1	201	250
A.thaliana_AT4G11400.1#1	~~~~~	~~~~~
G.hirsutum_ES795168#1	~~~~~	~~~~~
G.max_Glyma07g07530.1#1	~~~~~	~~~~~
P.persica_TC15862#1	~~~~~	~~~~~
V.vinifera_GSVIVT00026942001#1	~~~~~	~~~~~
G.max_Glyma09g39360.1#1	~~~~~	~~~~~
G.max_Glyma18g46950.1#1	~~~~~	~~~~~
M.truncatula_AC149204_12.4#1	~~~~~	~~~~~
P.trichocarpa_572279#1	~~~~~	~~~~~
Poptr_ELM2-related	~~~~~	~~~~~
P.trichocarpa_751302#1	~~~~~	~~~~~
P.trichocarpa_757535#1	~~~~~	~~~~~
V.vinifera_GSVIVT00002575001#1	~~~~~	~~~~~
O.sativa_LOC_Os04g08410.1#1	~~~~~	~~~~~
P.virgatum_TC11466#1	~~~~~	~~~~~
S.bicolor_Sb06g001680.1#1	~~~~~	~~~~~
Zea_mays_GRMZM2G424651_T01#1	~~~~~	~~~~~
G.max_Glyma08g01950.1#1	~~~~~	~~~~~
O.sativa_LOC_Os10g30944.2#1	~~~~~	~~~~~
S.bicolor_Sb01g020470.1#1	~~~~~	~~~~~
Zea_mays_GRMZM2G365731_T01#1	TTNREYGF GKPHPIQLPINGQKFKNKEQIHSP IPTQHTKMAGNPAPPLPL	~MAGSPAPPPPS
A.thaliana_AT2G46040.1#1	251	300
A.thaliana_AT4G11400.1#1	~~~~~	~~~~~
G.hirsutum_ES795168#1	~~~~~	~~~~~
G.max_Glyma07g07530.1#1	~~~~~	~~~~~
P.persica_TC15862#1	~~~~~	~~~~~
V.vinifera_GSVIVT00026942001#1	~~~~~	~~~~~
G.max_Glyma09g39360.1#1	~~~~~	~~~~~
G.max_Glyma18g46950.1#1	~~~~~	~~~~~
M.truncatula_AC149204_12.4#1	~~~~~	~~~~~
P.trichocarpa_572279#1	~~~~~	~~~~~
Poptr_ELM2-related	~~~~~	~~~~~
P.trichocarpa_751302#1	~~~~~	~~~~~
P.trichocarpa_757535#1	~~~~~	~~~~~
V.vinifera_GSVIVT00002575001#1	~~~~~	~~~~~
O.sativa_LOC_Os04g08410.1#1	~~~~~	~~~~~
P.virgatum_TC11466#1	~~~~~	~~~~~
S.bicolor_Sb06g001680.1#1	~~~~~	~~~~~
Zea_mays_GRMZM2G424651_T01#1	~~~~~	~~~~~
G.max_Glyma08g01950.1#1	~~~~~	~~~~~
O.sativa_LOC_Os10g30944.2#1	~~~~~	~~~~~
S.bicolor_Sb01g020470.1#1	~~~~~	~~~~~
Zea_mays_GRMZM2G365731_T01#1	KSNSSAPAFKSNASGSPSPGYEPAQSSSTSPAAAAQVPVPVPSDALVAVF	~~~~~
	KSSSSAPAFKSNASGSSRGHEPAKASSTSPAAAE..LPVP I PSDSLVAVL	

FIGURE 3 (continued)

4/70

A.thaliana_AT2G46040.1#1	301	350
A.thaliana_AT4G11400.1#1	LDANRSPESVNPESTGFEEKIKELISLFRPLLD SFLAEF.CSADGF	
G.hirsutum_ES795168#1	~~MGSFNDTSCSYVDVEIKYVDECEERLRRLFDQALLVFLEEE....GSI	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	DSGSIVGCSNDGSQR.RFEELVSEFTSEI.CGSDYF	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	SKGIVVGGGGISELRCSFDQFLGPFLKQI.RGHNSY	
G.max_Glyma18g46950.1#1	~~~~~	
M.truncatula_AC149204_12.4#1	~~~~~	
P.trichocarpa_572279#1	~~~~~	
Poptr_ELM2-related		
P.trichocarpa_751302#1	HHVKDDEAKDWIAIDNSDEDDDYECEDKVICLFDDELVSKFLNET.AEGGCI	
P.trichocarpa_757535#1	HPMK.DEIKECIVVDDSDEDDCDCEDKVVCLFDEVVSIFLNEA.ADGGCI	
V.vinifera_GSVIVT00002575001#1	..TGGSCETDGICVDIGNNNSLNGTDKLQVLF DQVLLIFMKEV.LGNECV	
O.sativa_LOC_Os04g08410.1#1	ILSILSKLQSLGFCADLRIGDAAA SGPPSDLFDTVLA AFLREVYPGGREV	
P.virgatum_TC11466#1	~~~~~	
S.bicolor_Sb06g001680.1#1	VLAILSKLQSLGFCADLRIPDAAA ASDPSEAFDAVLATFLREAYTGGREA	
Zea_mays_GRMZM2G424651_T01#1	VLAILSKLQSLGFCADLRVPDAVA ASDPSEAFDAVLATFLREAYTGGREA	
G.max_Glyma08g01950.1#1	FPGNGCGLGKHHVDNGADNGKLKQKYL FYQLFPAYLKDS.CSKGGV	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1	SVVGELHARGFLAGFEIPKA.EVTEEGASALFDGVLTAF LAELAGHPAVA	
Zea_mays_GRMZM2G365731_T01#1	GVVGELSARGLLQGFTIPEAGGLTDEGASAVFDGMLTAF LAEHGVDPAAA	
A.thaliana_AT2G46040.1#1	351	400
A.thaliana_AT4G11400.1#1	LPL...PAMTGEGRTVDLFNLFLNVTHKGGF DAVSEN.GSWDEVVQESGL	
G.hirsutum_ES795168#1	KPL...PAVIGDGKNVDL FKLFLVLREREGFDTVSRK.RLWEVVAEKLG F	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	LPF...PPMLGDGKPV DLYKLFLVVRGKGGYDAVCNC.KLWDIVGEESGL	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	RPL...PPMLGNGQCVDL FKLFLLVKEKGGYRTVSEN.VLWNLVAEESGL	
G.max_Glyma18g46950.1#1	~~~~~MMSNGDPLDLYKLFMVVKEKGGYDAVC KN.KLWDLVGEEYGL	
M.truncatula_AC149204_12.4#1	~~~~~MMSNGDPLDLYKLFMVVKEKGGYDAVC KN.KLWDLVGEEYGL	
P.trichocarpa_572279#1	~~~~~MLGSEQTLDLYKLFMVVKDKGGYDVVCK N.ELWDLVGEEYGL	
Poptr_ELM2-related	~~~~~MDGSFSNED.....VKGK.....	
P.trichocarpa_751302#1	MDGSFSNED.....VKGK.....	
P.trichocarpa_757535#1	RPI...PALLDGQSLDL FKLFWVVRKRGGFDLV..N.GFWSFVAKELGL	
V.vinifera_GSVIVT00002575001#1	RPI...PALLDGQSLDL FKLFWVVRKRGGFDLV..N.GFWSFVAKELGL	
O.sativa_LOC_Os04g08410.1#1	RPI...PAMLDGGRSVDL FKLFWVVRGKGGYEWVSDK.GLWGLVAEECGL	
P.virgatum_TC11466#1	RPL...PAKLGDGSRVDL LRLFSAIRAAGGYAATSSSPAVWASAAESVCL	
S.bicolor_Sb06g001680.1#1	~~~~~	
Zea_mays_GRMZM2G424651_T01#1	RPI...PVALGDGRCDL LRLFLAVRAAGGYAGVPNSPGGWAAAAAESAGV	
G.max_Glyma08g01950.1#1	RPI...PVALNGRCDL LRLFLAVRAAGGYAGVPNTPGGWAAAAESAGI	
O.sativa_LOC_Os10g30944.2#1	RPL....PVLVDGQWLDL YKLFFLKR.....LWGSVTKELGL	
S.bicolor_Sb01g020470.1#1	~~~~~	
Zea_mays_GRMZM2G365731_T01#1	SWLPIPPPPPLGDGRKVELVSLFLAVRARGGFSAV....ASWAAVAEAVGL	
	SCLPIPPPPPLGDGKKVELVSLFLAVRARGGFSAV....ESWAAVAEAVGL	
A.thaliana_AT2G46040.1#1	401	450
A.thaliana_AT4G11400.1#1	ESYDSASAKLIYVKYLD AFGRWL...NRVVAGD TDVSSVELSGISDA...	
G.hirsutum_ES795168#1	DCSLVPSLILIIYLKYL NRMEEKWAVEESRIVNWDNKDS.....	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	GVSVGSSVKLVYSKYLS ALDAWL...KKVAESKGDVAREEVKGRVDDDDD	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	DSGVGSALKLVYIKYLD LLDRWL...DRIFKDKKSHG.....	
G.max_Glyma18g46950.1#1	GVKVGSSVELAYSKHLSS LEACL...KNVADD.....	
M.truncatula_AC149204_12.4#1	GVKVGSSAELVYSKHLSS LEACL...KNATDD.....	
P.trichocarpa_572279#1	GVKVGSSVKLVYSKYLS GLETP...KNVVDGEFPKR.....	
Poptr_ELM2-related		
P.trichocarpa_751302#1	ELQFSPSVKLIYIKYLYE LEKCM...SRSCKEEKIRK.....	
P.trichocarpa_757535#1	DLRFSPSVKLIYIKYLYE LEKCM...KICSDEKLRK.....	
V.vinifera_GSVIVT00002575001#1	DVGVKTCCLKLIYFKYLD QLDQWL...LGILRDGRLIL.....	
O.sativa_LOC_Os04g08410.1#1	DATLAAPVKLIYHKYLA ALDRWI...QRLVEAHGPFL.....	
P.virgatum_TC11466#1	~~~~~	
S.bicolor_Sb06g001680.1#1	DSALAAPVKLVYAKYLG VLDRWI...QRLVEAHGPFL.....	
Zea_mays_GRMZM2G424651_T01#1	DSALAAPVKLVYAKYLG ALDRWI...QRLVEAHGPFL.....	
G.max_Glyma08g01950.1#1	NLVVCWSVKLVYDKYLN DFDRWL...KKTFEKLNLS.....	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1	DPTADTAVKLLYGKYL ALLEQNL...E.....VV.....	
Zea_mays_GRMZM2G365731_T01#1	DPTADTAVKLLYGKYL ALLEQSC...DKPLKQDHMVV.....	

FIGURE 3 (continued)

5/70

A.thaliana_AT2G46040.1#1	451	500
A.thaliana_AT4G11400.1#1	L.....V	
G.hirsutum_ES795168#1	EKKGCYSGMLHEL.....G	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	NVDGRKSCGDGNGGCLPDL.....E	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	...SLSVCGDTSGRLLMEL.....E	
G.max_Glyma18g46950.1#1	...GRELCG.....T	
M.truncatula_AC149204_12.4#1	...GRELCD.....T	
P.trichocarpa_572279#1	...DLVGDRVKFGERLTELQAELVLDDYGEDVGDEVKSVYGCRKKLCDT	
Poptr_ELM2-related		
P.trichocarpa_751302#1	...GKRQCDGNLSCSSLEL.....E	
P.trichocarpa_757535#1	...EMHQCDGNLSFSLLEL.....E	
V.vinifera_GSVIVT00002575001#1		
O.sativa_LOC_Os04g08410.1#1		
P.virgatum_TC11466#1	~~~~~	
S.bicolor_Sb06g001680.1#1		
Zea_mays_GRMZM2G424651_T01#1		
G.max_Glyma08g01950.1#1	G	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1		
Zea_mays_GRMZM2G365731_T01#1		
A.thaliana_AT2G46040.1#1	501	550
A.thaliana_AT4G11400.1#1	ARLNGFLSEV.....KKKYELRKGRPAKELGAEL	
G.hirsutum_ES795168#1	NGFKSLLDN.....	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	VPEGDMNES.....	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	TEFKGFLPEI.....LDQKMKDEE	
G.max_Glyma18g46950.1#1	NRVKGMSPE.....SNGDKKVQSEG	
M.truncatula_AC149204_12.4#1	NRVKGVSPPE.....SNGDKKVQSKG	
P.trichocarpa_572279#1	NRVKVVNPPELNASELEKVYEYIDGRKSCGTNKMKDTNLVSNMAKKVESEG	
Poptr_ELM2-related		
P.trichocarpa_751302#1		
P.trichocarpa_757535#1	MQLRSL.....	
V.vinifera_GSVIVT00002575001#1	MQFRSL.....	
O.sativa_LOC_Os04g08410.1#1		
P.virgatum_TC11466#1		
S.bicolor_Sb06g001680.1#1	~~~~~	
Zea_mays_GRMZM2G424651_T01#1		
G.max_Glyma08g01950.1#1	NHGALLCPNL.....	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1		
Zea_mays_GRMZM2G365731_T01#1		
A.thaliana_AT2G46040.1#1	551	600
A.thaliana_AT4G11400.1#1	KWFISKTKRRYDKHHVGKESASNDAVK.....	
G.hirsutum_ES795168#1	GKCQKRNRAVAFGCNHMEESCSEFDRSRKRFRESDDDDKGVG	
G.max_Glyma07g07530.1#1	~~~~~	
P.persica_TC15862#1	...VIGSGGEGNESVNGNSGEGNDAMEVGEEFDRGIMSKEQVVDVSDGVN	
V.vinifera_GSVIVT00026942001#1	~~~~~	
G.max_Glyma09g39360.1#1	YPHFDLAKSESSFSGVENL.....YCNDEVK	
G.max_Glyma18g46950.1#1	LVDLDMLDHGTDEHIIRNLCSLNTERDILEEFDGG...KILTVDASD.AE	
M.truncatula_AC149204_12.4#1	LVDLDMLDNGTDEHIIRNLCSLNTERNILEEFDGG...KILTVDASD.AE	
P.trichocarpa_572279#1	LVDVLMQDCKTNETSLRKLNVQNAAMEIMDDVSD.....VA	
Poptr_ELM2-related	DVDE...NLISDSRKSEKVEN.....EDEVK	
P.trichocarpa_751302#1	DVDE...NLISDSRKSEKVEN.....EDEVK	
P.trichocarpa_757535#1	...LHRCDRKQEDCKFASEVYEKNGRYVEMDTGKGKIGLLDTKAVHR.VH	
V.vinifera_GSVIVT00002575001#1	...LLRCDRKQGDGKFASLGYKVNGRYIETDTGKRKIGLLGTKDVHR.IH	
O.sativa_LOC_Os04g08410.1#1	...GGTGPKKDDGVFELESERTD.....	
P.virgatum_TC11466#1		
S.bicolor_Sb06g001680.1#1	~~~~~	
Zea_mays_GRMZM2G424651_T01#1		
G.max_Glyma08g01950.1#1	KDKDDHKLVKSKSINKKKNTDLDVQR.....	
O.sativa_LOC_Os10g30944.2#1	~~~~~	
S.bicolor_Sb01g020470.1#1		
Zea_mays_GRMZM2G365731_T01#1		

FIGURE 3 (continued)

6/70

A.thaliana_AT2G46040.1#1 A.thaliana_AT4G11400.1#1 G.hirsutum_ES795168#1 G.max_Glyma07g07530.1#1 P.persica_TC15862#1 V.vinifera_GSVIVT00026942001#1 G.max_Glyma09g39360.1#1 G.max_Glyma18g46950.1#1 M.truncatula_AC149204_12.4#1 P.trichocarpa_572279#1 Poptr_ELM2-related P.trichocarpa_751302#1 P.trichocarpa_757535#1 V.vinifera_GSVIVT00002575001#1 O.sativa_LOC_Os04g08410.1#1 P.virgatum_TC11466#1 S.bicolor_Sb06g001680.1#1 Zea_mays_GRMZM2G424651_T01#1 G.max_Glyma08g01950.1#1 O.sativa_LOC_Os10g30944.2#1 S.bicolor_Sb01g020470.1#1 Zea_mays_GRMZM2G365731_T01#1	601 EFQGSKLAERRLEQIMILE.SVTQECSSPG LSSVVIREETVVC.....AVEEGLSDFS ~~~~~ YSVAGLLSGGQKCGGND.....DKGLVDSOSSDR ~~~~~ SDVKVESEGGVKC.....VDDNDEVKS SNMPRLSDGSKSCDND.....DDSDVLILDPSGVDKEIFGR SNMPRLSDGSN.....DEVLILDPSGVDKEIFGR NSMPGLSDGSKSCANDDA.....NDSGDDVLM LDPSSVNRESFGR SVVVVETDGDEEGNKGDN.....SEIEELDLATFEESVSSR SVVVVETDGDEEGNKGDN.....SEIEELDLATFEESVSSR NGV.....GNRNC.DYDKKFHAERNNYFNDSDDDVILDP SIVKKEFNSR NDV.....GDRNGDDDDEKFHGEKENYCDDSDDDVVILDP SIAKKEFNSR NCVDLDKEKSIL DGNDGRKKPEPFFDSNG..RENEEPLLECNGGDLQHPIL ~~~~~ DVDMRKRQELFFGANGVEEEEEEGLLNCNGREQRHVML DGDMRKRQELFIGTDGV..EGEEGLLGYNGREQRHVML IEGDNDEKS ~~~~~ SSGNADLLVGSKKGRFLSSPTKRPTTTTSTGSAHL SSGNADVLMGSKKERFL.SPIKCLTSTAGSAS AHL	650
A.thaliana_AT2G46040.1#1 A.thaliana_AT4G11400.1#1 G.hirsutum_ES795168#1 G.max_Glyma07g07530.1#1 P.persica_TC15862#1 V.vinifera_GSVIVT00026942001#1 G.max_Glyma09g39360.1#1 G.max_Glyma18g46950.1#1 M.truncatula_AC149204_12.4#1 P.trichocarpa_572279#1 Poptr_ELM2-related P.trichocarpa_751302#1 P.trichocarpa_757535#1 V.vinifera_GSVIVT00002575001#1 O.sativa_LOC_Os04g08410.1#1 P.virgatum_TC11466#1 S.bicolor_Sb06g001680.1#1 Zea_mays_GRMZM2G424651_T01#1 G.max_Glyma08g01950.1#1 O.sativa_LOC_Os10g30944.2#1 S.bicolor_Sb01g020470.1#1 Zea_mays_GRMZM2G365731_T01#1	651 KRKREC.PLETLKWLSDVAKDPCDPSLGIVPDRSEWVSYGSEEPWKQLLL LEKRDD.LPGMLKWLALVATSPHDP AIGVIPHSSKWKQYNGNK CWLQVAR ~~~~~ KRMREF.ALDMLS WVTGA AKNPCDPEVGS IPEKAKWKSYSNQEVWKQVLL ~~~~~MAGMED..... SRKREY.MLGMLNWVT TIAKNPCDPSIGKLPERSKWKLTGPGELWKQVLL KRKRES.KSEMLSWITSI AKNPCDPTVGS IPEKSKWKSYSSQEIWKQALL KRKRES.KSEMLSWITSI AKNPCDSTVGS IPEKSKWKIFYSSQEIWKQVLL KRKRDDL MSEMQSWVIRTAKNPCDPVLGSMPEKSKWKSYGNQEIWKQVLL KRKRDS.IYRMLNWVTEI AKDPCDPVVGSLPEWSKWKSYGNEECWKQVLV KRKRDS.IYRMLNWVTEI AKDPCDPVVGSLPEWSKWKSYGNEECWKQVLV KRKREP.LTRMLNWVIQ IAKCPDDPSIGVRSPQFKWKNHKGNELWLQAIR KRKRES.LTRMLNWVIQ IAKCPDDPSIRVI.....HRGDELWSQAIR NLKRES.FSGMLNWVREI AKHPEEPYKGN.....KNGNEVFSTQILR KRKRED.MVGMLDWVRELAENGGEAGTMA.....AGSANGYYS LALA ~~~~~ KRKRGD.VVGMLGWVREI AEKAGNGVAVA.....AGTSDLYFSMALA KRKRGD.VVGMLGWVREMAENAGNDVAVS.....AGSSAECFSMALA TNGRDA.LSGMLNWMKHVAKHALDPLTQPIPKPSKWKEYKGQDFFGQFLR ~~~~~MALM KRKRDT.LAGMLDWVRLVAKNPGK.....QGKN.PGGHKRA.VIE KRKRET.LVGMLDWVHLVAKNPGK.....QGKNRAAGHKQAVVVE	700
A.thaliana_AT2G46040.1#1 A.thaliana_AT4G11400.1#1 G.hirsutum_ES795168#1 G.max_Glyma07g07530.1#1 P.persica_TC15862#1 V.vinifera_GSVIVT00026942001#1 G.max_Glyma09g39360.1#1 G.max_Glyma18g46950.1#1 M.truncatula_AC149204_12.4#1 P.trichocarpa_572279#1 Poptr_ELM2-related P.trichocarpa_751302#1 P.trichocarpa_757535#1 V.vinifera_GSVIVT00002575001#1 O.sativa_LOC_Os04g08410.1#1 P.virgatum_TC11466#1 S.bicolor_Sb06g001680.1#1 Zea_mays_GRMZM2G424651_T01#1 G.max_Glyma08g01950.1#1 O.sativa_LOC_Os10g30944.2#1 S.bicolor_Sb01g020470.1#1 Zea_mays_GRMZM2G365731_T01#1	701 FRAS...RTNND SACE..... AKNSLLVQRD..... ~~~~~ FRESAFLKRGSDPSSE..... VREALFLQRDV DLSAE..... FREAVFLKKDFETISE..... FREAVFLKKDFETTSE..... FREAAFLKKDFGSDCE..... TRESLFLKRNADSTSS..... TRESLFLKRNADSTSS..... AREALLHRRHFDPNIE..... AREALLQRRHVNP NIE..... VRELSPKRRPFSPRTE..... ARKAVFAKRARRSSMT..... ~~~~~ VREMVTGKRARRASM..... VRKMVTTKKARRASM..... AREFAPKRVSFSIMVSL LPCDS AIMSSSHGIGLFAHMGIMHVIPI SHIQS FRRQM FVDDG FSDKPH..... LRRQM FADKDC LARSELVGMLNWVRGVAASAGQPGIIGMNP PPP.....D LRRQM FADKDC LAREELVKTLNWVRLVAKSAGHPGIIGMN.NRP.....D	750

FIGURE 3 (continued)

851	Motif 3	900
.....KRGKTGNGSDIGSSDEEDRPCA	ALVGSKFQAKVPEWT	GITPES..
GLTAGTSRARKRNK.....AEI	PRRCIKVGHQHQAQVDEWTE	SGVDS..
....ECSDFTTSSLMF.....NYNVDI	QVPIGPLFQAEVLEWT	GVALES..
.....DPSDANSGL....DRCATVRI	PLGANHQAEIPEWT	GVISEG..
....QLLGTSDFTSVL....DRYP	PSHIPGSGNYQAHLP	PEWTGEASES..
.....DSHVRKR..PVGPAFQAT	IPEWTGVVSEI..	
....QLPDSCTARSSL....DICARVPI	PLGPNHQAEVPEWT	DITYEV..
....LLPDSGTAHSSL....DRCARVHI	PLGPNHQAEVPEWT	MITYEV..
....KLLDSCVPESIL....DAPPTVNI	PLGPNHQAEVPEWT	GTTHKS..
.....DSSTECVL....DLPI	TKRIPVGPVFQAEVPEWT	GMASES..
.....DSSTECVL....DLPI	TKRIPVGPVFQAEVPEWT	GMASES..
....LAVDLSSKNATFDGSGDG	HVRRHVAVGPLFQAEVPEWT	SVVSES..
KQELVIDDLSSKHHTTLGSGDK	HVRRHVAVGPLFQAEVPEWI	GVVSES..
....NEQEPENTNTLDDLFPDDL	FLKPIPVGPNFQAEVPEWT	GEVTES..
...QAKYESRKHHNSDNWLFTS	QQRNKIPVGSEFQAQVPQWT	GELPVS.Y
.....		
...EGKHGTRTHNSSNGWLFTS	HQKNEIPVGPDYQVHVPQWT	GEVPVN.Y
...EGKHESRTSNSSNGWL	FKSHQKNEIPVGPNYQIHVPQWT	SEVPTN.Y
.....	DESLEKQVSVGPRFQAEVPEWT	GVFSES..
.....	NWNSRANRSFG...LADIPEWT	GKPSLP.Y
.....	AASPRSGIPLRPGQADIPEWT	GKPSSTYY
.....	AASPOSGISIRPCGQADIPEWT	GKRSSP.Y

ELM2 domain

8/70

A.thaliana_AT2G46040.1#1	901	Motif 1	950
A.thaliana_AT4G11400.1#1	DSKWLGT	RIWPLTKEQT	KANLLIERDR
G.hirsutum_ES795168#1	DSKWLGT	RIWPPENSEA	L.DQTLGNDL
G.max_Glyma07g07530.1#1	DAKWLGT	LIWPLEKNEYS	..FLIEGDRI
P.persica_TC15862#1	YSKFSGT	PIWPLATANTR	..LLIERDPI
V.vinifera_GSVIVT00026942001#1	ELKWLGS	KFWPLEKPEHR	..YLIERDPI
G.max_Glyma09g39360.1#1	DSKWLGT	RVWPLDKIEHK	..FLIERDPI
G.max_Glyma18g46950.1#1	DSKWLGT	QIWPLKTVNSK	.HHLFERDPI
M.truncatula_AC149204_12.4#1	DSKWLGT	QIWPPKTVNSK	.CRLFERDPI
P.trichocarpa_572279#1	DSKWLGT	QIWPLQIVKSK	...LLEGEPI
Poptr_ELM2-related	DSKWLGT	RVWPLETSINK	..FVIEREPI
P.trichocarpa_751302#1	DSKWLGT	RVWPLETSINK	..FVIEREPI
P.trichocarpa_757535#1	DSKWLGT	RLWPLECENHN	..AVFAMDPI
V.vinifera_GSVIVT00002575001#1	ASKWLGT	RLWPLECENNN	..APVAMDPI
O.sativa_LOC_Os04g08410.1#1	DSKWLGT	QVWPLENGECS	..FRIEKDCV
P.virgatum_TC11466#1	DNAETRKWL	GTKVWPLENGNRK	..LSYFCNPV
S.bicolor_Sb06g001680.1#1	~~~~~	~~~~~MKTVK	..HWFVVSLL
Zea_mays_GRMZM2G424651_T01#1	DDPETLKW	LGTKVWPPENESCK	..TLFCGDPI
G.max_Glyma08g01950.1#1	DDPETLKW	LGTKVWPPENEGCK	..TLFCGDPI
O.sativa_LOC_Os10g30944.2#1	DSKWLGT	HVWLSKNDTEP	ATATDV
S.bicolor_Sb01g020470.1#1	EDPDMLR	FLGE...PILTPKNN	..EVFD.DT
Zea_mays_GRMZM2G365731_T01#1	DDPRTLRL	FLGE...PILLPGST	..EDPDGAS
	ELM2 domain		
A.thaliana_AT2G46040.1#1	951	Motif 1	1000
A.thaliana_AT4G11400.1#1	VKFHITAK	RDKLKLELGP	AFYMWCFDVMGEC
G.hirsutum_ES795168#1	TRLHIAEK	RMELKRELGD	DDFFHWRFNQMGE
G.max_Glyma07g07530.1#1	VKFHVAEK	RLEVKHELGP	AFNQWNFDKMGE
P.persica_TC15862#1	VRFHITEK	RAKVHLELG	AVFYDWNFDQVGE
V.vinifera_GSVIVT00026942001#1	VRFHISEK	RLRVKRELGP	AFYHWEFNQMGDE
G.max_Glyma09g39360.1#1	VQFHVAEK	RIRIKLELGS	AFQHWQFQDKMGE
G.max_Glyma18g46950.1#1	VRFHIAKK	RAKVKLELGE	AFHQWNLHKVGEE
M.truncatula_AC149204_12.4#1	VRFHIAKR	RAKVKLELGE	AFHQWNLHKVGEE
P.trichocarpa_572279#1	VRFHIAEK	SAKLKLEIG	VAFYQWNLDKAGE
Poptr_ELM2-related	VRFHITER	RSRVMRELG	KAFNQWRFDKMGE
P.trichocarpa_751302#1	VRFHITER	RSRVMRELG	KAFNQWRFDKMGE
P.trichocarpa_757535#1	VRFHIAEK	RIRIKLELGY	LFYHWQFDRMGEE
V.vinifera_GSVIVT00002575001#1	VRFHIAEK	RMKLKLELGL	PLFYHWRFDHMGEE
O.sativa_LOC_Os04g08410.1#1	VRFHIAEK	RMKLKLDLGS	LFYHWRFDHMGEE
P.virgatum_TC11466#1	VRFHVAER	RLQLRRELD	SAFYAWGFDRMGEE
S.bicolor_Sb06g001680.1#1	VRFHVAEQ	RFKLKRELGV	AFYAWGFDRMGEE
Zea_mays_GRMZM2G424651_T01#1	VRFHVAER	RLKLKRELGV	AFYAWGFNRMGEE
G.max_Glyma08g01950.1#1	VRFHVAEW	RLKLKHELGA	AFYAWGFNRMGEE
O.sativa_LOC_Os10g30944.2#1	VRFHVAEK	KTELKREMG	SSYYAMKFDEIGQD
S.bicolor_Sb01g020470.1#1	VRFHVAEK	KIKLKDELGS	AYFEMGFHRIGED
Zea_mays_GRMZM2G365731_T01#1	IRFHVAEK	KIKLKDELGS	AYYEMGFHRIGED
	1001	Motif 2	1050
A.thaliana_AT2G46040.1#1	SPPSLSPA	FIHQAKMILP	SKSRGKIVSYFY
A.thaliana_AT4G11400.1#1	DPQS....	FWTNAAKNF	PKKKREELVS
G.hirsutum_ES795168#1	NPP.LGKC	FWDEIYNF	FHDKSREELV
G.max_Glyma07g07530.1#1	NPPSPETY	FWDHIFRA	FPTKSRADLV
P.persica_TC15862#1	NPPSLGIT	FWDQIFKS	FPKKSRRELV
V.vinifera_GSVIVT00026942001#1	NPPSMGVR	FWDEIFKS	FSTKSREDLV
G.max_Glyma09g39360.1#1	NPASLEK	CFWDHLEK	TFPKKSREDLV
G.max_Glyma18g46950.1#1	NPASLDK	CFWDHLEK	TFPKKSREDLV
M.truncatula_AC149204_12.4#1	NPASLDRC	FWDDIFTF	FPKKSRESLV
P.trichocarpa_572279#1	NPLSLDK	CFWDEIFR	CFPSRRREDLV
Poptr_ELM2-related	NPLSLDK	CFWDEIFR	CFPSRRREDLV
P.trichocarpa_751302#1	NLLSAGK	CFWDNKKY	FPRKTREELV
P.trichocarpa_757535#1	NPQSAGK	CFWDNKKY	FPRKTREELI
V.vinifera_GSVIVT00002575001#1	NSSLQSP	SPFDNALR	IFPTKTREALV
O.sativa_LOC_Os04g08410.1#1	NAPSSGR	NFWKRLHML	FQSKGRKELV
P.virgatum_TC11466#1	NAQSSGR	NFWNRL....	QLKGRKELV
S.bicolor_Sb06g001680.1#1	SAASSGR	NFWNRLHL	FFQFKGRKELV
Zea_mays_GRMZM2G424651_T01#1	SAASSGR	NFWNRLHL	FFQFKGRKELV
G.max_Glyma08g01950.1#1	NISSKNKY	FWNPNPSKY	FPKKTRRNLV
O.sativa_LOC_Os10g30944.2#1	NLPSSKY	NFWDKLRAA	FRYKGERALV
S.bicolor_Sb01g020470.1#1	NLPSSTN	KFWDKLLAV	LHDKGRKGLV
Zea_mays_GRMZM2G365731_T01#1	NMPASKN	KLWDKLLAV	LHAKGKEGLV

FIGURE 3 (continued)

9/70

A.thaliana_AT2G46040.1#1	1051	1100
A.thaliana_AT4G11400.1#1	IDSDD..EGAFGSVGGSFGRDAVTSSGSDVMICAQNRQC.....	
G.hirsutum_ES795168#1	ITYDDE.EVVLGPESIGKG.IGPKSYTSILIPLRKSQKKSKDALPQFSTE	
G.max_Glyma07g07530.1#1	INSDDDE.DDEAGPLRNVFGQQTQNSRGSILLTPKKSQ.....	
P.persica_TC15862#1	IDSDDDE.GLESGSVTNGFGDEERKPSKSILKSPHKPH.....	
V.vinifera_GSVIVT00026942001#1	IDSDDD.ESEFGPLSNGFGHEAINLPSSILYAQNNPH.....	
G.max_Glyma09g39360.1#1	IDSDDD.ESECTPLRKVFGHHPPKARSFTLLSPKKATFKRENNK~~~~~	
G.max_Glyma18g46950.1#1	IDSDDD.ESEFTPLRKNF.....KSPLILGSKKILHASNFETSLLSVG	
M.truncatula_AC149204_12.4#1	IDSDDDE.ESEFTPLKGSFGHQTDKSNSTLLSPKKPQPTPHTKGRQLVNG	
P.trichocarpa_572279#1	INSDDDD.ESECELVNYGSRREAVNSPGSLISAKKQHK.....	
Poptr_ELM2-related	INSDDDD.ESECELVNYGSRREAVNSPGSLISAKKQHK.....	
P.trichocarpa_751302#1	IDSDDDD.ETEFGSFSDGYGHEVLMVPGAYMLICSENKQC.....	
P.trichocarpa_757535#1	IDSDDDD.ETKFGSLSDAYGHEALTVPGANMLICSENKQC.....	
V.vinifera_GSVIVT00002575001#1	IDSDDDD.ELEFGSLGGSFGHEVIKVPWSEFLTCTQNEQS.....	
O.sativa_LOC_Os04g08410.1#1	IDSDDDEDETEFGFLGNRLGHNATKYDSSKYTLCIESTHC.....	
P.virgatum_TC11466#1	IDSDDDEEETEFRLGNRLGHGAACYHNTKHTICIENTHS.....	
S.bicolor_Sb06g001680.1#1	IDSDDDEEETEFRLGNRLGHFAAKYHSTKHTICIENTHT.....	
Zea_mays_GRMZM2G424651_T01#1	IDSDDDEEETEFRLGNRLGHRAAKYHSTKHTACIENTHS.....	
G.max_Glyma08g01950.1#1	VDSDDD.EVEFGSFSDGFGMDAVKSLGDDFLECSLNKQC.....	
O.sativa_LOC_Os10g30944.2#1	VDSDDD.SIEPAFLRLRQDGGKSRSRSS..ASSRNQNQS~~~~~	
S.bicolor_Sb01g020470.1#1	ADSDDDD.SIEPGLLYARQQRVDQSRSKSA..GSSRKRRRS~~~~~	
Zea_mays_GRMZM2G365731_T01#1	ADSDDDD.SIEPGFLYFRQGGDRSRSRSA..ASSRKRRRS~~~~~	
A.thaliana_AT2G46040.1#1	1101	1142
A.thaliana_AT4G11400.1#1	~~~~~	
G.hirsutum_ES795168#1	EDY~~~~~	
G.max_Glyma07g07530.1#1	HFDKKMEK.....WKSI~~~~~	
P.persica_TC15862#1	TKKK~~~~~	
V.vinifera_GSVIVT00026942001#1	AKCR~~~~~	
G.max_Glyma09g39360.1#1	MNFR~~~~~	
G.max_Glyma18g46950.1#1	~~~~~	
M.truncatula_AC149204_12.4#1	SFDKGYLKLKEQHMEYIEQHFVIEFHEQMQ~~~~~	
P.trichocarpa_572279#1	LKS~~~~~	
Poptr_ELM2-related	YVK~~~~~	
P.trichocarpa_751302#1	YVK~~~~~	
P.trichocarpa_757535#1	TDFK~~~~~	
V.vinifera_GSVIVT00002575001#1	TDFT~~~~~	
O.sativa_LOC_Os04g08410.1#1	TELD~~~~~	
P.virgatum_TC11466#1	MDLNQ~~~~~	
S.bicolor_Sb06g001680.1#1	MDLDD~~~~~	
Zea_mays_GRMZM2G424651_T01#1	MDLDE~~~~~	
G.max_Glyma08g01950.1#1	MDLDE~~~~~	
O.sativa_LOC_Os10g30944.2#1	TDLKYTKLECKFDNWI	
S.bicolor_Sb01g020470.1#1	~~~~~	
Zea_mays_GRMZM2G365731_T01#1	~~~~~	

FIGURE 3 (continued)

Phylogenetic tree showing the relationships between various plant species and accessions, with bootstrap values indicated at the nodes. The tree is rooted at the bottom with *P. patens* TC37758#1 and TC43927#1. The tree is divided into several major clades, including *H. petiolaris*, *T. officinale*, *A. thaliana*, *B. oleracea*, *M. domestica*, *O. sativa*, *S. bicolor*, *Z. mays*, *A. lyrata*, *P. trichocarpa*, *V. vinifera*, *G. max*, *S. tuberosum*, *L. sativa*, *T. aestivum*, *O. sativa*, *S. bicolor*, *Z. mays*, *P. virgatum*, *S. bicolor*, *Z. mays*, *A. thaliana*, *G. hirsutum*, *V. vinifera*, *P. trichocarpa*, *Poptr_ELM2-related*, *G. max*, *M. truncatula*, *G. max*, *P. persica*, *G. max*, *P. trichocarpa*, *P. trichocarpa*, *V. vinifera*, and *A. thaliana*.

FIGURE 4

11/70

	1	2	3	4	5	6	7	8	9	10	11
1. A.thaliana_AT2G46040.1#1		28,8	19,7	36,2	28,3	34,6	31,4	30,9	27,4	17,3	20
2. A.thaliana_AT4G11400.1#1	49,4		19,1	33,1	31,5	30,4	30	29,9	28,8	18,8	20,4
3. G.hirsutum_ES795168#1	29,7	30		20,5	21,1	24,5	24,7	21,4	20	27,5	40,8
4. G.max_Glyma07g07530.1#1	54,8	51,8	28,5		31	47,7	42,7	42,7	29,5	18,3	22,9
5. G.max_Glyma08g01950.1#1	46,2	52,1	30,1	50,3		31,7	30,7	29	29,2	15,9	19,7
6. G.max_Glyma09g39360.1#1	52,7	52,4	34,3	62,6	48,3		77,3	55,3	29,9	19,8	23,6
7. G.max_Glyma18g46950.1#1	50,4	50,8	36	57,5	47,9	84,7		50	29,3	17,8	24,8
8. M.truncatula_AC149204_12.4#1	49,3	47,3	30,3	60	47,8	68,3	64		27,2	17,3	22,3
9. O.sativa_LOC_Os04g08410.1#1	46,1	47,8	30,7	47,7	48,8	48,6	49,1	43,8		20,7	22,6
10. O.sativa_LOC_Os10g30944.2#1	24,9	27,9	48,8	26	24,9	29,8	29,3	25,3	29,9		33
11. P.persica_TC15862#1	26,9	26,7	55,9	28,3	27	31,5	32	28	29,6	51,9	
12. P.trichocarpa_572279#1	43,4	44,5	40,9	49,4	42,4	56,2	54,3	46,8	43,1	32,7	37,6
13. P.trichocarpa_751302#1	49,4	53,2	27,3	51,5	54,6	50,3	48,5	51,4	47,5	23,1	24,7
14. P.trichocarpa_757535#1	46,7	54,1	25,8	49,5	53	48,5	46,8	50,2	47,3	23,6	25
15. P.virgatum_TC11466#1	17,6	18,8	36,3	18,7	19,9	21	22,3	18,7	27,3	39,3	47,4
16. Popttr_ARID	43,8	44,5	39,9	49,4	42,4	56,4	54,5	47	42,9	32,9	38,3
17. S.bicolor_Sb01g020470.1#1	43,6	44,2	27,4	45,7	47,4	43,6	42,5	45,3	50,6	32,3	26
18. S.bicolor_Sb06g001680.1#1	44,1	47,1	32,5	46,4	46,4	46,2	50,1	43,3	77,3	32,1	29,6
19. V.vinifera_GSVIVT00002575001#1	48,6	54,3	33,1	49,4	56,9	51,7	50,9	48,7	53,9	28,1	31,3
20. V.vinifera_GSVIVT00026942001#1	56,2	55,7	33,4	58,7	56,4	58,6	57,4	54,8	53	28,9	33,4
21. Zea_mays_GRMZM2G365731_T01#1	32,7	34,2	20,4	35,7	36,1	35	32,1	36,7	35,9	24,6	18,1
22. Zea_mays_GRMZM2G424651_T01#1	43,6	45,5	31,1	45,7	47,4	45,3	47,6	42,5	74,7	31,4	29,1

FIGURE 5

12/70

	12	13	14	15	16	17	18	19	20	21	22
1. A.thaliana_AT2G46040.1#1	29,6	32,1	30,8	11,8	30,3	26,5	28,4	31,4	40	21	28,2
2. A.thaliana_AT4G11400.1#1	29,8	36,9	35,9	14,2	29,8	26,6	29,7	35,8	36,7	20,4	28,8
3. G.hirsutum_ES795168#1	28,5	18,8	18,2	25,2	28,2	16,3	20,2	22,5	23,8	12,9	20,6
4. G.max_Glyma07g07530.1#1	35,8	34,6	32,7	13,2	35,8	27,4	28,1	33	41,5	23,2	29,4
5. G.max_Glyma08g01950.1#1	27,9	36,5	35,5	14,9	27,9	25,7	27,7	38,8	35,4	20,2	27,1
6. G.max_Glyma09g39360.1#1	38	33,1	32	16	38,2	27,2	29	33,6	37,2	22,1	29,1
7. G.max_Glyma18g46950.1#1	36,3	30,7	30	16	36,4	26,1	27,4	29,9	37,4	19,4	30,1
8. M.truncatula_AC149204_12.4#1	33,2	32,8	31	14,8	33,2	27,2	28,2	29,5	35,4	23	28,3
9. O.sativa_LOC_Os04g08410.1#1	26,9	30,9	29,4	23,8	26,7	31,6	64,1	36,8	33,6	23,1	61,4
10. O.sativa_LOC_Os10g30944.2#1	22,4	15,4	15,9	25,2	22,3	25,6	21,9	19,3	19,6	20,1	21,8
11. P.persica_TC15862#1	29,1	19,3	18,8	33,9	29	18,5	22,7	22,8	26	12,3	22,8
12. P.trichocarpa_572279#1		28,4	28,6	16,9	99,8	23,5	26,8	31,3	37,9	17,8	27,1
13. P.trichocarpa_751302#1	41,2		78	14,2	28,5	25,9	31,8	39,9	39,3	21,7	31,2
14. P.trichocarpa_757535#1	41,1	86,1		14	28,6	26,1	30,5	40	37,6	21,5	30,6
15. P.virgatum_TC11466#1	24,5	17,7	18		17,6	13	28,4	17	15,7	9	27,5
16. Poptr_ARID	99,8	41,4	41,1	25,4		23,5	26,6	32,6	37,9	17,8	27,2
17. S.bicolor_Sb01g020470.1#1	37,4	42,1	42,6	18,5	37,3		30,2	28,4	30	55,3	30,7
18. S.bicolor_Sb06g001680.1#1	41,1	48	45,3	30,7	40,7	47,3		35,4	34	21,6	88,9
19. V.vinifera_GSVIVT00002575001#1	45,3	54,6	54,9	22,5	45,5	45,9	52,1		39,9	21,4	35,7
20. V.vinifera_GSVIVT00026942001#1	50,8	55,1	54,9	21,2	50,8	48,9	51,9	58,3		24	34,3
21. Zea_mays_GRMZM2G365731_T01#1	30,5	36,4	35,4	12,9	30,6	61,7	33,9	33,2	35,6		21,6
22. Zea_mays_GRMZM2G424651_T01#1	41,4	46,1	44,9	30,1	41,8	47,1	92,8	52,4	50,1	33	

FIGURE 5 (continued)

13/70

	1	2	3	4	5	6	7	8	9	10	11
1. O.sativa_LOC_Os10g30944.2		49.4	48.8	25.6	9.6	27.9	29.1	29.3	31.1	29.1	33.8
2. S.bicolor_Sb01g020470.1	60.5		84.0	31.5	9.9	34.1	34.5	29.3	32.1	33.0	38.8
3. Zea_mays_GRMZM2G365731_T01	63.8	88.9		28.7	10.0	34.9	33.3	31.4	31.3	31.3	36.1
4. O.sativa_LOC_Os04g08410.1	45.3	52.3	44.2		10.5	60.9	58.6	30.2	37.2	33.7	32.6
5. P.virgatum_TC11466	17.8	12.3	11.3	15.1		9.3	8.1	13.0	17.2	5.2	12.2
6. S.bicolor_Sb06g001680.1	47.7	50.0	47.7	83.7	15.1		88.4	33.7	37.2	40.7	30.2
7. Zea_mays_GRMZM2G424651_T01	48.8	48.8	47.7	76.7	12.8	91.9		34.9	36.0	38.4	31.4
8. G.hirsutum_ES795168	48.1	43.2	50.0	50.0	15.6	51.2	47.7		51.9	51.3	46.8
9. V.vinifera_GSVIVT00026942001	41.1	40.7	42.5	51.2	20.3	46.5	45.3	63.6		51.9	51.4
10. P.persica_TC15862	53.2	51.9	46.3	51.2	15.6	55.8	55.8	70.1	66.2		54.5
11. G.max_Glyma07g07530.1	45.9	48.1	46.3	51.2	18.9	48.8	47.7	59.7	60.8	62.3	
12. G.max_Glyma09g39360.1	50.0	49.4	48.8	53.5	12.8	53.5	54.7	62.8	60.3	62.8	66.7
13. G.max_Glyma18g46950.1	55.1	45.7	45.0	54.7	16.7	58.1	60.5	60.3	59.0	61.5	69.2
14. M.truncatula_AC149204_12.4	43.4	51.9	52.5	51.2	15.8	53.5	53.5	66.2	63.2	71.4	64.5
15. P.trichocarpa_572279	45.9	45.7	46.3	52.3	13.5	50.0	50.0	71.4	73.0	70.1	70.3
16. Poptr_ELM2-related	45.9	45.7	46.3	52.3	13.5	50.0	50.0	71.4	73.0	70.1	70.3
17. P.trichocarpa_751302	43.5	44.7	43.5	53.5	9.4	53.5	53.5	58.8	52.9	52.9	50.6
18. P.trichocarpa_757535	41.3	41.3	37.0	48.9	15.2	50.0	46.7	50.0	47.8	52.2	45.7
19. V.vinifera_GSVIVT00002575001	51.9	48.1	45.7	57.0	14.8	50.0	52.3	67.9	59.3	63.0	58.0
20. A.thaliana_AT4G11400.1	50.0	46.3	47.6	44.2	15.9	50.0	50.0	53.7	47.6	53.7	47.6
21. G.max_Glyma08g01950.1	46.6	45.7	47.5	46.5	20.6	47.7	44.2	59.7	67.2	49.4	55.4
22. A.thaliana_AT2G46040.1	40.7	43.2	49.4	57.0	17.3	53.5	51.2	65.4	59.3	63.0	60.5

FIGURE 6

	12	13	14	15	16	17	18	19	20	21	22
1. O.sativa_LOC_Os10g30944.2	30.9	31.3	26.9	30.4	30.4	29.9	25.8	24.4	32.2	26.7	31.3
2. S.bicolor_Sb01g020470.1	33.3	30.6	34.6	30.5	30.5	26.1	24.7	23.8	30.7	26.8	32.6
3. Zea_mays_GRMZM2G365731_T01	29.8	26.2	35.4	30.0	30.0	28.4	22.3	21.4	29.9	27.2	30.3
4. O.sativa_LOC_Os04g08410.1	34.5	35.6	36.0	37.2	37.2	37.8	32.0	42.5	34.5	33.7	37.5
5. P.virgatum_TC11466	7.7	9.0	10.5	9.5	9.5	5.9	10.9	8.6	8.5	7.9	7.4
6. S.bicolor_Sb06g001680.1	35.6	38.6	33.7	37.2	37.2	37.8	33.0	36.0	35.9	27.9	40.7
7. Zea_mays_GRMZM2G424651_T01	37.9	44.3	36.0	38.4	38.4	40.0	31.0	38.4	34.8	29.1	36.3
8. G.hirsutum_ES795168	47.4	44.9	48.1	51.9	51.9	42.4	40.2	49.4	38.6	45.5	46.3
9. V.vinifera_GSVIVT00026942001	47.4	47.4	48.1	63.5	63.5	47.1	41.3	45.7	39.0	52.3	51.9
10. P.persica_TC15862	47.4	50.0	49.4	53.2	53.2	40.0	38.0	40.7	33.7	35.1	46.3
11. G.max_Glyma07g07530.1	59.0	59.0	53.2	56.8	56.8	40.0	34.8	40.7	34.1	39.2	49.4
12. G.max_Glyma09g39360.1		84.6	64.1	53.8	53.8	40.7	33.3	47.0	39.0	41.0	46.9
13. G.max_Glyma18g46950.1	87.2		61.5	53.8	53.8	43.0	36.6	43.4	36.6	38.5	46.9
14. M.truncatula_AC149204_12.4	75.6	73.1		57.1	57.1	37.6	33.7	44.7	39.0	39.5	44.4
15. P.trichocarpa_572279	67.9	66.7	71.1		100.0	47.1	43.5	51.9	38.6	47.3	48.1
16. Poptr_ELM2-related	67.9	66.7	71.1	100.0		47.1	43.5	51.9	38.6	47.3	48.1
17. P.trichocarpa_751302	55.3	58.8	54.1	61.2	61.2		73.9	44.7	35.6	41.2	44.8
18. P.trichocarpa_757535	50.0	54.3	48.9	55.4	55.4	79.3		41.3	29.8	34.8	39.4
19. V.vinifera_GSVIVT00002575001	61.7	59.3	64.2	65.4	65.4	56.5	53.3		39.0	44.4	42.9
20. A.thaliana_AT4G11400.1	58.5	56.1	54.9	53.7	53.7	51.8	44.6	52.4		34.1	41.7
21. G.max_Glyma08g01950.1	51.3	51.3	55.3	58.1	58.1	52.9	47.8	56.8	46.3		39.5
22. A.thaliana_AT2G46040.1	58.0	58.0	58.0	61.7	61.7	56.5	51.1	61.7	57.3	49.4	

FIGURE 6 (continued)

15/70

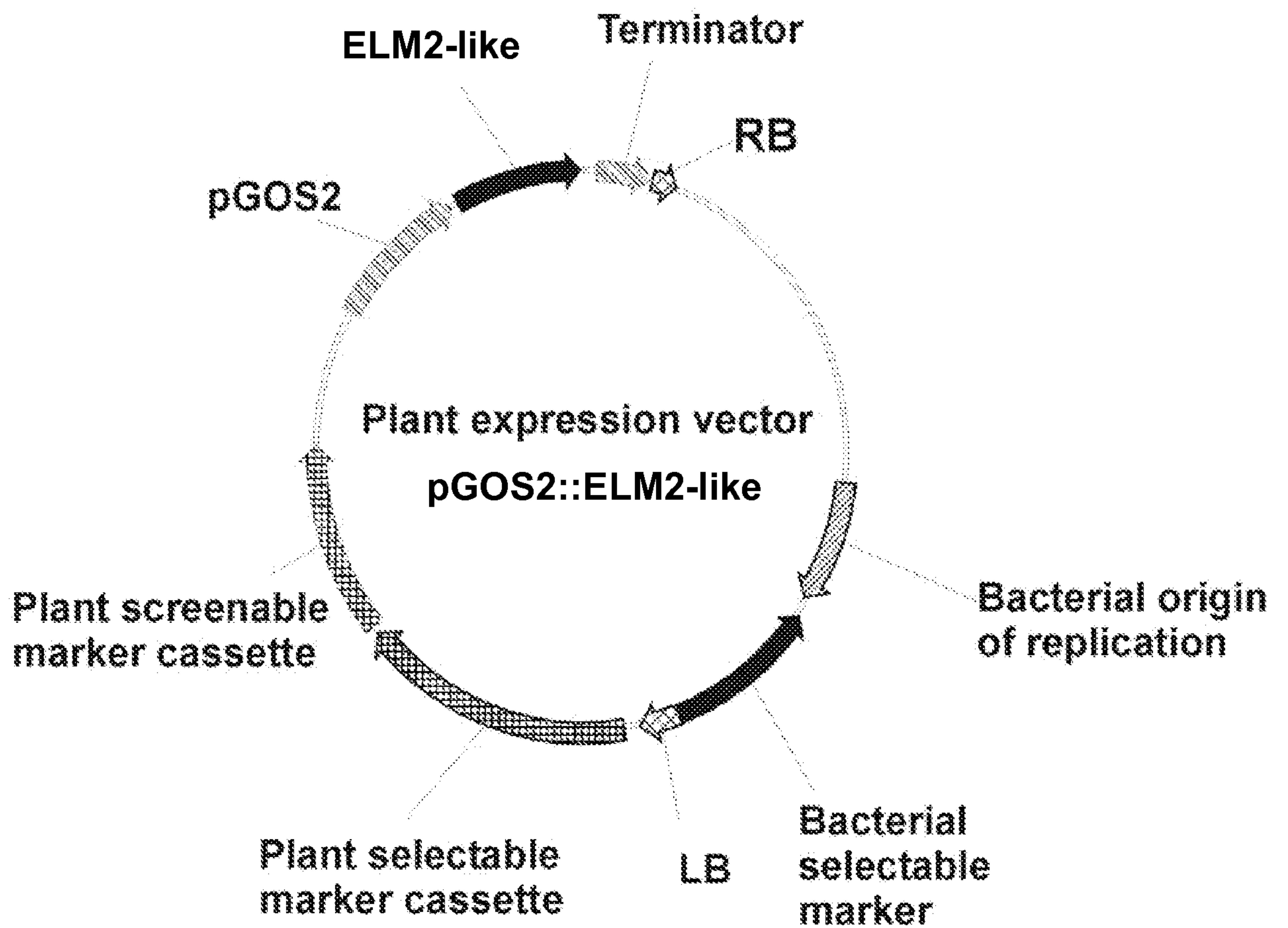


FIGURE 7

16/70

MECSSGRSGDGDGLKDDKRSEHSTGDDHESHKKKEIIAEEAPIASTERSIEAGPSTISSPKNDK
 VDEQLETTKAQMGEVREENNQRLKMCLNKIMTEYRALEMQFNNMVKQETKKNNNDNNHEEMNAESDLV
 SLSLGRVPSNNIPKNDQEKVNKVSKLALNNDEEFNKEELSLGLECKFETSKSGSTTEGLPNIPSPV
 NSSEVVPIKNDEVVETWPPSKTLNKTMRDAEDEVAQQTPAKKARVCVRARCDTPTMNDGCQWRKYG
QKIAKGNPCPRAYYRCTVAPSC_PVRKQVQRCVEDMSILITTYEGTHNHSLPLSATAMASTTSAAAS
MLLSGSSTSNSGSMPSAQTNNNLHGLNFYLPDGTKSNQLYLSNPALSSQHSHTITLDLTSNPSNS
 STSSPFVRFNSSYNNNNQLPRYPSSSTLSFSSPESNPMHWNSFLNYATTQNQPYSNNRNNNNLSTLN
 FGRQNTMESIYQTYMQKNNNSSNISQHVGLQDSTISAATKAITADPTFQSALAAALSSLIGNTTNQ
 GNQNSAGENLSQKMKWAEMFQVSSTSLPSSSSKVNGCASSFLNKTAPVNNTQNGSLMLLSPSLPF
 SATKSASTSPGGDNSDNTN

FIGURE 8

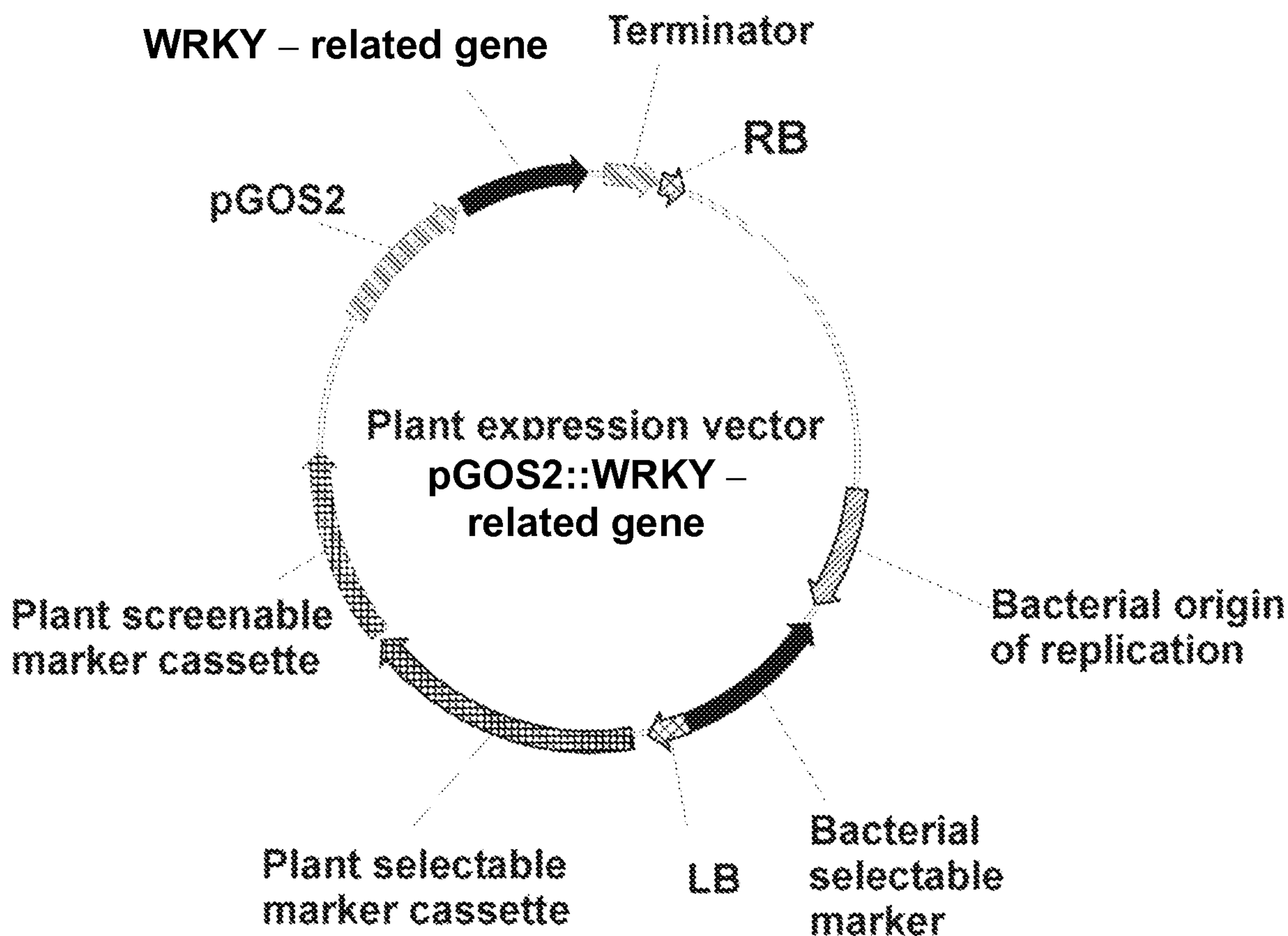


FIGURE 9

17/70

	1	2	3	4	5	6	7	8
1. SEQ ID NO: 103		38,80	38,60	41,60	37,40	34,20	38,20	40,00
2. AL6G14950	58,40		94,40	47,00	39,30	38,30	40,30	43,30
3. AT5G15130	56,90	96,00		47,10	41,20	39,20	40,70	42,80
4. CP00014G00310	60,40	63,00	62,10		43,40	40,40	44,90	44,20
5. GM09G37470	57,90	59,30	60,60	60,30		73,60	52,40	62,10
6. GM18G49140	49,80	52,10	53,80	54,50	78,60		48,30	55,50
7. GM19G02440	52,20	56,90	56,80	56,50	65,70	65,40		48,00
8. LJ1G030000	57,30	59,50	60,40	59,60	76,10	68,20	65,40	
9. MD00G204530	45,40	49,40	49,30	49,50	49,60	52,70	56,20	51,20
10. MD17G003750	50,70	54,40	54,40	54,00	54,90	57,30	62,80	56,90
11. ME04264G00010	60,40	63,20	63,30	67,60	66,60	62,90	64,40	68,20
12. MT7G02120	59,70	58,50	57,70	64,50	73,00	64,70	64,20	77,10
13. PT17G11190	62,00	65,00	65,10	66,90	65,10	60,60	62,10	66,50
14. RC29822G01580	63,30	64,80	64,10	69,10	65,20	60,20	60,50	64,50
15. VV14G01690	58,70	61,10	60,60	63,30	64,40	62,40	64,90	61,80

	9	10	11	12	13	14	15
1. SEQ ID NO: 103	34,10	37,80	44,70	37,60	44,20	44,80	44,40
2. AL6G14950	38,10	40,10	49,60	40,00	47,60	45,60	44,20
3. AT5G15130	37,90	39,20	48,40	40,00	46,90	46,30	44,10
4. CP00014G00310	38,80	42,00	52,50	45,10	52,60	53,90	49,20
5. GM09G37470	38,40	39,30	49,00	58,40	49,40	45,60	46,30
6. GM18G49140	35,40	37,20	48,20	53,70	47,60	44,90	42,60
7. GM19G02440	41,10	46,50	47,20	48,70	45,80	45,90	45,10
8. LJ1G030000	37,20	41,90	49,80	61,50	49,00	45,50	45,20
9. MD00G204530		70,30	44,20	35,70	42,30	40,00	47,70
10. MD17G003750	78,00		47,00	37,90	44,40	44,30	45,60
11. ME04264G00010	55,90	60,20		47,20	67,00	69,70	57,60
12. MT7G02120	47,80	52,80	63,50		47,10	43,80	42,50
13. PT17G11190	53,20	58,20	80,70	63,00		61,80	57,00
14. RC29822G01580	51,10	57,10	79,80	63,30	76,10		54,30
15. VV14G01690	58,00	59,20	73,10	59,70	71,40	68,90	

FIGURE 10

18/70

MVRPYGKGHKRKKGERYDKEEEEVEEEEQVEEEKADSETEMRAQDGEEK

TEEEKEMQLPGVEEGIPIVPSHQTGKKPGVIFVLEKASLEVA**KVGKSYQ**
motif 4

ILNSEEHANFLRRNKKNPADYRPDIYQALLSILDSPLNKAGRLRAVYV
EMG1 domain

*KTDK***GVLFEVKPHVRIPRTYKRFAGIMLQLLQKLSITAVGNREKLLRVI**
motif 5

KNPVTQYLPLNSRKIGFSHSSEKLVQMEKYVAGVGDDTDLVFVVGAMSH

GKIECDYVDDFISVSEYPLSAAWCIARVCEAVSKKW*RVL*
motif 6

FIGURE 11

19/70

	1	50
T.aestivum_TC329781	~~~~~	~~~~~
T.aestivum_TC342076	~~~~~	~~~~~
H.vulgare_TC168675	~~~~~	~~~~~
O.sativa_LOC_Os02g18830.1	~~~~~	~~~~~
P.virgatum_TC36436	~~~~~	~~~~~
Z.mays_GRMZM2G130339_T02	~~~~~	~~~~~
S.bicolor_Sb04g033060.1	~~~~~	~~~~~
S.bicolor_Sb04g011280.1	~~~~~	~~~~~
C.longa_TA2440_136217	~~~~~	~~~~~
Y.filamentosa_DT581455	~~~~~	~~~~~
A.sp_TC27214	~~~~~	~~~~~
P.patens_46413	~~~~~	~~~~~
P.patens_TC42118	~~~~~	~~~~~
S.moellendorffii_407397	~~~~~	~~~~~
A.trichopoda_CK754382	~~~~~	~~~~~
P.taeda_TA10927_3352	~~~~~	~~~~~
P.glauca_TA18327_3330	~~~~~	~~~~~
P.sitchensis_TA12989_3332	~~~~~	~~~~~
P.glauca_CO481519	~~~~~	~~~~~
C.japonica_TA2268_3369	~~~~~	~~~~~
M.truncatula_CU012050_19.4	~~~~~	~~~~~
L.japonicus_TC43316	~~~~~	~~~~~
G.max_Glyma12g02820.2	~~~~~	~~~~~
G.max_TC278049	~~~~~	~~~~~
G.max_Glyma12g02820.5	~~~~~	~~~~~
G.max_Glyma11g10520.1	~~~~~	~~~~~
G.max_Glyma12g02820.1	MLIHIIPLSAAYRNRHGPTGAGDDSVLPQLRPFPPARPLRGPLLGRHVVHP	~~~~~
P.coccineus_TA3737_3886	~~~~~	~~~~~
P.coccineus_TC15108	~~~~~	~~~~~
P.vulgaris_TC10901	~~~~~	~~~~~
C.sinensis_TC22219	~~~~~	~~~~~
G.hirsutum_ES833864	~~~~~	~~~~~
G.raimondii_TC1038	~~~~~	~~~~~
P.trichocarpa_667309	~~~~~	~~~~~
P.trichocarpa_831805	~~~~~	~~~~~
P.trichocarpa EMG1-like	~~~~~	~~~~~
P.trifoliata_TA8498_37690	~~~~~	~~~~~
G.hirsutum_TC158561	~~~~~	~~~~~
E.esula_TC3222	~~~~~	~~~~~
E.esula_DV156953	~~~~~	~~~~~
V.vinifera_GSVIVT00033904001	~~~~~	~~~~~
B.napus_TC70157	~~~~~	~~~~~
B.napus_TC76938	~~~~~	~~~~~
A.thaliana_AT3G57000.1	~~~~~	~~~~~
A.lyrata_486143	~~~~~	~~~~~
P.americana_CV001201	~~~~~	~~~~~
M.crystallinum_TC9042	~~~~~	~~~~~
V.carteri_106743	~~~~~	~~~~~
C.reinhardtii_178371	~~~~~	~~~~~
M.sp_77569	~~~~~	~~~~~
O.lucimarinus_39810	~~~~~	~~~~~
O.sp_18991	~~~~~	~~~~~
O.auri_37777	~~~~~	~~~~~
A.officinalis_TA1655_4686	~~~~~	~~~~~

FIGURE 12

20/70

	51	100
T.aestivum_TC329781	~~~~~	~~~~~
T.aestivum_TC342076	~~~~~	~~~~~
H.vulgare_TC168675	~~~~~	~~~~~
O.sativa_LOC_Os02g18830.1	~~~~~	~~~~~
P.virgatum_TC36436	~~~~~	~~~~~
Z.mays_GRMZM2G130339_T02	~~~~~	~~~~~
S.bicolor_Sb04g033060.1	~~~~~	~~~~~
S.bicolor_Sb04g011280.1	~~~~~	~~~~~
C.longa_TA2440_136217	~~~~~	~~~~~
Y.filamentosa_DT581455	~~~~~	~~~~~
A.sp_TC27214	~~~~~	~~~~~
P.patens_46413	~~~~~	~~~~~
P.patens_TC42118	~~~~~	~~~~~
S.moellendorffii_407397	~~~~~	~~~~~
A.trichopoda_CK754382	~~~~~	~~~~~
P.taeda_TA10927_3352	~~~~~	~~~~~
P.glaucā_TA18327_3330	~~~~~	~~~~~
P.sitchensis_TA12989_3332	~~~~~	~~~~~
P.glaucā_CO481519	~~~~~	~~~~~
C.japonica_TA2268_3369	~~~~~	~~~~~
M.truncatula_CU012050_19.4	~~~~~	~~~~~
L.japonicus_TC43316	~~~~~	~~~~~
G.max_Glyma12g02820.2	~~~~~	~~~~~
G.max_TC278049	~~~~~	~~~~~
G.max_Glyma12g02820.5	~~~~~	~~~~~
G.max_Glyma11g10520.1	~~~~~	~~~~~
G.max_Glyma12g02820.1	LAPRGAAHCLPFPRAQTASFTHNDGRSVNLLQADGTIPMTFQGVITYNVPV	~~~~~
P.coccineus_TA3737_3886	~~~~~	~~~~~
P.coccineus_TC15108	~~~~~	~~~~~
P.vulgaris_TC10901	~~~~~	~~~~~
C.sinensis_TC22219	~~~~~	~~~~~
G.hirsutum_ES833864	~~~~~	~~~~~
G.raimondii_TC1038	~~~~~	~~~~~
P.trichocarpa_667309	~~~~~	~~~~~
P.trichocarpa_831805	~~~~~	~~~~~
P.trichocarpa EMG1-like	~~~~~	~~~~~
P.trifoliata_TA8498_37690	~~~~~	~~~~~
G.hirsutum_TC158561	~~~~~	~~~~~
E.esula_TC3222	~~~~~	~~~~~
E.esula_DV156953	~~~~~	~~~~~
V.vinifera_GSVIVT00033904001	~~~~~	~~~~~
B.napus_TC70157	~~~~~	~~~~~
B.napus_TC76938	~~~~~	~~~~~
A.thaliana_AT3G57000.1	~~~~~	~~~~~
A.lyrata_486143	~~~~~	~~~~~
P.americana_CV001201	~~~~~	~~~~~
M.crystallinum_TC9042	~~~~~	~~~~~
V.carteri_106743	~~~~~	~~~~~
C.reinhardtii_178371	~~~~~	~~~~~
M.sp_77569	~~~~~	~~~~~
O.lucimarinus_39810	~~~~~	~~~~~
O.sp_18991	~~~~~	~~~~~
O.tauri_37777	~~~~~	~~~~~
A.officinalis_TA1655_4686	~~~~~	~~~~~

FIGURE 12 (continued)

21/70

	101	150
T.aestivum_TC329781	~~~~~MGRPYAVK.GKKRKLKLEG..GEASRA	
T.aestivum_TC342076	~~~~~MGRPYAVK.G.KKKRKLKLEG.GEASRA	
H.vulgare_TC168675	~~~~~MGRPYAVK.G.KKKRKLKLEG.GEASRA	
O.sativa_LOC_Os02g18830.1	~~~~~MGRPYAVK.GRKKRKLDD.GGASRE	
P.virgatum_TC36436	~~~~~MGRPYAVK.GRKKRKLKVE.AAASDA	
Z.mays_GRMZM2G130339_T02	~~~~~MGRPYAVK.GRKKRKLKLEEEAAASDA	
S.bicolor_Sb04g033060.1	~~~~~MVRPYAVK.GRKKRKLKLE.ASASGA	
S.bicolor_Sb04g011280.1	~~~~~MGRPYGVK.GRKKRKLKLE.AATSDA	
C.longa_TA2440_136217	~~~~~MVRPYAVK.GRKLIS....WRVEEP	
Y.filamentosa_DT581455	~~~~~	
A.sp_TC27214	~~~~~MVRPFAVK.GRKRKRDD..SYIVEK	
P.patens_46413	~~~~~	
P.patens_TC42118	~~~~~MG.....	
S.moellendorffii_407397	~~~~~	
A.trichopoda_CK754382	~~~~~	
P.taeda_TA10927_3352	~~~~~MDRGIKRK.R.....	
P.glaucata_TA18327_3330	~~~~~	
P.sitchensis_TA12989_3332	~~~~~MDGGIKRK.R.....	
P.glaucata_CO481519	~~~~~MDGGIKRK.R.....	
C.japonica_TA2268_3369	~~~~~MEESLKRK.R.....	
M.truncatula_CU012050_19.4	~~~~~MTRAYTLK.GKKRKNKDAATKHDHEQ	
L.japonicus_TC43316	~~~~~MTRPFGMK.GRKRKDKEA..KYDREE	
G.max_Glyma12g02820.2	~~~~~MTRPYGVKGGRKRKHAEP..KYD...	
G.max_TC278049	~~~~~MTRPYGVKGGRKRKHAEP..KYD...	
G.max_Glyma12g02820.5	~~~~~MTRPYGVKGGRKRKHAEP..KYD...	
G.max_Glyma11g10520.1	~~~~~MTRPFGVKGRKRKHAEP..KYDAHD	
G.max_Glyma12g02820.1	~~~~~VIWLAESYPRLPPCVYVNPTDMMIKRPHPHG.GRKRKHAEP..KYD...	
P.coccineus_TA3737_3886	~~~~~MTRPFGVK.GKKRKHQS..KQYSE.	
P.coccineus_TC15108	~~~~~MTRPFGVK.GKKRKHQS..KQYSE.	
P.vulgaris_TC10901	~~~~~MTRPFGVK.GKKRKHQS..KQYSE.	
C.sinensis_TC22219	~~~~~	
G.hirsutum_ES833864	~~~~~	
G.raimondii_TC1038	~~~~~	
P.trichocarpa_667309	~~~~~	
P.trichocarpa_831805	~~~~~MVRPYG.K.GHKRKKGE..RYDKEE	
P.trichocarpa EMG1-like	~~~~~MVRPYG.K.GHKRKKGE..RYDKEE	
P.trifoliata_TA8498_37690	~~~~~	
G.hirsutum_TC158561	~~~~~MVRPYAIK.GKKRKNRGE..KYDKKE	
E.esula_TC3222	~~~~~MVRLHGIK.GPKRKRREE..KYDKEE	
E.esula_DV156953	~~~~~	
V.vinifera_GSVIVT00033904001	~~~~~MVRPYAIK.GHKRKKREE..NYDREE	
B.napus_TC70157	~~~~~MVRPYGIK.VNKRKERRE..RYDKEE	
B.napus_TC76938	~~~~~MVRPYGIK.VNKRKERKE..RYDKEE	
A.thaliana_AT3G57000.1	~~~~~MVRPYGIK.VNKRKEREE..RYDKEE	
A.lyrata_486143	~~~~~MVRPYGIK.VNKRKERTE..RYDKEE	
P.americana_CV001201	~~~~~	
M.crystallinum_TC9042	~~~~~MPRPFVK.GKKRKKREE..NYDREE	
V.carteri_106743	~~~~~	
C.reinhardtii_178371	~~~~~	
M.sp_77569	~~~~~	
O.lucimarinus_39810	~~~~~	
O.sp_18991	~~~~~	
O.auri_37777	~~~~~	
A.officinalis_TA1655_4686	~~~~~MVRPYAVK.GRKKPKKRP..RVEEEE	

FIGURE 12 (continued)

22/70

T.aestivum_TC329781 T.aestivum_TC342076 H.vulgare_TC168675 O.sativa_LOC_Os02g18830.1 P.virgatum_TC36436 Z.mays_GRMZM2G130339_T02 S.bicolor_Sb04g033060.1 S.bicolor_Sb04g011280.1 C.longa_TA2440_136217 Y.filamentosa_DT581455 A.sp_TC27214 P.patens_46413 P.patens_TC42118 S.moellendorffii_407397 A.trichopoda_CK754382 P.taeda_TA10927_3352 P.glauca_TA18327_3330 P.sitchensis_TA12989_3332 P.glauca_CO481519 C.japonica_TA2268_3369 M.truncatula_CU012050_19.4 L.japonicus_TC43316 G.max_Glyma12g02820.2 G.max_TC278049 G.max_Glyma12g02820.5 G.max_Glyma11g10520.1 G.max_Glyma12g02820.1 P.coccineus_TA3737_3886 P.coccineus_TC15108 P.vulgaris_TC10901 C.sinensis_TC22219 G.hirsutum_ES833864 G.raimondii_TC1038 P.trichocarpa_667309 P.trichocarpa_831805 P.trichocarpa EMG1-like P.trifoliata_TA8498_37690 G.hirsutum_TC158561 E.esula_TC3222 E.esula_DV156953 V.vinifera_GSVIVT00033904001 B.napus_TC70157 B.napus_TC76938 A.thaliana_AT3G57000.1 A.lyrata_486143 P.americana_CV001201 M.crystallinum_TC9042 V.carteri_106743 C.reinhardtii_178371 M.sp_77569 O.lucimarinus_39810 O.sp_18991 O.auri_37777 A.officinalis_TA1655_4686	151 PPADAEAV.....EQEVKENEEVMEGEAAAAEA...FAEE PPVDAEAV.....EQEVKENEEVMEGEAAAAEE..AFAEE PPVDAEEV.....ELEGKENEEVTEGEAAPEA...FAEE PPVDEEAA.....EELPPE..GVEEEGKEEDVAAAA...AAGE APPAEEAE.....ELPPPEEAGGEEKEGEDGAAAVVE..GEEHA APPVEEAE.....ELPPPEVRSGEESSEEDAAAAGE...EEHA APPAEEAE.....ELPPPEVGS...EEDEAAAGK...EEHA APPAEEVE.....ELPPQEEAGGEEKGNEEDAAAAGE...EEHA APPEEEKV.....EDASDDGEREGDQAAFETGVEVEENEDERRTEEALD ~~~~~ KIKDEEEE.....VVVEEEEEEEEEQTPPKKHKGEEEDERIAE...EAVN ~~~~~ GKEEDRDPEMAES..SDEEI ~~~~~ SREEDEDAADSAT..QDIET ~~~~~ SREEDEDAAGAAT..QDIET SREEDEDAAGAGT..QDIET EDDDEEDEVLAG..... EEEEEQQ.....QIEPTPKKPNLHNDEPSA...PTEES .DEDEQVQ....PKR.....PMLQNAEV..AA..ANEED .EEDQETQ....PKKAVVE.....VEHNKEEQQ..QQ..TPEGD .EEDQETQ....PKKAVVE.....VEHNKEEQQ..QQ..TPEGD .EEDQETQ....PKKAVVE.....VEHNKEEQQ..QQ..TPEGD EEEEETQ....PKKTVE.....VEHNQEEEEETPAP..APEDD .EEDQETQ....PKKAVVE.....VEHNKEEQQ..QQ..TPEGD .EEHQPP....PKKAVAEESKEPSPEPS..PEPERESE..PA..PTEED .EEHQPP....PKKAVAEESKEPSPEPS..PEPERESE..PA..PTEED .EEHQPP....PKKAVAEESKEPSPEPSPEPERESE..PA..PTEED ~~~~~ ~~~~~ ~~~~~ ~~~~~ EEEEEEQV....EEEKADSETEMRAQDGEEKTEEEKEMQ..LP..GV... EEVEEEQV....EEEKADSETEMRAQDGEEKTEEEKEMQ..LP..GV... ~~~~~ QEHEEQVE....EEEEKEPAKQAAIEKPTEEVEGKEEEEGG....EAGTD DEADEQQN.....EPENEEQKDEME..QV..EEQSY ~~~~~ ~~~~~ V...EQLL....EEETAATATQD.....EK..KA..EEVTL DEEEEQPKFEKKQ..KERVKRAKKEVTAKAEEEDDEENEIAEE...AADE DEEEEQPKFEQKQKEKERVKRAKREVTAKAEEEDDEENEIAEE...AADE DEVVEEQPKFEQKQKARESSKKAKKESTSRAEEDNDEEEVTVEAT.AAAED EEVEEQPKFERKQKARESSKKAKKEITSRAEEGNEEEEITEEVTAAAED ~~~~~ FE.....QMVNEDNTDANIQTQAEDENLV.....DDLVR ~~~~~ ~~~~~ ~~~~~ ~~~~~ EEKEEQQ...EENDEAVSEKENEETLENEEKGESTAEDERKAE...EAIE	200
---	---	-----

FIGURE 12 (continued)

23/70

T.aestivum_TC329781 T.aestivum_TC342076 H.vulgare_TC168675 O.sativa_LOC_Os02g18830.1 P.virgatum_TC36436 Z.mays_GRMZM2G130339_T02 S.bicolor_Sb04g033060.1 S.bicolor_Sb04g011280.1 C.longa_TA2440_136217 Y.filamentosa_DT581455 A.sp_TC27214 P.patens_46413 P.patens_TC42118 S.moellendorffii_407397 A.trichopoda_CK754382 P.taeda_TA10927_3352 P.glauca_TA18327_3330 P.sitchensis_TA12989_3332 P.glauca_CO481519 C.japonica_TA2268_3369 M.truncatula_CU012050_19.4 L.japonicus_TC43316 G.max_Glyma12g02820.2 G.max_TC278049 G.max_Glyma12g02820.5 G.max_Glyma11g10520.1 G.max_Glyma12g02820.1 P.coccineus_TA3737_3886 P.coccineus_TC15108 P.vulgaris_TC10901 C.sinensis_TC22219 G.hirsutum_ES833864 G.raimondii_TC1038 P.trichocarpa_667309 P.trichocarpa_831805 P.trichocarpa EMG1-like P.trifoliata_TA8498_37690 G.hirsutum_TC158561 E.esula_TC3222 E.esula_DV156953 V.vinifera_GSVIVT00033904001 B.napus_TC70157 B.napus_TC76938 A.thaliana_AT3G57000.1 A.lyrata_486143 P.americana_CV001201 M.crystallinum_TC9042 V.carteri_106743 C.reinhardtii_178371 M.sp_77569 O.lucimarinus_39810 O.sp_18991 O.tauri_37777 A.officinalis_TA1655_4686	201 VGEPIVIP.....QPVDSKRRPGAIFVLERACLEVGKVGKGMQILNSD VGEPIVIP.....RPVDSKRRPGAIFVLERACLEVGKVGKGMQILNSD VGDGPIVIP.....RPLDSKRRPGAIFVLERACLEVGKVGKGMQILNSD VADGLPVVP.....RPVDGKRRPGAIFVLERACLEVGKVGKTMQILNSD AVEGLPVVP.....RPVDGKRRPGAIFVLERACLEVGKVGKTMQILNSD AADGLPILP.....RTVDGKRRPGAIFVLERACLEVGKVGKTMQILNSD AADGLPIIP.....RTVDGKRRPGAIFVLERACLEVGKVGKTMQILNSD AVDGIPIVP.....RTLKGKRRPGAIFVLERACLEVGKVGKTMQILNSD KIPVLPITG.....PVDRAKRPVIFVLEKACLEVGKVGKNFQILNSD ~~~~~ ELPGIPIVP...CSSSSSDKKSISGIIIFVLEKASLEVAKVGTKYQLLNSD ~~~~~ AEASADPAP.....VKSLTAPDGGVWFVLEKASLEVAKVGKNYQLLNCD AVQDASVGQ.....GDRDDDKSGVIVLERASLETAKVGKTYQLLNCD ~~~~~ ELVGLPAPP.....PIVKGGEVIFILEKAHLEVAKVGTKYQLLNCD ~~~~~ ELVGLPALP.....TIKKGGEVIFILEKASLEVAKVGTKYQLLNVD ELVGLPALP.....TIKKGGEVIFILEKASLEVAKVGTKYQLLNVD ..LGAACPP.....ANNNSDGGVVFVLLKASLEVGKVGKIYQLLNCD ELPGIPIAP.....LNEKNSEKQGVIFILEKASLEVAKVGTKYQLLNCD ELVGIPIMP.....SDHKSDNPVIFILERASLEVAKVGTKYQLLNCD ELSGIPITP.....SENNTNKPVIFILEKASLEVAKVGTKYQLLNCD ELSGIPITP.....SENNTNKPVIFILEKASLEVAKVGTKYQLLNCD ELSGIPITP.....SENNTNKPVIFILEKASLEVAKVGTKYQLLNCD GLSGIPIAP.....SANNTNKPVIFILEKASLEVAKVGTKYQLLNCD ELSGIPITP.....SENNTNKPVIFILEKASLEVAKVGTKYQLLNCD ELSGIPIVP.....REKNSTKPNVIFILEKASLEVAKVGTKYQLLNCD ELSGIPIVP.....SEKNSTKPNVIFILEKASLEVAKVGTKYQLLNCD ELSGIPIVP.....SEKNSTKPNVIFILEKASLEVAKVGTKYQLLNCD ~~~~~ ~~~~~ ~~~~~ ~~~~~ . EEGIPIVP.....SHQTGKKPGVIFVLEKASLEVAKVGSYQILNSE . EEGIPIVP.....SHQTGKKPGVIFVLEKASLEVAKVGSYQILNSE ~MEGIPIAP.....SDQNTKRPGVIFVLEKASLEVAKVGSYQILNSE ELEGIPIAP.....SDKKNKNGVIFVLEKASLEVAKVGSFQILNSE EMEGIPSPV.....SDQNAKKPGVIFVLEKASLEVAKVGIYQILNSE EMEGIPSPV.....SDQYAKKPGVIFVLEKASLEVAKVGIYQILNSE ELPGIPLVP.....SNLN.NKTGVTFILERACLEVAKVGSYQLLSSE LDVGPIVV.....TDNKKDQTGVIFVLEKASLEVAKVGTKYQLLNCD LDVGPIIIV.....SDDKKDQTGVIFVLEKASLEVAKVGTKYQLLNCD IVGGIPIVL.....NAPNKEKSGVIFVLEKASLEVAKVGTKYQLLNCD IVGGIPIVL.....SAPNKEKPGVVFVLEKASLEVAKVGTKYQLLNCD ~~~~~ DMPGIPIV.....KDQNNQSGAVFILERASLEVAKVGTKYQILNSE ~MVAGGLQPPPIEEEEALRAQGQGVIFILDDAQLEVAQVGKTFQLLNCD ~MVAGGLQPPSIDEEEAARAVGQGVIFILDDAQLEVAQVGKTYQLLNCD ~~~~~METPKLDED..SAAVPGEGYIFVLEKATLEAAKIGKAYQILNCD ~~~~~MEPPPPDDD...ANRKDGGYVFVLELATLETAKVGKGYAILNCD ~~~~~MEPPPPDPETPASALDGGGYVFVLELATLETAKVGKGYAILNCD ~~~~~MDPPPSDDG.STADARAGGYVFVLELATLETAKVGKGYAVLNCD EMPGIPIAP.....RVQDPKKKPGVIFVLERACLEFAKVGTKYQILNSD	250
---	---	--------------

FIGURE 12 (continued)

24/70

T.aestivum_TC329781 T.aestivum_TC342076 H.vulgare_TC168675 O.sativa_LOC_Os02g18830.1 P.virgatum_TC36436 Z.mays_GRMZM2G130339_T02 S.bicolor_Sb04g033060.1 S.bicolor_Sb04g011280.1 C.longa_TA2440_136217 Y.filamentosa_DT581455 A.sp_TC27214 P.patens_46413 P.patens_TC42118 S.moellendorffii_407397 A.trichopoda_CK754382 P.taeda_TA10927_3352 P.glauca_TA18327_3330 P.sitchensis_TA12989_3332 P.glauca_CO481519 C.japonica_TA2268_3369 M.truncatula_CU012050_19.4 L.japonicus_TC43316 G.max_Glyma12g02820.2 G.max_TC278049 G.max_Glyma12g02820.5 G.max_Glyma11g10520.1 G.max_Glyma12g02820.1 P.coccineus_TA3737_3886 P.coccineus_TC15108 P.vulgaris_TC10901 C.sinensis_TC22219 G.hirsutum_ES833864 G.raimondii_TC1038 P.trichocarpa_667309 P.trichocarpa_831805 P.trichocarpa EMG1-like P.trifoliata_TA8498_37690 G.hirsutum_TC158561 E.esula_TC3222 E.esula_DV156953 V.vinifera_GSVIVT00033904001 B.napus_TC70157 B.napus_TC76938 A.thaliana_AT3G57000.1 A.lyrata_486143 P.americana_CV001201 M.crystallinum_TC9042 V.carteri_106743 C.reinhardtii_178371 M.sp_77569 O.lucimarinus_39810 O.sp_18991 O.tauri_37777 A.officinalis_TA1655_4686	251 300 DSRKLFEAAESQSS~ DHANYLRKQNRNPADYRPDIHQALLAIFDSPLTKAGRLQAVYVRTEKGV DHANYLRKQNRNPADYRPDIHQALLAIFDSPLTKAGRLQAVYVRTEKGV DHANYLRKQNRNPADYRPDIHQALLAIFDSPLTKAGRLQAVYVRTEKGV DHANYLRKQNRNPADYRPDIHQALLAIFDSPLTKAGRLQAVYVRTEKGV DHANYLRKQNRNPADYRPDIHQALLAIFDSPLTKAGRLQAVYVRTEKGV DHANYLRKQNRNPADYRPDIHQALLAIFDSPLTKAGRLQAVYVRTEKGV DHANYLRKQNRNPADYRPDIHQALLAIFDSPLTKAGRLQAVYVRTEKGV DHTNYLKKQNRNPADYRPDIHQALLAIFDSPLAKAGRLQAVYVKTEKGV ~ DHSSFLRKHNRPDYPDIAHQALLAILDSPLNKAGRLRAVYVRTEKGV ~ DHANFLRKHKRDPAYRPDIHQALLAILDSPMNKAGRLKGLYVHTEKNV DHANFLRKHKRDPALYRPDIHQALLAILDSPLNKAGRLKALYVKTEKNV ~ DHRHFLRKHKQDLALFRPDIVHQALLAILDSPLNKSGRVS AVYVQTQKKV ~ DHGHFLRKHKQDLATFRPDIVHQALLAILDSPLNKAGRLSAVYVQTQKKV DHGHFLRKHKQDLATFRPDIVHQALLAILDSPLNKAGRLSAVYVQTQKKV DHVHFLKKNKQNPAYRPDIPVHQALLAILDSPLNKAGRLSAVYVQTQKNI EHSNFLRKHSKNPAGDYPDICHQALLSILDSPVNKAGRLKMYVIRTEKGV DHANFLRKNNKNPAGDYPDITHQALLSILDSPLNKAGRLQAVYIKTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV DHANFLRKNNKNPAGDYPDITHQSLLSILDSPLNKAGRLRSVYIRTEKGV ~ ~ ~ ~ EHANFLRRNKNPADYRPDIHQALLSILDSPLNKAGRLRAVYVKTDKGV EHANFLRRNKNPADYRPDIHQALLSILDSPLNKAGRLRAVYVKTDKGV DHSNFLRRNKNPADYRPDIPVHQALLSILDSPLNKAGRLQAVYVRTEKGV DHANFLRKNNKNPADYRPDITHQALLSILDSPVNKAGRLQAVYVRTEKGV DHSNFLRKHKKDPADYRPDIPVHQALLSILDSPLNKAGRLRAVYVKTDKGV DHSNFLRKHKKDPADYRPDIPVHQALLSILDSPLNKAGRLRAVYVKTDKGV DHANFLRKNNKNPAYRPDIHQALLTILDSPLNKAGRLRAVYVRTEKGV DHANFLRKNGRDPAYRPDITHQALLMILDSPVNKAGRLKAVYVRTEQGV DHANFLRKNGRDPAYRPDITHQALLMILDSPVNKAGRLKAVYVRTEQGV DHANFLKNNRNPADYRPDITHQALLMILDSPVNKAGRLKAVYVRTEKGV DHANFLRKNNRNPADYRPDITLQALLMILDSPVNKAGRLKAVYVRTEKGV ~ DHANFLRKNNKNPADYRPDIAHQALLMILDSPLNKAGRIKAVYVRTEKGV DHATYLRKHKKDPALYRPDICHQALLNILDSPLNKAAKIKAVYVRTHKNV DHATYLRKHKKDPALYRPDICHQALLNVLDSPLNKAGRIKAVYVRTHKNV DHANFLRRHGKDPANFRPDIPVHQELLAILDSPLNKAGRVNAVYVHTQKNV DHANFIRRHGKQPGDHRPDICHQALLAILDSPLNKAGMVKAIYVNTQKNV DHANFIRRHGKQPGDHRPDICHQALLAILDSPLNKAGRVKAIYVNTQKNV DHANFIRRHGKQPGDYPDICHQALLSILDSPLNKAGKVRGIYVNTQKNV DHANFLKKQNRNPADYRPDIHQALLAILDSPLNKAGRLQALYVRTEKGV
--	---

FIGURE 12 (continued)

	301	350
T.aestivum_TC329781	~~~~~	~~~~~
T.aestivum_TC342076	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTO.	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTO.
H.vulgare_TC168675	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTO.	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTO.
O.sativa_LOC_Os02g18830.1	LFEIKPYVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTR.	LFEIKPYVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTR.
P.virgatum_TC36436	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.
Z.mays_GRMZM2G130339_T02	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.
S.bicolor_Sb04g033060.1	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.
S.bicolor_Sb04g011280.1	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.	LFEIKPHVRMPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVVKNPVTR.
C.longa_TA2440_136217	LFEIKPHVRIPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTR.	LFEIKPHVRIPRTFKRFCGLMSQLLQKLSITAVGKREKLLNVIKPNVTR.
Y.filamentosa_DT581455	~~~~~MPRTFKRFCGLMLQLLQNLISITAAGKREKLLNVIKPNVTR.	~~~~~MPRTFKRFCGLMLQLLQNLISITAAGKREKLLNVIKPNVTR.
A.sp_TC27214	LFEVKPHVRIPTFKRFCGVMQLLQKLSISATTKREKLLRVIKPNVTO.	LFEVKPHVRIPTFKRFCGVMQLLQKLSISATTKREKLLRVIKPNVTO.
P.patens_46413	LIQINPHIRIPRTFKRFCGLMVQLLQKLIVIRATNGPDKLMKVVKQPVTR.	LIQINPHIRIPRTFKRFCGLMVQLLQKLIVIRATNGPDKLMKVVKQPVTR.
P.patens_TC42118	LIQINPHIRIPRTFKRFCGLMVQLLQKLIVIRATNGPDKLMKVVKQPVTR.	LIQINPHIRIPRTFKRFCGLMVQLLQKLIVIRATNGPDKLMKVVKQPVTR.
S.moellendorffii_407397	LIQVNPHIRLPRRTFKRFCGLMAQLLQKLSIRATNGPDKLLRVVKQPVTR.	LIQVNPHIRLPRRTFKRFCGLMAQLLQKLSIRATNGPDKLLRVVKQPVTR.
A.trichopoda_CK754382	~~~~~MVQLLQKLSIRATNGPDILLRVVKPNVTR.	~~~~~MVQLLQKLSIRATNGPDILLRVVKPNVTR.
P.taeda_TA10927_3352	LIQVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDKLLHVIKPNVTK.	LIQVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDKLLHVIKPNVTK.
P.glauca_TA18327_3330	~~~~~MVQLLQKLSIRATNGPDKILRVIKPNVTK.	~~~~~MVQLLQKLSIRATNGPDKILRVIKPNVTK.
P.sitchensis_TA12989_3332	LIQVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDKILRVIKPNVTK.	LIQVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDKILRVIKPNVTK.
P.glauca_CO481519	LIQVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDKILRVIKPNVTK.	LIQVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDKILRVIKPNVTK.
C.japonica_TA2268_3369	LIKVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDRLLRVIKPNVTK.	LIKVNPHVRLPRRTFKRFCGLMVQLLQKLSIRATNGPDRLLRVIKPNVTK.
M.truncatula_CU012050_19.4	LIEVKPYVRIPTFKRFAGVMLELLQKLSITAAGKREKLLRTIKPNVTO.	LIEVKPYVRIPTFKRFAGVMLELLQKLSITAAGKREKLLRTIKPNVTO.
L.japonicus_TC43316	LIEVKPYVRIPTFKRFSGVMQLLQKLSISAAGKHEKLLRTIKPNVTO.	LIEVKPYVRIPTFKRFSGVMQLLQKLSISAAGKHEKLLRTIKPNVTO.
G.max_Glyma12g02820.2	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
G.max_TC278049	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
G.max_Glyma12g02820.5	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
G.max_Glyma11g10520.1	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
G.max_Glyma12g02820.1	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
P.coccineus_TA3737_3886	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
P.coccineus_TC15108	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
P.vulgaris_TC10901	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.	LIEVKPFVRIPTFKRFAGVMLELLQKLSISAVGKREKLLRTIKPNVTO.
C.sinensis_TC22219	~~~~~MLQLLQKLSITAVGKREKLLRVIKPNVTO.	~~~~~MLQLLQKLSITAVGKREKLLRVIKPNVTO.
G.hirsutum_ES833864	~~~~~MLQLLQKLNITAVGKREKLLRVIKPNVTN.	~~~~~MLQLLQKLNITAVGKREKLLRVIKPNVTN.
G.raimondii_TC1038	~~~~~MLQLLQKLNITAVGKREKLLRVIKPNVTN.	~~~~~MLQLLQKLNITAVGKREKLLRVIKPNVTN.
P.trichocarpa_667309	~~~~~MLQLLQKLSIAAVGKREKLLRVIKPNVTO.	~~~~~MLQLLQKLSIAAVGKREKLLRVIKPNVTO.
P.trichocarpa_831805	LFEVKPHVRIPTYKRFAGIMLQLLQKLSITAVGNREKLLRVIKPNVTO.	LFEVKPHVRIPTYKRFAGIMLQLLQKLSITAVGNREKLLRVIKPNVTO.
P.trichocarpa EMG1-like	LFEVKPHVRIPTYKRFAGIMLQLLQKLSITAVGNREKLLRVIKPNVTO.	LFEVKPHVRIPTYKRFAGIMLQLLQKLSITAVGNREKLLRVIKPNVTO.
P.trifoliata_TA8498_37690	LFEVKPHVRLPRTYKRFAGIMLQLLQKLSITAVGKREKLLRVIKPNVTO.	LFEVKPHVRLPRTYKRFAGIMLQLLQKLSITAVGKREKLLRVIKPNVTO.
G.hirsutum_TC158561	LFEVKPHVRIPTYKRFSGIMLQLLQKLNITAVGKREKLLRVIKPNVTN.	LFEVKPHVRIPTYKRFSGIMLQLLQKLNITAVGKREKLLRVIKPNVTN.
E.esula_TC3222	LFEVKPHVRIPTYKRFAGVMLQLLQKLSISAMGKREKLLRVIKPNVTO.	LFEVKPHVRIPTYKRFAGVMLQLLQKLSISAMGKREKLLRVIKPNVTO.
E.esula_DV156953	LFEVKPHVRIPTYKRFAGVMLQLLQKLSISAMGKREKLLRVIKPNVTP.	LFEVKPHVRIPTYKRFAGVMLQLLQKLSISAMGKREKLLRVIKPNVTP.
V.vinifera_GSVIVT00033904001	LFEVKPHVRIPTYKRFSGIMLQLLQKLSITAVGKREKLLRVIKPNVTO.	LFEVKPHVRIPTYKRFSGIMLQLLQKLSITAVGKREKLLRVIKPNVTO.
B.napus_TC70157	LFEVKPHVRIPTYKRFAGIMLQLLQKLSISAAGKREKLMRVIKPNVTG.	LFEVKPHVRIPTYKRFAGIMLQLLQKLSISAAGKREKLMRVIKPNVTG.
B.napus_TC76938	LFEVKPHVRIPTYKRFAGIMLQLLQKLSISAAGKREKLMRVIKPNVTG.	LFEVKPHVRIPTYKRFAGIMLQLLQKLSISAAGKREKLMRVIKPNVTG.
A.thaliana_AT3G57000.1	LFEVKPHVRIPTFKRFAGIMLQLLQKLSITAVNSREKLLRCVKNP1EEH	LFEVKPHVRIPTFKRFAGIMLQLLQKLSISAVGSGEKLLRCVKNP1EEH
A.lyrata_486143	LFEVKPHVRIPTFKRFAGIMLQLLQKLSISAVGSGEKLLRCVKNP1EEY	LFEVKPHVRIPTFKRFAGIMLQLLQKLSISAVGSGEKLLRCVKNP1EEY
P.americana_CV001201</		

FIGURE 12 (continued)

26/70

	351	400
T.aestivum_TC329781	~~~~~	
T.aestivum_TC342076	YLPVGVRKIGLSYSAEKAVNLF DYVAKSN.DDEPLVF.VVGAMAHGKVDK	
H.vulgare_TC168675	YLPVGTRKIGLSYSAEKAVNLC DYVAKSN.DDEPLVF.VVGAMAHGKVDK	
O.sativa_LOC_Os02g18830.1	YLPVGAKKIGLSYSAEKSVNLF DYVAKSS.DDVPLVF.VVGAMAHGKIDN	
P.virgatum_TC36436	YLPVGARKIGLSFSAEKSVNLF DYVAKSS.DDEPLVF.VVGAMAHGKIDK	
Z.mays_GRMZM2G130339_T02	YLPVGARKIGLSFSAEKSVNLF DYVAKSN.DDEPLVF.VVGAMAHGKIDK	
S.bicolor_Sb04g033060.1	YLPVGARKIGLSFSAEKSVNLF DYVAKSS.DDEPLVF.VVGAMAHGKIDK	
S.bicolor_Sb04g011280.1	YLPVGARKIGLSFSAEKSVNLF DYVAKSS.DDEPLVF.VVGAMAHGKIDK	
C.longa_TA2440_136217	HLPVNSKKIGLSYSSDKLVNLS NYVAAAS.DDAPLVF.VVGAMAHGVISK	
Y.filamentosa_DT581455	HLPVNSRRIGLSYSSPKVVQMR DYVAAAS.DDVNLVF.VVGAMAHGKISN	
A.sp_TC27214	HLPVNSRKVGLSHSSEKLVQLQ DYVAAAK.DDMNLVF.VVGAMAHGKIDK	
P.patens_46413	HLPIGARRIGMSYSAPKVVQLK DYILTTLK.EDTPLVF.VVGAMSHGKIEV	
P.patens_TC42118	HLPIGARRIGMSYSAPKVVQLK DYILTTLK.EDTPLVF.VVGAMSHGKIEV	
S.moellendorffii_407397	HLPAGVRRIGLSYRAPKVVQLR DYVAASG.PEETLVF.VVGAMAHGKIEA	
A.trichopoda_CK754382	HLPIDSRRIGLSYSSPKVVQLH DYVTDAAK.DDVNLVF.VVGAMAHGKITA	
P.taeda_TA10927_3352	YLPSEAHRIIGLSYAPKVVKLH DYVAAST.DKIPLIF.VVGAMAHGKIDV	
P.glauca_TA18327_3330	HLPSEARRIGLSYAPKVVKLH DYVAAASS.DKIPLVF.VVGAMAHGKINA	
P.sitchensis_TA12989_3332	HLPSEARRIGLSYAPKVVKLH DYVAAASS.DKIPLVF.VVGAMAHGKINA	
P.glauca_CO481519	HLPSEARRIGLSYAPKVVKLH DYVAAASS.DKIPLVF.VVGAMAHGKINA	
C.japonica_TA2268_3369	YLPSEARRVGLSHSASKVVQLH DYVEAVR.DNVPLVF.VVGAMSHGKIDA	
M.truncatula_CU012050_19.4	YLPVNSRKTGLSYSSEKLVDMN NYLSTIP.SNQDLVF.VVGAMAHGKVET	
L.japonicus_TC43316	YLPVNSRKIGLSFSSEKLVNM NDYVSTVP.SDMNLVF.VVGAMAHGKVET	
G.max_Glyma12g02820.2	YLPINSRKIGLSYSSEKLVDM DDYVSTVP.SNMDLVF.V.....	
G.max_TC278049	YLPINSRKIGLSYSSEKLVDM DDYVSTVP.SNMDLVF.VVGAMAHGKIET	
G.max_Glyma12g02820.5	YLPINSRKIGLSYSSEKLVDM DDYVSTVP.SNMDLVF.VVGAMAHGKIET	
G.max_Glyma11g10520.1	YLPINSRKIGLSYSSEKLVDM DDYVSTVT.SNMDLVF.VVGAMAHGKIET	
G.max_Glyma12g02820.1	YLPINSRKIGLSYSSEKLVDM DDYVSTVP.SNMDLVF.VVGAMAHGKIET	
P.coccineus_TA3737_3886	YLPVNSRKIGLSYSSEN LVDMDDYVSTIS.SNTDLGFGGKEPMAHG~~~~	
P.coccineus_TC15108	YLPVNSRKIGLSYSSEN LVDMDDYVSTIS.SNTDLGFGGK~~~~~	
P.vulgaris_TC10901	YLPVNSRKIGLSYSSEN LVDMDDYVSTIS.SNTDLVF.VVGAMAHGKIET	
C.sinensis_TC22219	YLPVNSRKIGFSYSSEKLVKMR NCVAPIS.DDDNLVF.VVGAMAHGKIDC	
G.hirsutum_ES833864	YFPVNSRKIGFSYSSSDKLVKMR KYVDAVG.DDVNLVF.VVGAMAHGKIEV	
G.raimondii_TC1038	YFPVNSRKIGFSYSSSDKLAKMR KYVDAVG.DDVNLVF.VVGAMAHGKIEV	
P.trichocarpa_667309	HLPLNSRKIGFSHS SDKFVEMEKYVAGVG.GDSDLVF.VVGAMSHGKIEC	
P.trichocarpa_831805	YLPLNSRKIGFSHSSEKLVQME KYVAGVG.DDSDLVF.VVGAMSHGKIEC	
P.trichocarpa EMG1-like	YLPLNSRKIGFSHSSEKLVQME KYVAGVG.DDSDLVF.VVGAMSHGKIEC	
P.trifoliata_TA8498_37690	YLSVNSRKIGFSYSSEKLVKMR NYVASIS.DDDNLVF.VVGAMAHGKNRL	
G.hirsutum_TC158561	YFPVNSRKIGFSYSSSDKLAKMR KYVDAVG.DDGNLVF.VVGAMAHGKIEV	
E.esula_TC3222	YLPLNSHKIGFSFSXXXLVKM NKYVTTV~~~~~	
E.esula_DV156953	YLPLNSHKIGFSFSXXXLVKM NKYVTTVG.DDKDLVF.VVGAMAH~~~~~	
V.vinifera_GSVIVT00033904001	YLPVNSRKIGFSYSSEKS IPMRDYVAAVS.DDVDLVF.VVGAMAHGKISN	
B.napus_TC70157	HLPPNCRKIGFSYSSEKLVNMQ KHLSTVC.NETDTVF.VLGAMSHGKIEC	
B.napus_TC76938	HLPPNCRKIGFSYSSEKLVNMQ KHLSTVC.NETDTVF.VLGAMSHGKIEC	
A.thaliana_AT3G57000.1	HLPVNSHRIGFSHSSEKLVNMQ KHLATVCDDDRDTVF.VVGAMAHGKIDC	
A.lyrata_486143	HLPVNSHRTGFSHSSEKLVNMQ KHLATIS.DDSDLVF.VVGAMAHGKIEC	
P.americana_CV001201	HLPVNSRKIGFSYSSEKLVQLR ~~~~~	
M.crystallinum_TC9042	YLPLNSRKIGFSYSSEKVS ~~~~~	
V.carteri_106743	YLPLDCKRVGFSFAAAKRVAI HDYVAAMD.DSKPVVF.VVGAFAGHGRIEA	
C.reinhardtii_178371	HLPLNCKRVGFSFAADKHVAI HEYVAGLE.DSGPIVF.VVGAFAGHGRIEA	
M.sp_77569	YFPAGARRVGFSFSAADV KDFKSYINELP.SDACVIF.TLGAMSHGMIDS	
O.lucimarinus_39810	YFPAGARRVGFSFSAPEVKKL PEYVEALP.EDAAVVF.TIGAMAHGKVDV	
O.sp_18991	YFPAGARRVGFSFSAPEVKKL PDYVKS LP.EDATVVF.TVGAMAHGKVDV	
O.auri_37777	YFPAGARRVGFSFSAPEVKKL PEYVRALP.EDATVVF.TVGAMAHGKVDV	
A.officinalis_TA1655_4686	HLPCQFTK.....	

FIGURE 12 (continued)

27/70

	401	436
T.aestivum_TC329781	~~~~~	
T.aestivum_TC342076	EYSDDYIQISGYPLSAACCLNRICSALEQKWNIIQ~~	
H.vulgare_TC168675	EYSDDYIQISGYPLSAACCLNRICSALEQKWSIIQ~~	
O.sativa_LOC_Os02g18830.1	EYSDDYIQICNYPLSAACCLNRICSALEQKWNIIQ~~	
P.virgatum_TC36436	EYTDDYIQICNYPLSAACCLNRICSALEQKWNIR~~	
Z.mays_GRMZM2G130339_T02	EYTDDYIQICNYPLSAACCLNRICSALEQKWNIIQ~~	
S.bicolor_Sb04g033060.1	EYTDDYIQICNYPLSAACCLNRICSALEQKWNIIQ~~	
S.bicolor_Sb04g011280.1	EYTDDYIQICNYPLSAACCLNRICSALEQKWNIIQ~~	
C.longa_TA2440_136217	EYIDDFVSVSKYPLSAACCLNRICSAFEQKWKIT~~	
Y.filamentosa_DT581455	DYTDDFISISNFQLSAACCLGRICNALEQKWKIL~~	
A.sp_TC27214	DYTDDFISVSGYPLSAACCISRICNTLELKWKIV~~	
P.patens_46413	DYIDDLVAVSAYPLSAMWAIARICNALETQWNI~~~	
P.patens_TC42118	DYIDDLVAVSAYPLSAMWAIARICNALETQWNIV~~	
S.moellendorffii_407397	DYIDDLVAVSEYPLSAACCIGRICNALEQHWQII~~	
A.trichopoda_CK754382	DYIDDFISVSEYPLSAACCIGRICNVVERKWNIL~~	
P.taeda_TA10927_3352	DYIDDFISISEYPLSAACCIGRICNAVEQKWKIL~~	
P.glauca_TA18327_3330	DYIDDFISISEYPLSAACCIGRICNAVEQKWKIL~~	
P.sitchensis_TA12989_3332	DYIDDFISISEYPLSAACCIGRICNAVEQKWKIL~~	
P.glauca_CO481519	XYIDDFISISXYPLSAACCIGRICNAVXXXXXXXXXE	
C.japonica_TA2268_3369	DYLDDFISISQYPLSAACCIGRICNAIERKWKIL~~	
M.truncatula_CU012050_19.4	DYTEDYIAVSGYPLSAAYCITRITNAIEGKWNIL~~	
L.japonicus_TC43316	DYTEDYIAVSGYPLSAAYCITRITNAIEGKWKIL~~	
G.max_Glyma12g02820.2	EQWPMGRLRQIIQ.....KIM~~	
G.max_TC278049	DYTEDYVAISGYPLSAAYRITRITGALERKWKIL~~	
G.max_Glyma12g02820.5	DYTEDYVAISGYPLSAAYCITRITGALERKWKIL~~	
G.max_Glyma11g10520.1	DYTEDYVAISGYPLSAAYCITRITGALERKWKIL~~	
G.max_Glyma12g02820.1	DYTEDYVA.SVY~~~~~	
P.coccineus_TA3737_3886	~~~~~	
P.coccineus_TC15108	~~~~~	
P.vulgaris_TC10901	DYTEDYLAISGYPLSAAYCITRITGALERKWKIL~~	
C.sinensis_TC22219	DYTDDLIAISGYPLSAACCIARICEALEDKWNLL~~	
G.hirsutum_ES833864	DYIDDFIAISGYPLSAAMCIARITEALADKWSIL~~	
G.raimondii_TC1038	DYIDDFIAISGYPLSAAMCIARITEALADKWSIL~~	
P.trichocarpa_667309	DYIDDFISVSEYPLSAAWCIARICEAVSKKWRVL~~	
P.trichocarpa_831805	DYVDDFISVSEYPLSAAWCIARVCEAVSKKWRVL~~	
P.trichocarpa EMG1-like	DYVDDFISVSEYPLSAAWCIARVCEAVSKKWRVL~~	
P.trifoliata_TA8498_37690	~~~~~	
G.hirsutum_TC158561	DYIDDFIA.SKFEHISYLYAC~~~~~	
E.esula_TC3222	~~~~~	
E.esula_DV156953	~~~~~	
V.vinifera_GSVIVT00033904001	DYTDDFIAISGYPLSAAWCTYLICGALMQKWNIF~~	
B.napus_TC70157	DYIDDFVAISGYPLSAAYCISRICEALSTNWNII~~	
B.napus_TC76938	DYIDDFVAISGYPLSAAYCISRICEALSTNWNII~~	
A.thaliana_AT3G57000.1	NYIDEFVSVSEYPLSAAYCISRICEALATNWNII~~	
A.lyrata_486143	DYIDEFLSISGYPLSAAYCISRICEALATNWNII~~	
P.americana_CV001201	~~~~~	
M.crystallinum_TC9042	~~~~~	
V.carteri_106743	PWVDEELNVSEYPLSAAYCISRITNAFEMKWKIV~~	
C.reinhardtii_178371	PWVDEELNISEYPLSAAYCLARITQAFEMKWKIV~~	
M.sp_77569	TYTDDLISVSHYPLSGACCIARITNCLNKFGIL~~	
O.lucimarinus_39810	AYTDDFISVSRYPPLSAACCIGRICNALELKHDIIV~~	
O.sp_18991	AYTDDFISVSRYPPLSAACCIGRICNALELKHDIIV~~	
O.auri_37777	AYTDDFISVSKYPLSAACCLGRICNALELKHDIIV~~	
A.officinalis_TA1655_4686	NWLII~~	

FIGURE 12 (continued)

28/70

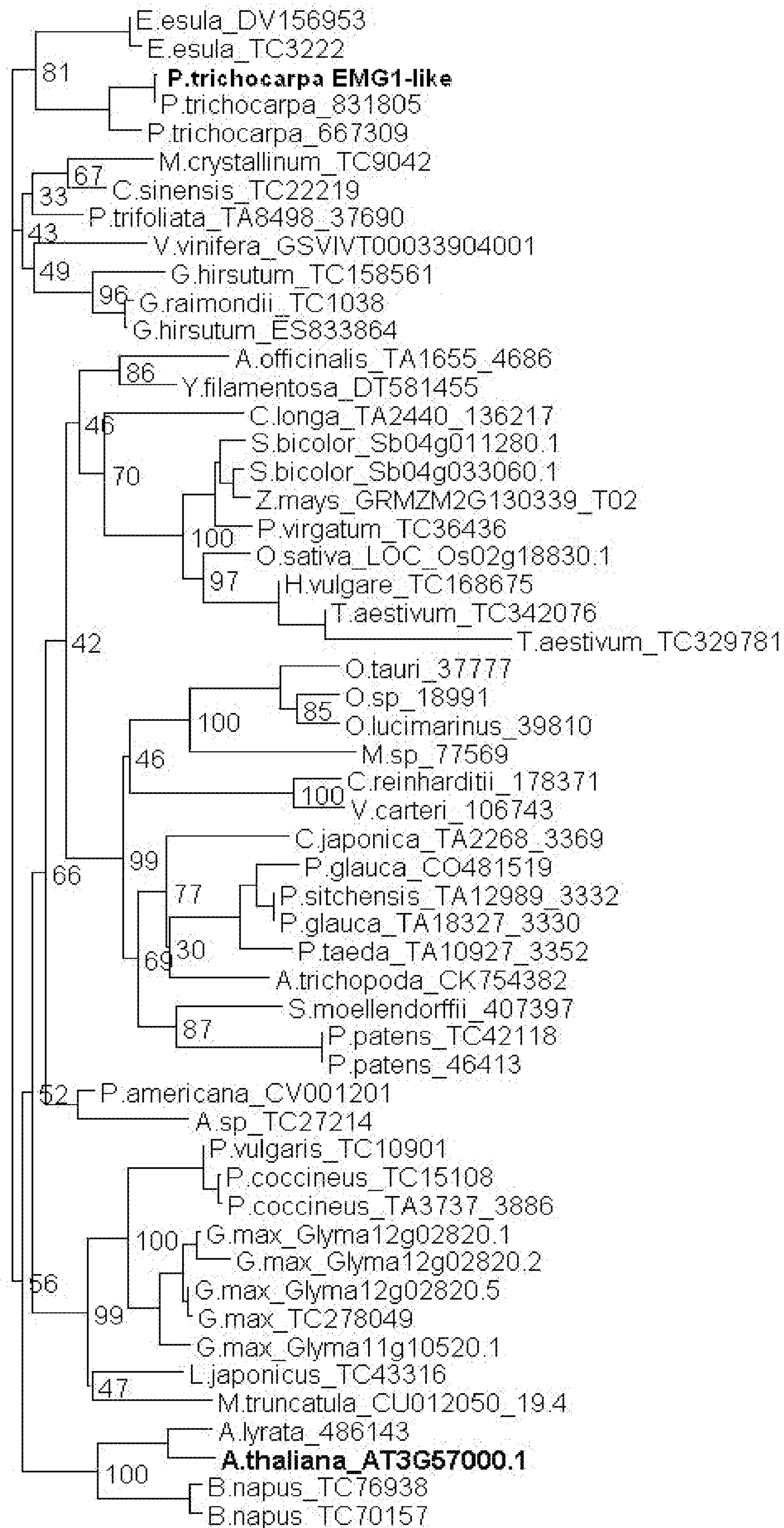


FIGURE 13

29/70

	1	2	3	4	5	6	7	8	9
1. T.aestivum_TC329781		37.5	35.1	28.9	27.3	26.2	26.1	26.7	19.8
2. T.aestivum_TC342076	38.9		96.1	85.6	82.5	84.4	84.0	84.0	69.3
3. H.vulgare_TC168675	37.3	97.1		84.9	81.8	82.5	82.9	82.9	69.3
4. O.sativa_LOC_Os02g18830.1	32.0	90.1	89.1		88.2	85.8	86.7	86.4	68.9
5. P.virgatum_TC36436	30.8	88.5	88.5	92.7		92.3	91.3	93.0	70.1
6. Z.mays_GRMZM2G130339_T02	28.3	89.5	87.8	92.3	95.5		95.1	94.4	69.7
7. S.bicolor_Sb04g033060.1	28.2	88.6	86.8	91.2	93.7	96.5		93.0	70.4
8. S.bicolor_Sb04g011280.1	28.8	87.7	86.3	90.5	95.8	96.2	94.7		69.9
9. C.longa_TA2440_136217	24.0	79.2	79.2	79.5	81.3	80.9	80.6	79.9	
10. A.officinalis_TA1655_4686	33.9	61.1	59.9	60.2	59.4	59.1	60.0	59.3	61.5
11. Y.filamentosa_DT581455	34.4	36.4	36.9	37.0	36.0	36.0	36.8	36.1	37.2
12. M.sp_77569	25.1	58.9	59.1	58.8	58.4	58.7	59.6	57.9	58.7
13. O.lucimarinus_39810	22.5	59.3	59.1	58.5	58.4	58.4	59.6	58.9	57.6
14. O.sp_18991	22.2	59.6	60.2	60.2	59.8	59.8	61.1	60.0	58.7
15. O.tauri_37777	21.4	60.0	59.9	59.5	58.7	58.7	59.3	60.0	59.0
16. V.carteri_106743	24.3	61.8	61.3	59.5	58.4	58.7	61.1	59.6	60.8
17. C.reinhardtii_178371	22.6	60.7	60.6	60.2	58.4	60.1	61.4	60.4	59.4
18. P.patens_46413	34.5	40.7	40.9	40.1	39.9	39.9	40.7	40.0	39.6
19. P.patens_TC42118	22.9	66.8	65.2	65.8	66.1	65.4	66.1	65.6	64.6
20. S.moellendorffii_407397	24.0	64.3	63.8	62.0	62.6	62.9	63.9	63.2	63.9
21. A.trichopoda_CK754382	32.1	30.0	30.5	29.9	29.4	29.4	30.4	29.5	30.9
22. P.taeda_TA10927_3352	22.0	68.2	67.7	68.3	67.8	67.1	68.2	68.1	68.8
23. P.glauca_TA18327_3330	34.8	31.1	31.2	31.0	30.4	30.4	31.4	30.5	32.3
24. P.sitchensis_TA12989_3332	22.4	68.6	67.4	69.0	66.4	66.4	68.9	67.7	68.1
25. P.glauca_CO481519	22.2	65.0	65.2	67.3	65.7	65.0	66.4	66.0	66.0
26. C.japonica_TA2268_3369	24.0	66.4	66.3	65.5	63.6	64.3	66.4	64.9	65.3
27. P.americana_CV001201	31.3	19.3	19.7	19.4	19.2	19.2	19.6	19.3	19.4

FIGURE 14

30/70

	1	2	3	4	5	6	7	8	9
28. A.sp_TC27214	24.9	75.4	76.5	76.8	75.8	76.1	75.1	76.1	79.5
29. M.truncatula_CU012050_19.4	22.0	73.8	74.1	71.8	72.4	73.4	74.5	74.4	72.9
30. L.japonicus_TC43316	23.5	72.9	73.5	74.3	72.7	74.8	74.3	73.0	73.3
31. G.max_Glyma12g02820.2	26.3	65.0	65.9	64.8	64.7	64.0	64.6	65.3	66.0
32. G.max_Glyma12g02820.1	18.1	51.6	52.1	52.1	52.7	52.7	52.1	53.2	53.7
33. G.max_TC278049	23.6	74.6	75.6	73.6	73.4	72.7	73.6	74.0	74.7
34. G.max_Glyma12g02820.5	23.6	75.0	76.0	73.9	73.8	73.1	73.9	74.4	75.0
35. G.max_Glyma11g10520.1	22.7	75.5	74.8	75.4	72.0	72.7	73.8	74.0	74.3
36. P.coccineus_TA3737_3886	27.0	64.6	63.8	62.7	65.0	64.0	63.6	62.8	62.5
37. P.coccineus_TC15108	27.3	63.2	62.4	61.6	64.0	62.9	62.1	61.8	60.8
38. P.vulgaris_TC10901	22.6	74.2	73.9	73.9	74.9	74.9	73.9	74.2	72.9
39. M.crystallinum_TC9042	32.5	59.6	60.9	59.2	58.4	58.7	60.7	58.9	59.4
40. C.sinensis_TC22219	33.9	32.9	33.3	31.7	31.8	31.8	32.9	31.9	31.9
41. P.trifoliata_TA8498_37690	29.1	59.6	60.2	58.8	58.7	58.7	60.0	58.9	58.7
42. G.hirsutum_TC158561	26.3	72.1	72.0	71.5	70.6	71.0	71.4	71.9	73.3
43. G.hirsutum_ES833864	34.8	30.7	31.2	30.3	29.7	29.7	30.7	29.8	30.9
44. G.raimondii_TC1038	33.0	30.4	30.8	29.9	29.4	29.4	30.4	29.5	30.6
45. E.esula_TC3222	30.5	58.9	59.9	59.9	58.4	59.1	59.6	59.6	60.4
46. E.esula_DV156953	28.9	56.8	57.3	55.6	55.6	56.3	56.4	56.1	57.3
47. P.trichocarpa_667309	35.7	30.7	31.2	30.3	29.7	29.7	30.7	29.8	31.6
48. P.trichocarpa_831805	23.9	74.6	74.6	75.7	73.4	75.9	74.3	75.1	75.0
49. P.trichocarpa EMG1-like	23.9	75.0	73.9	75.7	73.4	75.5	73.2	75.1	75.0
50. B.napus_TC70157	24.7	72.3	71.6	71.9	71.6	70.9	70.5	72.3	72.6
51. B.napus_TC76938	23.1	70.8	71.2	71.9	70.5	70.8	70.5	70.5	71.9
52. A.thaliana_AT3G57000.1	24.5	69.5	69.8	69.8	67.4	69.1	68.1	69.1	69.8
53. A.lyrata_486143	23.8	70.1	70.1	70.5	69.5	69.5	69.1	70.5	70.8
54. V.vinifera GSVIVT00033904001	24.4	77.1	77.4	76.1	74.8	75.5	75.7	75.8	75.0

FIGURE 14 (continued)

31/70

	10	11	12	13	14	15	16	17	18
1. T.aestivum_TC329781	28.6	13.8	16.1	12.2	14.2	12.9	11.3	11.7	14.0
2. T.aestivum_TC342076	52.9	31.8	45.0	45.9	46.8	46.4	45.7	46.8	31.4
3. H.vulgare_TC168675	51.5	31.9	45.2	46.1	47.7	46.6	45.9	47.0	31.2
4. O.sativa_LOC_Os02g18830.1	51.2	32.7	44.7	44.4	46.8	45.4	47.2	47.9	31.0
5. P.virgatum_TC36436	51.0	31.8	45.8	45.8	47.6	46.2	45.5	45.8	31.5
6. Z.mays_GRMZM2G130339_T02	52.0	31.5	45.5	46.2	47.6	46.5	45.8	47.2	31.5
7. S.bicolor_Sb04g033060.1	51.7	32.5	46.8	47.5	48.6	46.8	47.5	48.6	32.1
8. S.bicolor_Sb04g011280.1	52.4	31.9	45.3	46.3	48.1	46.7	46.3	46.7	31.6
9. C.longa_TA2440_136217	51.7	33.3	42.0	43.1	43.4	44.1	47.2	46.5	30.6
10. A.officinalis_TA1655_4686		13.8	31.4	31.4	33.1	31.1	32.8	32.8	17.5
11. Y.filamentosa_DT581455	22.8		29.1	32.0	32.0	32.6	28.3	29.1	49.7
12. M.sp_77569	56.3	39.9		70.0	68.4	66.1	55.7	55.2	40.4
13. O.lucimarinus_39810	53.6	43.7	81.2		91.1	87.9	56.5	55.7	40.5
14. O.sp_18991	52.4	42.7	81.8	96.0		87.6	55.7	54.3	40.4
15. O.tauri_37777	52.7	42.9	80.4	94.6	95.6		54.8	56.5	40.2
16. V.carteri_106743	53.0	39.1	73.0	75.7	75.2	75.2		90.0	38.3
17. C.reinhardtii_178371	53.9	38.3	72.6	74.8	74.8	75.2	95.7		38.3
18. P.patens_46413	28.1	63.5	50.2	55.0	54.7	54.9	48.7	49.1	
19. P.patens_TC42118	55.1	38.8	70.6	72.2	73.5	73.9	71.8	72.7	60.4
20. S.moellendorffii_407397	56.3	43.2	75.5	77.3	77.7	79.5	75.2	76.1	55.0
21. A.trichopoda_CK754382	22.3	76.2	36.8	40.5	40.0	40.2	37.8	37.8	61.5
22. P.taeda_TA10927_3352	56.0	42.0	70.4	70.8	70.8	70.4	72.8	72.4	50.4
23. P.glauca_TA18327_3330	22.8	77.9	37.2	41.9	41.3	41.5	40.0	38.7	61.5
24. P.sitchensis_TA12989_3332	56.4	42.4	72.4	70.8	71.6	70.8	73.2	72.4	50.4
25. P.glauca_CO481519	56.0	38.9	69.8	68.7	69.0	68.3	69.8	69.0	48.0
26. C.japonica_TA2268_3369	55.0	42.1	74.4	74.8	74.8	73.6	74.4	73.1	52.9
27. P.americana_CV001201	19.6	46.7	23.3	25.2	24.4	24.1	22.2	22.6	36.5

FIGURE 14 (continued)

32/70

	10	11	12	13	14	15	16	17	18
28. A.sp_TC27214	62.5	37.2	58.0	58.7	59.4	60.4	61.1	61.4	41.3
29. M.truncatula_CU012050_19.4	58.9	35.5	61.7	60.3	61.0	61.0	62.4	61.7	40.8
30. L.japonicus_TC43316	58.1	37.9	61.8	63.2	63.6	61.4	65.1	64.3	42.6
31. G.max_Glyma12g02820.2	66.4	30.4	59.5	58.3	60.3	59.1	60.3	62.3	35.2
32. G.max_Glyma12g02820.1	43.1	22.1	41.2	39.9	41.0	41.2	42.0	42.8	26.1
33. G.max_TC278049	60.1	36.6	61.2	59.8	61.6	60.5	63.4	64.5	40.9
34. G.max_Glyma12g02820.5	59.8	37.0	61.6	60.1	62.0	60.9	63.8	64.9	40.9
35. G.max_Glyma11g10520.1	57.8	36.5	59.2	59.2	60.6	58.9	62.8	62.8	40.1
36. P.coccineus_TA3737_3886	66.5	27.0	56.0	54.8	56.5	54.4	58.1	58.5	33.5
37. P.coccineus_TC15108	68.6	26.4	56.2	56.2	57.0	55.4	58.3	58.7	33.5
38. P.vulgaris_TC10901	57.5	35.2	58.9	58.5	59.6	57.8	61.3	61.3	39.4
39. M.crystallinum_TC9042	75.0	30.2	57.0	55.4	53.8	55.4	56.1	56.1	32.1
40. C.sinensis_TC22219	24.6	77.0	35.9	37.8	37.3	37.1	35.2	35.7	56.8
41. P.trifoliata_TA8498_37690	62.1	41.8	62.8	63.1	62.2	62.5	63.9	62.6	48.1
42. G.hirsutum_TC158561	63.5	32.1	57.3	59.1	58.0	58.8	59.1	59.9	38.7
43. G.hirsutum_ES833864	23.7	73.8	35.4	37.8	36.0	36.6	35.7	36.5	56.8
44. G.raimondii_TC1038	24.1	73.0	35.4	37.8	36.0	36.6	35.2	36.1	56.1
45. E.esula_TC3222	75.0	25.9	60.5	56.3	56.0	57.1	59.1	59.1	33.6
46. E.esula_DV156953	61.2	35.1	62.3	60.8	59.1	60.3	60.0	61.7	43.3
47. P.trichocarpa_667309	24.6	73.0	34.1	36.0	35.1	35.3	36.1	37.8	56.8
48. P.trichocarpa_831805	62.0	35.6	59.5	57.7	59.2	59.2	64.1	65.5	41.5
49. P.trichocarpa EMG1-like	61.6	35.6	59.5	57.7	59.2	59.2	64.1	65.8	41.5
50. B.napus_TC70157	60.3	32.9	56.8	56.8	56.5	56.2	59.9	60.6	39.0
51. B.napus_TC76938	59.7	32.5	57.6	56.3	55.9	55.6	59.3	61.0	38.6
52. A.thaliana_AT3G57000.1	55.7	30.9	55.0	55.0	55.4	55.0	59.7	60.4	38.3
53. A.lyrata_486143	55.4	30.9	53.7	53.7	54.4	54.7	57.4	58.7	37.9
54. V.vinifera_GSVIVT00033904001	62.0	38.7	60.9	60.9	60.1	60.5	62.4	63.1	41.3

FIGURE 14 (continued)

33/70

	19	20	21	22	23	24	25	26	27
1. T.aestivum_TC329781	11.8	14.8	11.5	12.6	12.9	13.0	12.9	11.6	16.1
2. T.aestivum_TC342076	51.8	53.2	24.3	55.0	26.4	55.7	53.2	53.2	17.1
3. H.vulgare_TC168675	50.9	53.0	24.0	55.6	26.5	55.2	52.7	52.7	17.2
4. O.sativa_LOC_Os02g18830.1	51.0	50.4	24.3	56.3	26.4	56.3	54.2	52.8	16.9
5. P.virgatum_TC36436	52.1	52.1	24.1	55.2	25.9	54.9	53.1	51.0	16.8
6. Z.mays_GRMZM2G130339_T02	51.9	51.7	24.1	54.1	25.5	54.9	52.6	53.1	16.4
7. S.bicolor_Sb04g033060.1	51.6	53.9	25.0	55.4	26.8	56.1	53.7	53.9	16.8
8. S.bicolor_Sb04g011280.1	51.7	51.6	24.2	54.0	26.0	55.1	52.8	53.3	16.5
9. C.longa_TA2440_136217	51.0	49.8	25.7	52.1	26.4	53.5	51.0	50.3	18.1
10. A.officinalis_TA1655_4686	35.8	36.9	11.9	37.3	12.8	37.5	36.9	35.2	16.7
11. Y.filamentosa_DT581455	30.2	37.1	66.4	35.6	66.4	36.8	33.3	35.5	42.6
12. M.sp_77569	57.1	56.8	26.9	56.4	26.9	55.2	53.2	56.2	15.7
13. O.lucimarinus_39810	55.9	60.7	29.7	56.0	31.1	54.8	53.2	57.4	15.8
14. O.sp_18991	56.7	61.1	30.2	56.0	30.2	55.2	53.6	56.6	16.0
15. O.tauri_37777	57.1	60.9	29.9	54.0	29.9	53.2	51.6	55.0	15.2
16. V.carteri_106743	57.1	59.3	28.7	58.0	31.3	58.0	55.2	57.4	14.8
17. C.reinhardtii_178371	58.3	59.4	27.4	57.6	30.9	58.8	56.0	56.6	15.7
18. P.patens_46413	60.4	49.8	52.3	39.6	49.7	40.8	38.9	40.5	27.7
19. P.patens_TC42118		69.0	31.8	62.1	30.2	63.5	61.4	62.7	16.7
20. S.moellendorffii_407397	76.3		36.7	66.4	37.6	68.4	63.8	64.0	19.2
21. A.trichopoda_CK754382	37.6	40.6		34.4	81.1	36.0	32.9	35.5	29.5
22. P.taeda_TA10927_3352	79.6	77.6	39.2		41.2	91.6	86.9	72.6	16.4
23. P.glauca_TA18327_3330	37.6	43.2	91.0	43.2		44.4	40.5	38.4	27.0
24. P.sitchensis_TA12989_3332	78.4	77.6	40.4	94.8	44.4		95.2	72.8	17.2
25. P.glauca_CO481519	75.4	75.0	36.9	90.1	40.5	95.2		69.8	17.1
26. C.japonica_TA2268_3369	80.8	78.9	41.7	83.2	43.8	84.0	80.2		16.9
27. P.americana_CV001201	22.0	22.7	40.5	21.2	39.6	22.0	21.8	22.7	

FIGURE 14 (continued)

34/70

	19	20	21	22	23	24	25	26	27
28. A.sp_TC27214	66.9	64.5	32.4	67.6	32.4	69.3	67.2	68.6	21.2
29. M.truncatula_CU012050_19.4	67.4	62.1	30.9	67.0	29.8	67.7	65.6	67.0	19.5
30. L.japonicus_TC43316	68.4	64.7	32.7	72.4	32.4	73.2	70.2	71.7	20.2
31. G.max_Glyma12g02820.2	62.8	60.7	26.3	66.0	24.7	67.2	64.3	66.0	23.5
32. G.max_Glyma12g02820.1	44.9	42.3	18.6	45.5	17.6	46.0	45.7	46.0	14.9
33. G.max_TC278049	65.6	62.0	31.2	68.8	30.8	69.6	66.3	68.1	20.3
34. G.max_Glyma12g02820.5	65.6	62.3	31.5	69.2	31.2	69.9	66.7	68.5	20.3
35. G.max_Glyma11g10520.1	66.0	61.3	30.9	68.1	30.9	68.8	66.0	69.1	19.9
36. P.coccineus_TA3737_3886	62.5	56.9	21.4	61.6	20.6	63.6	63.1	61.7	22.2
37. P.coccineus_TC15108	62.0	56.2	21.1	60.4	20.2	63.2	62.7	62.0	23.1
38. P.vulgaris_TC10901	64.8	59.6	29.6	66.2	29.6	67.6	64.8	65.5	19.2
39. M.crystallinum_TC9042	55.9	57.2	25.0	57.6	25.5	57.6	58.7	58.3	25.9
40. C.sinensis_TC22219	34.7	36.7	79.3	34.4	79.3	35.2	32.5	36.4	44.1
41. P.trifoliata_TA8498_37690	62.9	65.9	32.8	60.4	32.8	61.6	61.1	62.8	30.7
42. G.hirsutum_TC158561	64.6	61.7	27.0	65.7	26.3	66.8	65.7	65.3	20.8
43. G.hirsutum_ES833864	34.7	35.8	78.4	34.0	76.6	34.0	31.7	36.8	43.2
44. G.raimondii_TC1038	34.3	35.4	77.5	33.6	75.7	33.6	31.3	36.4	42.3
45. E.esula_TC3222	60.4	59.4	23.6	60.0	25.5	61.6	60.7	60.3	25.0
46. E.esula_DV156953	59.6	62.9	29.4	60.0	27.3	61.6	61.1	62.0	27.8
47. P.trichocarpa_667309	34.7	37.1	78.4	34.4	78.4	34.8	31.3	36.8	42.3
48. P.trichocarpa_831805	68.0	64.4	31.3	71.5	31.3	70.4	67.6	69.4	20.1
49. P.trichocarpa EMG1-like	68.0	64.4	31.3	71.5	31.3	70.8	68.0	68.7	20.1
50. B.napus_TC70157	64.4	60.3	28.8	64.4	28.4	65.8	64.4	63.7	18.8
51. B.napus_TC76938	64.1	60.0	28.5	63.4	28.1	65.1	63.7	63.7	18.6
52. A.thaliana_AT3G57000.1	62.8	59.1	28.5	64.4	28.2	64.4	62.4	62.4	18.1
53. A.lyrata_486143	64.1	58.4	27.2	64.1	27.2	62.8	62.1	61.7	17.1
54. V.vinifera_GSVIVT00033904001	67.9	64.6	32.1	71.2	32.1	72.3	69.7	69.0	21.0

FIGURE 14 (continued)

35/70

	28	29	30	31	32	33	34	35	36
1. T.aestivum_TC329781	20.8	14.7	16.1	19.1	13.0	17.2	17.2	16.5	20.9
2. T.aestivum_TC342076	63.8	56.3	61.1	50.5	37.3	59.3	59.6	59.9	51.7
3. H.vulgare_TC168675	64.5	56.3	60.7	50.0	37.2	58.7	59.1	58.9	50.5
4. O.sativa_LOC_Os02g18830.1	62.8	54.2	60.4	50.0	38.5	58.5	58.8	57.1	49.0
5. P.virgatum_TC36436	61.9	54.9	59.6	47.6	38.8	56.3	56.6	56.6	49.0
6. Z.mays_GRMZM2G130339_T02	62.9	54.9	60.6	47.6	39.0	56.3	56.6	58.4	48.3
7. S.bicolor_Sb04g033060.1	62.1	55.8	61.6	48.1	38.5	56.8	57.2	58.0	49.3
8. S.bicolor_Sb04g011280.1	63.9	56.8	61.2	48.8	39.3	57.5	57.9	58.9	48.8
9. C.longa_TA2440_136217	64.5	54.0	58.7	49.0	38.1	56.2	56.6	57.2	47.3
10. A.officinalis_TA1655_4686	50.7	43.9	44.9	48.7	32.6	43.9	43.9	44.2	50.6
11. Y.filamentosa_DT581455	32.4	28.7	32.0	20.7	16.5	29.7	30.1	29.4	19.8
12. M.sp_77569	44.4	44.3	45.6	37.3	27.1	44.2	44.6	42.6	33.8
13. O.lucimarinus_39810	45.7	42.6	47.1	34.4	25.3	41.7	42.0	41.5	32.8
14. O.sp_18991	46.8	44.0	47.8	35.9	26.8	43.1	43.5	42.9	34.8
15. O.tauri_37777	46.8	42.9	46.3	36.2	27.3	43.5	43.8	42.6	34.1
16. V.carteri_106743	46.8	45.4	47.8	38.4	28.1	46.0	46.4	46.5	36.5
17. C.reinhardtii_178371	46.1	44.3	47.8	39.1	28.6	46.7	47.1	46.5	35.8
18. P.patens_46413	30.7	29.1	31.3	21.5	18.1	30.4	30.4	29.8	19.9
19. P.patens_TC42118	52.2	49.3	50.7	40.6	32.3	49.6	49.6	50.9	40.3
20. S.moellendorffii_407397	51.9	47.2	53.7	40.9	32.1	49.6	50.0	49.6	37.8
21. A.trichopoda_CK754382	28.0	23.0	25.7	15.2	12.5	23.9	24.3	23.8	13.5
22. P.taeda_TA10927_3352	54.3	51.8	57.7	44.9	33.1	54.0	54.3	53.5	43.1
23. P.glauca_TA18327_3330	26.6	22.3	25.0	14.9	12.0	23.9	24.3	23.8	13.5
24. P.sitchensis_TA12989_3332	56.3	52.8	58.8	45.7	33.3	54.7	55.1	54.3	43.1
25. P.glauca_CO481519	54.2	50.3	55.5	44.6	33.4	51.8	52.2	51.4	42.8
26. C.japonica_TA2268_3369	56.0	50.5	55.9	45.1	32.5	53.4	53.8	52.7	40.3
27. P.americana_CV001201	18.8	17.7	18.0	20.6	13.0	17.8	17.8	17.4	19.8

FIGURE 14 (continued)

36/70

	28	29	30	31	32	33	34	35	36
28. A.sp_TC27214		64.5	66.7	55.9	42.3	64.7	65.1	65.4	54.4
29. M.truncatula_CU012050_19.4	80.2		75.5	61.3	46.8	73.0	73.4	73.9	62.0
30. L.japonicus_TC43316	78.8	87.6		65.6	48.3	77.8	78.1	78.0	61.6
31. G.max_Glyma12g02820.2	68.9	72.3	79.4		60.9	85.5	85.1	78.0	77.0
32. G.max_Glyma12g02820.1	57.7	57.7	59.3	62.2		61.9	61.9	57.8	51.4
33. G.max_TC278049	78.5	83.7	89.5	87.3	66.2		99.6	92.2	69.6
34. G.max_Glyma12g02820.5	78.8	84.0	89.9	87.3	66.2	99.6		92.6	69.6
35. G.max_Glyma11g10520.1	78.8	85.1	88.7	82.3	63.8	94.3	94.7		68.8
36. P.coccineus_TA3737_3886	65.2	72.7	73.9	85.5	56.4	77.5	77.5	75.9	
37. P.coccineus_TC15108	64.5	72.3	73.5	85.8	56.1	76.8	76.8	75.5	97.2
38. P.vulgaris_TC10901	77.5	85.0	84.0	76.0	60.1	87.8	88.2	88.9	84.3
39. M.crystallinum_TC9042	63.8	60.3	63.6	68.4	44.9	61.2	61.2	59.9	70.2
40. C.sinensis_TC22219	33.4	33.0	34.2	27.5	20.2	33.7	34.1	33.7	24.6
41. P.trifoliata_TA8498_37690	59.0	61.0	61.8	67.6	45.2	61.6	61.6	60.6	66.1
42. G.hirsutum_TC158561	77.8	75.9	78.5	71.5	57.2	77.2	77.5	77.3	72.6
43. G.hirsutum_ES833864	32.4	32.6	33.5	26.7	20.2	33.7	34.1	33.3	23.4
44. G.raimondii_TC1038	32.1	32.3	33.1	26.3	19.9	33.3	33.7	33.0	23.0
45. E.esula_TC3222	61.8	64.9	65.1	72.5	47.9	64.9	64.9	63.1	71.8
46. E.esula_DV156953	56.0	59.6	61.4	66.0	43.9	59.8	59.8	59.2	66.1
47. P.trichocarpa_667309	32.4	32.6	34.2	27.9	20.7	35.1	35.5	34.8	25.4
48. P.trichocarpa_831805	81.2	79.2	79.2	70.4	55.9	80.3	80.6	80.6	67.6
49. P.trichocarpa EMG1-like	81.2	78.9	79.2	70.4	55.9	80.3	80.6	80.6	67.3
50. B.napus_TC70157	80.2	77.7	78.8	66.8	55.3	76.0	76.4	76.0	64.4
51. B.napus_TC76938	80.3	76.6	78.0	66.4	55.6	75.6	75.9	75.3	63.7
52. A.thaliana_AT3G57000.1	77.5	75.2	74.8	65.8	54.0	74.2	74.5	74.8	62.4
53. A.lyrata_486143	76.5	74.8	74.8	66.4	53.5	75.5	75.8	77.2	62.8
54. V.vinifera_GSVIVT00033904001	79.2	78.0	79.4	70.5	54.5	79.3	79.7	79.1	69.7

FIGURE 14 (continued)

37/70

	37	38	39	40	41	42	43	44	45
1. T.aestivum_TC329781	21.4	17.4	21.9	12.8	16.4	21.2	13.8	14.9	23.1
2. T.aestivum_TC342076	50.9	60.2	47.9	28.9	51.8	56.6	26.4	26.1	46.1
3. H.vulgare_TC168675	49.7	59.5	46.8	28.7	51.6	56.5	26.9	26.5	45.5
4. O.sativa_LOC_Os02g18830.1	48.4	58.5	45.1	27.8	49.3	55.9	25.7	25.4	44.7
5. P.virgatum_TC36436	48.5	58.0	44.8	27.6	49.3	55.0	24.8	24.5	44.8
6. Z.mays_GRMZM2G130339_T02	47.8	57.7	45.3	27.3	49.0	55.4	24.8	24.5	44.8
7. S.bicolor_Sb04g033060.1	48.4	58.3	45.8	28.2	50.4	56.0	25.4	25.0	45.0
8. S.bicolor_Sb04g011280.1	48.3	58.6	45.1	27.7	49.8	55.4	24.9	24.6	46.3
9. C.longa_TA2440_136217	46.2	54.2	44.1	26.7	50.0	59.4	26.4	26.0	46.9
10. A.officinalis_TA1655_4686	51.0	44.4	57.4	12.3	45.2	49.1	12.3	12.8	55.6
11. Y.filamentosa_DT581455	18.9	29.3	15.0	67.2	30.9	26.1	65.6	64.8	16.4
12. M.sp_77569	34.4	41.8	34.5	25.1	45.3	37.7	24.2	24.2	37.2
13. O.lucimarinus_39810	33.3	40.4	34.9	26.1	47.5	40.1	26.1	26.1	35.0
14. O.sp_18991	34.4	42.5	34.5	25.8	47.6	39.5	25.8	25.8	35.4
15. O.tauri_37777	34.0	41.8	33.3	25.4	46.9	40.5	25.9	25.9	34.3
16. V.carteri_106743	35.8	45.3	37.8	23.5	44.8	41.4	24.3	23.9	37.6
17. C.reinhardtii_178371	35.4	44.9	36.7	24.3	45.2	41.0	25.2	24.8	38.0
18. P.patens_46413	19.7	28.6	18.6	40.3	27.5	25.7	40.9	40.3	17.6
19. P.patens_TC42118	40.0	49.8	39.5	24.5	44.9	46.0	24.9	24.5	40.5
20. S.moellendorffii_407397	37.2	47.0	38.5	28.8	49.8	44.0	28.8	28.4	38.0
21. A.trichopoda_CK754382	13.0	22.6	12.3	63.1	22.0	19.9	63.1	62.2	10.9
22. P.taeda_TA10927_3352	42.5	51.6	41.5	28.0	48.0	48.1	28.0	27.6	42.7
23. P.glauca_TA18327_3330	11.6	23.3	13.7	64.0	22.0	19.6	63.1	62.2	15.0
24. P.sitchensis_TA12989_3332	42.8	53.0	41.8	28.4	49.0	49.8	28.0	27.6	43.1
25. P.glauca_CO481519	42.5	50.2	41.9	26.2	48.2	48.8	25.8	25.4	42.8
26. C.japonica_TA2268_3369	40.1	50.5	37.8	27.7	49.6	46.6	28.1	27.7	40.7
27. P.americana_CV001201	20.2	17.1	21.4	36.9	27.5	17.9	34.4	33.6	21.4

FIGURE 14 (continued)

38/70

	37	38	39	40	41	42	43	44	45
28. A.sp_TC27214	54.3	64.2	51.0	28.3	50.3	62.5	26.6	26.3	51.2
29. M.truncatula_CU012050_19.4	61.9	74.4	47.7	28.7	51.1	59.0	27.3	27.0	51.6
30. L.japonicus_TC43316	61.9	74.7	51.8	30.1	53.3	63.3	29.4	29.0	55.5
31. G.max_Glyma12g02820.2	77.0	70.1	56.0	18.5	55.5	59.1	18.1	17.8	59.9
32. G.max_Glyma12g02820.1	51.6	52.6	37.1	15.5	38.3	44.5	15.5	15.3	39.4
33. G.max_TC278049	69.6	83.7	50.2	29.0	52.5	60.8	29.0	28.6	53.6
34. G.max_Glyma12g02820.5	69.6	84.0	50.2	29.3	52.5	61.0	29.3	29.0	53.6
35. G.max_Glyma11g10520.1	69.2	82.8	51.1	28.7	51.8	62.3	28.7	28.4	52.7
36. P.coccineus_TA3737_3886	97.2	83.4	56.5	18.9	54.7	57.5	17.0	16.7	59.3
37. P.coccineus_TC15108		84.0	57.9	17.9	55.0	58.1	16.5	16.1	61.2
38. P.vulgaris_TC10901	84.0		49.1	28.9	50.9	60.7	27.9	27.5	51.9
39. M.crystallinum_TC9042	71.5	60.6		13.0	53.3	53.6	11.8	11.8	64.0
40. C.sinensis_TC22219	24.0	32.8	24.5		32.3	24.6	81.1	80.2	13.5
41. P.trifoliata_TA8498_37690	67.4	59.6	69.3	38.1		57.7	28.7	28.3	58.3
42. G.hirsutum_TC158561	71.5	75.3	66.4	28.8	63.5		29.7	30.0	58.4
43. G.hirsutum_ES833864	23.1	32.1	26.9	88.3	36.5	32.1		99.1	12.2
44. G.raimondii_TC1038	22.7	31.7	26.9	87.4	36.0	32.5	99.1		11.8
45. E.esula_TC3222	74.0	62.7	77.7	24.1	70.5	67.5	23.6	23.2	
46. E.esula_DV156953	66.9	58.9	69.3	30.9	85.6	63.5	33.0	32.5	79.5
47. P.trichocarpa_667309	25.2	34.1	27.4	86.5	35.4	30.3	88.3	87.4	26.4
48. P.trichocarpa_831805	67.3	80.1	62.7	34.2	61.3	79.9	34.9	34.5	67.6
49. P.trichocarpa EMG1-like	66.9	77.7	62.7	34.2	61.3	79.9	34.9	34.5	67.6
50. B.napus_TC70157	64.0	76.4	59.9	32.2	56.5	77.1	32.9	32.5	63.0
51. B.napus_TC76938	63.4	75.6	60.3	31.9	55.9	76.3	32.5	32.2	62.0
52. A.thaliana_AT3G57000.1	62.1	74.5	60.7	31.2	56.4	74.8	31.5	31.2	60.7
53. A.lyrata_486143	62.4	74.5	59.1	31.2	56.0	73.8	31.2	30.9	60.7
54. V.vinifera_GSVIVT00033904001	69.4	78.0	67.9	35.4	64.6	79.9	34.7	34.3	66.4

FIGURE 14 (continued)

39/70

	46	47	48	49	50	51	52	53	54
1. T.aestivum_TC329781	17.0	16.5	18.9	18.9	18.5	18.6	17.3	17.9	18.1
2. T.aestivum_TC342076	46.4	24.6	59.8	60.1	56.6	56.4	54.5	56.2	64.7
3. H.vulgare_TC168675	46.6	24.7	59.6	59.4	55.4	56.1	54.5	55.1	64.4
4. O.sativa_LOC_Os02g18830.1	45.1	23.6	57.5	58.0	55.1	53.7	54.5	55.5	62.4
5. P.virgatum_TC36436	44.8	23.4	57.3	57.3	54.7	53.9	54.5	55.4	60.8
6. Z.mays_GRMZM2G130339_T02	44.8	23.4	59.1	59.1	54.1	55.1	55.5	55.2	61.5
7. S.bicolor_Sb04g033060.1	45.0	23.9	57.4	57.7	55.1	56.4	55.5	55.9	62.1
8. S.bicolor_Sb04g011280.1	45.3	23.5	60.1	60.1	56.1	55.1	56.0	57.3	63.5
9. C.longa_TA2440_136217	47.6	26.0	59.7	59.7	56.7	56.8	57.3	56.9	59.4
10. A.officinalis_TA1655_4686	45.7	11.4	48.6	48.6	48.1	47.5	45.2	45.8	47.4
11. Y.filamentosa_DT581455	24.5	60.7	28.9	28.9	25.7	25.4	25.2	25.5	34.7
12. M.sp_77569	44.3	23.8	43.3	43.3	41.1	41.0	39.6	38.6	43.5
13. O.lucimarinus_39810	43.6	24.3	42.6	42.6	40.8	40.3	40.6	39.6	43.9
14. O.sp_18991	43.8	24.4	44.4	44.4	40.8	40.3	41.9	41.6	45.0
15. O.tauri_37777	43.8	24.6	43.3	43.3	40.1	39.7	40.9	40.6	45.4
16. V.carteri_106743	45.9	24.3	46.5	45.8	43.5	43.4	42.6	43.6	47.2
17. C.reinhardtii_178371	45.8	26.1	45.8	45.8	45.5	45.8	44.3	45.0	48.3
18. P.patens_46413	25.0	42.3	29.6	29.6	30.5	30.2	28.5	28.2	30.6
19. P.patens_TC42118	45.3	25.7	49.6	49.6	53.4	53.2	49.7	50.2	51.7
20. S.moellendorffii_407397	47.6	28.8	48.2	48.2	50.0	49.2	47.7	45.6	51.3
21. A.trichopoda_CK754382	18.5	64.9	25.4	25.4	21.2	21.0	22.8	21.5	25.8
22. P.taeda_TA10927_3352	47.8	27.2	53.2	53.2	50.0	48.5	49.3	49.3	56.5
23. P.glauca_TA18327_3330	17.6	64.0	24.6	24.6	21.6	21.4	21.1	21.1	26.6
24. P.sitchensis_TA12989_3332	48.6	28.4	53.5	53.5	52.7	51.2	50.0	49.0	57.6
25. P.glauca_CO481519	48.2	26.2	51.4	51.4	51.4	49.8	48.3	48.7	54.9
26. C.japonica_TA2268_3369	47.1	28.1	52.5	52.8	48.3	48.1	48.7	48.3	53.9
27. P.americana_CV001201	24.2	32.8	17.6	17.6	16.8	16.6	15.4	14.4	18.8

FIGURE 14 (continued)

40/70

	46	47	48	49	50	51	52	53	54
28. A.sp_TC27214	49.1	26.6	63.6	63.6	63.5	63.1	61.6	62.5	66.6
29. M.truncatula_CU012050_19.4	49.6	25.5	60.4	60.1	61.0	60.7	58.5	59.1	62.1
30. L.japonicus_TC43316	52.9	27.9	63.2	63.3	64.2	63.5	59.5	60.9	67.3
31. G.max_Glyma12g02820.2	55.5	18.5	56.3	56.3	54.6	54.1	51.5	51.8	57.2
32. G.max_Glyma12g02820.1	37.8	15.3	41.4	41.4	40.6	39.9	39.4	40.5	41.9
33. G.max_TC278049	51.4	26.8	63.9	63.9	63.8	63.2	59.2	60.5	66.7
34. G.max_Glyma12g02820.5	51.4	27.2	64.2	64.2	64.2	63.5	59.5	60.9	67.0
35. G.max_Glyma11g10520.1	50.7	26.6	64.5	64.1	63.1	62.1	60.1	62.6	65.6
36. P.coccineus_TA3737_3886	54.8	17.0	54.3	54.0	51.9	50.7	49.8	52.2	53.8
37. P.coccineus_TC15108	54.9	16.8	54.9	54.5	52.1	51.2	49.7	52.3	54.0
38. P.vulgaris_TC10901	49.5	25.8	62.6	61.7	63.0	61.7	58.1	62.8	64.1
39. M.crystallinum_TC9042	53.2	14.0	51.6	51.6	50.7	49.8	48.0	48.0	56.7
40. C.sinensis_TC22219	23.2	73.0	29.6	29.6	26.4	26.1	24.8	25.5	32.8
41. P.trifoliata_TA8498_37690	75.4	26.0	56.0	56.0	46.9	46.8	46.6	47.7	56.3
42. G.hirsutum_TC158561	55.5	23.9	64.7	64.7	64.7	64.4	60.4	62.1	68.6
43. G.hirsutum_ES833864	24.5	75.7	29.6	29.6	26.4	26.1	24.2	24.8	32.1
44. G.raimondii_TC1038	24.0	74.8	29.2	29.2	26.0	25.8	23.8	24.5	32.1
45. E.esula_TC3222	75.3	13.1	58.2	58.2	53.4	52.9	50.3	50.0	54.4
46. E.esula_DV156953		24.0	57.0	57.0	48.3	47.1	46.0	47.7	52.6
47. P.trichocarpa_667309	32.0		35.9	35.9	27.1	26.8	25.2	24.8	29.5
48. P.trichocarpa_831805	62.3	38.0		99.6	65.5	64.1	63.3	64.8	69.5
49. P.trichocarpa EMG1-like	62.3	38.0	99.6		65.5	63.7	63.7	65.1	69.5
50. B.napus_TC70157	56.8	33.6	82.2	82.2		96.6	76.2	77.3	63.7
51. B.napus_TC76938	55.9	33.2	82.0	81.7	99.0		76.6	77.6	63.0
52. A.thaliana_AT3G57000.1	56.0	32.2	81.9	82.2	87.6	89.6		90.3	62.1
53. A.lyrata_486143	56.0	32.2	81.2	81.5	88.6	89.3	95.3		62.8
54. V.vinifera_GSVIVT00033904001	62.0	33.9	82.0	82.0	76.7	75.6	75.5	74.5	

FIGURE 14 (continued)

41/70

	1	2	3	4	5	6	7	8	9
1. T.aestivum_TC329781		16.2	16.2	16.2	16.2	15.7	16.2	16.2	14.1
2. T.aestivum_TC342076	16.2		99.0	93.9	94.4	94.9	94.4	94.4	83.8
3. H.vulgare_TC168675	16.2	99.0		93.4	93.9	94.4	93.9	93.9	83.8
4. O.sativa_LOC_Os02g18830.1	17.2	97.0	96.5		96.5	96.0	96.5	96.5	84.3
5. P.virgatum_TC36436	17.2	98.0	97.5	99.0		99.5	100.0	100.0	83.8
6.Z.mays_GRMZM2G130339_T02	15.7	98.0	97.5	99.0	100.0		99.5	99.5	83.3
7. S.bicolor_Sb04g033060.1	17.2	98.0	97.5	99.0	100.0	100.0		100.0	83.8
8. S.bicolor_Sb04g011280.1	17.2	98.0	97.5	99.0	100.0	100.0	100.0		83.8
9. C.longa_TA2440_136217	15.2	92.4	92.9	91.9	92.4	92.4	92.4	92.4	
10. A.officinalis_TA1655_4686	21.5	63.1	63.6	63.6	63.6	63.6	63.6	63.6	61.6
11. Y.filamentosa_DT581455	16.4	49.5	50.0	51.0	50.0	50.0	50.0	50.0	51.5
12. M.sp_77569	12.3	73.0	73.0	73.5	73.5	73.5	73.5	73.5	74.5
13. O.lucimarinus_39810	13.6	76.8	76.8	76.3	76.3	76.3	76.3	76.3	76.8
14. O.sp_18991	12.1	77.8	77.8	77.3	77.3	77.3	77.3	77.3	78.3
15. O.tauri_37777	13.1	77.8	77.8	77.3	77.3	77.3	77.3	77.3	77.8
16. V.carteri_106743	13.6	76.3	76.3	75.3	75.8	76.3	75.8	75.8	76.8
17. C.reinhardtii_178371	13.6	75.3	75.3	75.3	75.3	75.3	75.3	75.3	75.8
18. P.patens_46413	14.2	55.6	55.6	55.6	55.6	55.6	55.6	55.6	56.1
19. P.patens_TC42118	14.2	77.5	77.5	77.5	77.5	77.5	77.5	77.5	78.4
20. S.moellendorffii_407397	12.1	81.3	80.8	80.3	80.8	80.8	80.8	80.8	81.8
21. A.trichopoda_CK754382	18.0	39.9	40.4	40.4	39.9	39.9	39.9	39.9	42.4
22. P.taeda_TA10927_3352	13.1	81.3	81.3	82.8	82.3	81.8	82.3	82.3	83.8
23. P.glauca_TA18327_3330	17.1	41.9	41.9	42.4	41.9	41.9	41.9	41.9	43.9
24. P.sitchensis_TA12989_3332	13.6	81.8	81.8	82.8	82.3	82.3	82.3	82.3	84.3
25. P.glauca_CO481519	13.7	80.8	80.8	81.8	81.3	81.3	81.3	81.3	82.8
26. C.japonica_TA2268_3369	14.1	81.3	81.3	81.8	80.8	80.8	80.8	80.8	83.3
27. P.americana_CV001201	35.5	26.8	27.3	27.3	27.3	27.3	27.3	27.3	28.3

FIGURE 15

42/70

	1	2	3	4	5	6	7	8	9
28. A.sp_TC27214	14.2	86.3	86.8	86.3	85.8	85.8	85.8	85.8	87.3
29. M.truncatula_CU012050_19.4	12.7	82.8	83.3	82.4	82.8	82.8	82.8	82.8	82.8
30. L.japonicus_TC43316	14.2	82.8	83.3	83.3	82.8	82.8	82.8	82.8	83.3
31. G.max_Glyma12g02820.2	15.4	71.7	72.2	73.7	72.7	72.7	72.7	72.7	73.2
32. G.max_Glyma12g02820.1	15.5	78.3	78.8	80.3	79.3	79.3	79.3	79.3	78.3
33. G.max_TC278049	13.2	81.9	82.4	82.8	81.9	81.9	81.9	81.9	81.9
34. G.max_Glyma12g02820.5	13.2	82.4	82.8	83.3	82.4	82.4	82.4	82.4	82.4
35. G.max_Glyma11g10520.1	13.2	82.4	82.8	83.8	82.8	82.4	82.8	82.8	82.8
36. P.coccineus_TA3737_3886	16.2	69.7	70.2	71.2	70.7	70.7	70.7	70.7	71.2
37. P.coccineus_TC15108	16.8	68.7	69.2	70.2	69.7	69.7	69.7	69.7	69.7
38. P.vulgaris_TC10901	13.2	82.4	82.8	82.8	82.4	82.4	82.4	82.4	82.4
39. M.crystallinum_TC9042	19.9	64.6	65.2	65.7	65.7	65.7	65.7	65.7	65.7
40. C.sinensis_TC22219	16.5	43.4	43.9	42.4	42.9	42.9	42.9	42.9	43.4
41. P.trifoliata_TA8498_37690	16.5	77.3	77.8	77.8	78.3	78.3	78.3	78.3	78.8
42. G.hirsutum_TC158561	14.8	79.3	79.8	79.8	79.8	79.8	79.8	79.8	82.3
43. G.hirsutum_ES833864	16.2	41.4	41.9	41.4	40.9	40.9	40.9	40.9	42.9
44. G.raimondii_TC1038	16.2	40.9	41.4	40.9	40.4	40.4	40.4	40.4	42.4
45. E.esula_TC3222	18.7	64.6	65.2	65.2	65.2	65.2	65.2	65.2	67.2
46. E.esula_DV156953	17.0	70.7	71.2	70.7	71.2	71.2	71.2	71.2	72.7
47. P.trichocarpa_667309	22.3	40.9	41.4	40.9	40.4	40.4	40.4	40.4	42.9
48. P.trichocarpa_831805	15.3	84.3	84.8	84.8	84.3	84.3	84.3	84.3	86.9
49. P.trichocarpa EMG1-like	15.3	84.3	84.8	84.8	84.3	84.3	84.3	84.3	86.9
50. B.napus_TC70157	14.3	82.8	82.8	82.3	82.3	82.3	82.3	82.3	83.3
51. B.napus_TC76938	14.3	82.8	82.8	82.3	82.3	82.3	82.3	82.3	83.3
52. A.thaliana_AT3G57000.1	15.7	82.3	82.8	81.8	81.8	81.8	81.8	81.8	83.3
53. A.lyrata_486143	14.7	82.3	82.8	81.8	81.3	81.3	81.3	81.3	82.8
54. V.vinifera_GSVIVT00033904001	15.3	86.4	86.9	87.4	86.4	86.4	86.4	86.4	87.4

FIGURE 15 (continued)

43/70

	10	11	12	13	14	15	16	17	18
1. T.aestivum_TC329781	18.5	10.7	10.3	10.6	11.1	11.1	11.1	10.6	6.8
2. T.aestivum_TC342076	56.6	41.7	54.9	60.1	61.1	60.1	57.6	58.1	41.4
3. H.vulgare_TC168675	57.1	41.7	54.9	60.1	61.1	60.1	57.6	58.1	41.4
4. O.sativa_LOC_Os02g18830.1	57.1	43.6	54.9	58.6	60.1	59.1	58.6	59.1	41.4
5. P.virgatum_TC36436	57.6	42.6	56.9	60.6	62.1	61.1	58.1	58.6	42.4
6. Z.mays_GRMZM2G130339_T02	57.6	42.2	56.9	60.6	62.1	61.1	58.1	58.6	42.4
7. S.bicolor_Sb04g033060.1	57.6	42.6	56.9	60.6	62.1	61.1	58.1	58.6	42.4
8. S.bicolor_Sb04g011280.1	57.6	42.6	56.9	60.6	62.1	61.1	58.1	58.6	42.4
9. C.longa_TA2440_136217	55.6	44.1	53.4	56.6	57.1	57.6	58.1	58.6	41.9
10. A.officinalis_TA1655_4686		22.1	40.7	41.4	42.4	40.9	40.4	41.4	24.6
11. Y.filamentosa_DT581455	36.3		31.4	33.3	33.8	34.3	29.4	30.4	49.7
12. M.sp_77569	54.4	43.6		71.6	70.6	68.1	58.3	56.9	43.6
13. O.lucimarinus_39810	53.5	47.5	81.9		93.9	90.9	60.1	59.1	43.3
14. O.sp_18991	53.5	47.0	82.4	98.5		90.9	59.1	57.6	43.8
15. O.tauri_37777	53.5	47.0	81.9	97.0	99.0		58.6	58.6	43.3
16. V.carteri_106743	52.5	42.9	74.0	79.3	79.8	79.8		90.9	41.9
17. C.reinhardtii_178371	53.5	41.9	73.5	78.8	79.3	79.3	97.0		41.9
18. P.patens_46413	42.6	63.5	54.9	60.1	60.6	60.6	55.1	55.6	
19. P.patens_TC42118	53.9	46.6	79.9	80.9	81.4	81.4	77.0	77.5	72.5
20. S.moellendorffii_407397	55.6	47.5	78.9	84.3	86.4	85.9	80.8	81.3	62.1
21. A.trichopoda_CK754382	35.6	76.2	40.2	43.4	43.4	43.4	41.9	41.9	61.5
22. P.taeda_TA10927_3352	57.6	50.0	77.5	82.3	82.8	82.3	82.3	81.3	62.1
23. P.glauca_TA18327_3330	35.6	77.9	40.7	45.5	45.5	45.5	43.9	42.4	61.5
24. P.sitchensis_TA12989_3332	57.1	50.5	78.9	82.3	82.8	82.8	81.8	80.8	62.1
25. P.glauca_CO481519	56.3	49.7	78.4	81.8	81.8	81.8	80.8	79.8	61.4
26. C.japonica_TA2268_3369	55.1	48.5	80.9	84.8	84.8	83.8	82.3	80.8	63.1
27. P.americana_CV001201	32.6	46.7	25.5	27.3	26.3	26.3	25.8	26.3	36.5

FIGURE 15 (continued)

44/70

	10	11	12	13	14	15	16	17	18
28. A.sp_TC27214	60.3	53.4	77.9	78.9	78.9	78.9	77.0	77.5	59.3
29. M.truncatula_CU012050_19.4	57.8	49.0	78.4	76.0	77.0	76.5	77.0	76.0	56.4
30. L.japonicus_TC43316	57.8	50.5	78.4	77.5	77.9	77.5	76.5	76.5	56.9
31. G.max_Glyma12g02820.2	66.9	42.9	66.7	66.7	67.2	66.7	66.2	67.7	49.7
32. G.max_Glyma12g02820.1	65.7	45.9	69.6	71.7	72.2	71.7	70.2	72.2	54.1
33. G.max_TC278049	57.4	49.5	77.5	75.5	76.0	75.5	76.0	76.5	55.4
34. G.max_Glyma12g02820.5	57.8	50.0	77.9	76.0	76.5	76.0	76.5	77.0	55.4
35. G.max_Glyma11g10520.1	57.8	50.5	77.5	75.5	76.0	75.5	76.5	77.0	55.4
36. P.coccineus_TA3737_3886	68.3	40.1	63.2	63.1	63.6	63.1	64.1	65.7	49.7
37. P.coccineus_TC15108	72.0	39.8	61.8	64.1	62.6	63.6	63.1	64.1	50.3
38. P.vulgaris_TC10901	57.8	49.5	77.9	75.0	75.5	75.0	76.0	76.5	55.4
39. M.crystallinum_TC9042	84.4	37.6	56.4	56.6	56.6	56.6	58.6	59.6	45.9
40. C.sinensis_TC22219	31.1	72.1	36.3	39.4	39.4	38.9	37.9	38.4	52.7
41. P.trifoliata_TA8498_37690	71.8	46.5	66.2	66.7	67.2	66.7	69.2	68.7	53.5
42. G.hirsutum_TC158561	63.0	46.6	71.1	75.8	74.7	75.3	73.7	75.3	56.1
43. G.hirsutum_ES833864	31.1	73.8	38.7	40.9	39.4	39.9	39.4	40.4	56.8
44. G.raimondii_TC1038	31.1	73.0	38.7	40.9	39.4	39.9	38.9	39.9	56.1
45. E.esula_TC3222	80.0	38.0	59.3	58.6	58.6	58.6	60.1	61.1	49.3
46. E.esula_DV156953	71.5	41.2	61.8	63.6	63.6	63.6	65.2	65.7	50.9
47. P.trichocarpa_667309	31.9	67.2	35.3	38.4	37.9	37.9	39.4	41.4	54.1
48. P.trichocarpa_831805	61.2	48.0	74.5	77.3	77.3	77.8	79.8	82.3	58.2
49. P.trichocarpa EMG1-like	61.2	48.0	74.5	77.3	77.3	77.8	79.8	82.3	58.2
50. B.napus_TC70157	60.7	46.9	74.5	75.8	75.8	75.8	77.8	79.8	55.6
51. B.napus_TC76938	60.7	46.9	74.5	75.8	75.8	75.8	77.8	79.8	55.6
52. A.thaliana_AT3G57000.1	58.1	44.4	72.5	75.3	75.3	75.3	78.8	80.3	55.1
53. A.lyrata_486143	57.9	44.7	71.1	73.7	73.2	73.2	75.8	77.8	54.8
54. V.vinifera_GSVIVT00033904001	62.8	50.5	73.0	75.8	75.8	76.3	76.8	77.3	54.6

FIGURE 15 (continued)

45/70

	19	20	21	22	23	24	25	26	27
1. T.aestivum_TC329781	10.8	10.1	9.0	9.1	7.2	10.6	10.7	11.1	17.7
2. T.aestivum_TC342076	61.3	67.7	31.4	66.7	34.3	67.2	66.7	66.7	24.2
3. H.vulgare_TC168675	61.3	67.2	31.4	66.7	34.3	67.2	66.7	66.7	24.2
4. O.sativa_LOC_Os02g18830.1	61.3	66.7	31.9	67.2	34.8	67.7	67.2	66.7	24.2
5. P.virgatum_TC36436	62.3	68.2	31.9	66.7	34.3	67.7	67.2	66.7	23.7
6. Z.mays_GRMZM2G130339_T02	62.3	68.2	31.9	66.7	33.8	67.2	66.7	66.7	23.7
7. S.bicolor_Sb04g033060.1	62.3	68.2	31.9	66.7	34.3	67.7	67.2	66.7	23.7
8. S.bicolor_Sb04g011280.1	62.3	68.2	31.9	66.7	34.3	67.7	67.2	66.7	23.7
9. C.longa_TA2440_136217	61.8	65.7	34.3	65.2	34.3	67.2	66.7	65.7	25.8
10. A.officinalis_TA1655_4686	44.1	47.5	17.9	43.4	17.2	44.4	44.7	44.9	27.1
11. Y.filamentosa_DT581455	35.8	39.7	66.4	40.2	66.4	41.7	40.7	39.2	42.6
12. M.sp_77569	64.7	60.8	29.4	62.7	29.4	60.8	60.3	61.3	16.7
13. O.lucimarinus_39810	62.7	66.7	31.4	66.2	32.8	64.6	64.1	64.1	17.2
14. O.sp_18991	63.2	68.7	32.4	65.7	32.4	64.1	63.6	63.6	16.7
15. O.tauri_37777	62.7	67.7	31.9	63.6	31.9	62.6	62.1	62.6	16.7
16. V.carteri_106743	62.3	65.7	30.9	66.7	33.3	66.2	65.2	64.1	16.7
17. C.reinhardtii_178371	62.3	65.2	29.4	65.7	32.8	66.2	65.2	63.1	17.7
18. P.patens_46413	72.5	54.7	52.3	47.3	49.7	48.8	47.8	46.8	27.7
19. P.patens_TC42118		78.9	38.2	69.1	36.3	70.6	70.1	68.1	19.6
20. S.moellendorffii_407397	85.3		40.2	76.8	40.7	79.3	77.8	73.7	21.7
21. A.trichopoda_CK754382	45.1	45.5		40.2	81.1	42.2	40.7	39.7	29.5
22. P.taeda_TA10927_3352	84.8	89.4	47.0		47.5	92.9	91.4	81.8	20.2
23. P.glauca_TA18327_3330	45.1	48.0	91.0	51.5		51.5	50.0	43.1	27.0
24. P.sitchensis_TA12989_3332	84.8	89.9	48.5	96.5	53.0		98.5	83.3	21.2
25. P.glauca_CO481519	83.8	88.4	47.2	94.9	51.8	98.5		82.3	21.3
26. C.japonica_TA2268_3369	86.3	88.4	48.5	91.9	50.5	93.4	91.9		20.2
27. P.americana_CV001201	26.5	26.3	40.5	26.8	39.6	27.8	27.9	27.8	

FIGURE 15 (continued)

46/70

	19	20	21	22	23	24	25	26	27
28. A.sp_TC27214	83.8	81.9	46.6	82.4	46.6	83.8	82.8	83.3	30.4
29. M.truncatula_CU012050_19.4	80.9	77.0	42.6	77.9	41.2	78.9	77.9	79.4	27.0
30. L.japonicus_TC43316	81.4	78.4	43.6	79.9	43.1	80.9	79.9	80.9	27.0
31. G.max_Glyma12g02820.2	67.2	68.2	37.1	68.2	34.9	69.7	69.0	69.7	33.1
32. G.max_Glyma12g02820.1	72.1	72.2	38.7	73.7	36.5	74.7	74.6	75.3	30.9
33. G.max_TC278049	79.9	76.0	42.2	77.5	41.7	78.4	77.5	78.9	27.5
34. G.max_Glyma12g02820.5	79.9	76.5	42.6	77.9	42.2	78.9	77.9	79.4	27.5
35. G.max_Glyma11g10520.1	79.9	76.5	42.6	78.4	42.6	79.4	78.4	79.4	27.5
36. P.coccineus_TA3737_3886	64.7	64.1	31.7	65.2	30.5	66.7	67.0	66.2	32.9
37. P.coccineus_TC15108	63.7	63.1	31.7	64.1	30.4	65.7	66.0	65.2	34.8
38. P.vulgaris_TC10901	79.9	76.0	41.7	77.5	41.7	78.4	77.5	78.4	27.0
39. M.crystallinum_TC9042	57.4	58.6	29.8	59.1	32.6	60.6	60.9	60.1	39.0
40. C.sinensis_TC22219	38.2	39.9	73.0	40.4	73.9	41.4	41.1	41.4	47.6
41. P.trifoliata_TA8498_37690	70.1	71.2	36.5	70.7	36.5	72.2	72.6	71.7	34.1
42. G.hirsutum_TC158561	77.0	77.3	39.2	77.3	38.1	77.8	76.6	79.3	30.2
43. G.hirsutum_ES833864	41.7	39.9	78.4	40.9	76.6	40.9	40.6	42.9	43.2
44. G.raimondii_TC1038	41.2	39.4	77.5	40.4	75.7	40.4	40.1	42.4	42.3
45. E.esula_TC3222	61.3	61.6	33.3	62.1	33.3	62.6	63.5	63.1	36.7
46. E.esula_DV156953	66.2	66.7	34.5	67.2	32.1	67.7	68.0	68.7	32.7
47. P.trichocarpa_667309	39.2	39.9	73.0	39.9	72.1	40.4	39.6	41.4	45.6
48. P.trichocarpa_831805	81.9	82.8	42.3	82.3	41.8	83.3	82.7	83.8	29.1
49. P.trichocarpa EMG1-like	81.9	82.8	42.3	82.3	41.8	83.3	82.7	83.8	29.1
50. B.napus_TC70157	77.9	79.3	40.3	79.8	40.3	80.8	80.7	79.8	28.1
51. B.napus_TC76938	77.9	79.3	40.3	79.8	40.3	80.8	80.7	79.8	28.1
52. A.thaliana_AT3G57000.1	77.5	78.8	40.4	80.8	40.4	81.3	80.3	81.3	27.3
53. A.lyrata_486143	76.5	76.8	38.6	79.3	39.1	79.3	79.2	79.3	25.9
54. V.vinifera_GSVIVT00033904001	77.5	78.8	40.8	80.3	41.3	81.8	81.7	80.8	29.1

FIGURE 15 (continued)

47/70

	28	29	30	31	32	33	34	35	36
1. T.aestivum_TC329781	12.7	9.8	11.8	12.6	12.2	10.8	10.8	10.8	13.2
2. T.aestivum_TC342076	74.0	65.7	71.1	58.6	62.6	67.6	68.1	68.1	57.7
3. H.vulgare_TC168675	74.0	65.7	71.1	58.6	62.6	67.6	68.1	68.1	57.7
4. O.sativa_LOC_Os02g18830.1	72.1	64.2	69.6	58.1	62.6	66.7	67.2	67.2	57.7
5. P.virgatum_TC36436	73.5	63.7	70.1	57.6	62.6	66.7	67.2	67.2	57.2
6. Z.mays_GRMZM2G130339_T02	73.5	63.7	70.1	57.6	62.6	66.7	67.2	67.2	56.7
7. S.bicolor_Sb04g033060.1	73.5	63.7	70.1	57.6	62.6	66.7	67.2	67.2	57.2
8. S.bicolor_Sb04g011280.1	73.5	63.7	70.1	57.6	62.6	66.7	67.2	67.2	57.2
9. C.longa_TA2440_136217	75.0	64.2	67.6	57.6	62.6	65.2	65.7	65.7	57.2
10. A.officinalis_TA1655_4686	52.5	47.1	49.5	56.0	53.0	47.5	48.0	48.0	58.1
11. Y.filamentosa_DT581455	46.1	39.2	42.2	27.5	31.9	39.7	40.2	40.2	27.1
12. M.sp_77569	58.8	55.4	57.4	45.6	47.5	54.9	55.4	54.9	44.2
13. O.lucimarinus_39810	61.3	54.4	57.8	43.9	48.5	52.5	52.9	52.5	43.5
14. O.sp_18991	61.8	54.9	58.8	45.5	50.0	53.4	53.9	53.4	44.5
15. O.tauri_37777	61.8	55.4	58.8	46.5	51.0	54.4	54.9	54.4	45.5
16. V.carteri_106743	58.8	54.9	56.9	46.5	50.5	54.4	54.9	54.9	45.5
17. C.reinhardtii_178371	57.8	54.4	55.9	47.5	50.0	54.9	55.4	55.4	46.0
18. P.patens_46413	43.6	39.7	41.2	28.6	35.0	40.7	40.7	40.7	27.2
19. P.patens_TC42118	66.7	60.8	62.3	49.5	54.9	61.8	61.8	61.8	47.8
20. S.moellendorffii_407397	69.1	60.8	65.2	52.0	58.6	62.3	62.7	62.7	50.2
21. A.trichopoda_CK754382	39.7	31.9	34.3	20.6	24.5	32.4	32.8	32.8	18.8
22. P.taeda_TA10927_3352	68.6	59.8	62.3	51.0	56.6	60.8	61.3	61.8	50.7
23. P.glauca_TA18327_3330	37.7	30.9	33.3	20.1	23.5	32.4	32.8	32.8	18.8
24. P.sitchensis_TA12989_3332	71.1	61.8	64.2	53.0	58.6	62.7	63.2	63.2	52.7
25. P.glauca_CO481519	70.6	60.8	63.2	53.8	58.9	61.8	62.3	62.3	53.0
26. C.japonica_TA2268_3369	70.6	60.3	63.2	53.5	57.1	61.8	62.3	62.3	50.7
27. P.americana_CV001201	26.5	24.0	23.5	28.6	26.5	23.5	23.5	23.5	28.7

FIGURE 15 (continued)

48/70

	28	29	30	31	32	33	34	35	36
28. A.sp_TC27214		76.0	80.4	64.7	69.1	77.5	77.9	77.9	62.8
29. M.truncatula_CU012050_19.4	89.7		86.8	70.6	77.0	86.8	87.3	86.8	69.1
30. L.japonicus_TC43316	91.2	96.1		72.1	78.4	88.7	89.2	88.7	69.1
31. G.max_Glyma12g02820.2	76.5	79.9	81.9		88.4	80.4	79.9	79.4	88.0
32. G.max_Glyma12g02820.1	80.9	85.8	86.8	90.6		87.3	87.3	86.8	85.3
33. G.max_TC278049	90.2	95.6	97.1	82.8	87.7		99.5	99.0	75.8
34. G.max_Glyma12g02820.5	90.7	96.1	97.5	82.8	87.7	99.5		99.5	75.8
35. G.max_Glyma11g10520.1	90.7	95.6	97.1	82.4	87.3	99.0	99.5		75.8
36. P.coccineus_TA3737_3886	72.1	76.0	77.0	89.7	87.8	77.9	77.9	78.4	
37. P.coccineus_TC15108	71.1	75.0	76.0	89.1	86.7	77.0	77.0	77.5	96.4
38. P.vulgaris_TC10901	89.7	95.1	96.1	81.4	86.3	98.0	98.5	99.0	79.4
39. M.crystallinum_TC9042	66.2	64.7	65.2	76.6	74.0	65.7	65.7	65.7	79.6
40. C.sinensis_TC22219	45.1	42.2	42.6	36.6	42.0	42.6	43.1	43.6	37.1
41. P.trifoliata_TA8498_37690	77.9	76.0	76.0	87.4	86.2	76.5	76.5	77.0	88.2
42. G.hirsutum_TC158561	83.3	80.9	81.9	79.4	86.8	81.9	82.4	82.4	77.8
43. G.hirsutum_ES833864	46.6	45.1	44.6	37.7	42.0	45.6	46.1	46.1	34.7
44. G.raimondii_TC1038	46.1	44.6	44.1	37.1	41.4	45.1	45.6	45.6	34.1
45. E.esula_TC3222	65.7	66.2	66.7	78.3	75.7	67.2	67.2	67.2	82.0
46. E.esula_DV156953	71.1	72.1	72.1	83.4	81.8	72.5	72.5	72.5	86.2
47. P.trichocarpa_667309	43.6	42.6	42.6	38.3	43.6	44.1	44.6	44.6	38.3
48. P.trichocarpa_831805	87.3	85.3	85.8	78.1	84.2	86.3	86.8	86.8	76.5
49. P.trichocarpa EMG1-like	87.3	85.3	85.8	78.1	84.2	86.3	86.8	86.8	76.5
50. B.napus_TC70157	84.8	85.3	87.3	75.5	82.1	85.3	85.8	85.8	74.5
51. B.napus_TC76938	84.8	85.3	87.3	75.5	82.1	85.3	85.8	85.8	74.5
52. A.thaliana_AT3G57000.1	84.8	85.3	85.8	76.3	82.3	85.8	86.3	86.3	73.7
53. A.lyrata_486143	83.8	85.3	85.8	75.1	81.7	85.3	85.8	86.3	74.1
54. V.vinifera_GSVIVT00033904001	86.8	85.3	86.8	77.0	84.7	86.8	87.3	87.7	76.0

FIGURE 15 (continued)

49/70

	37	38	39	40	41	42	43	44	45
1. <i>T.aestivum</i> _TC329781	13.7	10.8	16.3	9.7	15.3	13.8	10.8	10.8	18.0
2. <i>T.aestivum</i> _TC342076	57.6	68.1	55.1	38.4	67.2	67.0	34.8	34.3	52.5
3. <i>H.vulgare</i> _TC168675	57.6	68.1	55.1	38.4	67.2	67.0	34.8	34.3	52.5
4. <i>O.sativa</i> _LOC_Os02g18830.1	57.6	67.6	54.0	37.4	66.2	66.0	34.3	33.8	51.5
5. <i>P.virgatum</i> _TC36436	57.1	67.6	54.0	37.4	66.2	66.0	33.3	32.8	52.0
6. <i>Z.mays</i> _GRMZM2G130339_T02	56.6	67.2	54.0	36.9	65.7	66.0	33.3	32.8	52.0
7. <i>S.bicolor</i> _Sb04g033060.1	57.1	67.6	54.0	37.4	66.2	66.0	33.3	32.8	52.0
8. <i>S.bicolor</i> _Sb04g011280.1	57.1	67.6	54.0	37.4	66.2	66.0	33.3	32.8	52.0
9. <i>C.longa</i> _TA2440_136217	56.1	65.7	53.0	37.4	66.7	70.0	35.8	35.3	54.5
10. <i>A.officinalis</i> _TA1655_4686	59.6	48.0	71.6	17.2	60.6	54.0	16.9	16.9	65.3
11. <i>Y.filamentosa</i> _DT581455	26.0	40.7	20.6	63.1	33.3	36.4	65.6	64.8	21.6
12. <i>M.sp</i> _77569	44.1	54.9	43.1	26.0	47.5	49.5	26.5	26.5	45.1
13. <i>O.lucimarinus</i> _39810	44.4	52.5	43.4	27.8	50.0	53.5	27.5	27.5	43.9
14. <i>O.sp</i> _18991	43.9	53.4	43.9	27.3	51.5	54.0	27.0	27.0	44.4
15. <i>O.tauri</i> _37777	46.0	54.4	42.9	26.8	51.0	55.0	27.0	27.0	44.4
16. <i>V.carteri</i> _106743	44.4	54.9	48.0	25.8	49.0	53.8	26.0	25.5	45.5
17. <i>C.reinhardtii</i> _178371	44.9	55.4	48.0	26.8	50.0	54.3	27.0	26.5	46.0
18. <i>P.patens</i> _46413	27.1	39.7	24.6	37.8	29.6	36.0	40.9	40.3	23.2
19. <i>P.patens</i> _TC42118	48.0	60.8	45.6	27.5	50.0	57.8	29.9	29.4	46.1
20. <i>S.moellendorffii</i> _407397	50.0	61.8	48.5	31.8	55.1	60.0	31.4	30.9	48.5
21. <i>A.trichopoda</i> _CK754382	18.1	31.9	14.9	58.6	24.0	27.7	63.1	62.2	13.9
22. <i>P.taeda</i> _TA10927_3352	50.5	60.8	48.0	33.3	55.6	59.5	32.4	31.9	49.5
23. <i>P.glauca</i> _TA18327_3330	18.1	32.8	18.1	60.4	24.0	27.2	63.1	62.2	16.1
24. <i>P.sitchensis</i> _TA12989_3332	52.5	63.2	49.0	33.8	58.1	61.0	32.4	31.9	50.0
25. <i>P.glauca</i> _CO481519	52.8	62.3	49.2	33.5	58.4	60.9	31.9	31.4	50.8
26. <i>C.japonica</i> _TA2268_3369	51.0	61.3	46.5	31.8	56.1	60.0	31.9	31.4	49.0
27. <i>P.americana</i> _CV001201	29.8	23.5	31.3	39.5	30.0	25.4	34.4	33.6	30.7

FIGURE 15 (continued)

50/70

	37	38	39	40	41	42	43	44	45
28. A.sp_TC27214	62.7	77.5	56.9	38.7	67.2	69.9	38.2	37.7	58.3
29. M.truncatula_CU012050_19.4	69.1	87.3	55.4	37.3	65.2	67.3	37.7	37.3	56.9
30. L.japonicus_TC43316	69.1	87.7	56.9	37.7	65.7	71.6	39.2	38.7	58.8
31. G.max_Glyma12g02820.2	87.4	77.5	66.9	26.5	71.4	67.7	24.5	24.0	67.4
32. G.max_Glyma12g02820.1	85.6	84.3	65.2	32.7	72.9	74.1	30.4	29.9	65.2
33. G.max_TC278049	76.0	96.6	57.4	36.8	65.2	69.6	39.2	38.7	57.8
34. G.max_Glyma12g02820.5	76.0	97.1	57.4	37.3	65.2	69.8	39.7	39.2	57.8
35. G.max_Glyma11g10520.1	76.0	97.1	57.4	37.3	65.2	69.8	39.7	39.2	57.8
36. P.coccineus_TA3737_3886	96.4	78.3	69.5	27.8	74.6	66.1	23.7	23.2	70.1
37. P.coccineus_TC15108		78.4	72.0	26.0	74.7	66.1	23.0	22.5	72.7
38. P.vulgaris_TC10901	78.4		57.4	38.2	66.2	69.3	39.2	38.7	57.4
39. M.crystallinum_TC9042	82.6	65.7		21.4	70.6	62.4	15.4	15.4	77.3
40. C.sinensis_TC22219	36.0	43.1	34.0		36.2	36.2	76.6	75.7	22.4
41. P.trifoliata_TA8498_37690	87.1	76.5	80.0	42.4		76.7	31.4	30.9	73.5
42. G.hirsutum_TC158561	76.7	81.9	73.0	41.8	84.1		42.0	42.4	64.0
43. G.hirsutum_ES833864	34.8	45.1	34.8	82.0	40.6	46.6		99.1	21.6
44. G.raimondii_TC1038	34.2	44.6	34.8	81.1	40.0	47.1	99.1		21.1
45. E.esula_TC3222	85.1	67.2	88.0	32.7	81.8	73.0	33.3	32.7	
46. E.esula_DV156953	85.5	72.5	79.4	36.4	88.8	80.4	38.8	38.2	89.1
47. P.trichocarpa_667309	37.9	44.6	33.3	88.3	40.0	43.9	82.9	82.0	33.3
48. P.trichocarpa_831805	75.0	86.8	69.9	46.9	81.6	89.3	47.4	46.9	71.9
49. P.trichocarpa EMG1-like	75.0	86.8	69.9	46.9	81.6	89.3	47.4	46.9	71.9
50. B.napus_TC70157	73.0	85.8	67.9	45.4	79.1	86.2	45.9	45.4	68.9
51. B.napus_TC76938	73.0	85.8	67.9	45.4	79.1	86.2	45.9	45.4	68.9
52. A.thaliana_AT3G57000.1	72.7	85.8	68.2	44.4	79.3	86.4	44.4	43.9	68.2
53. A.lyrata_486143	73.1	86.3	67.5	44.7	78.7	84.8	44.2	43.7	68.5
54. V.vinifera_GSVIVT00033904001	74.5	86.8	70.4	45.9	82.1	87.8	44.9	44.4	69.9

FIGURE 15 (continued)

51/70

	46	47	48	49	50	51	52	53	54
1. T.aestivum_TC329781	16.4	10.7	12.2	12.2	13.3	13.3	13.6	13.2	11.2
2. T.aestivum_TC342076	58.1	33.8	69.7	69.7	66.7	66.7	67.0	67.3	74.2
3. H.vulgare_TC168675	58.1	33.8	69.7	69.7	66.7	66.7	67.0	67.3	74.2
4. O.sativa_LOC_Os02g18830.1	57.6	32.8	68.2	68.2	65.7	65.7	67.0	67.3	74.2
5. P.virgatum_TC36436	58.1	32.8	68.7	68.7	65.7	65.7	68.0	68.3	73.7
6. Z.mays_GRMZM2G130339_T02	58.1	32.8	68.7	68.7	65.7	65.7	68.0	67.8	73.2
7. S.bicolor_Sb04g033060.1	58.1	32.8	68.7	68.7	65.7	65.7	68.0	68.3	73.7
8. S.bicolor_Sb04g011280.1	58.1	32.8	68.7	68.7	65.7	65.7	68.0	68.3	73.7
9. C.longa_TA2440_136217	60.6	36.9	71.2	71.2	67.7	67.7	71.5	69.3	72.2
10. A.officinalis_TA1655_4686	59.4	16.9	51.0	51.0	51.5	51.5	51.5	51.3	52.6
11. Y.filamentosa_DT581455	27.5	58.2	38.2	38.2	34.8	34.8	34.5	35.1	43.1
12. M.sp_77569	47.1	25.5	55.9	55.9	53.9	53.9	52.4	51.7	52.9
13. O.lucimarinus_39810	48.0	26.8	57.6	57.6	54.5	54.5	55.5	54.8	56.1
14. O.sp_18991	49.0	26.8	58.1	58.1	54.5	54.5	56.0	55.3	57.1
15. O.tauri_37777	49.0	26.8	58.6	58.6	54.0	54.0	55.5	54.8	57.6
16. V.carteri_106743	50.5	27.3	57.6	57.6	57.6	57.6	58.0	57.8	58.1
17. C.reinhardtii_178371	50.5	29.3	58.6	58.6	59.1	59.1	58.0	58.8	58.6
18. P.patens_46413	28.1	41.9	40.4	40.4	40.9	40.9	38.5	38.2	38.4
19. P.patens_TC42118	50.5	30.4	60.8	60.8	64.2	64.2	60.7	60.5	58.8
20. S.moellendorffii_407397	53.0	32.8	64.1	64.1	66.2	66.2	64.5	63.3	64.6
21. A.trichopoda_CK754382	21.1	61.3	32.8	32.8	28.4	28.4	31.1	29.3	31.9
22. P.taeda_TA10927_3352	54.5	32.3	63.1	63.1	61.6	61.6	62.5	62.3	64.6
23. P.glauca_TA18327_3330	20.1	60.4	31.9	31.9	29.4	29.4	29.1	29.3	32.8
24. P.sitchensis_TA12989_3332	55.6	33.8	65.2	65.2	64.6	64.6	64.0	64.3	66.7
25. P.glauca_CO481519	55.8	33.0	64.5	64.5	64.5	64.5	63.8	64.1	66.5
26. C.japonica_TA2268_3369	54.0	32.8	66.7	66.7	62.6	62.6	63.5	62.8	65.7
27. P.americana_CV001201	27.9	35.1	25.0	25.0	24.5	24.5	22.7	21.3	25.5

FIGURE 15 (continued)

52/70

	46	47	48	49	50	51	52	53	54
28. A.sp_TC27214	63.2	37.3	74.0	74.0	72.1	72.1	73.8	74.1	75.0
29. M.truncatula_CU012050_19.4	61.8	33.8	70.1	70.1	71.6	71.6	68.4	70.7	70.6
30. L.japonicus_TC43316	63.7	35.8	72.5	72.5	72.5	72.5	69.9	71.7	75.5
31. G.max_Glyma12g02820.2	70.3	25.9	64.0	64.0	64.3	64.3	61.8	62.4	65.8
32. G.max_Glyma12g02820.1	70.7	31.6	70.4	70.4	70.4	70.4	67.2	68.0	73.0
33. G.max_TC278049	62.7	34.8	71.6	71.6	73.5	73.5	69.4	71.2	75.0
34. G.max_Glyma12g02820.5	62.7	35.3	72.1	72.1	74.0	74.0	69.9	71.7	75.5
35. G.max_Glyma11g10520.1	62.7	35.3	72.1	72.1	74.0	74.0	69.9	71.7	75.5
36. P.coccineus_TA3737_3886	72.8	25.0	63.8	63.8	63.3	63.3	60.7	63.0	65.3
37. P.coccineus_TC15108	72.7	24.5	63.3	63.3	62.8	62.8	60.6	62.9	64.8
38. P.vulgaris_TC10901	62.3	34.8	71.6	71.6	73.0	73.0	68.9	73.7	75.5
39. M.crystallinum_TC9042	69.7	15.4	61.2	61.2	60.7	60.7	57.1	57.9	64.3
40. C.sinensis_TC22219	27.6	75.7	41.3	41.3	37.8	37.8	35.9	37.1	43.4
41. P.trifoliata_TA8498_37690	80.0	29.6	74.0	74.0	65.3	65.3	66.7	67.0	72.4
42. G.hirsutum_TC158561	70.9	35.2	76.5	76.5	74.0	74.0	72.2	72.6	79.7
43. G.hirsutum_ES833864	27.9	73.0	39.7	39.7	36.3	36.3	33.0	34.1	40.7
44. G.raimondii_TC1038	27.5	72.1	39.2	39.2	35.8	35.8	32.5	33.7	40.7
45. E.esula_TC3222	90.3	23.0	65.3	65.3	60.7	60.7	57.1	57.9	59.2
46. E.esula_DV156953		28.6	71.4	71.4	64.8	64.8	63.1	64.0	65.3
47. P.trichocarpa_667309	37.6		48.0	48.0	39.3	39.3	37.4	37.1	39.8
48. P.trichocarpa_831805	78.6	51.0		100	77.0	77.0	76.8	77.2	80.6
49. P.trichocarpa EMG1-like	78.6	51.0	100		77.0	77.0	76.8	77.2	80.6
50. B.napus_TC70157	75.0	47.4	92.9	92.9		100	82.3	83.8	75.5
51. B.napus_TC76938	75.0	47.4	92.9	92.9	100		82.3	83.8	75.5
52. A.thaliana_AT3G57000.1	74.2	46.0	92.9	92.9	92.4	92.4		91.9	73.2
53. A.lyrata_486143	74.1	46.2	92.4	92.4	92.4	92.4	96.0		74.6
54. .vinifera_GSVIVT00033904001	76.0	43.9	90.8	90.8	88.3	88.3	87.4	87.8	

FIGURE 15 (continued)

53/70

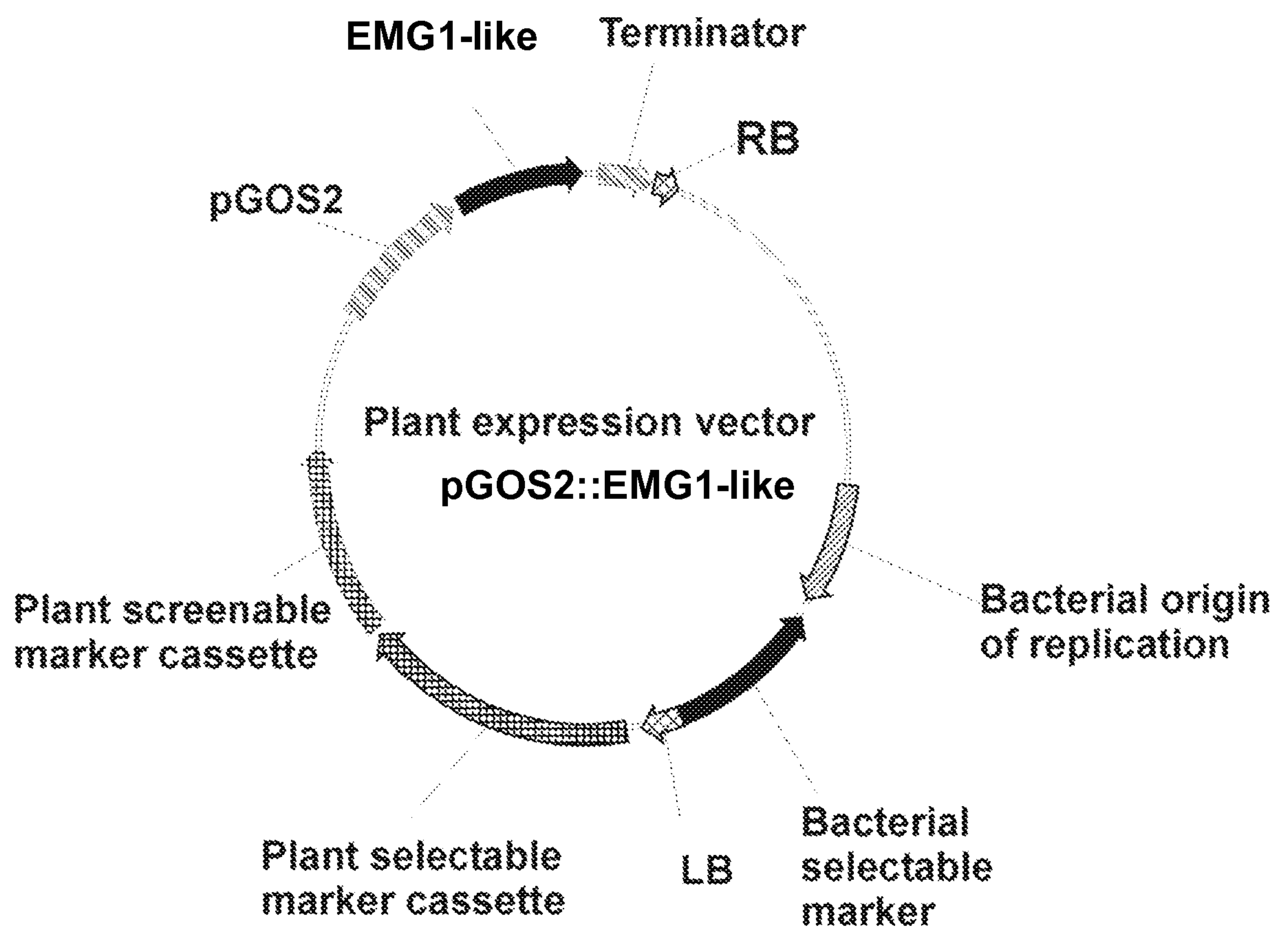


FIGURE 16

54/70

MPSRTAPFISRLRLLRRTAAALSPFSSSAPPRHSPRILRAIPVGAPHPPR

VSAAVSPLVRPVAAGFALFSMATAASSAASVHDFTVKDASGKDVDLSTF

-----18-----

KGKVLLIVNVASQCGLTNSNYTELSQLYEKYKDQGFEILAFPCNQFGGQ

---7---

-----10-----

-----13-----

-----18-----

-----16-----

-----8---

EPGTNEEIVQFACTRFKAEYPIFDKVDVNGDNTAPIYKFLKSSKGGLFG

-----10-----11-----

-----14-----

-----16-----17-----

DNIKWNFSKFLVDKEGRVVERYAPTTSPLSMEKDIKKLLGSS

-----12-----

-----15-----

-----17-----

--9---

FIGURE 17

55/70

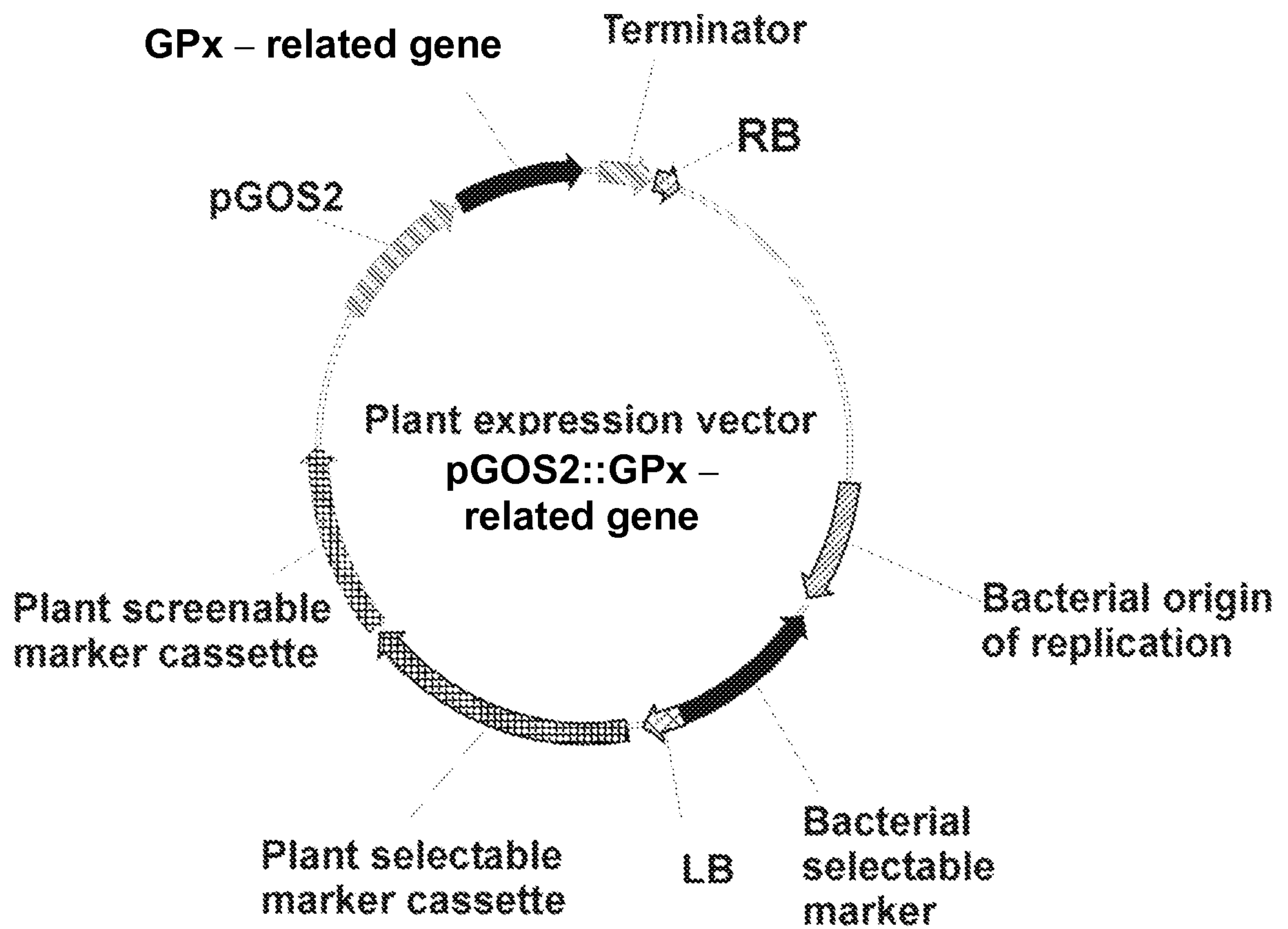


FIGURE 18

56/70

	1	2	3	4	5	6	7	8	9	10	11	12
1. Zea_mays_ZmGPx03		62	64	63	66	64	64	51	90	65	52	79
2. Zea_mays_ZmGPx01-2	65		93	92	89	91	94	72	60	95	71	63
3. Zea_mays_ZmGPx01	66	98		92	92	94	96	72	62	98	74	66
4. Triticum_monococcum_TmGPx01	66	97	97		92	99	94	72	61	95	73	64
5. Triticum_aestivum-TaGPx03	68	95	95	96		92	93	73	63	94	74	66
6. Triticum_aestivum-TaGPx01	66	97	97	100	96		94	72	61	95	74	65
7. Setaria_italica_SiGPx01	66	97	98	96	95	96		72	63	99	74	66
8. Sorghum_bicolor_SbGPx06	59	85	84	85	85	85	85		50	73	82	51
9. Sorghum_bicolor_SbGPx03	92	64	65	64	65	64	65	58		63	52	77
10. Sorghum_bicolor_SbGPx01	67	98	99	98	96	98	99	85	65		74	67
11. Oryza_sativas_OsGPx06	59	84	84	86	85	86	86	91	58	86		53
12. Oryza_sativa_OsGPx03	84	68	69	68	69	68	69	62	82	69	61	
13. Oryza_sativas_OsGPx01	66	97	98	96	95	96	98	84	65	99	84	69
14. Hordeum_vulgare_HvGPx06	59	83	85	84	84	84	85	88	57	85	90	60
15. Hordeum_vulgare_HvGPx03	82	68	68	69	71	69	68	62	82	69	62	89
16. Hordeum_vulgare_HvGPx01	65	95	95	98	95	98	94	84	63	95	84	66
17. Pinus_taeda_PtaGPx06-2	62	90	89	91	88	91	89	85	60	89	85	64
18. Pinus_taeda_PtaGPx06-1	59	87	85	87	84	87	85	81	56	86	81	61
19. Pinus_pinaster_PpGPx06-2	60	87	87	88	85	88	86	83	58	87	83	62
20. Gnetum_gnemon_GgGPx06	60	88	87	89	86	89	88	81	58	88	83	63
21. Zea_mays_ZmGPx03-2	61	85	87	86	88	86	87	76	58	88	77	62
22. Vitis_vinifera_VvGPx06	61	92	91	91	88	90	91	85	60	92	85	65
23. Solanum_tuberosum_StGPx06	61	89	89	91	89	90	89	83	59	90	83	64
24. Spinacia_oleracea_SoGPx06	61	88	90	90	88	90	88	83	59	90	85	64

12. Oryza_sativa_OsGPx03 == SEQ ID NO: 293

FIGURE 19

57/70

	1	2	3	4	5	6	7	8	9	10	11	12
25. Populus_trichocarpa_PtGPx06-2	71	65	65	66	63	66	65	61	73	66	63	76
26. Populus_trichocarpa_PtGPx06	62	92	92	92	89	92	92	84	61	93	87	66
27. Nicotiana_tabacum_NtGPx06	61	89	88	90	88	89	89	83	59	89	84	64
28. Nicotiana_sylvestris_NsGPx06	61	89	88	90	88	89	89	83	59	89	84	64
29. Medicago_truncatula_MtGPx06	66	55	54	54	54	55	54	54	64	55	51	65
30. Malus_domestica_MdGPx06	63	92	93	92	91	91	94	84	61	94	86	66
31. Mesembryanthemum_crystallinum_McGPx06	61	90	91	90	87	90	90	82	59	91	84	65
32. Lotus_japonicus_LjGPx06	72	64	63	64	64	63	64	60	67	64	60	72
33. Lycopersicon_esculentum_LeGPx06	61	88	88	89	87	89	88	81	59	89	83	63
34. Hevea_brasiliensis_HbGPx06	60	82	84	84	82	84	84	78	58	84	80	63
35. Helianthus_annuus_HaGPx06-3_ANN1312	58	86	86	86	83	86	86	80	57	87	80	61
36. Helianthus_annuus_HaGPx06-1	59	88	88	87	85	87	88	83	58	88	80	62
37. Gossypium_hirsutum_GhGPx06-3	63	90	90	91	91	91	90	84	61	91	86	65
38. Gossypium_hirsutum_GhGPx06-2	63	92	92	92	92	92	92	86	61	92	86	65
39. Gossypium_hirsutum_GhGPx06-1	61	88	88	90	88	89	89	82	60	89	84	64
40. Citrus_sinensis_CsGPx06	63	93	92	93	91	92	92	83	61	93	83	65
41. Cucumis_melo_CmGPx06	71	65	65	65	64	65	65	62	71	66	62	74
42. Capsicum_chinense_CcGPx06	61	89	88	91	88	90	89	81	59	89	82	64
43. Brassica_napus_BnGPx06-2	64	92	93	93	92	92	94	85	62	94	86	66
44. Brassica_napus_BnGPx06-1	63	92	91	93	91	92	92	86	61	93	86	66
45. Arabidopsis_thaliana_AtGPx06	74	67	66	67	66	67	67	64	71	67	64	77
46. Populus_trichocarpa_PtGPx SE	67	62	62	62	62	62	62	58	66	63	56	73
47. Populus_trichocarpa_PtGPx01	67	62	62	62	62	62	62	59	65	63	56	72
48. Solanum_tuberosum_StGPx08-2	61	82	84	84	84	84	84	78	59	85	79	63

26. Populus_trichocarpa_PtGPx06 == SEQ ID NO: 295

46. Populus_trichocarpa_PtGPx == SEQ ID NO: 297

FIGURE 19 (continued)

58/70

	13	14	15	16	17	18	19	20	21	22	23	24
1. Zea_mays_ZmGPx03	63	51	78	62	54	50	51	52	53	58	54	54
2. Zea_mays_ZmGPx01-2	91	70	64	89	78	73	74	76	76	82	78	78
3. Zea_mays_ZmGPx01	93	70	66	90	78	74	74	77	79	85	80	80
4. Triticum_monococcum_TmGPx01	92	71	65	96	79	74	74	76	76	83	79	77
5. Triticum_aestivum-TaGPx03	90	72	70	91	77	71	73	74	78	84	79	77
6. Triticum_aestivum-TaGPx01	92	70	65	96	79	74	74	76	76	84	80	78
7. Setaria_italica_SiGPx01	94	72	66	91	79	74	74	77	79	85	80	79
8. Sorghum_bicolor_SbGPx06	72	80	53	72	73	66	70	66	66	71	66	70
9. Sorghum_bicolor_SbGPx03	61	50	76	60	53	49	50	50	51	56	53	53
10. Sorghum_bicolor_SbGPx01	95	72	67	92	79	74	75	78	80	86	81	80
11. Oryza_sativas_OsGPx06	76	82	53	73	73	66	68	68	65	73	67	72
12. Oryza_sativa_OsGPx03	64	50	81	63	55	53	52	54	55	59	56	56
13. Oryza_sativas_OsGPx01		73	64	90	77	71	72	78	76	84	79	78
14. Hordeum_vulgare_HvGPx06	84		51	70	70	65	66	68	62	70	66	71
15. Hordeum_vulgare_HvGPx03	68	61		64	56	52	53	54	54	60	56	58
16. Hordeum_vulgare_HvGPx01	94	84	68		78	72	72	77	75	82	80	76
17. Pinus_taeda_PtaGPx06-2	89	84	63	89		85	89	78	66	79	76	79
18. Pinus_taeda_PtaGPx06-1	85	81	60	85	94		81	74	63	75	72	74
19. Pinus_pinaster_PpGPx06-2	86	83	61	86	97	94		73	66	76	70	75
20. Gnetum_gnemon_GgGPx06	88	83	62	88	90	89	88		65	78	77	77
21. Zea_mays_ZmGPx03-2	88	74	62	86	79	77	78	81		73	68	66
22. Vitis_vinifera_VvGPx06	91	85	63	89	89	87	88	90	83		82	82
23. Solanum_tuberosum_StGPx06	89	83	64	90	89	87	86	89	85	93		77
24. Spinacia_oleracea_SoGPx06	89	85	64	89	89	86	87	89	79	90	91	

12. Oryza_sativa_OsGPx03 == SEQ ID NO: 293

FIGURE 19 (continued)

59/70

	13	14	15	16	17	18	19	20	21	22	23	24
25. Populus_trichocarpa_PtGPx06-2	65	62	74	64	66	63	66	63	58	68	66	66
26. Populus_trichocarpa_PtGPx06	93	86	64	91	90	88	88	88	82	95	91	93
27. Nicotiana_tabacum_NtGPx06	89	83	63	89	89	88	87	90	83	94	98	91
28. Nicotiana_sylvestris_NsGPx06	89	83	63	89	89	88	87	90	83	94	98	91
29. Medicago_truncatula_MtGPx06	54	51	63	54	55	53	54	54	50	57	57	57
30. Malus_domestica_MdGPx06	94	85	65	90	90	86	87	88	84	94	90	90
31. Mesembryanthemum_crystallinum_McGPx06	90	84	63	88	89	87	87	87	79	91	91	94
32. Lotus_japonicus_LjGPx06	64	59	69	62	64	62	63	64	58	65	64	64
33. Lycopersicon_esculentum_LeGPx06	89	82	62	89	89	87	87	89	83	93	100	91
34. Hevea_brasiliensis_HbGPx06	84	80	62	82	81	80	80	82	76	85	84	86
35. Helianthus_annuus_HaGPx06-3_ANN1312	86	79	60	85	82	81	81	84	81	88	86	83
36. Helianthus_annuus_HaGPx06-1	88	81	61	87	83	82	81	84	82	87	87	84
37. Gossypium_hirsutum_GhGPx06-3	90	85	65	89	89	85	88	87	85	91	90	90
38. Gossypium_hirsutum_GhGPx06-2	92	85	65	92	89	86	88	89	85	92	90	90
39. Gossypium_hirsutum_GhGPx06-1	89	83	64	89	89	88	87	89	84	93	97	91
40. Citrus_sinensis_CsGPx06	93	83	65	92	91	87	88	88	84	92	90	88
41. Cucumis_melo_CmGPx06	66	61	75	64	65	62	64	62	58	64	64	65
42. Capsicum_chinense_CcGPx06	89	81	64	89	89	87	87	88	84	92	99	91
43. Brassica_napus_BnGPx06-2	93	85	66	91	89	85	87	89	85	92	92	93
44. Brassica_napus_BnGPx06-1	92	85	65	91	89	85	85	87	83	91	91	93
45. Arabidopsis_thaliana_AtGPx06	67	63	75	66	66	63	65	64	60	66	66	68
46. Populus_trichocarpa_PtGPx	63	57	70	62	60	58	59	59	55	61	61	60
47. Populus_trichocarpa_PtGPx01	63	57	71	62	60	59	59	59	55	61	61	60
48. Solanum_tuberosum_StGPx08-2	85	78	62	82	84	82	83	83	79	88	84	85

26. Populus_trichocarpa_PtGPx06 == SEQ ID NO: 295

46. Populus_trichocarpa_PtGPx == SEQ ID NO: 297

FIGURE 19 (continued)

60/70

	25	26	27	28	29	30	31	32	33	34	35	36
1. Zea_mays_ZmGPx03	63	57	54	54	55	60	54	61	55	52	50	52
2. Zea_mays_ZmGPx01-2	58	83	78	78	48	85	76	55	78	72	74	77
3. Zea_mays_ZmGPx01	59	85	79	80	49	89	81	56	79	76	76	79
4. Triticum_monococcum_TmGPx01	58	83	78	79	48	86	77	56	79	74	75	79
5. Triticum_aestivum-TaGPx03	59	83	78	78	49	86	77	57	79	75	73	77
6. Triticum_aestivum-TaGPx01	58	84	79	79	48	86	78	56	79	74	76	79
7. Setaria_italica_SiGPx01	60	85	79	80	49	89	79	57	79	77	76	79
8. Sorghum_bicolor_SbGPx06	49	74	66	67	45	72	71	54	65	66	64	69
9. Sorghum_bicolor_SbGPx03	59	55	52	52	53	57	52	58	53	51	49	51
10. Sorghum_bicolor_SbGPx01	61	86	80	81	50	89	79	57	80	77	76	80
11. Oryza_sativas_OsGPx06	53	74	69	70	44	76	70	53	67	66	66	69
12. Oryza_sativa_OsGPx03	63	59	56	57	55	61	56	62	57	54	52	54
13. Oryza_sativas_OsGPx01	59	84	78	79	50	85	78	58	78	74	76	79
14. Hordeum_vulgare_HvGPx06	51	72	65	65	44	72	69	51	65	64	63	67
15. Hordeum_vulgare_HvGPx03	63	60	56	57	53	61	57	58	56	55	52	54
16. Hordeum_vulgare_HvGPx01	58	82	79	79	48	85	75	55	79	74	74	79
17. Pinus_taeda_PtaGPx06-2	59	82	78	79	47	82	80	58	75	73	71	75
18. Pinus_taeda_PtaGPx06-1	55	75	75	74	45	78	74	53	71	66	69	71
19. Pinus_pinaster_PpGPx06-2	56	78	72	72	44	76	75	54	69	68	69	70
20. Gnetum_gnemon_GgGPx06	55	76	79	78	48	76	74	56	75	69	72	73
21. Zea_mays_ZmGPx03-2	49	70	67	68	42	73	67	48	67	65	67	69
22. Vitis_vinifera_VvGPx06	61	87	82	83	51	85	82	57	80	77	75	77
23. Solanum_tuberosum_StGPx06	57	79	95	94	48	79	76	54	97	73	75	76
24. Spinacia_oleracea_SoGPx06	57	84	77	77	48	81	83	55	75	74	70	73

12. Oryza_sativa_OsGPx03 == SEQ ID NO: 293

FIGURE 19 (continued)

61/70

	25	26	27	28	29	30	31	32	33	34	35	36
25. Populus_trichocarpa_PtGPx06-2		63	58	59	57	62	57	65	57	56	53	55
26. Populus_trichocarpa_PtGPx06	69		80	81	49	87	82	56	79	78	75	77
27. Nicotiana_tabacum_NtGPx06	66	92		99	49	81	77	55	92	73	76	78
28. Nicotiana_sylvestris_NsGPx06	66	92	100		49	82	77	55	91	74	76	78
29. Medicago_truncatula_MtGPx06	66	56	57	57		51	49	64	48	45	44	45
30. Malus_domestica_MdGPx06	67	95	90	90	56		81	58	79	77	77	79
31. Mesembryanthemum_crystallinumMcGPx06	66	94	91	91	56	91		57	75	74	72	75
32. Lotus_japonicus_LjGPx06	79	65	64	64	72	65	64		54	52	52	53
33. Lycopersicon_esculentum_LeGPx06	66	91	98	98	57	89	91	64		72	75	75
34. Hevea_brasiliensis_HbGPx06	64	85	85	85	54	86	87	61	83		68	70
35. Helianthus_annuus_HaGPx06-3_ANN1312	62	88	86	86	52	87	85	60	87	77		88
36. Helianthus_annuus_HaGPx06-1	62	88	86	86	52	88	85	59	87	78	93	
37. Gossypium_hirsutum_GhGPx06-3	66	92	91	91	54	89	90	64	89	84	85	87
38. Gossypium_hirsutum_GhGPx06-2	66	94	92	92	55	91	91	64	90	83	87	90
39. Gossypium_hirsutum_GhGPx06-1	66	92	99	99	57	90	91	64	97	86	85	86
40. Citrus_sinensis_CsGPx06	66	94	90	90	56	93	91	66	91	84	87	87
41. Cucumis_melo_CmGPx06	78	65	64	64	69	66	63	76	64	61	61	61
42. Capsicum_chinense_CcGPx06	65	90	98	98	56	88	91	64	99	83	86	87
43. Brassica_napus_BnGPx06-2	66	92	92	92	57	94	92	64	91	86	84	86
44. Brassica_napus_BnGPx06-1	67	91	90	90	57	91	91	64	90	84	83	85
45. Arabidopsis_thaliana_AtGPx06	79	67	66	66	68	68	66	78	65	64	60	62
46. Populus_trichocarpa_PtGPx	68	61	60	60	60	61	61	66	61	59	59	62
47. Populus_trichocarpa_PtGPx01	69	61	60	60	62	61	61	66	61	59	59	62
48. Solanum_tuberosum_StGPx08-2	63	88	85	85	52	89	86	60	84	81	82	82

26. Populus_trichocarpa_PtGPx06 == SEQ ID NO: 295

46. Populus_trichocarpa_PtGPx == SEQ ID NO: 297

FIGURE 19 (continued)

62/70

	37	38	39	40	41	42	43	44	45	46	47	48
1. Zea_mays_ZmGPx03	58	59	54	58	65	54	58	58	64	57	56	50,4
2. Zea_mays_ZmGPx01-2	82	83	77	82	60	77	82	83	60	52	52	68,2
3. Zea_mays_ZmGPx01	84	86	79	84	61	78	87	86	63	53	53	71,8
4. Triticum_monococcum_TmGPx01	83	85	78	83	61	77	83	83	60	53	53	71,8
5. Triticum_aestivum-TaGPx03	83	86	78	82	60	77	85	85	62	55	55	71,8
6. Triticum_aestivum-TaGPx01	84	86	79	84	61	78	83	84	60	53	53	72,4
7. Setaria_italica_SiGPx01	85	86	79	84	62	78	86	86	62	53	53	72,4
8. Sorghum_bicolor_SbGPx06	71	76	66	72	53	66	71	73	55	48	48	61,3
9. Sorghum_bicolor_SbGPx03	56	56	52	56	63	53	57	59	63	55	55	48,6
10. Sorghum_bicolor_SbGPx01	86	87	80	85	62	79	87	87	63	54	54	72,4
11. Oryza_sativas_OsGPx06	72	75	69	74	54	67	73	73	54	47	47	62,8
12. Oryza_sativa_OsGPx03	59	60	57	59	65	56	59	60	66	57	57	52,5
13. Oryza_sativas_OsGPx01	83	85	78	85	63	77	83	84	60	54	54	71,8
14. Hordeum_vulgare_HvGPx06	70	73	64	72	52	66	69	71	52	47	47	61,6
15. Hordeum_vulgare_HvGPx03	59	61	56	59	67	56	61	62	66	58	58	51,9
16. Hordeum_vulgare_HvGPx01	84	86	79	83	60	78	82	82	59	53	53	70,6
17. Pinus_taeda_PtaGPx06-2	76	80	78	81	59	75	81	78	60	52	52	71,8
18. Pinus_taeda_PtaGPx06-1	72	73	74	76	54	71	77	74	56	48	49	68,8
19. Pinus_pinaster_PpGPx06-2	73	77	72	76	55	69	74	74	55	51	51	67,6
20. Gnetum_gnemon_GgGPx06	74	76	78	76	56	77	75	75	56	50	50	66,5
21. Zea_mays_ZmGPx03-2	73	73	67	70	51	67	75	73	53	46	46	62,6
22. Vitis_vinifera_VvGPx06	81	85	82	85	61	80	84	84	60	52	52	74,1
23. Solanum_tuberosum_StGPx06	78	80	94	80	56	95	80	80	57	51	51	69,4
24. Spinacia_oleracea_SoGPx06	78	80	76	77	59	77	81	82	60	50	50	67,8

12. Oryza_sativa_OsGPx03 == SEQ ID NO: 293

FIGURE 19 (continued)

63/70

	37	38	39	40	41	42	43	44	45	46	47	48
25. Populus_trichocarpa_PtGPx06-2	58	59	58	59	69	55	59	60	67	54	54	53,4
26. Populus_trichocarpa_PtGPx06	82	87	79	85	60	78	83	84	60	50	50	71,2
27. Nicotiana_tabacum_NtGPx06	78	81	99	81	57	92	81	80	59	51	51	70,6
28. Nicotiana_sylvestris_NsGPx06	79	81	99	81	57	91	82	80	59	52	52	71,2
29. Medicago_truncatula_MtGPx06	48	49	49	50	62	50	51	51	58	48	49	43
30. Malus_domestica_MdGPx06	83	85	81	86	61	78	89	86	64	54	54	75,9
31. Mesembryanthemum_crystallinum_McGPx06	78	83	77	81	56	76	81	80	59	52	52	69,4
32. Lotus_japonicus_LjGPx06	55	57	55	61	67	56	57	58	65	53	54	48,9
33. Lycopersicon_esculentum_LeGPx06	78	79	91	79	56	94	79	79	57	49	49	68,8
34. Hevea_brasiliensis_HbGPx06	72	75	73	75	54	73	78	76	57	48	48	66,9
35. Helianthus_annuus_HaGPx06-3_ANN1312	76	75	76	77	54	74	76	75	53	49	49	65,9
36. Helianthus_annuus_HaGPx06-1	77	80	78	80	55	77	79	78	56	52	52	67,6
37. Gossypium_hirsutum_GhGPx06-3		89	78	81	59	75	82	81	59	53	53	71,8
38. Gossypium_hirsutum_GhGPx06-2	94		81	86	60	78	82	83	61	54	54	71,8
39. Gossypium_hirsutum_GhGPx06-1	91	91		81	57	91	81	80	59	51	51	70,2
40. Citrus_sinensis_CsGPx06	91	94	89		60	79	82	83	60	52	52	70,6
41. Cucumis_melo_CmGPx06	64	65	64	64		57	62	63	72	56	56	51,9
42. Capsicum_chinense_CcGPx06	90	89	98	89	64		80	80	58	51	51	68,4
43. Brassica_napus_BnGPx06-2	92	92	92	91	66	91		94	69	53	53	73,1
44. Brassica_napus_BnGPx06-1	91	92	90	89	67	90	97		67	52	52	70,2
45. Arabidopsis_thaliana_AtGPx06	66	68	67	66	81	65	72	72		56	56	52,2
46. Populus_trichocarpa_PtGPx	61	61	60	60	67	62	61	60	67		99	48,5
47. Populus_trichocarpa_PtGPx01	61	61	60	60	66	62	61	60	67	99		48,5
48. Solanum_tuberosum_StGPx08-2	87	86	85	85	61	84	88	85	64	59	59	

26. Populus_trichocarpa_PtGPx06 == SEQ ID NO: 295

46. Populus_trichocarpa_PtGPx == SEQ ID NO: 297

FIGURE 19 (continued)

64/70

CLUSTAL 2.0.11 multiple sequence alignment

```

Oryza_sativa_OsGPx06 -----
Hordeum_vulgare_HvGPx06 -----
Sorghum_bicolor_SbGPx06 -----
Zea_mays_ZmGPx03 -----MPPRPVPLFSRLRFRLPSCAAVSISTPPSSAAP
Sorghum_bicolor_SbGPx03 -----MPPSPAPLFSRLRFRLRTCAAVSISTRPSSAAP
Sorghum_bicolor_SbGPx01 -----
Zea_mays_ZmGPx01 -----
Setaria_italica_SiGPx01 -----
Oryza_sativa_OsGPx03 (SEQIDNO: 2) -----MPSRTAPFISRLRL-LRTAAALS---PFSSSAP
Zea_mays_ZmGPx01-2 -----
Triticum_monococcum_TmGPx01 -----
Triticum_aestivum-TaGPx01 -----
Hordeum_vulgare_HvGPx01 -----
Triticum_aestivum-TaGPx03 -----
Hordeum_vulgare_HvGPx03 -----MPSRPAPFIHRLRLLRSAP-----VSPSSA
Oryza_sativa_OsGPx01 -----
Cucumis_melo_CmGPx06 -----MFCSSSIRSLLTRNFFFSVRSLSSSPLLSNTRF
Gossypium_hirsutum_GhGPx06-3 -----
Gossypium_hirsutum_GhGPx06-2 -----
Populus_trichocarpa_PtGPx06-2 -----MLTSRSLRILSQKYL--FASLSASFLLSKQ
P_trichocarpa_PtGPx06 (SEQIDNO: 4) -----
Vitis_vinifera_VvGPx06 -----
Hevea_brasiliensis_HbGPx06 -----
Brassica_napus_BnGPx06-2 -----
Arabidopsis_thaliana_AtGPx06 -----MLRSSIRLLYIRRTSPLLRLSSSSSSSSSS
Brassica_napus_BnGPx06-1 -----
Malus_domestica_MdGPx06 -----
Spinacia_oleracea_SoGPx06 -----
Mesembryanthemum_crystallinum_ -----
Gnetum_gnemon_GgGPx06 -----
Medicago_truncatula_MtGPx06 MELLGVKYRQIGTWTREEASTRNLRYSVHTHNCRWLSEQSTSTTTTRIRFI
Solanum_tuberosum_StGPx06 -----
Lycopersicon_esculentum_LeGPx0 -----
Capsicum_chinense_CcGPx06 -----
Nicotiana_sylvestris_NsGPx06 -----
Gossypium_hirsutum_GhGPx06-1 -----
Nicotiana_tabacum_NtGPx06 -----
Lotus_japonicus_LjGPx06 -----MLCTRILFFS---RTIRFAAPLSSSSLHSHFVFS
Citrus_sinensis_CsGPx06 -----
Pinus_taeda_PtaGPx06-2 -----
Pinus_pinaster_PpGPx06-2 -----
Pinus_taeda_PtaGPx06-1 -----
Solanum_tuberosum_StGPx08-2 -----
Helianthus_annuus_HaGPx06-3_AN -----
Helianthus_annuus_HaGPx06-1 -----
Zea_mays_ZmGPx03-2 -----
P_trichocarpa_PtGPx (SEQIDNO: 6) -----MASLPFSCSSVPG
Populus_trichocarpa_PtGPx01 -----MASLPFSCSSVPG

```

FIGURE 20

65/70

Oryza_sativa_OsGPx06	-----MG
Hordeum_vulgare_HvGPx06	-----MG
Sorghum_bicolor_SbGPx06	-----MG
Zea_mays_ZmGPx03	LHYGHRTPLASPFPLSPPCPRILRAAVAV---SSLDLRPITGGFALFS--
Sorghum_bicolor_SbGPx03	LRY-RRRLSIPFPLSPPCPRILRAAAAAAVSSSLDLRAIAGGFALFS--
Sorghum_bicolor_SbGPx01	-----
Zea_mays_ZmGPx01	-----
Setaria_italica_SiGPx01	-----
Oryza_sativa_OsGPx03 (SEQIDNO: 2)	PRHSPRILRAI--PVGAPHP--PRVSAA----VSPLVRPVAAGFALFS-M
Zea_mays_ZmGPx01-2	-----
Triticum_monococcum_TmGPx01	-----
Triticum_aestivum-TaGPx01	-----
Hordeum_vulgare_HvGPx01	-----
Triticum_aestivum-TaGPx03	-----M
Hordeum_vulgare_HvGPx03	PLPCSRPRLAVPFSQPPIRAPAFAAAA-----SPLLRPVGARFSLSSNM
Oryza_sativa_OsGPx01	-----
Cucumis_melo_CmGPx06	FPDSKQTLHLHIPQFPSVTRFVSPINSRS-----SLIASFLTRLDH
Gossypium_hirsutum_GhGPx06-3	-----
Gossypium_hirsutum_GhGPx06-2	-----
Populus_trichocarpa_PtGPx06-2	SSFNSKQTLPLSLHNSPVSLYSQSIKAGVSRRLLGSVRFNHS-----
P_trichocarpa_PtGPx06 (SEQIDNO: 4)	-----
Vitis_vinifera_VvGPx06	-----
Hevea_brasiliensis_HbGPx06	-----
Brassica_napus_BnGPx06-2	-----
Arabidopsis_thaliana_AtGPx06	KRFDSAKPLFNSHR-----IISLPISTTGAKLSRSEHS-----
Brassica_napus_BnGPx06-1	-----MA-----
Malus_domestica_MdGPx06	-----
Spinacia_oleracea_SoGPx06	-----
Mesembryanthemum_crystallinum_	-----
Gnetum_gnemon_GgGPx06	-----
Medicago_truncatula_MtGPx06	STTKRLLTAPLSSLLRFFSISTTLPNKPIIHKPLFTTLTPSLYFTLRRTD
Solanum_tuberosum_StGPx06	-----
Lycopersicon_esculentum_LeGPx0	-----
Capsicum_chinense_CcGPx06	-----
Nicotiana_sylvestris_NsGPx06	-----
Gossypium_hirsutum_GhGPx06-1	-----
Nicotiana_tabacum_NtGPx06	-----
Lotus_japonicus_LjGPx06	NSPITLSRSYHSSLLTTTSFP-----IKSLVSTSTTPFSFTLR---
Citrus_sinensis_CsGPx06	-----MASQSK-----
Pinus_taeda_PtaGPx06-2	-----
Pinus_pinaster_PpGPx06-2	-----
Pinus_taeda_PtaGPx06-1	-----
Solanum_tuberosum_StGPx08-2	-----
Helianthus_annuus_HaGPx06-3_AN	-----
Helianthus_annuus_HaGPx06-1	-----
Zea_mays_ZmGPx03-2	-----
P_trichocarpa_PtGPx (SEQIDNO: 6)	LKIHPNPTKMAFFVTSIKSSLGPSKSAFLQRGFSLQSPNLPGFASKARSF
Populus_trichocarpa_PtGPx01	LKIHPNPTKMASFVISIKSSLGPSKSAFLQRGFSLQSPNLPGFASKARSF

FIGURE 20 (continued)

66/70

Oryza_sativa_OsGPx06	AAESSSKLAGSIHEFTVKDARGSDVELSRYKKGKVLLIVNNAASRCGLTNYN
Hordeum_vulgare_HvGPx06	AAESSSKLGGSVHDFVVKDVRGNDVELSRYKKGKVLLIVNVASRCGLANSN
Sorghum_bicolor_SbGPx06	AAESSSKLASSVHGFSVKDARGNDVELSRYKKGKVLLIVNVAS-QGLTNSN
Zea_mays_ZmGPx03	MAAASS--ATSVHDFTVKDASGKDVDLSTYKKGKVLLIVNVASQCGLTNSN
Sorghum_bicolor_SbGPx03	MAAASS--ATSVHDFTVKDASGKDVDLSTYKKGKVLLIVNVASQCGLTNSN
Sorghum_bicolor_SbGPx01	MAAASS--ATSVHDFTVKDASGKDVDLSTYKKGKVLLIVNVASQCGLTNSN
Zea_mays_ZmGPx01	MAASST--ATSVHDFIVKDASGKDVDLSTYKKGKVLLIVNVASQCGLTNSN
Setaria_italica_SiGPx01	MAAASS--TTSVHDFTVKDASGKDVDLSTYKKGKVLLIVNVASQCGLTNSN
Oryza_sativa_OsGPx03 (SEQIDNO: 2)	ATAASS--AASVHDFTVKDASGKDVDLSTFKGKVLLIVNVASQCGLTNSN
Zea_mays_ZmGPx01-2	MAAASS--ATSVHDFTVKDSSGKDVDLSVYRGKVLLIVNVASQCGLTNSN
Triticum_monococcum_TmGPx01	MAAASS--ATSVHDFTVKDSSGKDVDLSVYKKGKVLLIVNVASQCGLTNSN
Triticum_aestivum-TaGPx01	MAAASS--ATSVHDFTVKDASGKDVDLSVYKKGKVLLIVNVASQCGLTNSN
Hordeum_vulgare_HvGPx01	MAAASS--ATSVHDFTVKDASGKDVDLSVYKKGKVLLIVNVASQCGLTNSN
Triticum_aestivum-TaGPx03	AAAASS--ASSVHDFTVKDASGKDVDLSTYKKGKVLLIVNVASQCGLTNSN
Hordeum_vulgare_HvGPx03	AAAASS--ASSVHDFTVKDASGKDVDLSTYKKGKVLLIVNVASQCGLTNSN
Oryza_sativa_OsGPx01	MAAAPS--ATSVHDFTVKDASGKDVLNLSYKKGKVLLIVNVASQCGLTNSN
Cucumis_melo_CmGPx06	TMATPS--KTSVHDFTVKDAKGNVDLSDYKKGKVLLIVNVASQCGLTNSN
Gossypium_hirsutum_GhGPx06-3	MAAQSS--KGSVHDFTVKDAKGNVDLSIYKKGKVLLIVNVASQCGLTNSN
Gossypium_hirsutum_GhGPx06-2	MASQSS--KGSVHDFTVKDARGNDVDLSIYKKGKVLLIVNVASQCGLTNSN
Populus_trichocarpa_PtGPx06-2	MASQSS--PQSAHDFTVKDAKGNVDLSIYKKGKVLLIVNVASQCGLTDSN
P_trichocarpa_PtGPx06 (SEQIDNO: 4)	MASQSS--AQSVHDFTVKDARENDVDLSIYKKGKVLLIVNVASQCGLTNSN
Vitis_vinifera_VvGPx06	MASQSS--PQSVHSFTVKDARGNDVDLSIYKKGKALLIVNVASQCGLTNSN
Hevea_brasiliensis_HbGPx06	MASQSE--PKSVYDFTVKDARGNDVDLSTYKKGKVLLIVNVASQCGLTNSN
Brassica_napus_BnGPx06-2	MAASSE--PKSIYDFTVKDAKGNVDLSTYKKGKVLLIVNVASQCGLTNSN
Arabidopsis_thaliana_AtGPx06	MAASSE--PKSLYDFTVKDAKGNVDLSIYKKGKVLLIVNVASQCGLTNSN
Brassica_napus_BnGPx06-1	AATSSS--PKSVHDFTVKDAKNEVDLSIYKKGKVLLIVNVASQCGLTNSN
Malus_domestica_MdGPx06	MAGHSE--SKSIHDFTVKDAKGNVDLSTYKKGKVLLIVNVASQCGLTNSN
Spinacia_oleracea_SoGPx06	MASDSSAQPKSVHEFVVRDARGNDVDLSIYKKGKVLLIVNVASQCGLTNSN
Mesembryanthemum_crystallinum_	MASQSTDQPKSIHDFIVKDARGNDVDLSIYKKGKVLLIVNVASQCGLTNSN
Gnetum_gnemon_GgGPx06	-MSSSSAKPQSVHEFTVKDIKGNVDLSLADYKKGKVLLIVNVASQCGLTNSN
Medicago_truncatula_MtGPx06	HTMASASNPQSIHDFTVKDAKGNVDNLGDYKKGKVLIIIVNVASQCGLTNSN
Solanum_tuberosum_StGPx06	-MASQSSNPQSVYDFTVKDAKGDVDLSIYKKGKVLIIIVNVASQCGLTNSN
Lycopersicon_esculentum_LeGPx0	-MATQTSNPQSVYDFTVKDAKGDVDLSIYKKGKVLIIIVNVASQCGLTNSN
Capsicum_chinense_CcGPx06	-MASQSNRPQSVYDFTVKDAKGDVDLSIYKKGKVLIIIVNVASQCGLTNSN
Nicotiana_sylvestris_NsGPx06	-MASQSSKPQSIYDFTVKDAKGNVDLSIYKKGKVLIIIVNVASQCGLTNSN
Gossypium_hirsutum_GhGPx06-1	-MASQSSKPQSIYDFTVKDAKGNVDLSIYKKGKVLIIIVNVASQCGLTNSN
Nicotiana_tabacum_NtGPx06	-MASQSSKPQSIYDFTVKDAKGNVDLSIYKKGKVLIIIVNVASQCGLTNSN
Lotus_japonicus_LjGPx06	-PDHTMAAPTSVYDFTVKDARGNDVNLGDYKKGKVLLIVNVASQCGLTNSN
Citrus_sinensis_CsGPx06	-----TSVHDFTVKDAKGQDVDLSIYKGLLLIVNVASQCGLTNSN
Pinus_taeda_PtaGPx06-2	MTGSSSSKQSSSIYDFTVKDIRGNDVDLSVYKKGKVLLIVNVASQCGLTNSN
Pinus_pinaster_PpGPx06-2	MTGTSSKQYSSVHDLTVKDIRGNDVDLSLYRGKVLLIVNVASQCGLTNSN
Pinus_taeda_PtaGPx06-1	MTESSSEQYSSSIYDFTIKDIKGNVDLSVYKKGKILLIVNVASQCGLTNSN
Solanum_tuberosum_StGPx08-2	MAGQPEKKPESVYDFTLKDAKGNVDLSTYKKGKVLLIVNVASKCGLTDSN
Helianthus_annuus_HaGPx06-3_AN	MAT-SK---KSVYDFTVKDAKGQEVLSKYRGKVLLIVNVASQCGLTNSN
Helianthus_annuus_HaGPx06-1	MASDSK---KTVYDFTVKDVKGQDVELSKYKKGKVLLIVNVASQCGLTNSN
Zea_mays_ZmGPx03-2	MAAASS-----VHGFTVKDASGEDVHLSTFKGKVLLIVNVASQCGLTNSN
P_trichocarpa_PtGPx (SEQIDNO: 6)	GVFARAATEKSVHDFTVKDINGKDVALSKFKGKALLIVNVASKCGLTSSN
Populus_trichocarpa_PtGPx01	GVFARAATEKSVHDFTVKDINGKDVALSKFKGKALLIVNVASKCGLTSSN
	: : : * . : * * . : * * : * * * . * * : *

FIGURE 20 (continued)

67/70

Oryza_sativa_OsGPx06	YTELGQLYGKYKETG-LEILAFPCNQFAGQEPGSNEQ-VVEFACTRFKAE
Hordeum_vulgare_HvGPx06	YTEMGQLYEKYREKG-LEILAFPCNQFAGQEPDSDEK-IVEFACDRFQAQ
Sorghum_bicolor_SbGPx06	YTELGSLHKKYGEKGGFEILAFPCNQFVGQEPGTNEQ-IAEFACTRFKAD
Zea_mays_ZmGPx03	YTELSQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Sorghum_bicolor_SbGPx03	YTELSQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Sorghum_bicolor_SbGPx01	YTELAQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Zea_mays_ZmGPx01	YTELAQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Setaria_italica_SiGPx01	YTELAQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Oryza_sativa_OsGPx03 (SEQIDNO: 2)	YTELSQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Zea_mays_ZmGPx01-2	YTQQAQLYEKYKNQG-FEILAFPCNQFGGQEPGTNEE-IAQFACTRFKAD
Triticum_monococcum_TmGPx01	YTELSQLYPKYKDQG-FEILAFPCNQFGGQEPGTNDE-IVQFACTRFKAE
Triticum_aestivum-TaGPx01	YTELSQLYPKYKDQG-FEILAFPCNQFGGQEPGTNDE-IVQFACTRFKAE
Hordeum_vulgare_HvGPx01	YTELSQLYPKYKDQG-FEILAFPCNQFGGQEPGTNDE-IVQFACTRFKAE
Triticum_aestivum-TaGPx03	YTELAQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Hordeum_vulgare_HvGPx03	YTELAQLYEKYKDQG-FEILAFPCNQFGGQEPGTNEE-IVQFACTRFKAE
Oryza_sativa_OsGPx01	YTELSQLYEKYKVQG-FEILAFPCNQFGGQEPGSNEE-IVQFACTRFKAE
Cucumis_melo_CmGPx06	YTELSQLYEKYKGHG-FEILAFPCNQFGGQEPGSNEE-IVQFACTRFKAE
Gossypium_hirsutum_GhGPx06-3	YTELSKLYEKYKDQG-FEILAFPCNQFGGQEPGNNEE-ILEFACTRLKAE
Gossypium_hirsutum_GhGPx06-2	YTELSKLYEQYKDQG-FEILAFPCNQFGGQEPGNNEE-ILEFACTRFKAE
Populus_trichocarpa_PtGPx06-2	YTELTQLYAKYKDQG-LEILAFPCNQFGSQEPGSSEE-IVEFACTRFKAE
P_trichocarpa_PtGPx06 (SEQIDNO: 4)	YTELTQLYDKYRDQG-LEILAFPCNQFGSQEPGNNEE-IVEFACTRFKAD
Vitis_vinifera_VvGPx06	YTELHQLYEKYKDQG-LEILAFPCNQFGAQEPGSNEE-IEKFVCTRFKAE
Hevea_brasiliensis_HbGPx06	YTELTQLYQKYKDQG-LEILAFPCNQFGSQEPGTNEE-IVEFACTRFKAE
Brassica_napus_BnGPx06-2	YTELAQLYQKYKDHG-FEILAFPCNQFGNQEPGSNEE-IVQFACTRFKAE
Arabidopsis_thaliana_AtGPx06	YTELAQLYEKYKGHG-FEILAFPCNQFGNQEPGTNEE-IVQFACTRFKAE
Brassica_napus_BnGPx06-1	YTELAQLYQKYKDHG-FEILAFPCNQFGNQEPGSNEE-IVQFACTRFKAE
Malus_domestica_MdGPx06	YTELAQLYEKYKTQG-LEILAFPCNQFGAQEPGTNDE-IVEFACTRFKAE
Spinacia_oleracea_SoGPx06	YTEMTELYEKYRELG-LEILAFPCNQFGNQEPGSNEE-VLEFACTRFKAE
Mesembryanthemum_crystallinum_	YPELTKLYEQYKDKG-LEILAFPCNQFGNQEPGDNEE-IMEFACTRFKAE
Gnetum_gnemon_GgGPx06	YKEMNELYSKYKDQG-LEILAFPCNQFAGQEPGSNEE-IAQTVCTRFKAE
Medicago_truncatula_MtGPx06	YTELSQLYEKYKSKG-LEILAFPCNQFGAQEPGSVEE-IQNFCVCTRFKAE
Solanum_tuberosum_StGPx06	YTDMTELYKKYKDQG-LEILAFPCNQFGGQEPGSIED-IQQMVCTRFKAE
Lycopersicon_esculentum_LeGPx0	YTDMTELYKKYKDQG-LEILAFPCNQFGGQEPGNIED-IQQMVCTRFKAE
Capsicum_chinense_CcGPx06	YTDMTELYRKYKDKG-LEILAFPCNQFGGQEPGSIED-IQQMVCTRFKAE
Nicotiana_sylvestris_NsGPx06	YTDLTEIYKKYKDQG-LEILAFPCNQFGGQEPGSIIE-IQNMVCTRFKAE
Gossypium_hirsutum_GhGPx06-1	YTDLTEIYKKYKDQG-LEILAFPCNQFGGQEPGSIIEESIQNMVCTRFKAE
Nicotiana_tabacum_NtGPx06	YTDMTETIYKKYKDQG-LEILAFPCNQFGGQEPGSIIE-IQNMVCTRFKAE
Lotus_japonicus_LjGPx06	YTELSQLYEKYKSKG-LEILGFPCNQFGAQEPGDNEE-IQEFVCTRFKAE
Citrus_sinensis_CsGPx06	YTELSQLYDKYKNQG-LEILAFPCNQFGAQEPGDNEE-IQEFVCTRFKAE
Pinus_taeda_PtaGPx06-2	YTELNEVYTKYKDQG-LEILAFPCNQFGGQEPGDNAQ-IAEVACTRFKAE
Pinus_pinaster_PpGPx06-2	YTELNEVYAKYKDQG-LQILAFPCNQFGGQEPGDNAQ-IAEVACTRFKAE
Pinus_taeda_PtaGPx06-1	YKEMSDLYAKYKDQG-LEILAFPCNQFGEEEPGDNAE-IAEKICTRFKAE
Solanum_tuberosum_StGPx08-2	YKDLNHLHEKYKDQG-LEILAFPCNQFGAQEPGSNDE-IVEFACTRFKSE
Helianthus_annuus_HaGPx06-3_AN	YPELTTLTYQKYKDQG-LEILGFPSNQFNGQEPGSNEE-IQEFVCTRFKAE
Helianthus_annuus_HaGPx06-1	YPELTTLTYQKYKDQG-FEILAFPCNQFGGQEPGSNEE-IQEFVCTRFKAE
Zea_mays_ZmGPx03-2	YTELAQLHEMYKDQD-FEILAFPCNQFGGQEPGTSEE-IVQLVCARFKAK
P_trichocarpa_PtGPx (SEQIDNO: 6)	YSELTHIYEKYKTQG-FEILAFPCNQFGGQEPGSNPE-IKQFACTRYKAE
Populus_trichocarpa_PtGPx01	YSELTHIYEKYKTQG-FEILAFPCNQFGGQEPGSNPE-IKQFACTRYKAE
	* : : : * . : : * . * . * . : * . . : : * * : .

FIGURE 20 (continued)

68/70

Oryza_sativa_OsGPx06	YPILGKVDVNGGNAAPLYKFLKSERG----	GLFGERIKWNFTKFLVDKEG
Hordeum_vulgare_HvGPx06	FPIFRKVDVNGNNAAPLYKFLKSERG----	GLFGERIKWNFTKFLVDKEG
Sorghum_bicolor_SbGPx06	FPVFGKVDVNGGKAAPLYKFLKSERG----	GLFGERIKWNFTKFLVDKDG
Zea_mays_ZmGPx03	YPIFDKVDVNGDNTAPIYKFLKSSKG----	SLFGDNIKWNFSKFLVDKEG
Sorghum_bicolor_SbGPx03	YPIFDKVDVNGDNTAPIYKFLKSSKG----	SLFGENIKWNFSKFLVDKEG
Sorghum_bicolor_SbGPx01	YPIFDKVDVNGNNAAPIYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Zea_mays_ZmGPx01	YPIFDKVDVNGSNAAPIYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Setaria_italica_SiGPx01	YPIFDKVDVNGNNAAPIYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Oryza_sativa_OsGPx03 (SEQIDNO: 2)	YPIFDKVDVNGDNTAPIYKFLKSSKG----	GLFGDNIKWNFSKFLVDKEG
Zea_mays_ZmGPx01-2	YPIFDKVDVNGNNAAPIYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Triticum_monococcum_TmGPx01	YPIFDKVDVNGNNVSPLYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Triticum_aestivum-TaGPx01	YPIFDKVDVNGSNVSPLYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Hordeum_vulgare_HvGPx01	YPIFDKVDVNGNNVSPLYKFLKSSKG----	GFFGDSIKWNFSKFLVDKEG
Triticum_aestivum-TaGPx03	YPIFDKVDVNGDNVAPVYKFLKSSKG----	SLFGDNIKWNFSKFLVDKEG
Hordeum_vulgare_HvGPx03	YPIFDKVDVNGDNVAPVYKFLKSSKG----	SLFGDNIKWNFSKFLVDKDG
Oryza_sativa_OsGPx01	YPIFDKVDVNGNNAAPLYKYLKSNKG----	GLFGDSIKWNFSKFLVDKEG
Cucumis_melo_CmGPx06	YPIFDKVDVNGNNAAPLYKFLKSSKG----	GLFGDAIKWNFSKFLVDKDG
Gossypium_hirsutum_GhGPx06-3	YPIFDKVDVNGENAAPIYKFLKASKG	GALGGLLGDDIKWNFAKFLVDKEG
Gossypium_hirsutum_GhGPx06-2	YPIFDKVDVNGEKAAPVYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Populus_trichocarpa_PtGPx06-2	YPIFDKVEVNGNNAAPIYKYLKSSKG----	GLFGDNIKWNFSKFLVDKEG
P_trichocarpa_PtGPx06 (SEQIDNO: 4)	YPIFDKVDVNGKNAAPIYKFLKSNKG----	GLFGDSIKWNFSKFLVDKDG
Vitis_vinifera_VvGPx06	YPIFDKIDVNGDSAAPLYKFLKSSKG----	GLFGDNIKWNFSKFLVDKDG
Hevea_brasiliensis_HbGPx06	YPIFDKVDVNGNNAAPLYKFLKSSKG----	GIFGDNIKWNFSKFLVDKDG
Brassica_napus_BnGPx06-2	YPIFDKVDVNGDSAAPLYKFLKSSKG----	GLFGDGIKWNFAKFLVDKDG
Arabidopsis_thaliana_AtGPx06	YPIFDKVDVNGDKAAPVYKFLKSSKG----	GLFGDGIKWNFAKFLVDKDG
Brassica_napus_BnGPx06-1	YPIFDKVDVNGDKAAPLYKFLKSSKG----	GLFGSGIKWNFSKFLVDKDG
Malus_domestica_MdGPx06	YPIFDKVDVNGDNAAPLYKFLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Spinacia_oleracea_SoGPx06	YPIFDKVDVNGSNAAPLYKFLKSSKG----	GLFGDGLKWNFTKFLVDKDG
Mesembryanthemum_crystallinum_	FPIFDKVDVNGSNAAPVYKYLKSSKG----	GLFGDGIKWNFTKFLVDRDG
Gnetum_gnemon_GgGPx06	FPIFDKVDVNGNNAAPLYKYMRSVKG----	GIFGDSIKWNFSKFLVDQEG
Medicago_truncatula_MtGPx06	FPVFDKVDVNGATAAPLYKYLKSSKG----	GLFGDGIKWNFSKFLVDKNG
Solanum_tuberosum_StGPx06	YPIFDKVDVNGDNAAPLYKFLKSSKG----	GFFGDGIKWNFSKFLIDKEG
Lycopersicon_esculentum_LeGPx0	YPIFDKVDVNGDNAAPLYRFLKSSKG----	GFFGDGIKWNFSKFLIDKEG
Capsicum_chinense_CcGPx06	YPVFDKVDVNGDNAAPLYKFLKSSKG----	GFFGDGIKWNFSKFLIDKEG
Nicotiana_sylvestris_NsGPx06	YPIFDKVDVNGDNAAPLYKFLKSSKG----	GFFGDSIKWNFSKFLVDKEG
Gossypium_hirsutum_GhGPx06-1	YPIFDKVDVNGDNAAPLYKFLKSSKG----	GFFGDSIKWNFSKFLVDKEG
Nicotiana_tabacum_NtGPx06	YPIFDKVDVNGDNAAPLYKFLKSSKG----	GFFGDSIKWNFSKFLVDKEG
Lotus_japonicus_LjGPx06	FPVFDKVDVNGDSAAPLYKYLKSSKG----	GLFGDKIKWNFSKFLVDKEG
Citrus_sinensis_CsGPx06	FPIFDKVDVNGDNAAPLYKHLKSSKG----	GLFGDSIKWNFSKFLVDKEG
Pinus_taeda_PtaGPx06-2	FPIFDKVEVNGNNAAPLYKFLKSSKG----	GLFGDGIKWNFTKFLVDKDG
Pinus_pinaster_PpGPx06-2	FPIFDKVEVNGSSAAPMYKFLKSSKR----	GIFGEDIKWNFTKFMVDKDG
Pinus_taeda_PtaGPx06-1	FPIFEKIEVNGSNAAPLYKFLKSSKG----	GTFGDSIKWNFTKFLVDKDG
Solanum_tuberosum_StGPx08-2	FPIFDKIEVNGENTSPFYKFLKSAKW----	GLVGDNIQWNFAKFLVDKNG
Helianthus_annuus_HaGPx06-3_AN	YPVFAKVDVNGKNAPLYSFLKASKG----	GILGDSIKWNFTKFLVDREG
Helianthus_annuus_HaGPx06-1	YPVFSKVNNGKEADPLYKFLKSSKG----	GFLGDSIKWNFTKFLVDREG
Zea_mays_ZmGPx03-2	YPILHKVDVNGEDAAPLYKFLKSSKT----	GPMGEDIKWNFAKFLVDRQG
P_trichocarpa_PtGPx (SEQIDNO: 6)	FPIFDKVDVNGPSTAPVYQFLKSSAG----	GFLGDLIKWNFEKFLVDKNG
Populus_trichocarpa_PtGPx01	FPIFDKVDVNGPSTAPVYQFLKSSAG----	GFLGDLIKWNFEKFLVDKNG
	:*:: *::*** . *. * .:::	. *. *::*** **::*:*

FIGURE 20 (continued)

69/70

Oryza_sativa_OsGPx06	HVVNRYAPTSSPLSIENDIKNLLGA-----
Hordeum_vulgare_HvGPx06	HVMNRYAPTWSPLGIENDIKKLLV-----
Sorghum_bicolor_SbGPx06	HVVSRYAPTSSPLSIENDIQKLLA-----
Zea_mays_ZmGPx03	HVVERYAPTTSPLSIEKDIKKLLG--SY-----
Sorghum_bicolor_SbGPx03	RVVERYAPXTSPLSIEKDIKKLLG--SPSTL---
Sorghum_bicolor_SbGPx01	RVVDRYAPTTSPLSIEKDIKKLLG--SS-----
Zea_mays_ZmGPx01	RVVDRYAPTTSPLSIEKDIKKLLG--SS-----
Setaria_italica_SiGPx01	RVVDRYAPTTSPLSIEKDIKKLLG--SS-----
Oryza_sativa_OsGPx03 (SEQIDNO: 2)	RVVERYAPTTSPLSMEKDIKKLLG--SS-----
Zea_mays_ZmGPx01-2	RVVDRYAPTTSPLSIEKDIKKLLG--SS-----
Triticum_monococcum_TmGPx01	HVVDRYAPTTSPLSIEKDIKKLLG--SS-----
Triticum_aestivum-TaGPx01	HVVDRYAPTTSPLSIEKDIKKLLG--SS-----
Hordeum_vulgare_HvGPx01	HVVDRYAPTTSPMSIEKDIKKLL-----
Triticum_aestivum-TaGPx03	HVVDRYAPTTSPLSIEKDIKKLLA--SY-----
Hordeum_vulgare_HvGPx03	NVVDRYAPTTSPLSIEKDIKKLLA--SS-----
Oryza_sativa_OsGPx01	RVVDRYAPTTSPLSIEKDIKKLLG--SS-----
Cucumis_melo_CmGPx06	NVVDRYAPTTSPLSIEKDVKKLLG--IA-----
Gossypium_hirsutum_GhGPx06-3	HVVDRYAPTTSPISIEKDIKKLLG-----
Gossypium_hirsutum_GhGPx06-2	HVFDYAPTTSPLSIEKDIKKLLA-----
Populus_trichocarpa_PtGPx06-2	KVVDRYAPTTSPLSIEKEVKKLLG--IA-----
P_trichocarpa_PtGPx06 (SEQIDNO: 4)	KVVDRYAPTTSPLSIEKDVKKLLG--IA-----
Vitis_vinifera_VvGPx06	KVVDRYAPTTSPLSIEKDIKKLLG--IS-----
Hevea_brasiliensis_HbGPx06	NVVDRYAPTTVPSQHREGCEETAGDCMSAQGRPN
Brassica_napus_BnGPx06-2	NVVDRYAPTTSPLSIEKDLKKLLG--VTA-----
Arabidopsis_thaliana_AtGPx06	NVVDRFAPTTSPLSIEKDVKKLLG--VTA-----
Brassica_napus_BnGPx06-1	NVVDRYAPTTSPLSIEKDLKKLLG--VTA-----
Malus_domestica_MdGPx06	KVVDRYAPTTSPLSIEKDVKKLLG--VA-----
Spinacia_oleracea_SoGPx06	NVVDRYAPTTSPKSIKDVKKLLG--IQK-----
Mesembryanthemum_crystallinum_	KVVDRYAPTTSPASIEKDIKKLLG--TS-----
Gnetum_gnemon_GgGPx06	NVVDRYAPTTSPKSIKDVKKLLG-----
Medicago_truncatula_MtGPx06	NVVDRYAPTTSPLSIEKDLKLLD--A-----
Solanum_tuberosum_StGPx06	HVVDRYSPTTTPASMEKDIKKLLG--VA-----
Lycopersicon_esculentum_LeGPx0	HVVDRYSPTTSPASMEKDIKKLLG--VA-----
Capsicum_chinense_CcGPx06	NVVDRYSPTTSPASMEKDIKKLLG--VAA-----
Nicotiana_sylvestris_NsGPx06	NVVDRYSPTTTPASMEKDIKKLLG--VA-----
Gossypium_hirsutum_GhGPx06-1	NVVDRYSPTTTPASMEKDIKKLLG--VA-----
Nicotiana_tabacum_NtGPx06	NVVDRYSPTTTPASMEKDIKKLLG--VA-----
Lotus_japonicus_LjGPx06	NVVERYAPTTSPLSIEKDLVKLLG--A-----
Citrus_sinensis_CsGPx06	NVVERYAPTTSPLSIEKDIKKLLE--TA-----
Pinus_taeda_PtaGPx06-2	NVVERYAPTTSPLSIEKDVKKLLG--IA-----
Pinus_pinaster_PpGPx06-2	NVVERYAPTTSPLSIEKDVKKLLG--IA-----
Pinus_taeda_PtaGPx06-1	NVVERYAPTTSPLSFEKDVKKLLG--VA-----
Solanum_tuberosum_StGPx08-2	QVSDRYPTTSPLSMERDIKVLLG--IE-----
Helianthus_annuus_HaGPx06-3_AN	KVVDRYAPTTTPLSIEKDIKKLLG--VA-----
Helianthus_annuus_HaGPx06-1	KVVDRYAPTTSPLSIEKDIKKLLN--VA-----
Zea_mays_ZmGPx03-2	HVAERYAPTTYPLSIQKDIKKLLG--GSY-----
P_trichocarpa_PtGPx (SEQIDNO: 6)	KVVERYQPTTSPFQIEKDIQKLLAA-----
Populus_trichocarpa_PtGPx01	KVVERYQPTTSPFQIEKDIQKLLAA-----
	. * . * : * * .

FIGURE 20 (continued)

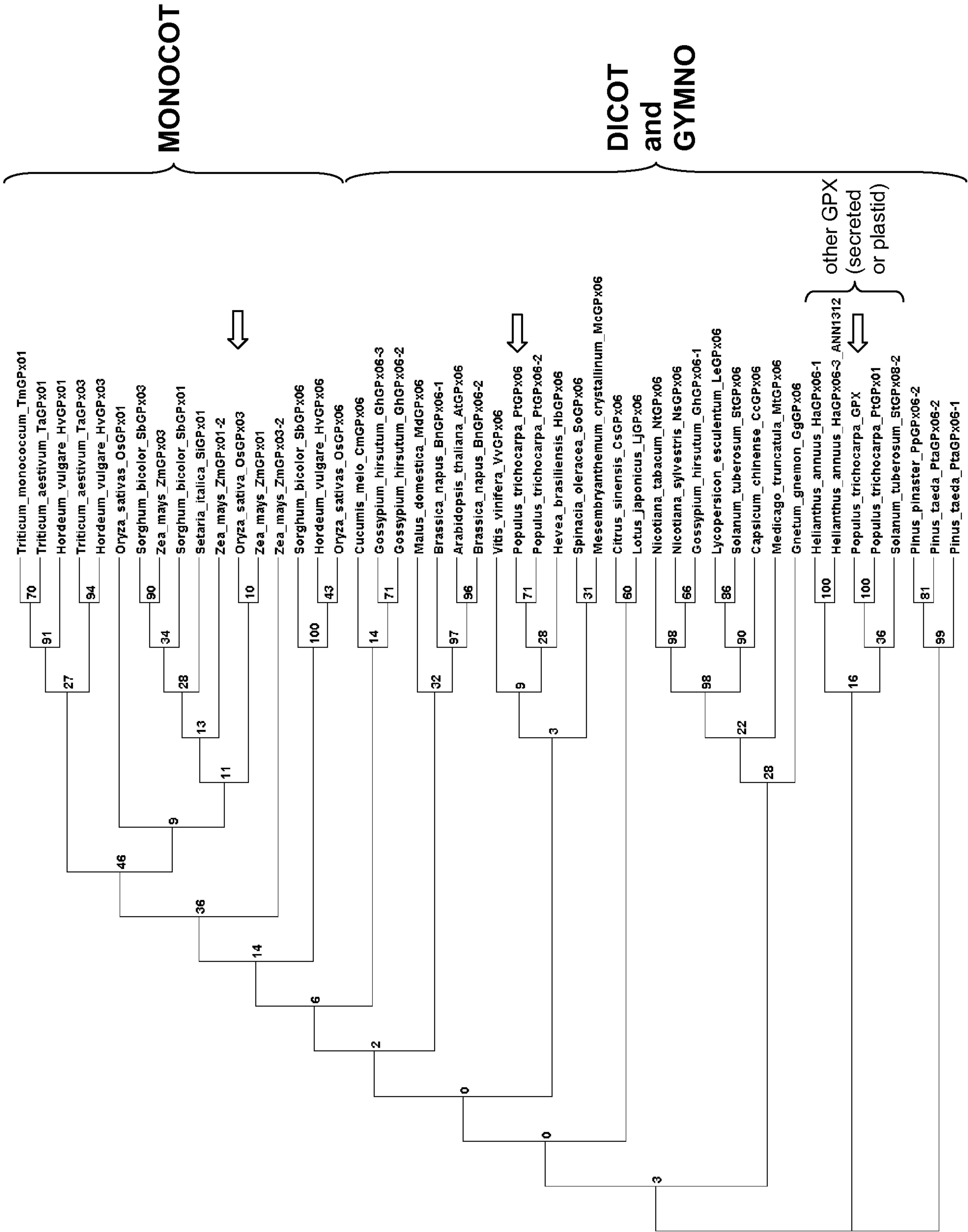


FIGURE 21