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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 May 2011 (19.05.2011)

PCT

(10) International Publication Number
WO 2011/058029 A1

(51) **International Patent Classification:**
C12N 15/82 (2006.01) ***C07K 14/415*** (2006.01)
C12N 15/29 (2006.01) ***A01H 5/00*** (2006.01)

(21) **International Application Number:**
PCT/EP2010/067164

(22) **International Filing Date:**
10 November 2010 (10.11.2010)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**

61/260935	13 November 2009 (13.11.2009)	US
61/260939	13 November 2009 (13.11.2009)	US
61/260946	13 November 2009 (13.11.2009)	US
09175964.7	13 November 2009 (13.11.2009)	EP
09175922.5	13 November 2009 (13.11.2009)	EP
09175997.7	13 November 2009 (13.11.2009)	EP
09178693.9	10 December 2009 (10.12.2009)	EP
61/285208	10 December 2009 (10.12.2009)	US
61/285219	10 December 2009 (10.12.2009)	US
09178742.4	10 December 2009 (10.12.2009)	EP

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

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only):* **HATZFELD, Yves** [FR/FR]; rue des Dondaines 18C, F-59000 Lille (FR). **REUZEAU, Christophe** [FR/BE]; Volkenslaan 437, B-24350 9000 Gent (BE). **FRANKARD, Valerie** [BE/BE]; Avenue des Croix du Feu 84, 1410 Waterloo (BE).

(74) **Agent: POPP, Andreas;** BASF SE, GVX/B - C 006, 67056 Ludwigshafen (DE).

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

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

Published:

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

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

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

WO 2011/058029 A1

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

(57) **Abstract:** The present invention relates generally to the field of molecular biology and concerns a method for enhancing various economically important yield-related traits in plants. More specifically, the present invention concerns a method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides hitherto unknown an O-FUT, or By-Pass (BPS) polypeptide, or SIZ1, or bZIP-S, or SPA15-like - encoding nucleic acids, and constructs comprising the same, useful in performing the methods of the invention.

Plants having enhanced yield-related traits and a method for making the same

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 a fucose protein O-fucosyltransferase (O-FUT) polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid encoding an O-FUT 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.

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.

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.

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.

One approach to increasing yield (seed yield and/or biomass) in plants may be through modification of the inherent growth mechanisms of a plant, such as the cell cycle or various signalling pathways involved in plant growth or in defense mechanisms.

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 a fucose protein O-fucosyltransferase (O-FUT) polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, in a plant.

Background

Plant small ubiquitin-like modifier (SUMO) E3 ligase is a focal controller of Pi starvation-dependent responses. Said polypeptide is also required for SA and PAD4-mediated R gene signaling, which in turn confers innate immunity in the plant. SUMO E3 ligases of the PIAS/SIZ family facilitate SUMO conjugation to lysine (K) residues in the SUMO consensus motif, YKXE/D (Y, a large hydrophobic residue; K, the acceptor lysine; X, any amino acid; E/D, glutamate or aspartate), located in protein substrates (Jin et al., 2008).

SUMO modification of target proteins in yeast and metazoans has been implicated in the regulation of innate immunity, cell-cycle progression and mitosis, DNA repair, chromatin stability, nucleocytoplasmic trafficking, subnuclear targeting, ubiquitination antagonism and transcriptional regulation (Johnson, 2004; Gill, 2005). Sumoylation in plants is reported to be involved in biotic and abiotic stress responses, flowering and development (Chosed et al., 2006; Downes and Vierstra, 2005; Kurepa et al., 2003; Lee et al., 2007; Miura et al., 2005, 2007; Novatchkova et al., 2004; Yoo et al., 2006).

Growth and development of all organisms depend on proper regulation of gene expression. The control of transcription initiation rates by transcription factors (TF) represents one of the most important means of modulating gene expression. TFs can be grouped into different protein families according to their primary and/or three-dimensional structure similarities in the DNA-binding and multimerization domains. Transcription factors (TFs) play crucial roles in almost all biological processes. Structurally, the basic region/leucine zipper (bZIP) class of TFs are usually classified by their DNA-binding domains, a basic region, and a leucine zipper dimerisation motif. Dimerisation may occur in homo or hererodimerisation. A common partner in dimarisation of bZIP TFs are TFs of the bHLH family. Proteins with bZIP domains are present in all eukaryotes analysed to date. Some, such as Jun/Fos or CREB, have been studied extensively in animals and serve as models for understanding TF-DNA interactions, ternary complex formation and TF post-translational modifications (Jakoby et al. 2002 TRENDS in Plant Science Vol.7 .No.3 106_111). In plants, basic region/leucine zipper motif (bZIP) transcription factors regulate processes including pathogen defence, light and stress signaling, seed maturation and flower development. The Arabidopsis genome sequence contains more than 75 distinct members of the bZIP family. Using phylogenetic analysis and common domains, the flowering plants bZIP TFs family has been subdivided into thirteen homologous groups. In Arabidopsis, rice and black cottonwood members of Group S of bZIPs TFs share two characteristics: they harbor a long leucine zipper (eight to nine heptads) and are encoded by intron-less genes.

Genes associated with leaf senescence have been studied since late 90s with the purpose to better understand the molecular mechanisms which are in the basis of leaf senescence. A wide range of senescence-associated genes (SAGs) were therefore identified, cloned and characterised from different plant origin such as A. thaliana, B. napus, tomato, maize, barley, sweet potato, rice, etc. However, many other SAGs remain unknown. Several SAGs genes, including SPA15 gene were cloned and characterised (Huang, Y.-J. et al. (2001) -

Cloning and characterization of leaf senescence up-regulated genes in sweet potato. Physiolog. Plantarum, 113: 384-391). Expression patterns of SPA15 suggest that it is highly specifically expressed in senescing leaves and SPA15 protein is a cell wall-associated protein Said expression is not influenced by growth-enhancing hormones, such as auxin, cytokinin, gibberllin, but is strongly induced by ethylene. (Yap M.N. et al. (2003) - Molecular characterization of a novel senescence-associated gene SPA15 induced during leaf senescence in sweet potato. Plant Molecular Biology 51: 471-481).

Summary

1. O-FUT- like polypeptides

Surprisingly, it has now been found that modulating expression of a nucleic acid encoding an O-FUT polypeptide gives plants having enhanced yield-related traits, in particular increased yield relative to control plants.

According one embodiment, there is provided a method for improving yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an O-FUT polypeptide.

2. By-Pass-like polypeptides

Surprisingly, it has now been found that modulating expression of a nucleic acid encoding a BPS polypeptide gives plants having enhanced yield-related traits, in particular increased seed yield relative to control plants.

According one embodiment, there is provided a method for improving yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a BPS polypeptide.

3. SIZ1 -like polypeptides

Surprisingly, it has now been found that modulating expression of a nucleic acid encoding a SIZ1 polypeptide gives plants having enhanced yield-related traits, in increased seed yield relative to control plants.

According one embodiment, there is provided a method for improving yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a SIZ1 polypeptide.

4. bZIP-S- like polypeptides

Surprisingly, it has now been found that modulating expression of a nucleic acid encoding a bZIP-S polypeptide gives plants having enhanced yield-related traits, relative to control plants.

According one embodiment, there is provided a method for improving yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a bZIP-S polypeptide.

5. SPA15-like polypeptide

Surprisingly, it has now been found that modulating expression of a nucleic acid encoding a SPA15-like polypeptide gives plants having enhanced yield-related traits, in particular increased seed yield relative to control plants.

According one embodiment, there is provided a method for improving yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a SPA15-like polypeptide.

Definitions

Polypeptide(s)/Protein(s)

The terms "polypeptide" and "protein" are used interchangeably herein and refer to amino acids in a polymeric form of any length, linked together by peptide bonds.

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

The terms "polynucleotide(s)", "nucleic acid sequence(s)", "nucleotide sequence(s)", "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)

"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.

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 a-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; insertions will usually be of the order of about 1 to 10 amino acid residues. 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 well 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.

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).

Orthologue(s)/Paralogue(s)

Orthologues and paralogues 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.

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.

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).

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 30(1): 276-280 (2002)). 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.

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.

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.

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 (Tm) 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.

The Tm 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 Tm decreases about 1°C per % base mismatch. The Tm 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):

Tm= 81.5°C + 16.6xlogio[Na^+][a] + 0.41x%[G/C [b]] - 500x[L [c]]$^{-1}$ - 0.61x% formamide

2) DNA-RNA or RNA-RNA hybrids:

Tm= 79.8 + 18.5 (logio[Na^+][a]) + 0.58 (%G/C [b]) + 11.8 (%G/C [b])^2 - 820/L [c]

3) oligo-DNA or oligo-RNA [d] hybrids:

For <20 nucleotides: T_m= 2 (l_n)

For 20-35 nucleotides: T_m= 22 + 1.46 (l_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; l_n, = effective length of primer = 2x(no. of G/C)+(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.

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 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. 1xSSC is 0.15M NaCI and 15mM sodium citrate; the hybridisation 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 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 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 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 (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 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

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

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 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 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 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 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 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 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.

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 1/500,0000 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.

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.

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.

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.

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

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.

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 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:

Table 2b: Examples of root-specific promoters

Gene Source	Reference
RCc3	Plant Mol Biol. 1995 Jan;27(2):237-48
Arabidopsis PHT1	Kovama et al., 2005; Mudge et al. (2002, Plant J. 31:341)
Medicago phosphate transporter	Xiao et al., 2006
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
SbPRPI	Suzuki et al., Plant Mol. Biol. 21: 109-1 19, 1993.
LRX1	Baumberger et al. 2001 , Genes & Dev. 15:1 128
BTG-26 Brassica napus	US 20050044585
LeAMTI (tomato)	Lauter et al. (1996, PNAS 3:8139)
The LeNRTM (tomato)	Lauter et al. (1996, PNAS 3:8139)

class 1patatin gene (potato)	Liu et al., Plant Mol. Biol. 153:386-395, 1991 .
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;1 Np (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	EP991 06056.7
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 a-globulin REB/OHP-1	Nakase et al. Plant Mol. Biol. 33: 513-522, 1997
rice ADP-glucose pyrophos-phorylase	Trans Res 6:157-68, 1997
maize ESR gene family	Plant J 12:235-46, 1997
sorghum a-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
PRO01 17, 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
a-amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-21 1, 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 Itr1 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(1 4):91 75-82
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

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.

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., 2001
Maize Phosphoenolpyruvate carboxylase	Leaf specific	Kausch et al., 2001
Rice Phosphoenolpyruvate carboxylase	Leaf specific	Liu et al., 2003
Rice small subunit Rubisco	Leaf specific	Nomura et al., 2000
rice beta expansin EXBP9	Shoot specific	WO 2004/070039

Pigeonpea small subunit Rubisco	Leaf specific	Panguluri et al., 2005
Pea RBCS3A	Leaf specific	

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.

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 et al. (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

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

"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 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 nptll 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 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/luceferase 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 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, 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 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).

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 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.

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. The term "modulating the activity" 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.

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.

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., W09322443), 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 1:1 183-1200). 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.

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 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 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.

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 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 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. U.S. Patent No. 4,987,071 ; and Cech et al. U.S. Patent No. 5,1 16,742). 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 , 141 1-1418). 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 95/03404; Lutziger et al. (2000) WO 00/00619; Prinsen et al. (1997) WO 97/13865 and Scott et al. (1997) WO 97/381 16).

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).

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.

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 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, 1121-1 133, 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.

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, 1099-1 102); 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 A 1, 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) 871 1). 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 Hofgen 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 208:274-289; 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: 1194-1 199], 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 is 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 maker 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 Hofgen and Willmitzer.

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 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 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 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) 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 transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

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 Traits

Yield related traits comprise one or more of yield, biomass, seed yield, early vigour, greenness index, increased growth rate, improved agronomic traits (such as improved Water Use Efficiency (WUE), Nitrogen Use Efficiency (NUE), etc.).

Yield

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. The term "yield" of a plant may relate to vegetative biomass (root and/or shoot biomass), to reproductive organs, and/or to propagules (such as seeds) of that plant.

Taking corn as an example, a yield increase 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 seeds divided by the total number of seeds and multiplied by 100), among others. 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 (florets) per panicle, increase in the seed filling rate (which is the number of filled seeds divided by the total number of seeds and multiplied by 100), increase in thousand kernel weight, among others. In rice, submergence tolerance may also result in increased yield.

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 dry mature seed up to the stage where the plant has produced dry 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. Mild stresses are the everyday biotic and/or abiotic (environmental) stresses to which a plant is exposed. Abiotic stresses may be due to drought or excess water, anaerobic stress, salt stress, chemical toxicity, oxidative stress and hot, cold or freezing temperatures. The abiotic stress may be an osmotic stress caused by a water stress (particularly due to drought), salt stress, oxidative stress or an ionic stress. Biotic stresses are typically those stresses caused by pathogens, such as bacteria, viruses, fungi, nematodes and insects.

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

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.

The term salt stress is not restricted to common salt (NaCl), but may be any one or more of: NaCl, KCl, LiCl, $MgCl_2$, $CaCl_2$, amongst others.

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 (filled) seeds; d) increased seed filling rate (which is expressed as the ratio between the number of filled seeds divided by the total number of seeds); e) increased harvest index,

which is expressed as a ratio of the yield of harvestable parts, such as seeds, divided by the total biomass; and f) increased thousand kernel weight (TKW), which is extrapolated from the number of filled 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.

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. Increased yield may also result in modified architecture, or may occur because of modified architecture.

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 colour) 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.

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.

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 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 rimpaui, 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., Vicia spp., Vigna spp., Viola odorata, Vitis spp., Zea mays, Zizania palustris, Ziziphus spp., amongst others.

Control plant(s)

The choice of suitable control plants is a routine part of an experimental setup and may 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 are individuals missing the transgene by segregation. 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

Surprisingly, it has now been found that modulating expression in a plant of a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide gives plants having enhanced yield-related traits relative to control plants. 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 O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide and optionally selecting for plants having enhanced yield-related traits.

A preferred method for modulating (preferably, increasing) expression of a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide is by introducing and expressing in a plant a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide.

Concerning O-FUT polypeptides, any reference hereinafter to a "protein useful in the methods of the invention" is taken to mean an O-FUT 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 O-FUT polypeptide. 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 "O-FUT nucleic acid" or "O-FUT gene".

An "O-FUT polypeptide" as defined herein refers to any polypeptide comprising a fucosyltransferase domain with an accession PFam number PF10250 or IPR019378

denomination (earlier IPR004348, DUF246 and PF03138). O-FUT polypeptides are involved in the biosynthesis of oligosaccharides, polysaccharides and glycoconjugates. O-FUT polypeptides belong to Enzyme Classification Number EC 2.4.1.221.

Preferably, a PF10250 domain has at least, in increasing order of preference, 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% or more sequence identity to the sequence SEQ ID NO 22.

Preferably, an O-FUT polypeptide has at least, in increasing order of preference, 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% or more sequence identity to the sequence SEQ ID NO 2.

Additionally or alternatively, the O-FUT polypeptide useful in the methods of the invention comprises one or more sequence motifs having at least, in increasing order of preference 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% or more sequence identity to any one or more of motifs 1 to 3:

The amino acids indicated herein in square brackets represent alternative amino acids for a particular position.

Motif 1: HYIALHLRYEKDM (SEQ ID NO: 261)

Motif 2: IYIVAGEIYGGHSMD (SEQ ID NO: 262)

Motif 3: ALDYNVAVQSDVFVYTYDGNMAKAVQGH (SEQ ID NO: 263)

Motifs 1 to 3 are typically found in any O-FUT polypeptide of any origin.

In a preferred embodiment of the present invention the O-FUT polypeptide of the invention may comprise a conserved Arginine residue in Motif 1.

In another preferred embodiment of the present invention, the O-FUT polypeptide of the invention comprises a conserved Arginine residue in Motif 1 and comprises in addition to Motif 1, at least Motif 2 or Motif 3 as defined above.

In a most preferably embodiment of the present invention, the O-FUT polypeptide of the invention comprises a conserved Arginine residue in Motif 1 and comprises in addition to Motif 1, Motif 2 and Motif 3 as defined above.

Motifs 1 to 3 were derived from an alignment obtained with AlignX from Vector NTI (Invitrogen).

Additionally or alternatively, the homologue of a O-FUT 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 amino acid represented by SEQ ID NO: 2, 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). 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 O-FUT 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: 261 to SEQ ID NO: 263 (Motifs 1 to 3).

Concerning By-Pass (BPS) polypeptides, any reference hereinafter to a "protein useful in the methods of the invention" is taken to mean a BPS 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 a BPS polypeptide. 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 "BPS nucleic acid" or "BPS gene".

A "BPS polypeptide" as defined herein refers to any plant specific polypeptide comprising a single transmembrane domain and the at least one of the following three motifs:

Motif 4: SWM[KT][LQ]A[MI]ESLC[EA][TI]H[TN]DIKTLIT[DE]LELP (SEQ ID NO: 341)

Motif 5: D[IL]C[IN]AFSSE[LI][ST]RLNQGHL[LY]L[QK]C[AV]LHNL[DE][SG]SS (SEQ ID NO: 342)

Motif 6: GKVLM[RQ]A[ML]YGV[KR]V[VQ]TV[FY][IV]CS[VI]FA[AV]AFSGS (SEQ ID NO: 343)

Preferably, the Motifs 4, 5 and 6 of a BPS polypeptide has at least, in increasing order of preference, 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% or more sequence identity to the sequence of SEQ ID NO: 341, 342 and 343 (Motif4, Motif 5 and Motif 6).

Motifs 4, 5 and 6 correspond to a consensus sequences which represent conserved protein regions in BPS polypeptide of any plant origin.

Additionally or alternatively, the BPS polypeptide useful in the methods of the invention comprises one or more sequence motifs having at least, in increasing order of preference 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% or more sequence identity to any one or more of motifs 7 to 9:

Motif 7: SWM[KT][LQ]A[MI]ESLC[EA][TI]H[NT]D[IV]KTLIT[DE]LELPVSDW[DE][ED]KW[IV]DVYLD[IN]SVKL (SEQ ID NO: 344)

Motif 8: SL[ND]LPK[VI]KNSAKGKVLM[RQ]A[ML]YGV[KR]V[QV]TV[FY][IV]CSVFA[AV]AFSGS (SEQ ID NO: 345)

Motif 9: PQ[ED]P[HP]R[PS]F[FL]PFGNPF (SEQ ID NO: 346)

Motifs 7, 8 and 9 correspond to consensus sequences which represent conserved protein regions in BPS polypeptide of Trees, Fabales, Solanales, Brassicales and Other Dicots clusters as defined in Figure 6.

In a preferred embodiment of the present invention the BPS polypeptide of the invention may comprise Motifs 7, 8 and 9 in addition to Motif 4, Motif 5 and Motif 6 as defined above, or may comprise a motif having, in increasing order of preference at least 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% or more sequence identity to any one or more of Motifs 10 to 12:

Motif 10: [VM]PK[EDN]K[SDN][DQ]ILT[LV]SWM[KS][QL]AM[EA]SLC[EQ]TH[KN][NAS]I[KNR]TL[IV]TDL[EQ]LPVSD[WL]E[ED][KN][WF][VI][DY][IV]Y (SEQ ID NO: 347)

Motif 11: LPK[VK]KNSAKGKVL[ML]RA[LF]YGVKV[KQ]T[LV]YI[CS][SG]VF[AT]AA[FW]S [GD]S[ST][NQK][ND]L[FL][YD][LV][TP][VI][SP][NE][EK] (SEQ ID NO: 348)

Motif 12: [PL]WA[KQP][SVA]F[MT][DE][MLV]Q[NS][TV][VM]N[AGPS]EI[KR][ND][IM][FL][LS] S[DG][GR][LFS]T[VI][LIM]K[ED]LE[AS]V[DE][AS][GS]V[KE][KQ]L[YA][PT][AM][IV]Q[DQE]G [SV] (SEQ ID NO: 349)

Motifs 10, 11 and 12 correspond to consensus sequences which represent conserved protein regions in BPS polypeptide of Brassicales cluster as defined in Figure 6.

More preferably, the BPS polypeptide comprises in increasing order of preference at least 3, at least 4, at least 5, at least 6, at least 7, at least 8 or all 9 motifs.

Motifs 4 to 12 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 with a frequency higher than 0.2. Residues within square brackets represent alternatives.

Additionally or alternatively, the homologue of a BPS 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 amino acid represented by SEQ ID NO: 268, 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). 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 BPS 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: 341 to SEQ ID NO: 349 (Motifs 4 to 12).

Concerning SIZ1 polypeptides, any reference hereinafter to a "protein useful in the methods of the invention" is taken to mean a SIZ1 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 a SIZ1 polypeptide. 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 "SIZ1 nucleic acid" or "SIZ1 gene".

A "SIZ1 polypeptide" as defined herein refers to any small ubiquitin-like modifier (SUMO) E3 ligase comprising at least one of the three following domains with PFam accession numbers: a "SAP" binding-DNA domain - PF02037; a "PHD Zn finger domain" domain PF00628 and a "MIZ SP/RING Zn finger" domain - PF02891 , respectively with an average length of 34, 54 and 49 amino acids.

Preferably, the "SAP" domain of a SIZ1 polypeptide has at least, in increasing order of preference, 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% or more sequence identity to the sequence located between amino acid 11 and 45 of SEQ ID NO 354.

Preferably, the "PHD Zn finger domain'Of a SIZ1 polypeptide has at least, in increasing order of preference, 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% or more sequence identity to the sequence located between amino acid 114 and 148 of SEQ ID NO 354.

Preferably, the "MIZ SP/RING Zn finger" domain of a SIZ1 polypeptide has at least, in increasing order of preference, 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% or more sequence identity to the sequence located between amino acid 359 and 408 of SEQ ID NO 354.

Additionally or alternatively, the SIZ1 polypeptide useful in the methods of the invention comprises one or more sequence motifs having at least, in increasing order of preference 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% or more sequence identity to any one or more of motifs 13 to 15:

Motif 13: FYCEICRLTRADPF (SEQ ID NO: 412)

Motif 14: FCFGVRLVKRR (SEQ ID NO: 413)

Motif 15: SDIEVVADFFGVNLRCPMSG (SEQ ID NO: 414)

Motifs 13 to 15 are typically found in any SIZ1 polypeptide of any origin.

In another preferred embodiment of the present invention the SIZ1 polypeptide of the invention may comprise Motifs 16, 17 and 18 in addition to Motif 13, Motif 14 and Motif 15 as defined above, or may comprise a motif having, in increasing order of preference at least 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% or more sequence identity to any one or more of Motifs 16 to 18:

Motif 16: RKWQCPICLKN (SEQ ID NO: 415)

Motif 17: VIVLSDSDDEND (SEQ ID NO: 416)

Motif 18: PSLQIFLP (SEQ ID NO: 417)

Motifs 16, 17 and 18 correspond to a consensus sequences which represent conserved protein regions in an SIZ1 polypeptide of II class origin, to which O. sativa and H. vulgare and A. thaliana belong.

The motifs were designed with 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.)

Additionally or alternatively, the homologue of a SIZ1 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 amino acid represented by SEQ ID NO: 354, 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). 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 SIZ1 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: 409 to SEQ ID NO: 414 (Motifs 13 to 18).

Concerning bZIP-S polypeptides, any reference hereinafter to a "protein useful in the methods of the invention" is taken to mean a bZIP-S 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 a bZIP-S polypeptide. 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 "bZIP-S nucleic acid" or "bZIP-S gene".

A "bZIP-S polypeptide" as defined herein refers to any transcription factor (TF) of the basic leucine zipper (bZIP) family comprising a Basic Leucine Zipper domain (bZIP domain, Pfam accession number PF0170 and InterPro entry IPR011616) and one or more of motifs 19 to 21 as described below.

A bZIP-S TF is characterized by a long conserved domain (bZIP domain) typically having 40- to 80-amino-acids that is composed of two regions: a basic region involved in the binding of the TF to its target DNA, and a leucine zipper required for multimerization, typically dimerisation of the bZIP-S. A preferred bZIP polypeptide of the invention comprises a bZIP domain (bZIP domain, Pfam accession number PF0170 and InterPro entry IPR011616) 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%, 99% or 100% sequence identity to the sequence of any of the bZIP domains of the polypeptides of Table A4, preferably of the domain located between amino acids 28 and 89 of SEQ ID NO: 422 (bZIP domain in SEQ ID NO: 422). Methods to determine the presence of a bZIP domain in a polypeptide are well known in the art. The Examples section herein gives further details on such methods.

Further preferred, the bZIP domain in the bZIP-S polypeptide useful in the methods of the invention comprises a basic region 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%, 99% or 100% sequence identity to the to the sequence located between amino acids 33 and 43 of SEQ ID NO: 422 (Basic region of the bZIP domain in SEQ ID NO: 422 having SMART accession number SM00036).

In addition or alternatively the bZIP polypeptide useful in the invention has a sequence which when used in the construction of a phylogenetic tree of bZIP transcription factors such as those of Arabidopsis, black cottonwood and rice described on Figure 3 of Guedes Correa et al. PLoS ONE, 2008, Volume 3, Issue 8, e2944), herein incorporated by reference, clusters with the bZIPs of group S, preferably of group SE2, most preferably with any one of AtbZIP2 (AT2g18160), AtbZIPH (At4g34590) and AtbZIP14 (At1g75390) rather than with any other group or bZIP TF. Methods to perform phylogenetic analysis and draw a phylogenetic tree are well known in the art, as for examples described herein or in Guedes Correa et al. 2008.

The bZIP polypeptide useful in the invention comprises one or more of the following conserved motifs:

Motif 19: a protein motif 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%, 99% or 100% sequence identity to any motif selected from Table 3a, preferably to SEQ ID NO: 522 (KQKHLDDLAVQLSQLRNENQQILTSVNLTTQ);

Motif 20: a protein motif 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%, 99% or 100% sequence identity to any motif selected from Table 3b, preferably to SEQ ID NO: 557 (VEAENSVLRAQMGELSNRLESLNEIV);

Motif 21: a protein motif 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%, 99% or 100% sequence identity to any motif selected from Table 3c, preferably to SEQ ID NO: 600 (KRMISNRESARRSRM);

Alternative Motifs 19 to 21 may be defined as follows:

Motif 19: a protein motif having in increasing order of preference at least 15, 16, 17, 18, 19,20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33 amino acid residues identical to any of the motifs of Table 3a, preferably to the motif represented by SEQ ID NO: 522 (KQKHLDDLAVQLSQLRNENQQILTSVNLTTQ), preferably the motif has sequence sharing at least 15, 16, 17, 18, 19,20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33 amino acids in common with the sequence of any of the motifs of Table 3a;

Motif 20: a protein motif having in increasing order of preference at least 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 amino acid residues identical to any of the motifs of Table 3b, preferably to the motif represented by SEQ ID NO: 557 (VEAENSVLRAQMGELSNRLE SLNEIV), preferably the motif has sequence sharing at least 15, 16, 17, 18, 19,20, 21, 22, 23, 24, 25, 26 amino acids in common with the sequence of any of the motifs of Table 3b;

Motif 21: a protein motif having in increasing order of preference at least 8, 9, 10, 11, 12, 13, 14, 15 amino acid residues identical to any of the motifs of Table 3c, preferably to the motif represented by SEQ ID NO: 600 (KRMISNRESARRSRM);

Examples of Motifs 19 to 21 are given in Tables 3a to 3c herein.

Table 3a. Motif 20 as present in the polypeptides of Table A4.

Polypeptide Name as in Table A4	Amino acid coordinate at Start	Amino acid coordinate at end	Sequence
Populus_trichocarpa_EF14	49	80	KQKHLDDLVAQVAQLKKENHQIITSINITTQ
Nicotiana_tabacum_AY0455	50	81	KQKHLDDLMAQVATLRKENNQILTSMNVTTQ
P.trichocarpa_818112#1	49	80	KQKHLDDLMAQVSQLRKENHQIITGINITTQ
G.max_GM06MC32426_sk55g0	49	80	KQKHLDDLVSQVAQLRKENQQILTSVNITTQ
Glycine_max_AF532621#1	51	82	KQKHLDDLVSQVAQLRKENQQILTSVNITTQ
P.trichocarpa_715285#1	50	81	KQKHLDDLMAQVTQLRKDNNQILTTINVTTQ
Capsicum_chinense_AF1277	49	80	KQKHLDDLMAQVSTLRKENDQILTSMNVTTQ
Capsicum_chinense_AF4303	49	80	KQKHLDDLMAQVSTLRKENDQILTSMNVTTQ
V.vinifera_GSVIVT0003633	49	80	KQKHLDDLMAQAAQLRKENSQIITSMNVTTQ
Medicago_truncatula_BT05	49	80	KQKHLDDLVSQVSKLRKENQEILTSVNITTQ
Mt_bZIP	49	80	KQKHLDDLVSQVSKLRKENQEILTSVNITTQ
A.thaliana_AT1G75390.1#1	60	91	KQKHLDDLTAQVTHLRKENAQIVAGIAVTTQ
Arabidopsis_thaliana_BT0	54	85	KQKHLDDLTAQVTHLRKENAQIVAGIAVTTQ
Solanum_lycopersicum_FJ6	49	80	KQKHLDDLMSQVTNLRKENNQILTSMNVTTQ
Capsicum_annuum_AY789639	49	80	KQKHLNDLMAQVSTLRKENDQILTSMNVTTQ
Arabidopsis_thaliana_AF0	46	77	KQKLLDDLTAQVNHLKKENTEIVTSVSITTQ
Arabidopsis_thaliana_AT4	46	77	KQKLLDDLTAQVNHLKKENTEIVTSVSITTQ
G.max_GM06MC17143_596542	52	83	KQKHLDDLASQVTQLRNENHQILTSVNLTTQ
V.vinifera_GSVIVT0001455	39	70	KQKHLDDLMAQVAQLRKENNEILSSINITNQ
B.napus_BN06MC17829_4559	46	77	KQKLLDDLTAQVNQLRKENSEIVTSVSITTQ
Oryza_sativa_Japonica_Gr	56	87	KQRHLDDLTAQVAHLRRENAHVATALGLTTQ
Mt_bZIP2	49	80	KQKHLDDLAVQLSQLRNENQQILTSVNLTTQ
V.vinifera_GSVIVT0003689	49	80	KQKHLDDMMAQMVHLRKENNRILTTMNVTTQ

P.trichocarpa_710131#1	48	79	KQKYLGDLMAQVAQLRTDNNQILTTINVTTQ
Petroselinum_crispum_AJ2	47	78	KQNHLDELMAQAAQLRKENNQIITTTNLTTQ
S.bicolor_Sb04g002700.1#	56	87	KQRHLDELTAQAAHLRRENAHVATALGLTTQ
B.napus_BN06MC15489_4421	35	66	KQKLLDDLTAQVNRLKEQNNEILTSVSITTQ
B.napus_BN06MC23287_4905	47	78	KQKHLDDLTAQIAHFLEENSHIVAGISVTTQ
Zea_mays_BT067356#1	56	87	KQRHLDELTAQAAHLRRENAHVATALGLTAQ
Z.mays_ZM07MC37862_BFb03	56	87	KQRHLDELTAQAAHLRRENAHVATALGLTAQ
Zea_mays_BT018074#1	55	86	KQRHLDELTAQAAHLRRENAHVATALGLTAQ
Zea_mays_EU976771#1	56	87	KQRHLDELTAQAAHLRRENAHVATALGLTAQ
T.erecta_SIN_01b-CS_Scar	45	76	KQKHLDDLMTQLSQLKKENNEIMAHVSITMQ
P.trichocarpa_719591#1	50	81	KQKHLDDLMGQLGQLSKENNEILKRMNVTSQ
Arabidopsis_thaliana_AT2	50	81	KQKHVDDLTAQINQLSNDNRQILNSLTVTSQ
H.vulgare_c62949710hv270	57	88	KQRHLDDLAAQAAHLRRENAHVAAALGLTAR
G.max_GM06MC32046_sj86f0	43	74	KQKHLDDLASQLTQLRSQNQQLLTSVNLTSH
Arabidopsis_thaliana_AF0	50	81	KQEHVDDLTAQINQLSNDNRQILNSLTVTSQ
S.bicolor_Sb07g015450.1#	58	89	KQQHLDDLNLQVDKLRTTKQQLMTALNITTQ
G.max_GM06MC33749_sm67c0	54	85	KRNHLDQLTKQLSQLAKNNGEILATIDITTQ
G.max_GM06MC01804_489153	51	82	KQQHLEGLSAQLDQLKKENTQMNTNIGISTQ
Tamarix_hispida_FJ752700	40	71	KQQHLGDLLNQVSKLQAENSQFVAKINSASQ

Table 3b. Motif 20 as present in the polypeptides of Table A4.

Polypeptide Name as in Table A4	Amino acid coordinate at Start	Amino acid coordinate at end	Sequence
B.napus_BN06MC17829_4559	81	107	VEAENSVLRAQLDELSHRLESLNDII
Nicotiana_tabacum_AY0455	85	111	VEAENSILRAQLAELNHRLESLNEII
Capsicum_annuum_AY789639	84	110	VEAENSILRAQLSELSHRLESLNEII

Capsicum_chinense_AF1277	84	110	VEAENSILRAQLSELSHRLESLNEII
Capsicum_chinense_AF4303	84	110	VEAENSILRAQLSELSHRLESLNEII
G.max_GM06MC32426_sk55g0	84	110	VEAENSVLRAQVGELSHRLESLNEIV
Glycine_max_AF532621#1	86	112	VEAENSVLRAQVGELSHRLESLNEIV
Arabidopsis_thaliana_AF0	81	107	VEAENSVLRAQLDELNHRLQSLNDII
Arabidopsis_thaliana_AT4	81	107	VEAENSVLRAQLDELNHRLQSLNDII
G.max_GM06MC17143_596542	87	113	VEAENSVLRAQVNELSHWLESLNEII
P.trichocarpa_719591#1	85	111	IEAENSILRAQMAELSHRLNSLNEII
P.trichocarpa_710131#1	83	109	VEAENSILRAQMMELNHRLDSLNEIL
P.trichocarpa_715285#1	85	111	VEAENSILRAQMMELNHRLDSLNEIL
G.max_GM06MC32046_sj86f0	78	104	VEAENSVLRAQVNELSHRLDSLNQII
Mt_bZIP	84	110	VEAENSVLRAQMGELSNRLESLNEIV
G.max_GM06MC33749_sm67c0	89	115	VEAENSILRAQMGELSQRLQSLNDIV
Solanum_lycopersicum_FJ6	84	110	VEAENSILRAQLSELSRRLESLNEII
B.napus_BN06MC15489_4421	70	96	VEAENSVLKAQLDELSHRLESLNGII
A.thaliana_AT1G75390.1#1	95	121	IEAENDILRAQVLELNHRLQSLNEIV
Arabidopsis_thaliana_BT0	89	115	IEAENDILRAQVLELNHRLQSLNEIV
Populus_trichocarpa_EF14	84	110	VEADNSILRAQVSELSHRLEFLNGII
Petroselinum_crispum_AJ2	82	108	VEAENSVLRAQMDELTQRLQSLNDIL
H.vulgare_c62949710hv270	92	118	VDAENAVLRTQAAELAARLQSLNDII
V.vinifera_GSVIVT0001455	74	100	VEADNSILRAQAMELSHRYQSLNDIL
P.trichocarpa_818112#1	84	110	VEADNSILRVQISELSNRLESLNEII
G.max_GM06MC01804_489153	86	112	VEAENAILRAQMEELSKRLNSLNEMI
S.bicolor_Sb04g002700.1#	91	117	VDAENAVLRTQAAELAARLGSLNDIL
Z.mays_ZM07MC37862_BFb03	91	117	VDAENAVLRTQAAELAARLGSLNDIL
Mt_bZIP2	84	110	VESENSVLRAQLNELNSRFESLNEII

Zea_mays_BT067356#1	91	117	VDAENAVLRTQTAELAARLGSLNDIL
V.vinifera_GSVIVT0003689	84	110	VEAENAILRAQMAELTLRLQTLNEIM
V.vinifera_GSVIVT0003633	84	110	IEAENSVLRAQFSELSNRLQYLVEII
Zea_mays_BT018074#1	90	116	VDADNAVLRTQAAELAARLGSLNDIL
Zea_mays_EU976771#1	91	117	VDADNAVLRTQAAELAARLGSLNDIL
Oryza_sativa_Japonica_Gr	91	117	VDAENAVLRTQAAELAARLASLNDIL
B.napus_BN06MC23287_4905	82	108	IEGENSVLRAQFLELNQRLNSLNEIV
Arabidopsis_thaliana_AT2	85	111	IQAENSVLTAQMEELSTRLQSLNEIV
Arabidopsis_thaliana_AF0	85	111	IQAENSVLTAQMEELSTRLQSLNEIV
T.erecta_SIN_01b-CS_Scar	80	106	MEAENLVLRAQVAELSHRLESLDQIK
S.bicolor_Sb07g015450.1#	93	119	AEAQNSVLRTQMMELESRLCALREII
Tamarix_hispida_FJ752700	75	101	VESENNVLRAQLMELTDRLNSLNSLL
Medicago_truncatula_BT05	5	31	SGTSSGSSLLQNSGSEEDLQALMDQR

Table 3c. Motif 3 as present in the polypeptides of Table A4.

Polypeptide Name as in Table A4	Amino acid coordinate at Start	Amino acid coordinate at end	Sequence
Tamarix_hispida_FJ752700	24	39	KRMISNRESARRSRM
G.max_GM06MC32046_sj86f0	27	42	KRMISNRESARRSRM
Mt_bZIP2	33	48	KRMISNRESARRSRM
V.vinifera_GSVIVT0003633	33	48	KRMISNRESARRSRM
G.max_GM06MC32426_sk55g0	33	48	KRMISNRESARRSRM
Medicago_truncatula_BT05	33	48	KRMISNRESARRSRM
Populus_trichocarpa_EF14	33	48	KRMISNRESARRSRM
P.trichocarpa_818112#1	33	48	KRMISNRESARRSRM
Nicotiana_tabacum_AY0455	34	49	KRMISNRESARRSRM

Capsicum_annuum_AY789639	33	48	KRMISNRESARRSRM
Capsicum_chinense_AF1277	33	48	KRMISNRESARRSRM
Capsicum_chinense_AF4303	33	48	KRMISNRESARRSRM
Solanum_lycopersicum_FJ6	33	48	KRMISNRESARRSRM
Glycine_max_AF532621#1	35	50	KRMISNRESARRSRM
G.max_GM06MC17143_596542	36	51	KRMISNRESARRSRM
Mt_bZIP	33	48	KRMISNRESARRSRM
V.vinifera_GSVIVT0003689	33	48	KRMLSNRESARRSRM
Arabidopsis_thaliana_AF0	30	45	KRMLSNRESARRSRM
Arabidopsis_thaliana_AT4	30	45	KRMLSNRESARRSRM
B.napus_BN06MC15489_4421	19	34	KRMLSNRESARRSRM
Arabidopsis_thaliana_AT2	34	49	KRMLSNRESARRSRM
H.vulgare_c62949710hv270	41	56	KRMLSNRESARRSRM
S.bicolor_Sb04g002700.1#	40	55	KRMLSNRESARRSRM
Zea_mays_BT067356#1	40	55	KRMLSNRESARRSRM
Z.mays_ZM07MC37862_BFb03	40	55	KRMLSNRESARRSRM
Zea_mays_BT018074#1	39	54	KRMLSNRESARRSRM
Zea_mays_EU976771#1	40	55	KRMLSNRESARRSRM
Oryza_sativa_Japonica_Gr	40	55	KRMLSNRESARRSRM
P.trichocarpa_710131#1	32	47	KRMLSNRESARRSRM
G.max_GM06MC01804_489153	35	50	KRMLSNRESARRSRM
P.trichocarpa_715285#1	34	49	KRMQSNRESARRSRM
T.erecta_SIN_01b-CS_Scar	29	44	KRMVSNRESARRSRM
G.max_GM06MC33749_sm67c0	38	53	KRKQSNRESARRSRM
A.thaliana_AT1G75390.1#1	44	59	KRKQSNRESARRSRM
Arabidopsis_thaliana_BT0	38	53	KRKQSNRESARRSRM

B.napus_BN06MC23287_4905	31	46	KRKESNRESARRSRM
Arabidopsis_thaliana_AF0	34	49	KRMLSNRESARRSRV
P.trichocarpa_719591#1	34	49	KRMLSNRESARRSRV
B.napus_BN06MC17829_4559	30	45	KRMLSNRESARRSRK
Petroselinum_crispum_AJ2	31	46	KRMQSNRESARRSRQ
S.bicolor_Sb07g015450.1#	42	57	RRMESNRESAKRSRQ
V.vinifera_GSVIVT0001455	24	39	KKKEENAIKSGIRKA

Motifs 19 to 21 may be 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 with a frequency higher than 0.2. Residues within square brackets represent alternatives.

More preferably, the bZIP-S polypeptide useful in the methods of the invention comprises 2, preferably 3 motifs selected from motifs 19 to 21.

Additionally or alternatively, the homologue of a bZIP-S 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 amino acid of any of the polypeptides of Table A4 preferably to the bZIP-S polypeptide represented by SEQ ID NO: 422, provided that the homologous protein comprises the bZIP domain and any one or more 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). Compared to overall sequence identity, the sequence identity will generally be higher when only conserved domains or motifs are considered.

Concerning SPA 15-like polypeptides, any reference hereinafter to a "protein useful in the methods of the invention" is taken to mean a SPA15-like 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 a SPA15-like polypeptide. 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 "SPA15-like nucleic acid" or "SPA15-like gene".

A "SPA15-like polypeptide" as defined herein refers to any polypeptide comprising an Armadillo-type fold domain with an InterPro accession number IPR016024 and SuperFamily accession number SSF48371, close to the C-terminal end, and a "winged helix" DNA-binding domain with a SuperFamily accession number SSF46785. SPA15-like polypeptides are found associated with plant leaf cell wall of various cell types and may play a significant role during leaf senescence phase.

Preferably, a winged helix" DNA-binding domain of a SPA15-like polypeptide has at least, in increasing order of preference, 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% or more sequence identity to the sequence located between amino acid 37 and 106 of SEQ ID NO 634.

Preferably, the Armadillo-type fold domain of a SPA15-like polypeptide has at least, in increasing order of preference, 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% or more sequence identity to the sequence located between amino acid 308 and 421 of SEQ ID NO 634.

Additionally or alternatively, the SPA15-like polypeptide useful in the methods of the invention comprises one or more sequence motifs having 1, 2, 3 or 4 mismatches that are allowed and at least, in increasing order of preference 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% or more sequence identity to any one or more of motifs 22 to 24:

The amino acids indicated herein in square brackets represent alternative amino acids for a particular position.

Motif 22: AAD[KQR]HWSDGALEADLR[RL]AD[FS][RV][AV][KR][QR]RAMEDA[LF]MAL[EK]F[VI][KR][ND][IV]HDMM[AV][SN][KR][ML][YQ][KE] (SEQ ID NO: 691)

Motif 23: RA[RC]QDVA[IV]LGS[GE]FLKLDARAR[EK]DT[EK]KID[RHN] (SEQ ID NO: 692)

Motif 24: L[SA]EA[DC]GIDY[TN]D[PA]E[EF][LV] (SEQ ID NO: 693)

Motifs 22 to 24 are typically found in any SPA15-like polypeptide of any plant origin.

In another preferred embodiment, the SPA15-like polypeptide useful in the methods of the invention comprises one or more sequence motifs having 1, 2, 3 or 4 mismatches that are allowed and at least, in increasing order of preference 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% or more sequence identity to any one or more of motifs 25 to 27:

Motif 25: EADGIDYTDPEELELLV[AT]TLIDLDAMDGK[SG]S[VA]SLLAECSSSPDVNTR[KQ]AL (SEQ ID NO: 694)

Motif 26: APSMW[TI]LGNAGMGALQRLA[EQ]DSN[PY]A[IV]A[AR]A (SEQ ID NO: 695)

Motif 27: FP[HG]EVS[TA]D[RQ]ITAI[QE][QE]AYW[SD]MA (SEQ ID NO: 696)

Motifs 25, 26 and 27 correspond to consensus sequences which represent conserved protein regions in a SPA15-like polypeptides class origin, to which Ipomoea_batatas_AF234536 and H.annuus_TC31796 belong, in other words, motifs 25, 26 and 27 correspond to consensus sequences which represent conserved protein regions in SPA15-like polypeptides having sequences that would cluster within the group of SPA-like polypeptides depicted in Figure 16.

In a most preferred embodiment, the SPA15-like polypeptide useful in the methods of the invention comprises one or more sequence motifs having 1, 2, 3 or 4 mismatches that are allowed and at least, in increasing order of preference 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% or more sequence identity to any one or more of motifs 28 to 30:

Motif 28: DGIDYTDPEELELLV[AT]TLIDLDAMDGK[KSR]S[VA]SL[LI]AECSSSPDVNTRKA LAN (SEQ ID NO: 697)

Motif 29: PSMW[TI]LGNAGMGALQRLA[QE]D[SP]N[YP]A[VI]A[RA]AA[ST]RAI[ND][EA]L [KT]KQWE[LV]EEGDSLRF (SEQ ID NO: 698)

Motif 30: [GL][SV][ST]S[PER][AT][NG][ST][TR][SDG][FR]I[TS]LEKNG[NKI][TA][LF][EG][LF] FP[GH]EVS[TSA]D[QR]I[TSY]AIE[EQ]AY[WKQ]SMASA[LF]SEA (SEQ ID NO: 699)

Motifs 28, 29 and 30 correspond to a consensus sequences which represent conserved protein regions in a SPA15-like polypeptides class origin, to which Os_SPA15-like and B.napus_TC82749 belong, in other words, motifs 28, 29 and 30 correspond to consensus sequences which represent conserved protein regions in SPA15-like polypeptides having sequences that would cluster within group A of SPA-like polypeptides depicted in Figure 16.

It is understood that Motif 22, 23, 24, 25, 26, 27, 28, 29 and 30 as referred to herein represent the consensus sequence of the motifs as present in SPA15-like polypeptides represented in Table A5, especially in SEQ ID NO: 634. However, it is to be understood that Motifs as defined herein are not limited to their respective sequence but they encompass the corresponding motifs as present in any SPA15-like polypeptide.

More preferably, the SPA15-like polypeptide useful in the methods of the invention 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 or all 9 motifs.

Motifs 22 to 30 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 with a frequency higher than 0.2. Residues within square brackets represent alternatives.

Additionally or alternatively, the homologue of a SPA15-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 amino acid represented by SEQ ID NO: 634, 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). 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 SPA15-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 of the motifs represented by SEQ ID NO: 691 to SEQ ID NO: 699 (Motifs 22 to 30).

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

Concerning O-FUT polypeptides, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 3, preferably clusters with the group of O-FUT polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group.

Furthermore, O-FUT polypeptides (at least in their native form) typically have peptide-O-fucosyltransferase activity. Tools and techniques for measuring peptide-O-fucosyltransferase activity are well known in the art.

In addition, O-FUT polypeptides, when expressed in rice according to the methods of the present invention as outlined in the Examples Section, give plants having increased yield

related traits, in particular total seed weight, fill rate, harvest index and number of filled seeds.

Concerning By-Pass (BPS) polypeptides, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 6, preferably clusters with the group of BPS polypeptides comprising the amino acid sequence represented by SEQ ID NO: 268 rather than with any other group.

Furthermore, BPS polypeptides (at least in their native form) seem to play a role in the regulation of the accumulation of a signal molecule, which circulates from roots to shoots. Tools and techniques for measuring its activity are well known in the art. Further details are provided in the Examples Section.

In addition, BPS polypeptides, when expressed in rice according to the methods of the present invention as outlined in the Examples Section, give plants having increased yield related traits, in particular harvest index, seeds fill rate and total seed yield per plant.

Concerning SIZ1 polypeptides, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 10, preferably clusters with the group of SIZ1 polypeptides comprising the amino acid sequence represented by SEQ ID NO: 354 rather than with any other group.

Furthermore, SIZ1 polypeptides (at least in their native form) typically have a SUMO E3 ligase activity. Tools and techniques for measuring ligase activity are well known in the art.

In addition, SIZ1 polypeptides, when expressed in rice according to the methods of the present invention as outlined in the Examples Section, give plants having increased yield related traits, in particular seed yield, number of filled seeds, fill rate, number of flowers per panicle, harvest index, thousand kernel weight, centre of gravity of the canopy and proportion of the thick root in the root system.

Concerning bZIP-S polypeptides, bZIP-S polypeptides, additionally, typically have DNA binding activity. Tools and techniques for measuring DNA bidning activity are well known in the art. (Izawa, T. et al. (1993), J. Mol. Biol. 230, 1131-1 144 ; Choi, H. et al. (2000) J. Biol.Chem. 275, 1723-1730). Preferably , bZIP-S polypeptides bind to a promoter sequence (in vivo and/or in vitro) comprising the ACGT core sequence. Further preferably, a bZIP-S polypeptide bind to a DNA fragment comprising any one or more of an A-box (TACGTA), a C-box (GACGTC and a G-Box (CACGTG) as represented by SEQ ID NO: 630, SEQ ID NO: 631 , SEQ ID NO: 632 respectively.

In addition, bZIP-S polypeptides, when expressed in rice according to the methods of the present invention as outlined in the Examples section, give plants having increased yield related traits, in particular increase seed yield.

Concerning SPA15-like polypeptides, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 16, preferably, clusters with the group of SPA15-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 634 rather than with any other group.

In addition, SPA15-like polypeptides, when expressed in rice according to the methods of the present invention as outlined in the Examples Section, give plants having increased yield related traits, in particular total seed weight, harvest index, number of filled seeds, fill rate and flower per panicle.

Concerning O-FUT 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. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any O-FUT -encoding nucleic acid or O-FUT polypeptide as defined herein.

Examples of nucleic acids encoding O-FUT polypeptides are given in Table A1 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table A1 of the Examples section are example sequences of orthologues and paralogues of the O-FUT 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.

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

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

(i) a nucleic acid represented by SEQ ID NO: 1;

(ii) the complement of a nucleic acid represented by SEQ ID NO: 1;

(iii) a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 21, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be derived from a polypeptide sequence as represented by any one of SEQ ID NO: 2 and further preferably confers enhanced yield-related traits relative 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%, 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 table A1 and further preferably conferring enhanced yield-related traits relative to control plants;

(v) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;

(vi) a nucleic acid encoding a O-FUT 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 SEQ ID NO: 2 and any of the other amino acid

(vii) sequences in Table A1 and preferably conferring 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 represented by any one of SEQ ID NO: 2;

(ii) 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 amino acid sequence represented by any one of SEQ ID NO: 2 or 22 and any of the other amino acid sequences in Table A1 and preferably conferring enhanced yield-related traits relative to control plants.

(iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.

Concerning By-Pass (BPS) polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 267, encoding the polypeptide sequence of SEQ ID NO: 268. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any BPS-encoding nucleic acid or BPS polypeptide as defined herein.

Examples of nucleic acids encoding BPS polypeptides are given in Table A2 of the Examples section herein. Such nucleic acids are useful in performing the methods of the

invention. The amino acid sequences given in Table A2 of the Examples section are example sequences of orthologues and paralogues of the BPS polypeptide represented by SEQ ID NO: 268, 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.

Concerning SIZ1 polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 353, encoding the polypeptide sequence of SEQ ID NO: 354. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any SIZ1-encoding nucleic acid or SIZ1 polypeptide as defined herein.

Examples of nucleic acids encoding SIZ1 polypeptides are given in Table A3 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table A3 of the Examples section are example sequences of orthologues and paralogues of the SIZ1 polypeptide represented by SEQ ID NO: 355, 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: 354 or SEQ ID NO: 355, the second BLAST (back-BLAST) would be against rice sequences.

The invention also provides hitherto unknown SIZ1-encoding nucleic acids and SIZ1 polypeptides useful for conferring enhanced yield-related traits in plants relative to control plants.

Concerning bZIP-S polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 421, encoding the polypeptide sequence of SEQ ID NO: 422. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any bZIP-S-encoding nucleic acid or bZIP-S polypeptide as defined herein.

Examples of nucleic acids encoding bZIP-S polypeptides are given in Table A4 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table A4 of the Examples section are example sequences of orthologues and paralogues of the bZIP-S polypeptide represented by SEQ ID NO: 422, 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: 421 or SEQ ID NO: 422, the second BLAST (back-BLAST) would be against Medicago truncatula sequences.

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

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

(i) a nucleic acid represented by any of the nucleic acids of Table A4;

(ii) the complement of a nucleic acid represented by any of the nucleic acids of Table A4;

(iii) a nucleic acid encoding a bZIP-S 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 of the polypeptides of Table A4 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: 501 to SEQ ID NO: 626, and further preferably conferring enhanced yield-related traits relative to control plants.

(iv) 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 represented by any of the polypeptides of Table A4;

(ii) 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 amino acid sequence represented by any of the polypeptides of Table A4, 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: 501 to SEQ ID NO: 626, and further preferably conferring enhanced yield-related traits relative to control plants;

(iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.

Concerning SPA15-like polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 633, encoding the

polypeptide sequence of SEQ ID NO: 634. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any SPA15-like-encoding nucleic acid or SPA15-like polypeptide as defined herein.

Examples of nucleic acids encoding SPA15-like polypeptides are given in Table A5 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table A5 of the Examples section are example sequences of orthologues and paralogues of the SPA15-like polypeptide represented by SEQ ID NO: 634, 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: 633 or SEQ ID NO: 634, the second BLAST (back-BLAST) would be against rice sequences.

The invention also provides hitherto unknown SPA15-like-encoding nucleic acids and SPA15-like polypeptides useful for conferring enhanced yield-related traits in plants relative to control plants.

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

(i) a nucleic acid represented by any one of SEQ ID NO: 633;

(ii) the complement of a nucleic acid represented by any one of SEQ ID NO: 633;

(iii) a nucleic acid encoding the polypeptide as represented by any one of SEQ ID NO: 634, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be derived from a polypeptide sequence as represented by any one of SEQ ID NO: 634, and further preferably confers enhanced yield-related traits relative 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%, 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 table A5 and further preferably conferring enhanced yield-related traits relative to control plants;

(v) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;

(vi) a nucleic acid encoding a SPA15-like 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: 634, and any of the other amino acid sequences in Table A5 and preferably conferring 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 represented by any one of SEQ ID NO: 634;

(ii) 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 amino acid sequence represented by any one of SEQ ID NO: 634, and any of the other amino acid sequences in Table A5 and preferably conferring enhanced yield-related traits relative to control plants.

(iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.

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 A5 of the Examples section, the terms "homologue" and "derivative" being as defined herein. Also useful in the methods 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 A5 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.

Further nucleic acid variants useful in practising the methods of the invention include portions of nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides, nucleic acids hybridising to nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides, splice variants of nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides, allelic variants of nucleic acids encoding an O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides and variants of nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides obtained by gene shuffling. The terms

hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

Nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like 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 A5 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 A5 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.
portion.

Concerning O-FUT-like polypeptides, portions useful in the methods of the invention, encode an O-FUT 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 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 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. 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 3, clusters with the group of O-FUT polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group and/or comprises motifs 1 to 3 and/or has a peptide-O-fucosyltransferase biological activity and/or has at least 50% sequence identity to SEQ ID NO: 2.

Concerning By-Pass (BPS) polypeptides, portions useful in the methods of the invention, encode a BPS 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 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: 267.

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 6, clusters with the group of BPS polypeptides comprising the amino acid sequence represented by SEQ ID NO: 268 rather than with any other group and/or comprises motifs 4 to 12 and/or has at least 40% sequence identity to SEQ ID NO: 268.

Concerning SIZ1 polypeptides, portions useful in the methods of the invention, encode a SIZ1 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 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: 353. 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 10, clusters with the group of SIZ1 polypeptides comprising the amino acid sequence represented by SEQ ID NO: 354 rather than with any other group and/or comprises motifs 13 to 18 and/or has biological activity of a SUMO E3 ligase and/or has at least 40% sequence identity to SEQ ID NO: 354.

Concerning bZIP-S polypeptides, portions useful in the methods of the invention, encode a bZIP-S polypeptide as defined herein, 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 100, 200, 300, 400, 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: 421. Preferably, the portion encodes a fragment of an amino acid sequence comprising a bZIP domain and one or more of Motifs 19 to 21 as defined herein.

Concerning SPA15-like polypeptides, portions useful in the methods of the invention, encode a SPA15-like polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A5 of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A5 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 A5 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 A5 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 A5 of the Examples section. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 633. 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 16, clusters with the group of SPA15-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 634 rather than with any other group and/or comprises one or more of the motifs 22 to 30 and/or has at least 30% sequence identity to SEQ ID NO: 634.

Concerning O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides, another nucleic acid variant useful in the methods of the invention is a nucleic acid capable of hybridising, under reduced stringency conditions, preferably under stringent conditions, with a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like 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 any one of the nucleic acids given in Table A1 to A5 of the Examples section, or comprising introducing and expressing in a plant a nucleic acid capable of hybridising to a nucleic acid encoding an orthologue, paralogue or homologue of any of the nucleic acid sequences given in Table A1 to A5 of the Examples section.

Concerning O-FUT-like polypeptides, hybridising sequences useful in the methods of the invention encode an O-FUT polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in Table A1 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 A1 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 A1 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: 1 or to a portion thereof.

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 3, clusters with the group of O-FUT polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group and/or comprises motifs 1 to 3 and/or has a peptide-O-fucosyltransferase biological activity and/or has at least 50% sequence identity to SEQ ID NO: 2.

Concerning By-Pass (BPS) polypeptides, hybridising sequences useful in the methods of the invention encode a BPS polypeptide as defined herein, having substantially the same biological activity as the amino acid 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: 267 or to a portion thereof.

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 9, clusters with the group of BPS polypeptides comprising the amino acid sequence represented by SEQ ID NO: 268 rather than with any other group and/or comprises motifs 4 to 12 and/or has at least 40% sequence identity to SEQ ID NO: 268.

Concerning SIZ1 polypeptides, hybridising sequences useful in the methods of the invention encode a SIZ1 polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in Table A3 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 A3 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 A3 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: 353 or to a portion thereof.

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 10, clusters with the group of SIZ1 polypeptides SUMO E3 ligases comprising the amino acid sequence represented by SEQ ID NO: 354 rather than with any other group and/or comprises motifs 13 to 18 and/or has biological activity of a SUMO E3 ligase and/or has at least 40% sequence identity to SEQ ID NO: 354.

Concerning bZIP-S polypeptides, hybridising sequences useful in the methods of the invention encode a bZIP-S polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in Table A4 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 A4 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 A4 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: 421 or to a portion thereof.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence comprising a bZIP domain and one or more of Motifs 19 to 21 as defined herein.

Concerning SPA15-like polypeptides, hybridising sequences useful in the methods of the invention encode a SPA15-like polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in Table A5 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 A5 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 A5 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: 633 or to a portion thereof.

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 16, clusters with the group of SPA15-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 634 rather than with any other group and/or comprises one or more of the motifs 22 to 30 and/or has at least 30% sequence identity to SEQ ID NO: 634.

Another nucleic acid variant useful in the methods of the invention is a splice variant encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or

a bZIP-S polypeptide, or a SPA15-like polypeptide as defined hereinabove, a splice variant being 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 splice variant of any one of the nucleic acid sequences given in Table A1 to A5 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 A5 of the Examples section.

Concerning O-FUT-like 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 3, clusters with the group of O-FUT polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group and/or comprises motifs 1 to 3 and/or has a peptide-O-fucosyltransferase biological activity and/or has at least 50% sequence identity to SEQ ID NO: 2.

Concerning By-Pass (BPS) polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 267, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 268. 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 6, clusters with the group of BPS polypeptides comprising the amino acid sequence represented by SEQ ID NO:268 rather than with any other group and/or comprises motifs 4 to 12 and/or has at least 40% sequence identity to SEQ ID NO: 268.

Concerning SIZ1 polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 353, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 354. 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 10, clusters with the group of SIZ1 polypeptides SUMO E3 ligases comprising the amino acid sequence represented by SEQ ID NO: 354 rather than with any other group and/or comprises motifs 13 to 18 and/or has biological activity of a SUMO E3 ligase and/or has at least 40% sequence identity to SEQ ID NO: 353.

Concerning bZIP-S polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 421, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 422. Preferably, the amino acid sequence encoded by the splice variant comprises a bZIP domain and one or more of Motifs 19 to 21 as defined herein.

Concerning SPA15-like polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 633, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 634. 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 16, clusters with the group of SPA15-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 634 rather than with any other group and/or comprises one or more of the motifs 22 to 30, and/or has at least 30% sequence identity to SEQ ID NO: 634.

Another nucleic acid variant useful in performing the methods of the invention is an allelic variant of a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, as defined hereinabove, an allelic variant being 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 an allelic variant of any one of the nucleic acids given in Table A1 to A5 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 A5 of the Examples section.

Concerning O-FUT-like polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the O-FUT 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 3, clusters with the group of O-FUT polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group and/or comprises motifs 1 to 3 and/or has a peptide-O-fucosyltransferase biological activity and/or has at least 50% sequence identity to SEQ ID NO: 2.

Concerning By-Pass (BPS) polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the BPS polypeptide of SEQ ID NO: 267 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: 266 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 267. 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 6, clusters with the group of BPS polypeptides comprising the amino acid sequence represented by SEQ ID NO: 267 rather than with any other group and/or comprises motifs 4 to 12 and/or has at least 40% sequence identity to SEQ ID NO: 267.

Concerning SIZ1 polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the SIZ1 polypeptide of SEQ ID NO: 354 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: 353 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 354. 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 10, clusters with the SIZ1 polypeptides SUMO E3 ligases comprising the amino acid sequence represented by SEQ ID NO: 354 rather than with any other group and/or comprises motifs 13 to 18 and/or has biological activity SUMO E3 ligase and/or has at least 40% sequence identity to SEQ ID NO: 354.

Concerning bZIP-S polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the bZIP-S polypeptide of SEQ ID NO: 422 and any of the amino acids 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: 421 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 422. Preferably, the amino acid sequence encoded by the allelic comprises a bZIP domain and one or more of Motifs 19 to 21 as defined herein.

Concerning SPA15-like polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the SPA15-like polypeptide of SEQ ID NO: 633 and any of the amino acids depicted in Table A5 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: 632 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 633. 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 16, clusters with the group of SPA15-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 633 rather than with any other group and/or comprises one or more of the motifs 22 to 30, and/or has at least 30% sequence identity to SEQ ID NO: 633.

Gene shuffling or directed evolution may also be used to generate variants of nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides, as defined above; the term "gene shuffling" being 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 variant of any one of the nucleic acid sequences given in Table A1 to a5 of the Examples section, or comprising 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 A5 of the Examples section, which variant nucleic acid is obtained by gene shuffling.

Concerning O-FUT-like polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling, when used in the construction of a phylogenetic tree such as the one depicted in Figure 3, preferably clusters with the group of O-FUT polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group and/or comprises motifs 1 to 3 and/or has a peptide-O-fucosyltransferase biological activity and/or has at least 50% sequence identity to SEQ ID NO: 2.

Concerning By-Pass (BPS) polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling, when used in the construction of a phylogenetic tree such as the one depicted in Figure 6, preferably clusters with the group of BPS polypeptides comprising the amino acid sequence represented by SEQ ID NO: 268 rather than with any other group and/or comprises motifs 4 to 12 and/or has at least 40% sequence identity to SEQ ID NO: 268.

Concerning SIZ1 polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling, when used in the construction of a phylogenetic tree such as the one depicted in Figure 10, preferably clusters with the group of SIZ1 polypeptides SUMO E3 ligases comprising the amino acid sequence represented by SEQ ID NO: 354 rather than with any other group and/or comprises motifs 13 to 18 and/or has biological activity SUMO E3 ligase and/or has at least 40% sequence identity to SEQ ID NO: 354.

Concerning bZIP-S polypeptides, preferably, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling comprises a bZIP domain and one or more of Motifs 19 to 21 as defined herein.

Concerning SPA15-like polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling, when used in the construction of a phylogenetic tree such as the one depicted in Figure 16, clusters with the group of SPA15-like

polypeptides comprising the amino acid sequence represented by SEQ ID NO: 634 rather than with any other group and/or comprises one or more of the motifs 22 to 30 and/or has at least 30% sequence identity to SEQ ID NO: 634.

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.).

Concerning O-FUT-like polypeptides, nucleic acids encoding O-FUT 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 O-FUT 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.

Concerning By-Pass (BPS) polypeptides, nucleic acids encoding BPS 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 BPS polypeptide-encoding nucleic acid is from a plant, further preferably from a dicotyledonous plant, more preferably from the family Brassicaceae, most preferably the nucleic acid is from Arabidopsis thaliana.

Concerning SIZ1 polypeptides, nucleic acids encoding SIZ1 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 SIZ1 polypeptide-encoding nucleic acid is from a plant, further preferably from a dicotyledonous plant, more preferably from the family Brassicaceae, most preferably the nucleic acid is from Arabidopsis thaliana.

Concerning bZIP-S polypeptides, nucleic acids encoding bZIP-S 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 bZIP-S polypeptide-encoding nucleic acid is from a plant, further preferably from a monocotyledonous plant, more preferably from the family Fabaceae, most preferably the nucleic acid is from Medicago truncatula.

Concerning SPA15-like polypeptides, nucleic acids encoding SPA15-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 SPA15-like 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.

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 yield, especially increased seed yield relative to control plants. The terms "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 early vigour and/or in biomass (weight) of one or more parts of a plant, which may include aboveground (harvestable) parts and/or (harvestable) parts 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 the seed yield of control plants.

Concerning O-FUT-like polypeptides, the present invention provides a method for increasing yield, especially seed yield of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding an O-FUT polypeptide as defined herein.

Concerning By-Pass (BPS) polypeptides, the present invention provides a method for increasing yield-related traits, especially seed yield of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding a BPS polypeptide as defined herein.

Concerning SIZ1 polypeptides, the present invention provides a method for increasing yield, especially seed yield of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding a SIZ1 polypeptide as defined herein.

Concerning bZIP-S polypeptides, the present invention provides a method for increasing yield-related traits, especially seed yield of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding a bZIP-S polypeptide as defined herein.

Concerning SPA15-like polypeptides, the present invention provides a method for increasing yield-related traits, especially seed yield of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding a SPA15-like polypeptide as defined herein.

Since the transgenic plants according to the present invention have increased yield, it is likely that these plants exhibit an increased growth rate (during at least part of their life cycle), relative to the growth rate of control plants at a corresponding stage in their life cycle.

According to a preferred feature of the present invention, performance of the methods of the 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 a fucose protein O-fucosyltransferase (O-FUT) polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like 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 relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing 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 O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide.

Performance of the methods of the invention gives plants grown under conditions of nutrient deficiency, particularly under conditions of nitrogen deficiency, increased yield relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield in plants grown under conditions of nutrient deficiency, which method comprises modulating expression in a plant of a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide.

Performance of the methods of the invention gives plants grown under conditions of salt stress, increased yield relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield in plants grown under conditions of salt stress, which method comprises modulating expression in a plant of a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide.

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZI polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides.

The gene constructs may be inserted into vectors, which may be commercially available, suitable for transforming into plants 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.

More specifically, the present invention provides a construct comprising:

(a) a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide as defined above;

(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 O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide is as defined above. The term "control sequence" and "termination sequence" are as defined herein.

Plants are transformed with a vector comprising any of the nucleic acids described above. The skilled artisan is well aware of the genetic elements that must be present on the vector 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).

Concerning O-FUT polypeptides, or By-Pass (BPS) polypeptides, 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. Preferably the constitutive promoter is a ubiquitous constitutive promoter of medium strength. See the "Definitions" section herein for definitions of the various promoter types. Also useful in the methods of the invention is a root-specific promoter.

Concerning SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides, 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. Preferably the constitutive promoter is a ubiquitous constitutive promoter of medium strength. See the "Definitions" section herein for definitions of the various promoter types.

Concerning O-FUT-like polypeptides, it should be clear that the applicability of the present invention is not restricted to the O-FUT polypeptide-encoding nucleic acid represented by SEQ ID NO: 1, nor is the applicability of the invention restricted to expression of a O-FUT polypeptide-encoding nucleic acid when driven by a constitutive promoter.

The constitutive promoter is preferably a medium strength promoter, more preferably selected from a plant derived promoter, such as a GOS2 promoter, more preferably is the promoter GOS2 promoter from rice. Further preferably the constitutive promoter is

represented by a nucleic acid sequence substantially similar to SEQ ID NO: 264, most preferably the constitutive promoter is as represented by SEQ ID NO: 264. See the "Definitions" section herein for further examples of constitutive promoters.

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: 264, and the nucleic acid encoding the O-FUT polypeptide. Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

Concerning By-Pass (BPS) polypeptides, it should be clear that the applicability of the present invention is not restricted to the BPS polypeptide-encoding nucleic acid represented by SEQ ID NO: 267, nor is the applicability of the invention restricted to expression of a BPS polypeptide-encoding nucleic acid when driven by a constitutive promoter.

The constitutive promoter is preferably a medium strength promoter, more preferably selected from a plant derived promoter, such as a GOS2 promoter, more preferably is the promoter GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 350, most preferably the constitutive promoter is as represented by SEQ ID NO: 350. See the "Definitions" section herein for further examples of constitutive promoters.

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: 350, and the nucleic acid encoding the BPS polypeptide. Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

Concerning SIZ1 polypeptides, it should be clear that the applicability of the present invention is not restricted to the SIZ1 polypeptide-encoding nucleic acid represented by SEQ ID NO: 353, nor is the applicability of the invention restricted to expression of a SIZ1 polypeptide-encoding nucleic acid when driven by a constitutive promoter.

The constitutive promoter is preferably a medium strength promoter, more preferably selected from a plant derived promoter, such as a GOS2 promoter, more preferably is the promoter GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 418, most preferably the constitutive promoter is as represented by SEQ ID NO: 418. See the "Definitions" section herein for further examples of constitutive promoters.

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: 418, and the nucleic acid encoding the SIZ1 polypeptide.

Concerning bZIP-S polypeptides, it should be clear that the applicability of the present invention is not restricted to the bZIP-S polypeptide-encoding nucleic acid represented by SEQ ID NO: 421 , nor is the applicability of the invention restricted to expression of a bZIP-S polypeptide-encoding nucleic acid when driven by a constitutive promoter.

The constitutive promoter is preferably a medium strength promoter, more preferably selected from a plant derived promoter, such as a GOS2 promoter, more preferably is the promoter GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 629, most preferably the constitutive promoter is as represented by SEQ ID NO: 629. See the "Definitions" section herein for further examples of constitutive promoters.

Preferably the bZIP-S nucleic acid used in the invention is any of the nucleic acids of Table A linked to a GOS2 promoter.

Concerning SPA15-like polypeptides, it should be clear that the applicability of the present invention is not restricted to the SPA15-like polypeptide-encoding nucleic acid represented by SEQ ID NO: 633, nor is the applicability of the invention restricted to expression of a SPA15-like polypeptide-encoding nucleic acid when driven by a constitutive promoter.

The constitutive promoter is preferably a medium strength promoter, more preferably selected from a plant derived promoter, such as a GOS2 promoter, more preferably is the promoter GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 700, most preferably the constitutive promoter is as represented by SEQ ID NO: 700. See the "Definitions" section herein for further examples of constitutive promoters.

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 rice promoter, substantially similar to SEQ ID NO: 700, and the nucleic acid encoding the SPA15-like polypeptide. Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

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 encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or

a bZIP-S polypeptide, or a SPA15-like polypeptide is by introducing and expressing in a plant a nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, however the effects of performing the method, i.e. enhancing yield-related traits may also be achieved using other 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 enhanced yield-related traits relative to control plants, comprising introduction and expression in a plant of any nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, as defined hereinabove.

More specifically, the present invention 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 an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like 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 O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, as defined herein.

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 by transformation. The term "transformation" is described in more detail in the "definitions" section herein.

The present invention clearly 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 parts thereof comprise a nucleic acid transgene encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, as defined above. 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.

The invention also includes host cells containing an isolated nucleic acid encoding an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide, as defined hereinabove. Preferred host cells according to the invention are plant cells. Host plants for the nucleic acids or the vector used in the method according to the invention, the expression cassette or construct or vector are, in principle, advantageously all plants, which are capable of synthesizing the polypeptides used in the inventive method.

The methods of the invention are advantageously applicable to any plant. 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 a preferred embodiment of the present invention, the plant is a crop plant. Examples of crop plants include soybean, sugarbeet, sunflower, canola, alfalfa, rapeseed, linseed, cotton, tomato, potato and tobacco. Further preferably, the plant is a monocotyledonous plant. Examples of monocotyledonous plants include sugarcane. More preferably the plant is a cereal. Examples of cereals include rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, secale, einkorn, teff, milo and oats.

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 O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide. The invention furthermore relates to products derived, preferably directly derived, from a harvestable part of such a plant, such as dry pellets or powders, oil, fat and fatty acids, starch or proteins.

The present invention also encompasses use of nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides as described herein and use of these O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides in enhancing any of the aforementioned yield-related traits in plants. For example, nucleic acids encoding O-FUT polypeptide, or By-Pass (BPS) polypeptide, or SIZ1 polypeptide, or bZIP-S polypeptide, or SPA15-like polypeptide described herein, or the O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like polypeptides themselves, may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide -encoding gene. The nucleic acids/genes, or the

O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like 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 hereinabove in the methods of the invention. Furthermore, allelic variants of an O-FUT polypeptide, or a By-Pass (BPS) polypeptide, or a SIZ1 polypeptide, or a bZIP-S polypeptide, or a SPA15-like polypeptide - encoding nucleic acid/gene may find use in marker-assisted breeding programmes. Nucleic acids encoding O-FUT polypeptides, or By-Pass (BPS) polypeptides, or SIZ1 polypeptides, or bZIP-S polypeptides, or SPA15-like 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 phenotypes.

Items .

1. O-FUT polypeptides

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 O-FUT polypeptide, wherein said O-FUT polypeptide comprises a domain with a PFam accession number PF1 0250.

2. Method, according to item 1, wherein said O-FUT polypeptide comprises one or more of the following motifs:
   (i) Motif 1: HYIALHLRYEKDM (SEQ ID NO: 261),
   (ii) Motif 2: IYIVAGEIYGGHSMD (SEQ ID NO: 262),
   (iii) Motif 3: ALDYNVAVQSDVFVYTYDGNMAKAVQGH (SEQ ID NO: 263)

3. Method, according to item 1 or 2, wherein said O-FUT polypeptide may comprise a conserved Arginine residue in Motif 1.

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

5. Method according to any one of items 1 to 4, wherein said nucleic acid encoding an O-FUT polypeptide encodes any one of the proteins listed in Table A or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

6. Method according to any one of items 1 to 5, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A 1.

7. Method according to any preceding item, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

8. Method according to any one of items 1 to 7, wherein said enhanced yield-related traits are obtained under non-stress conditions.

9. Method according to any one of items 1 to 7, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

10. Method according to any one of items 1 to 9, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

11. Method according to any one of items 1 to 10, wherein said nucleic acid encoding a O-FUT polypeptide is of any origin, preferably of plant origin, more preferably from a monocotyledonous plant, further preferably from the family Poaceae, particularly preferably from the genus Oryza, most preferably from Oryza sativa.

12. Plant or part thereof, including seeds, obtainable by a method according to any one of items 1 to 11, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a O-FUT polypeptide.

13. Construct comprising:
    (i) nucleic acid encoding a O-FUT polypeptide as defined in any of the items 1 to 3;
    (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
    (iii) a transcription termination sequence.

14. Construct according to item 13, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

15. Use of a construct according to item 13 or 14 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

16. An isolated nucleic acid molecule selected from:
    (i) a nucleic acid represented by SEQ ID NO: 1;
    (ii) the complement of a nucleic acid represented by SEQ ID NO: 1;
    (iii) a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 2, preferably as a result of the degeneracy of the genetic code, said isolated nucleic

acid can be derived from a polypeptide sequence as represented by SEQ ID NO: 2 and further preferably confers enhanced yield-related traits relative 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%, 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 table A1 and further preferably conferring enhanced yield-related traits relative to control plants;

(v) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;

(vi) a nucleic acid encoding a O-FUT 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 SEQ ID NO: 2 and any of the other amino acid sequences in Table A1 and preferably conferring enhanced yield-related traits relative to control plants.

17. An isolated polypeptide selected from:

(i) an amino acid sequence represented by SEQ ID NO: 2;

(ii) 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 amino acid sequence represented by any one of SEQ ID NO: 2 or 22 and any of the other amino acid sequences in Table A1 and preferably conferring enhanced yield-related traits relative to control plants.

(iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.

18. Plant, plant part or plant cell transformed with a construct according to item 13 or 14.

19. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:

(i) introducing and expressing in a plant a nucleic acid encoding an O-FUT polypeptide as defined in any of the items 1 to 3; and

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

20. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding an O-FUT polypeptide as defined in any of the items 1 to 3, or a transgenic plant cell derived from said transgenic plant.

21. Transgenic plant according to item 12, 18 or 20, or a transgenic plant cell derived thereof, wherein said plant is a crop plant such as sugarbeet, 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.

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

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

24. Use of a nucleic acid encoding a O-FUT polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

2. By-Pass (BPS) polypeptides

25. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a BPS polypeptide.

26. A method, according with item 25, wherein said BPS polypeptide further comprises at least one of the following motifs:
(i) Motif 4: SWM[KT][LQ]A[MI]ESLC[EA][TI]H[TN]DIKTLIT[DE]LELP (SEQ ID NO: 341)
(ii) Motif 5: D[IL]C[IN]AFSSE[LI][ST]RLNQGHL[LY]L[QK]C[AV]LHNL[DE][SG]SS (SEQ ID NO: 342)
(iii) Motif 6: GKVLM[RQ]A[ML]YGV[KR]V[VQ]TV[FY][IV]CS[VI]FA[AV]AFSGS (SEQ ID NO: 343)

27. Method according to any of the items 25 or 26, wherein said BPS polypeptide further comprises at least one or more of the following motifs:
(i) Motif 7: SWM[KT][LQ]A[MI]ESLC[EA][TI]H[NT]D[IV]KTLIT[DE]LELPVSDW[DE][ED]KW[IV]DVYLD[IN]SVKL (SEQ ID NO: 344)
(ii) Motif 8: SL[ND]LPK[VI]KNSAKGKVLM[RQ]A[ML]YGV[KR]V[QV]TV[FY][IV]CSVFA[AV]AFSGS (SEQ ID NO:345)

(iii) Motif 9: PQ[ED]P[HP]R[PS]F[FL]PFGNPF (SEQ ID NO: 346)

28. Method according to any of the items 25 to 27, wherein said BPS polypeptide further comprises one or more of the following motifs:

(i) Motif 10: [VM]PK[EDN]K[SDN][DQ]ILT[LV]SWM[KS][QL]AM[EA]SLC[EQ]TH[KN][NAS]I[KNR]TL[IV]TDL[EQ]LPVSD[WL]E[ED][KN][WF][VI][DY][IV]Y (SEQ ID NO: 347)

(ii) Motif 11: LPK[VK]KNSAKGKVL[ML]RA[LF]YGVKV[KQ]T[LV]YI[CS][SG]VF[AT]AA[FW]S[GD]S[ST][NQK][ND]L[FL][YD][LV][TP][VI][SP][NE][EK] (SEQ ID NO: 348)

(iii) Motif 12: [PL]WA[KQP][SVA]F[MT][DE][MLV]Q[NS][TV][VM]N[AGPS]EI[KR][ND][IM][FL][LS]S[DG][GR][LFS]T[VI][LIM]K[ED]LE[AS]V[DE][AS][GS]V[KE][KQ]L[YA][PT][AM][IV]Q[DQE]G[SV] (SEQ ID NO: 349)

29. Method according to any of the items 25 to 28, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a BPS polypeptide.

30. Method according to any one of items 25 to 29, wherein said nucleic acid encoding a BPS polypeptide encodes any one of the proteins listed in Table A2 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

31. Method according to any one of items 25 to 30, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A2.

32. Method according to any preceding items, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

33. Method according to any one of items 25 to 32, wherein said enhanced yield-related traits are obtained under non-stress conditions.

34. Method according to any one of items 25 to 32, wherein said enhanced yield-related traits are obtained under conditions of a type of stress affecting the plant fertility.

35. Method according to any one of items 25 to 34, wherein said nucleic acid is operably linked to a promoter active in roots.

36. Method according to any one of items 25 to 34, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

37. Method according to any one of items 25 to 36, wherein said nucleic acid encoding a BPS polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Brassicaceae, more preferably from the genus Arabidopsis, most preferably from Arabidopsis thaliana.

38. Plant or part thereof, including seeds, obtainable by a method according to any one of items 25 to 37, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a BPS polypeptide.

39. Construct comprising:
(i) nucleic acid encoding a BPS polypeptide as defined in any of the items 25 to 27;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

40. Construct according to item 39, wherein one of said control sequences is a promoter active in roots.

41. Construct according to item 39, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

42. Use of a construct according to any of the items 39 to 41 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

43. Plant, plant part or plant cell transformed with a construct according to any of the items 39 to 41.

44. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding a BPS polypeptide as defined in any of the items 25 to 28; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

45. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding a BPS polypeptide as defined in any of the items 25 to 28, or a transgenic plant cell derived from said transgenic plant.

46. Transgenic plant according to item 38, 43 or 45, or a transgenic plant cell derived thereof, wherein said plant is a crop plant such as sugarbeet, 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.

47. Harvestable parts of a plant according to item 46, wherein said harvestable parts are preferably shoot biomass and/or seeds.

48. Products derived from a plant according to item 46 and/or from harvestable parts of a plant according to item 47.

49. Use of a nucleic acid encoding a BPS polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

3. SIZ1 polypeptides

50. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a SIZ1 polypeptide, wherein said SIZ1 polypeptide comprises a DUF206 domain.

51. Method according to item 50, wherein said SIZ1 polypeptide comprises one or more of the following motifs:
    (i) Motif 13: MSCNGCRXLRKGCX (SEQ ID NO: 409),
    (ii) Motif 14: QXXATXFXAKFXGR (SEQ ID NO: 410),
    (iii) Motif 15: FXSLLXEAXG (SEQ ID NO: 411)

52. Method according to item 50 or 51, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a SIZ1 polypeptide.

53. Method according to any one of items 50 to 52, wherein said nucleic acid encoding a SIZ1 polypeptide encodes any one of the proteins 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.

54. Method according to any one of items 50 to 53, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A3.

55. Method according to any preceding item, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

56. Method according to any one of items 50 to 55, wherein said enhanced yield-related traits are obtained under non-stress conditions.

57. Method according to any one of items 50 to 55, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

58. Method according to any one of items 52 to 57, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

59. Method according to any one of items 50 to 58, wherein said nucleic acid encoding a SIZ1 polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Brassicaceae, more preferably from the genus Arabidopsis, most preferably from Arabidopsis thaliana.

60. Plant or part thereof, including seeds, obtainable by a method according to any one of items 50 to 59, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a SIZ1 polypeptide.

61. Construct comprising:
(i) nucleic acid encoding a SIZ1 polypeptide as defined in items 50 or 51;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

62. Construct according to item 61, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

63. Use of a construct according to item 61 or 62 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

64. Plant, plant part or plant cell transformed with a construct according to item 61 or 62.

65. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding a SIZ1 polypeptide as defined in item 50 or 51; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

66. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding a SIZ1 polypeptide as defined in item 50 or 51, or a transgenic plant cell derived from said transgenic plant.

67. Transgenic plant according to item 60, 64 or 66, or a transgenic plant cell derived thereof, wherein said plant is a crop plant such as sugarbeet, 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.

68. Harvestable parts of a plant according to item 67, wherein said harvestable parts are preferably shoot biomass and/or seeds.

69. Products derived from a plant according to item 67 and/or from harvestable parts of a plant according to item 68.

70. Use of a nucleic acid encoding a SIZ1 polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

4. bZIP-S polypeptides

71. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a bZIP-S polypeptide.

72. Method according to item 71, wherein said bZIP-S polypeptide comprises one or more of the following motifs:
    (i) Motif 19 as represented by SEQ ID NO: 522;
    (ii) Motif 20 as represented by SEQ ID NO: 587;
    (iii) Motif 21 as represented by SEQ ID NO: 600.

73. Method according to item 71 or 72, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a bZIP-S polypeptide.

74. Method according to any one of items 71 to 73, wherein said nucleic acid encoding a bZIP-S polypeptide encodes any one of the proteins 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.

75. Method according to any one of items 71 to 74, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A4.

76. Method according to any preceding item, wherein said enhanced yield-related traits comprise increased seed yield relative to control plants.

77. Method according to any one of items 71 to 76, wherein said enhanced yield-related traits are obtained under non-stress conditions.

78. Method according to any one of items 71 to 76, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

79. Method according to any one of items 73 to 78, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

80. Method according to any one of items 71 to 79, wherein said nucleic acid encoding a bZIP-S polypeptide 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.

81. Plant or part thereof, including seeds, obtainable by a method according to any one of items 71 to 80, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a bZIP-S polypeptide.

82. Construct comprising:
(i) nucleic acid encoding a bZIP-S polypeptide as defined in items 71 or 72;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

83. Construct according to item 82, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

84. Use of a construct according to item 82 or 83 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

85. Plant, plant part or plant cell transformed with a construct according to item 82 or 83.

86. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding a bZIP-S polypeptide as defined in item 71 or 72; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

87. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of

a nucleic acid encoding a bZIP-S polypeptide as defined in item 71 or 72, or a transgenic plant cell derived from said transgenic plant.

88. Transgenic plant according to item 81, 85 or 87, or a transgenic plant cell derived thereof, wherein said plant is a crop plant, such as beet or sugarbeet, 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.

89. Harvestable parts of a plant according to item 88, wherein said harvestable parts are preferably shoot biomass and/or seeds.

90. Products derived from a plant according to item 88 and/or from harvestable parts of a plant according to item 89.

91. Use of a nucleic acid encoding a bZIP-S polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

5. SPA15-like polypeptides

92. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a SPA15-like polypeptide, wherein said SPA15-like polypeptide comprises an Armadillo-type fold domain with an InterPro accession number IPR016024 and SuperFamily accession number SSF48371 and a "winged helix" DNA-binding domain with a SuperFamily accession number SSF46785.

93. Method according to item 92, wherein said SPA15-like polypeptide comprises one or more of the following motifs:

(i) Motif 22: AAD[KR]HWSDGALEADLR[RL]ADF[RV][AV][KR][QR]RAMEDA[LF]MAL [EK]F[VI]K[ND][IV]HDMMV[SN][KR][ML][YQ][KE] (SEQ ID NO: 691);

(ii) Motif 23: RA[RC]QDVA[IV]LGS[GE]FLKLDARAR[EK]DTEKID[RHN] (SEQ ID NO: 692);

(iii) Motif 24: L[SA]EA[DC]GIDY[TN]D[PA]E[EF][LV] (SEQ ID NO: 693).

94. Method according to any of the previous items, wherein said SPA15-like polypeptide comprises one or more of the following motifs:

(i) Motif 25: EADGIDYTDPEELELLV[AT]TLIDLDAMDGK[SG]S[VA]SLLAECSSSPD VNTR[KQ]AL (SEQ ID NO: 694);

(ii) Motif 26: APSMW[TI]LGNAGMGALQRLA[EQ]DSN[PY]A[IV]A[AR]A (SEQ ID NO: 695);

(iii) Motif 27: FPGEVS[TA]D[RQ]ITAI[QE]EAYW[SD]MA (SEQ ID NO: 696).

95. Method according to any of the previous items, wherein said SPA15-like polypeptide comprises one or more of the following motifs:
    (i) Motif 28: DGIDYTDPEELELLV[AT]TLIDLDAMDGK[KSR]S[VA]SL[LI]AECSSSPD VNTRKALAN (SEQ ID NO: 697);
    (ii) Motif 29: PSMW[TI]LGNAGMGALQRLA[QE]D[SP]N[YP]A[VI]A[RA]AA[ST]RAI [ND][EA]L[KT]KQWE[LV]EEGDSLRF (SEQ ID NO: 698);
    (iii) Motif 30: [GL][SV][ST]S[PER][AT][NG][ST][TR][SDG][FR]I[TS]LEKNG[NKI][TA] [LF][EG][LF]FP[GH]EVS[TSA]D[QR]I[TSY]AIE[EQ]AY[WKQ]SMASA[LF]SEA (SEQ ID NO: 699).

96. Method according to any of the previous items, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a SPA15-like polypeptide.

97. Method according to any of the previous items, wherein said nucleic acid encoding a SPA15-like polypeptide encodes any one of the proteins listed in Table A5 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

98. Method according to any of the previous items, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A5.

99. Method according to any of the previous items, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

100. Method according to item 99, wherein said enhanced yield-related traits are obtained under non-stress conditions.

101. Method according to item 99, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

102. Method according to any one of items 92 to 98, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

103. Method according to item 102, wherein said nucleic acid encoding a SPA15-like 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.

104. Plant or part thereof, including seeds, obtainable by a method according to any one of items 92 to 101, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a SPA15-like polypeptide.

105. Construct comprising:
(i) nucleic acid encoding a SPA15-like polypeptide as defined in any of the items 92 to 95;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

106. Construct according to item 105, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

107. Use of a construct according to item 105 or 106 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

108. Plant, plant part or plant cell transformed with a construct according to item 105 or 106.

109. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding a SPA15-like polypeptide as defined in any of the items 92 to 95; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

110. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding a SPA15-like polypeptide as defined in any of the items 92 to 95, or a transgenic plant cell derived from said transgenic plant.

111. Transgenic plant according to any of the items 104, 108 or 110, or a transgenic plant cell derived thereof, wherein said plant is a crop plant, such as beet or sugarbeet, 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.

112. Harvestable parts of a plant according to item 111, wherein said harvestable parts are preferably shoot biomass and/or seeds.

113. Products derived from a plant according to item 111 and/or from harvestable parts of a plant according to item 112.

114. Use of a nucleic acid encoding a SPA15-like polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

115. An isolated nucleic acid molecule selected from:

(i) a nucleic acid represented by SEQ ID NO: 633;

(ii) the complement of a nucleic acid represented by SEQ ID NO: 633;

(iii) a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 634, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be derived from a polypeptide sequence as represented by SEQ ID NO: 634, and further preferably confers enhanced yield-related traits relative 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%, 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 table A5 and further preferably conferring enhanced yield-related traits relative to control plants;

(v) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;

(vi) a nucleic acid encoding a SPA15-like 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 SEQ ID NO: 634, and any of the other amino acid sequences in Table A5 and preferably conferring enhanced yield-related traits relative to control plants.

116. An isolated polypeptide selected from:

(i) an amino acid sequence represented by SEQ ID NO: 634;

(ii) 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 amino acid

sequence represented by SEQ ID NO: 634, and any of the other amino acid sequences in Table A5 and preferably conferring enhanced yield-related traits relative to control plants.

(iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.

Description of figures

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

Figure 1 represents an O-FUT polypeptide as represented by SEQ ID NO: 22 (full length), which comprises the following features: a Subcellular Targteing Sequence (STS), a TMHMM predicted transmembrane (TM) domain, a GDP-fucose protein O-fucosyltransferase with InterPro accession number IPR019378. The bold vertical illustrates the truncation site, the STS and TM being deleted in SEQ ID NO: 2.

Figure 2 represents a multiple alignment of various O-FUT polypeptides. The InterPro IPR019378 domain is marked with XXX. These alignments can be used for defining further motifs, such as motifs 1 to 3 (boxed), when using conserved amino acids.

Figure 3 shows phylogenetic tree of O-FUT polypeptides according to the method of Yves Van De Peer et al. (2009) - Plaza, a resource for plant comparative genomics (www.vib.gent.be).

Figure 4 represents the binary vector used for increased expression in Oryza sativa of an O-FUT -encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2)

Figure 5 represents the gene structure of BPS.

Figure 6 shows phylogenetic tree of selected BPS polypeptides, where the several clusters are identified: Trees, Fabales, Other Dicots, Solanales, Coniferales, Poales and Brassicales to which BPS belongs.

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

Figure 8 represents the represents the overall structure of the SIZ1 polypeptides.

Figure 9 shows a multiple sequence alignment of SIZ1 polypeptides.

Figure 10 shows a phylogenetic tree of SIZ1 polypeptides. Class I includes organisms of any origin; Class II includes organisms such as H. vulgare TA46195 4513 f, O. sativa 0s05g0125000; Class III includes organisms such as A. thaliana AT5G60410.5 f and arabidopsisECsequence; Class IV includes organisms such as C. vulgaris 83729 f and AT5G41580 NP 198973.SEMB3001 .

Figure 11 represents the binary vector used for increased expression in Oryza sativa of a SIZ1 -encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

Figure 12 represents a multiple alignment of various bZIP-S polypeptides. The region indicated with interrupted line of squared boxes corresponds to the bZIP domain. Boxed regions flanking the bzip domain comprised conserved sequences in polypeptides of the bZIP-S group. The name of SEQID NO 422 boxed. These alignments can be used for defining further motifs, when using conserved amino acids.

Figure 13 represents the binary vector used for increased expression in Oryza sativa of a bZIP-S-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

Figure 14 represents the domain structure of SEQ ID NO: 634 with conserved domains underlined: the Armadillo-type fold domain is double-underlined and the "winged helix" DNA-binding domain is once underlined.
Figure 15 represents a multiple alignment of various SPA15-like polypeptides. These alignments can be used for defining further motifs, when using conserved amino acids. The conserved domains like the Armadillo-type fold domain, the "winged helix" DNA-binding domain and the conserved domain described in YAP et al. (2003) are indicated.
Figure 16 shows phylogenetic tree of SPA15-like polypeptides.
Figure 17 represents the binary vector used for increased expression in Oryza sativa of a SPA15-like-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

Examples
The present invention will now be described with reference to the following examples, which are by way of illustration alone. The following examples are not intended to completely define or otherwise limit the scope of the invention.

DNA manipulation: unless otherwise stated, recombinant DNA techniques are performed 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 Ltd (UK) and Blackwell Scientific Publications (UK).

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

1.1 O-FUT polypeptides
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 sequences related to SEQ ID NO: 1 and SEQ ID NO: 2.

Table A1: Examples of O-FUT nucleic acids and polypeptides:

Name	Nucleic acid SEQ ID NO:	Polypeptide SEQ ID NO:
Oryza sativa	1	2
OS01G63230	3	4
AT1G04910	5	6
AT1G11990	7	8
AT1G14020	9	10
AT1G14970	11	12
AT1G20550	13	14
AT1G22460	15	16
AT2G01480	17	18
AT2G03280	19	20
AT2G37980	21	22
AT2G44500	23	24
AT3G02250	25	26
AT3G03810	27	28
AT3G07900	29	30
AT3G26370	31	32
AT3G30300	33	34
AT3G54100	35	36
AT4G16650	37	38
AT4G24530	39	40
AT4G38390	41	42
AT5G01100	43	44
AT5G15740	45	46
AT5G35570	47	48
AT5G63390	49	50
AT5G64600	51	52
AT5G65470	53	54
CP00001G01680	55	56
CP00036G00140	57	58
CP00036G01730	59	60

CP00051G01430	61	62
CP00055G00270	63	64
CP00056G00770	65	66
CP00065G01500	67	68
CP00152G00510	69	70
CP00280G00020	71	72
CP00289G00010	73	74
CP00289G00050	75	76
CP00382G00010	77	78
CP00458G00010	79	80
CP32528G00010	81	82
CP33915G00010	83	84
OS01G07410	85	86
OS01G62390	87	88
OS02G04590	89	90
OS02G06400	91	92
OS02G49460	93	94
OS03G07310	95	96
OS03G21090	97	98
OS08G42550	99	100
OS09G24570	101	102
OS09G27080	103	104
OS09G29940	105	106
OS09G37590	107	108
OS11G07510	109	110
OS11G29120	111	112
OS12G07540	113	114
OS12G08820	115	116
OS12G23760	117	118
PP00008G00940	119	120
PP00010G00860	121	122
PP00011G00510	123	124
PP00029G00760	125	126
PP00044G01740	127	128
PP00046G01420	129	130
PP00047G01610	131	132
PP00066G00650	133	134
PP00075G00140	135	136
PP00077G00040	137	138
PP00077G00400	139	140

PP00114G00970	141	142
PP00131G00330	143	144
PP00173G00050	145	146
PP00204G00120	147	148
PP00215G00250	149	150
PP00284G00340	151	152
PP00316G00300	153	154
PP00376G00300	155	156
PP00452G00050	157	158
PT00G00080	159	160
PT00G02000	161	162
PT04G03650	163	164
PT04G14110	165	166
PT05G07220	167	168
PT05G08390	169	170
PT05G15970	171	172
PT06G07850	173	174
PT07G11820	175	176
PT08G08310	177	178
PT08G12170	179	180
PT14G09920	181	182
PT15G02530	183	184
PT15G06760	185	186
PT16G09540	187	188
PT19G03390	189	190
PT19G06190	191	192
SB00G01560	193	194
SB01G036540	195	196
SB01G045900	197	198
SB02G024540	199	200
SB02G025960	201	202
SB02G027470	203	204
SB02G031900	205	206
SB03G004640	207	208
SB03G039450	209	210
SB03G040020	211	212
SB04G004090	213	214
SB04G029280	215	216
SB10G007565	217	218
SB10G008680	219	220

SB10G010460	221	222
SB10G027900	223	224
VV00G04590	225	226
VV00G09305	227	228
VV00G09320	229	230
VV01G07570	231	232
VV03G02980	233	234
VV04G15360	235	236
VV05G12490	237	238
VV08G10900	239	240
VV09G06300	241	242
VV09G10190	243	244
VV10G01620	245	246
VV11G01110	247	248
VV11G01120	249	250
VV13G05630	251	252
VV13G13260	253	254
VV14G09190	255	256
VV14G09220	257	258
VV17G01580	259	260

Sequences have been tentatively assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR; beginning with TA). 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, such as by the Joint Genome Institute. Furthermore, access to proprietary databases, has allowed the identification of novel nucleic acid and polypeptide sequences.

1.2 By-Pass (BPS) polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 267 and SEQ ID NO: 268 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: 267 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 sequences related to SEQ ID NO: 267 and SEQ ID NO: 268.

Table A2: Examples of BPS nucleic acids and polypeptides:

Name	Nucleic acid SEQ ID NO:	Polypeptide SEQ ID NO:
A.thaliana_AT1G01 550.1	267	268
B.napus_TC75033	269	270
P_sitchensis_EF677456	271	272
C.endivia_TA154_1 14280	273	274
A.thaliana_AT2G46080.1	275	276
A.thaliana_AT3G61 500.1	277	278
A.thaliana_AT4G01 360.1	279	280
B.napus_TC69302	281	282
B.napus_TC72556	283	284
B.napus_TC76712	285	286
B.oleracea_TA51 03_371 2	287	288
G.max_Glyma07g07240.1	289	290
G.max_Glyma09g39170.1	291	292
G.max_Glyma1 8g471 60.1	293	294
M_truncatula_BT052402	295	296
T.pratense_TA1487_57577	297	298
J.hindsii_x_regia_TA339_432290	299	300
P.trichocarpa_scaff_l 1.1515	301	302
P.trichocarpa_scaff_XIV.343	303	304
G.hirsutum_TC1 14048	305	306
Aquilegia_sp_TC23399	307	308
M.domestica_TC8317	309	310
C.sinensis_TC8893	311	312
P.trifoliata_TA81 07_37690	313	314
C.annuum_TC9590	315	316

I.nil_TC49	317	318
I.nil_TC6085	319	320
N.benthamiana_NP1 3050546	321	322
N.benthamiana_TC1 1467	323	324
N.tabacum_TC16925	325	326
S.lycopersicum_TC1 92275	327	328
S.lycopersicum_TC1 95035	329	330
S.tuberosum_TC1 73984	331	332
V.vinifera_GSVIVT0002691 8001	333	334
O.sativa_LOC_Os10g36950.1	335	336
S.bicolor_Sb01g01 7226.1	337	338
Zea_mays_EU960931	339	340

Sequences have been tentatively assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR; beginning with TA). 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, such as by the Joint Genome Institute. Furthermore, access to proprietary databases, has allowed the identification of novel nucleic acid and polypeptide sequences.

1.3 SIZ1 polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 353 and SEQ ID NO: 354 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: 353 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 A3 provides a list of nucleic acid sequences related to SEQ ID NO: 353 and SEQ ID NO: 354.

Table A3: Examples of SIZ1 nucleic acids and polypeptides:

Name	Nucleic acid SEQ ID NO:	Polypeptide SEQ ID NO:
Arabidopsis EC sequence	353	354
sumo ligase, [Ricinus communis]	355	356
hypothetical protein [Vitis vinifera]	357	358
Zea mays ZM_BFb0169I21	359	360
hypothetical protein LOC100272532 [Zea mays]	361	362
A.thaliana_AT5G60410 SIZ1	363	364
A thaliana NP_001032109 ATSIZ1/SIZ1 "AT5G60410"AT5G60420	365	366
AT5G41580" NP_198973.3 EMB3001like (EMBRYO DEFECTIVE 3001)	367	368
AT1G08910.1 EMB3001protein	369	370
H.vulgare_TA46195_4513#1	371	372
M.truncatula_AC150891_19.5#1	373	374
M.truncatula_AC152176_5.4#1	375	376
O.sativa_Os05g0125000#1	377	378
O.sativa_Os03g0719100#1	379	380
P.patens_159214#1	381	382
P.patens_165698#1	383	384
P.patens_159935#1	385	386
3	387	388
P.trichocarpa_scaff_66.246#1	389	390
P.trichocarpa_scaff_IX.1493#1	391	392
P.trichocarpa_scaff_X.2133#1	3923	394
C.vulgaris_83729#1	395	396
S.bicolor_5288383#1	397	398
S.bicolor_5285414#1	399	400
sorghum bicolor MIZ XP_002439205.1	401	402
GRMZM2G007288_aa	403	404
AK244805Soybean translation	405	406
Z.mays_ZM07MSbpsHQ_62031266.f01@48407#1	407	408

Sequences have been tentatively assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR; beginning with TA). 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, such as by the Joint Genome Institute. Furthermore, access to proprietary databases, has allowed the identification of novel nucleic acid and polypeptide sequences.

1.4 bZIP-S polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 421 and SEQ ID NO: 422 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: 421 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 sequences related to SEQ ID NO: 421 and SEQ ID NO: 422.

Table A4: Examples of bZIP-S nucleic acids and polypeptides:

Name of bZIP-S	Nucleic acid SEQ ID NO	Polypeptide SEQ ID NO
Mt_bZIP	421	422
A.thaliana_AT1G75390.1#1	423	424
Arabidopsis_thaliana_AF053939#1	425	426
Arabidopsis_thaliana_AT2G1 8162. 1#1	427	428
Arabidopsis_thaliana_AT4G34590.1#1	429	430
B.napus_BN06MC1 5489_4421 5029@1 5438#1	431	432
B.napus_BN06MC1 7829_45597483@1 7769#1	433	434
B.napus_BN06MC23287_49055078@23201#1	435	436
Capsicum_annuum_AY789639#1	437	438
Capsicum_chinense_AF1 27797#1	439	440
Capsicum_chinense_AF430372#1	441	442

G.max_GM06MC01804_48915393@1791#1	443	444
G.max_GM06MC1 7143_59654278@1 6848#1	445	446
G.max_GM06MC32046_sj86f07@3131 1#1	447	448
G.max_GM06MC32426_sk55g01@31681#1	449	450
G.max_GM06MC33749_sm67c05@32966#1	451	452
Glycine_max_AF532621 # 1	453	454
H.vulgare_c62949710hv270303@7333#1	455	456
Medicago_truncatula_BT053497#1	457	458
Mt_bZIP2	459	460
Nicotiana_tabacum_AY045570#1	461	462
Oryza_sativa_Japonica_Group_AK070887#1	463	464
P.trichocarpa_710131#1	465	466
P.trichocarpa_715285#1	467	468
P.trichocarpa_719591#1	469	470
P.trichocarpa_8181 12#1	471	472
Petroselinum_crispum_AJ292745#1	473	474
Populus_trichocarpa_EF1 4731 5#1	475	476
S.bicolor_Sb04g002700.1#1	477	478
S.bicolor_Sb07g015450.1#1	479	480
Solanum_lycopersicum_FJ6471 90#1	481	482
T.erecta_SIN_01 b-CS_Scarletade-12-L23.b1@917#1	483	484
Tamarix_hispida_FJ752700#1	485	486
V.vinifera_GSVIVT00014558001#1	487	488
V.vinifera_GSVIVT00036338001#1	489	490
V.vinifera_GSVIVT00036899001#1	491	492
Z.mays_ZM07MC37862_BFb0368K21 @37736#1	493	494
Zea_mays_BT018074#1	495	496
Zea_mays_BT067356#1	497	498
Zea_mays_EU976771#1	499	500

Sequences have been tentatively assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR; beginning with TA). 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, such as by the Joint Genome Institute. Furthermore, access to proprietary databases, has allowed the identification of novel nucleic acid and polypeptide sequences.

1.5 SPA15-like polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 633 and SEQ ID NO: 634 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: 633 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 A5 provides a list of nucleic acid sequences related to SEQ ID NO: 633 and SEQ ID NO: 634.

Table A5: Examples of SPA15-like nucleic acids and polypeptides:

Name	Nucleic acid SEQ ID NO:	Polypeptide SEQ ID NO:
Os_SPA15 like	633	634
A.thaliana_AT1G66330.1#1	635	636
Arabidopsis_thaliana_AY086709#1	637	638
B.napus_TC82749#1	639	640
LOC_Os05g05600.1#1	641	642
LOC_Os05g05600.5#1	643	644
LOC_Os05g05600.6#1	645	646
S.bicolor_Sb05g026090.1#1	647	648
Zea_mays_EU956861#1	649	650
C.solstitialis_TA1 343_347529#1	651	652
G.max_Glyma14g39620.1#1	653	654
H.annuus_TC31796#1	655	656
Ipomoea_batatas_AF234536#1	657	658
L.sativa_TC17902#1	659	660
M.truncatula_AC1 55282_1 7.4#1	661	662
P.euphratica_TA2890_75702#1	663	664

P.patens_124589#1	665	666
P.patens_138180#1	667	668
P.trifoliata_TA5575_37690#1	669	670
P.trifoliata_TA5576_37690#1	671	672
Populus_trichocarpa_EF1 47825#1	673	674
Solanum_lycopersicum_BT01 3792#1	675	676
V.vinifera_GSVIVT00022467001#1	677	678
C.clementina_DY280874#1	679	680
C.clementina_DY297038#1	681	682
C.clementina_TC487#1	683	684
C.tinctorius_TA1 847_4222#1	685	686
G.hirsutum_TC91868#1	687	688
L.saligna_TA1 747_75948#1	689	690

Sequences have been tentatively assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR; beginning with TA). 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, such as by the Joint Genome Institute. Furthermore, access to proprietary databases, has allowed the identification of novel nucleic acid and polypeptide sequences.

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

2.1 O-FUT-like polypeptides

Alignment of polypeptide sequences was performed using the AlignX programme from the Vector NTI (Invitrogen). Minor manual editing was done to further optimise the alignment. The O-FUT polypeptides are aligned in Figure 2.

A phylogenetic tree of O-FUT polypeptides (Figure 3) was reproduced from the PLAZA web site, according to the method of Yves Van De Peer et al. (2009) - Plaza, a resource for plant comparative genomics (www.vib.gent.be).

2.2 By-Pass (BPS) polypeptides

The alignment was generated using MAFFT (Katoh and Toh (2008) Briefings in Bioinformatics 9:286-298). A neighbour-joining tree was calculated using QuickTree (Howe et al. (2002), Bioinformatics 18(1 1): 1546-7), 100 bootstrap repetitions. The circular phylogram was drawn using Dendroscope (Huson et al. (2007), BMC Bioinformatics 8(1):460). Confidence for 100 bootstrap repetitions is indicated for major branching. Minor manual editing was done to further optimise the alignment.

2.3 SIZ1 polypeptides

Alignment of 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 SIZ1 polypeptides are aligned in Figure 9.

A phylogenetic tree of SIZ1 polypeptides (Figure 10) was constructed using a neighbour-joining clustering algorithm as provided in the AlignX programme from the Vector NTI (Invitrogen).

2.4 bZIP-S polypeptides

A multiple alignment of the bZIP-S polypeptides of Table A (Figure 12) was made using an alignment program based on the algorithm ClustalW as provided in the AlignX programme from the Vector NTI (Invitrogen). Default parameters were used corresponding to Blosum 62 matrix (gap opening penalty 10, gap extension penalty: 0.2). Minor manual editing was done to further enhance the alignment.

2.5 SPA15-like polypeptides

Alignment of polypeptide sequences was performed using the MAFFT (mafft (version 6.624, L-INS-I method): MAFFT: Katoh and Toh (2008) - Briefings in Bioinformatics 9:286-298). Minor manual editing was done to further optimise the alignment. The SPA15-like polypeptides are aligned in Figure 15.

A phylogenetic tree of SPA15-like polypeptides (Figure 16) was constructed using Dendroscope (Dendroscope : Huson et al. (2007), BMC Bioinformatics 8(1):460).

Example 3: Calculation of global percentage identity between polypeptide sequences useful in performing the methods of the invention

3.1 O-FUT-like polypeptides

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention are determined using one of the methods available in the art, the 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 software 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 (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosum 62 (for polypeptides), and then places the results in a distance matrix.

Parameters to be used in the comparison are: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2.

3.2 By-Pass (BPS) polypeptides

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention were determined using one of the methods available in the art, the 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 software 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 (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosum 62 (for polypeptides), and then places the results in a distance matrix.

Parameters used in the comparison were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2.

Results of the software analysis are shown in Table B2 for the global similarity and identity over the full length of the polypeptide sequences. The sequence identity (in %) between the BPS polypeptide sequences useful in performing the methods of the invention is generally higher than 55% compared to SEQ ID NO: 268.

Table B2: MatGAT results for global similarity and identity over the full length of the polypeptide sequences.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. P_sitchensis_EF085295		52.30	28.60	31.30	33.90	24.30	26.20	30.90	25.50	34.60	34.20	30.50	33.70	36.40
2. P_sitchensis_EF677456			29.30	34.70	37.70	24.50	27.60	33.90	28.70	37.50	37.50	33.80	39.80	40.00
3. C.endivia_TA154_114280				40.10	43.10	29.90	27.90	38.90	30.20	43.20	42.30	38.50	37.50	36.50
4. A.thaliana_AT1G01550.1					59.30	34.40	52.80	86.00	54.80	58.00	56.60	86.60	53.40	54.50
5. A.thaliana_AT2G46080.1						41.60	38.20	56.10	44.30	88.50	86.50	55.70	60.20	60.20
6. A.thaliana_AT3G61500.1							28.40	32.60	30.50	40.60	40.50	32.50	35.90	34.60
7. A.thaliana_AT4G01360.1								48.10	66.80	38.90	37.60	47.80	37.00	35.40
8. B.napus_TC69302									52.10	54.30	53.80	99.10	52.50	53.50
9. B.napus_TC72556										44.90	43.20	52.20	41.40	39.20
10. B.napus_TC75033											86.80	54.10	59.70	59.20
11. B.napus_TC76712												53.30	58.10	58.50
12. B.oleracea_TA5103_3712													52.50	54.50
13. G.max_Glyma07g07240.1														72.00
14. G.max_Glyma09g39170.1														
15. G.max_Glyma18g47160.1														
16. M_truncatula_BT052402														
17. T.pratense_TA1487_57577														
18.J.hindsii_x_regia_TA339_432290														
19. P.trichocarpa_scaff_II.1515														
20. P.trichocarpa_scaff_XIV.343														
21. G.hirsutum_TC114048														
22. Aquilegia_sp_TC23399														

23. M.domestica_TC8317														
24. C.sinensis_TC8893														
25. P.trifoliata_TA8107_37690														
26. C.annuum_TC9590														

	15	16	17	18	19	20	21	22	23	24	25	26	27
1. P_sitchensis_EF085295	36.40	34.00	35.00	37.60	34.70	35.10	32.70	37.30	36.90	36.50	36.80	34.80	32.90
2. P_sitchensis_EF677456	39.70	40.60	39.20	40.70	40.10	38.00	35.90	38.60	39.70	40.10	39.60	37.90	35.00
3. C.endivia_TA154_114280	37.10	37.90	35.40	42.90	40.10	44.20	40.30	40.30	40.50	39.50	39.50	41.90	38.50
4. A.thaliana_AT1G01550.1	53.90	51.80	54.80	57.90	55.90	55.30	56.40	51.00	55.30	57.80	57.00	53.10	46.90
5. A.thaliana_AT2G46080.1	60.80	57.50	58.20	61.80	61.50	61.00	57.10	54.10	59.90	61.50	61.50	53.40	51.10
6. A.thaliana_AT3G61500.1	34.60	36.00	33.00	34.60	32.80	31.90	34.60	29.70	35.40	33.80	33.10	28.20	26.50
7. A.thaliana_AT4G01360.1	35.90	36.90	36.60	40.30	38.90	39.90	39.90	35.60	36.10	39.20	38.40	37.90	34.40
8. B.napus_TC69302	53.00	51.50	54.40	55.10	54.30	53.50	53.50	48.60	54.10	55.70	54.50	51.30	46.20
9. B.napus_TC72556	39.80	39.70	41.30	45.00	42.30	42.00	40.30	36.90	40.70	43.60	42.80	40.20	36.60
10. B.napus_TC75033	59.80	57.10	58.10	61.90	60.80	62.00	57.50	55.90	58.90	60.20	59.90	53.50	51.00
11. B.napus_TC76712	59.00	56.00	57.30	60.30	58.60	59.60	56.50	54.00	58.50	59.20	58.90	51.40	50.00
12. B.oleracea_TA5103_3712	54.00	51.50	54.10	54.80	53.70	53.00	53.00	48.50	54.40	54.80	54.00	52.00	46.00
13. G.max_Glyma07g07240.1	71.20	79.30	69.50	61.70	59.90	57.90	56.20	51.10	60.70	64.10	63.80	54.20	49.70
14. G.max_Glyma09g39170.1	95.50	67.90	77.60	63.70	61.60	59.60	58.80	55.60	63.60	64.50	63.40	54.70	49.40
15. G.max_Glyma18g47160.1		67.90	78.40	63.70	62.80	60.50	58.80	56.30	63.10	64.50	63.40	55.50	50.00
16. M_truncatula_BT052402			65.90	62.10	59.80	57.80	54.40	50.70	61.70	61.70	61.70	52.20	47.80

17. T.pratense_TA1487_57577				61.50	61.10	60.20	57.30	56.50	61.60	64.80	64.20	52.80	50.30
18. J.hindsii_x_regia_TA339_432290					68.30	66.40	63.90	56.30	71.40	67.10	66.90	61.50	55.80
19. P.trichocarpa_scaff_II.1515						80.50	61.90	54.40	68.20	65.50	65.20	58.00	54.50
20. P.trichocarpa_scaff_XIV.343							61.50	56.10	65.80	65.50	64.70	56.50	55.40
21. G.hirsutum_TC114048								51.70	61.00	65.30	64.40	54.20	50.60
22. Aquilegia_sp_TC23399									55.90	57.20	56.60	53.90	49.20
23. M.domestica_TC8317										68.50	68.50	61.10	54.50
24. C.sinensis_TC8893											97.70	58.50	55.40
25. P.trifoliata_TA8107_37690												58.80	54.50
26. C.annuum_TC9590													58.40

3.3 SIZ1 polypeptides

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention are determined using one of the methods available in the art, the 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 software 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 (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosum 62 (for polypeptides), and then places the results in a distance matrix.

Parameters to be used in the comparison are: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2.

3.4 bZIP-S polypeptides

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention were determined using one of the methods available in the art, the 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 software 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 (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosum 62 (for polypeptides), and then places the results in a distance matrix. 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 comparison were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2.

Results of the software analysis are shown in Table B4 for the global similarity and identity over the full length of the polypeptide sequences. The sequence identity (in %) between the bZIP-S polypeptide sequences useful in performing the methods of the invention can be is generally higher than 43% compared to SEQ ID NO: 354.

Table B4: MatGAT results for global similarity and identity over the full length of the polypeptide sequences.

	1	31	32	33	34	35	36	37	38	39	40	41	42
1. Mt_bZIP		60,1	60,1	58,5	61,1	53,3	50,3	46,2	50,9	48,8	43,8	62	43
31. B.napus_BN06MC17829_ 45597483@17769#1	81,4		80	79,5	47,8	49,1	47,1	44,6	47,4	46,9	36,5	51,3	37
32. Arabidopsis_thaliana_ AT4G34590.1#1	79,2	89,9		99,4	50,9	51,7	42,5	45,3	50	42,5	36,2	48,1	34,4
33. Arabidopsis_thaliana_ AF053940#1	77,5	89,4	99,4		51,2	50,9	42,5	45	49,2	42,3	36	47,8	34,2
34. V.vinifera_ GSVIVT00036899001#1	77,7	70,7	71,7	71,9		55	48,1	42,9	53,1	43,6	32,8	46,5	39,6
35. Arabidopsis_thaliana_ BT004768#1	68,9	68,9	71,9	70,1	70,1		41,9	64,1	96,5	48	33	43,7	41,7
36. V.vinifera_ GSVIVT00014558001#1	69,9	68,6	63,5	63,1	67,5	59,3		36,4	42	41,3	30,1	44,2	33,6
37. B.napus_BN06MC23287_ 49055078@23201#1	64,2	63	62,4	62,4	59,4	79,6	56,4		66,5	35,4	31,3	38,2	41,2
38. A.thaliana_AT1G75390.1#1	67,1	67,1	69,4	67,6	67,6	96,5	58,4	79,8		48,3	33	42,2	39,7
39. G.max_GM06MC33749_ sm67c05@32966#1	68,6	64,9	60,4	60	65	64,1	59,1	56,4	62,4		36,7	44,6	38
40. S.bicolor_Sb07g015450.1#1	57,4	58	55,7	55,7	53,4	52,8	44,9	48,3	53,4	56,3		37,5	31,3
41. G.max_GM06MC32046 _sj86f07@31311#1	76,3	69,9	66,7	66,3	66,9	63,5	65	57,6	61,3	63,6	54,5		43,2
42. Tamarix_hispida_FJ752700#1	66,7	62,7	57,2	56,9	59,9	58,1	59,7	53,9	56,1	54,5	48,3	66	

3.5 SPA15-like polypeptides

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention were determined using one of the methods available in the art, the 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 software 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 (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosum 62 (for polypeptides), and then places the results in a distance matrix. 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 comparison were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2.

Results of the software analysis are shown in Table B5 for the global similarity and identity over the full length of the polypeptide sequences. The sequence identity (in %) between the SPA15-like polypeptide sequences useful in performing the methods of the invention is generally higher than 30% compared to SEQ ID NO: 634.

Table B5: MatGAT results for global similarity and identity over the full length of the polypeptide sequences.

1. Os_SPA15like		53,90	54,60	27,60	99,80	94,10	65,20	71,40
2. A.thaliana_AT1G66330.1	67,00		97,80	39,70	54,10	56,30	48,10	54,80
3. Arabidopsis_thaliana_AY086709	67,00	99,30		39,20	54,80	57,00	49,10	55,30
4. B.napus_TC82749	32,20	41,20	41,00		27,80	29,50	41,70	28,50
5. LOC_Os05g05600.1	99,80	67,20	67,20	32,40		94,30	65,40	71,40
6. LOC_Os05g05600.5	94,10	69,40	69,40	34,30	94,30		69,40	72,50
7. LOC_Os05g05600.6	65,40	58,00	58,30	49,00	65,60	69,60		55,20
8. S.bicolor_Sb05g026090.1	82,50	67,80	68,90	33,60	82,50	81,70	60,00	
9. Zea_mays_EU956861	83,40	67,40	68,10	33,50	83,40	82,20	59,70	93,80
10. C.clementina_DY280874	41,40	47,50	48,40	32,30	41,40	43,90	36,40	39,50
11. C.clementina_DY297038	40,90	47,70	48,20	33,10	40,90	43,40	36,40	39,50
12. C.clementina_TC487	42,90	50,10	50,60	32,30	42,90	43,90	37,00	41,30
13. C.solstitialis_TA1343_347529	71,30	71,90	72,10	36,50	71,10	72,60	56,40	70,40
14. C.tinctorius_TA1847_4222	44,20	44,40	44,80	28,20	44,20	46,20	40,40	42,20
15. G.hirsutum_TC91868	40,50	44,40	44,80	31,50	40,50	40,60	34,40	40,00
16. G.max_Glyma14g39620.1	69,60	68,90	69,80	35,30	69,80	71,00	57,80	68,20
17. H.annuus_TC31796	69,80	72,60	73,50	34,80	69,60	72,40	57,30	68,90

18. Ipomoea_batatas_AF234536	66,50	70,00	70,00	31,70	66,70	67,70	52,40	64,70
19. L.saligna_TA1747_75948	43,30	46,30	46,80	31,40	43,30	44,50	40,70	41,90
20. L.sativa_TC17902	57,30	62,10	62,60	25,80	57,30	59,40	52,80	57,00
21. M.truncatula_AC155282_17.4	68,30	70,00	71,00	34,90	68,50	69,60	56,20	64,90
22. P.euphratica_TA2890_75702	71,30	72,20	72,80	34,60	71,50	70,40	54,10	70,00
23. P.patens_124589	43,50	51,60	51,30	41,70	43,50	46,20	60,90	44,20

1. Os_SPA15like	72,20	27,90	27,40	28,70	53,40	31,20	29,70	50,10
2. A.thaliana_AT1G66330.1	54,80	35,70	35,70	36,60	59,30	33,60	35,50	56,80
3. Arabidopsis_thaliana_AY086709	55,30	36,30	35,70	36,60	60,20	34,20	35,50	57,70
4. B.napus_TC82749	28,90	18,80	19,30	17,50	29,90	16,80	19,90	29,90
5. LOC_Os05g05600.1	72,20	27,90	27,40	28,70	53,20	31,20	29,70	50,30
6. LOC_Os05g05600.5	73,10	29,00	28,60	29,70	54,80	31,00	29,90	50,90
7. LOC_Os05g05600.6	55,10	18,90	18,60	17,50	45,70	21,20	18,80	46,20
8. S.bicolor_Sb05g026090.1	91,70	28,40	28,10	29,80	54,80	30,20	29,30	51,50
9. Zea_mays_EU956861		27,40	27,10	28,70	53,10	29,30	28,80	51,50
10. C.clementina_DY280874	39,20		95,90	91,10	34,10	50,30	62,40	33,00
11. C.clementina_DY297038	39,20	97,20		89,10	33,40	49,70	61,20	32,30
12. C.clementina_TC487	41,60	92,70	91,70		34,50	50,90	58,60	34,90
13. C.solstitialis_TA1343_347529	69,40	45,90	45,00	46,80		61,40	33,10	53,20
14. C.tinctorius_TA1847_4222	41,90	68,40	67,20	66,70	63,20		49,80	29,80
15. G.hirsutum_TC91868	39,00	73,30	71,40	69,60	43,60	68,00		32,00
16. G.max_Glyma14g39620.1	67,40	47,20	46,50	49,50	70,10	42,20	40,30	
17. H.annuus_TC31796	68,70	47,00	46,30	48,20	87,70	59,70	45,30	70,00
18. Ipomoea_batatas_AF234536	63,20	49,00	50,20	51,90	74,00	51,70	46,90	67,50
19. L.saligna_TA1747_75948	42,10	66,60	66,60	64,40	59,40	87,20	66,60	42,60
20. L.sativa_TC17902	57,00	54,50	54,80	56,70	75,60	72,80	54,50	57,40
21. M.truncatula_AC155282_17.4	67,20	46,80	46,60	49,20	70,50	43,80	44,30	86,90
22. P.euphratica_TA2890_75702	69,80	51,30	50,90	53,30	76,70	47,80	46,30	78,00
23. P.patens_124589	44,30	40,70	41,40	41,60	47,50	38,10	38,70	47,70

1. Os_SPA15like	52,50	52,60	30,30	42,90	50,40	53,00	33,00	33,50
2. A.thaliana_AT1G66330.1	58,40	56,10	35,50	49,40	55,60	61,60	37,80	38,20
3. Arabidopsis_thaliana_AY086709	59,50	57,20	36,40	50,30	55,90	62,10	37,60	38,90
4. B.napus_TC82749	30,20	25,60	17,50	19,00	29,10	29,30	27,90	28,50
5. LOC_Os05g05600.1	52,30	52,80	30,30	42,90	50,60	53,20	33,00	33,50
6. LOC_Os05g05600.5	53,70	53,00	30,90	43,90	49,90	53,00	35,00	35,50
7. LOC_Os05g05600.6	44,50	41,70	21,80	35,10	45,10	43,30	43,80	45,60
8. S.bicolor_Sb05g026090.1	52,10	51,00	30,50	44,00	51,10	54,70	32,90	33,00

9. Zea_mays_EU956861	52,00	51,20	29,40	42,80	49,90	54,50	32,60	32,30
10. C.clementina_DY280874	33,90	38,90	48,10	39,70	32,60	44,30	18,00	16,30
11. C.clementina_DY297038	32,80	38,50	48,40	39,40	31,90	43,70	18,30	17,90
12. C.clementina_TC487	34,20	40,30	48,40	40,50	34,60	44,80	18,20	15,60
13. C.solstitialis_TA1343_347529	80,60	60,80	54,20	69,40	52,30	60,60	35,50	36,20
14. C.tinctorius_TA1847_4222	51,80	42,40	78,50	65,40	31,40	37,10	21,40	21,50
15. G.hirsutum_TC91868	33,80	38,80	49,50	40,10	31,80	38,30	19,40	19,90
16. G.max_Glyma14g39620.1	53,50	54,50	30,60	43,00	80,20	64,10	35,90	35,30
17. H.annuus_TC31796		60,40	53,50	68,80	53,00	59,10	37,10	38,30
18. Ipomoea_batatas_AF234536	75,50		41,20	55,70	52,80	56,10	34,10	33,30
19. L.saligna_TA1747_75948	59,40	50,20		79,50	30,30	35,70	20,60	20,40
20. L.sativa_TC17902	75,90	66,40	80,10		42,80	48,70	26,80	27,40
21. M.truncatula_AC155282_17.4	71,00	67,20	43,80	58,80		62,00	35,60	36,20
22. P.euphratica_TA2890_75702	72,40	67,00	46,30	60,90	75,90		33,60	33,50
23. P.patens_124589	49,90	46,90	38,70	44,40	48,20	46,10		79,90

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

4.1 O-FUT-like polypeptides

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.

The results of the InterPro scan of the polypeptide sequence as represented by SEQ ID NO: 2 are presented in Table B1.

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

Database	Accession number	Accession name	Size SEQ ID NO 2	Size Full Length ortholog SEQ ID NO 21
PFam	PF03138	O-FucT	355	574

4.2 By-Pass (BPS) polypeptides

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.

4.3 SIZ1 polypeptides

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.

The results of the InterPro scan of the polypeptide sequence as represented by SEQ ID NO: 353 are presented in Table C3.

Table C3: InterPro scan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 354.

Database	Accession number	Accession name	Average size on SEQ ID NO 354
PFam	PF02037	SAP	34
PFam	PF00628	PHD	54
PFam	PF02891	zf-MIZ	49

4.4 bZIP-S polypeptides

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.

The results of the InterPro scan of the polypeptide sequence as represented by SEQ ID NO: 422 are presented in Table C4.

Table C4: InterPro scan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 422.

			start	stop	E-value		
PatternScan	PS00036	BZIP_BASIC	33	48	NA	IPR004827	Basic-leucine zipper (bZIP), transcriptionfactor
HMMSmart	SM00338	BRLZ	26	90	1,80E-17	IPR004827	Basic-leucine zipper (bZIP), transcriptionfactor
HMMPfam	PF00170	bZIP_1	28	89	1,20E-08	IPR011616	Basic-leucine zipper (bZIP), transcriptionfactor
ProfileScan	PS50217	BZIP	28	91	10,53	IPR004827	Basic-leucine zipper (bZIP), transcriptionfactor
HMMPanther	PTHR19304:SF2	ATF31-YEAST	38	75	0,00054	NULL	NULL

4.5 SPA15-like polypeptides

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, PROS ITE, 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.

The results of the InterPro scan of the polypeptide sequence as represented by SEQ ID NO: 634 are presented in Table C5.

Table C5: InterPro scan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 634

Database	Accession number	Accession name	Amino acid coordinates on SEQ ID NO 634	
			start	stop
superfamily	SSF46785	"Winged helix" DNA-binding domain	37	106
superfamily	SSF48371	ARM repeat	308	421

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

5.1 O-FUT-like polypeptides & 5.2 By-Pass (BPS) polypeptides & 5.3 SIZ1 polypeptides

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. TargetP is maintained at the server of the Technical University of Denmark.

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

A number of parameters are selected, 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:

- 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 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).

5.4 bZIP-S polypeptides

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. TargetP is maintained at the server of the Technical University of Denmark.

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

A number of parameters are selected, 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).

The "plant" organism group is selected, no cutoffs defined, and the predicted length of the transit peptide requested.

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

- 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 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).

5.5 SPA15-like polypeptides

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. TargetP is maintained at the server of the Technical University of Denmark.

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

A number of parameters are selected, such as organism group, cutoff sets (none, predefined set of cutoffs, or user-specified set of cutoffs), and the calculation of prediction of cleavage sites (yes or no).

The "plant" organism group is 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: 634 may be the cell wall, no transit peptide is predicted.

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

- 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 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).

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

Reference is made to Van Norman et al. (2004) - BYPASS1 Negatively Regulates a Root-Derived Signal that Controls Plant Architecture. Current Biology, Vol. 14, 1739-1746, October 15, 2004

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

7.1 O-FUT-like polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made Oryza sativa seedlings cDNA library (in pCMV Sport 6.0; Invitrogen, Paisley, UK). PCR was performed using Hifi Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm1403 (SEQ ID NO: 265; sense, start codon in bold): 5'-ggggacaagtttgtacaaaaaagcaggcttaaacaatggaccaatcactcaagtgg-3' and prm 14039 (SEQ ID NO: 266; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggttcctcttcataacaa atcagcg-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", pO-FUT. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 1 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: 264) for constitutive specific expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::0-FUT (Figure 4) was transformed into Agrobactehum strain LBA4044 according to methods well known in the art.

7.2 By-Pass (BPS) polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made Arabidopsis thaliana seedlings cDNA library (in pCMV Sport 6.0; Invitrogen, Paisley, UK). PCR was performed using Hifi Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix.

The primers used were prm13550 (SEQ ID NO: 351 sense, start codon in bold): 5'-ggggac aagtttgtacaaaaaagcaggcttaaacaatggctcgtccacaagac-3' and prm13551 (SEQ ID NO: 352; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggtgaagtaaaaccatctgtacaaaca-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", pBPS. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 367 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: 350) for constitutive specific expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::BPS (Figure 7) was transformed into Agrobacterium strain LBA4044 according to methods well known in the art.

7.3 SIZ1 polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made Arabidopsis thaliana seedlings cDNA library (in pCMV Sport 6.0; Invitrogen, Paisley, UK). PCR was performed using Hifi Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm13568 (SEQ ID NO: 419; sense, start codon in bold): 5'-ggggacaagtttgtacaaaaaagcaggcttaaacaatggatttggaagctaattgt-3' and prm13569 (SEQ ID NO: 420; reverse, complementary): 5'-ggggaccactttgtacaagaaagctg ggtcaacagaacagacaaatcagg-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", pSIZ1 . Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 353 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: 418) for constitutive specific expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::SIZ1 (Figure 10) was transformed into Agrobacterium strain LBA4044 according to methods well known in the art.

7.4 bZIP-S polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made Medicago truncatula seedlings cDNA library (in pCMV Sport 6.0; Invitrogen, Paisley, UK). PCR was performed using Hifi Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were as represented by SEQ ID NO: 627 (sense) and SEQ ID NO: 628 (reverse, complementary 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", pbZIP-S. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 421 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: 629) for constitutive specific expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::bZIP-S (Figure 13) was transformed into Agrobacterium strain LBA4044 according to methods well known in the art.

7.5 SPA15-like polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made Oryza sativa seedlings cDNA library (in pCMV Sport 6.0; Invitrogen, Paisley, UK). PCR was performed using Hifi Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm12099 (SEQ ID NO: 701 ; sense, start codon in bold): 5'-ggggacaagtttgtacaaaaaagcaggcttaaacaatggctactcgcattcctg-3' and prm12100 (SEQ ID NO: 702; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggtttcttatttgcacatgatcacc-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", pSPA15-like. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 633 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: 700) for constitutive specific expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::SPA15-like (Figure 17) was transformed into Agrobacterium strain LBA4044 according to methods well known in the art.

Example 8: Plant transformation

Rice transformation

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 minutes in 0.2% $HgCl_2$, followed by a 6 times 15 minutes wash with sterile distilled water. The sterile seeds were then germinated on a medium containing 2,4-D (callus induction medium). After incubation in the dark for four weeks, embryogenic, scutellum-derived calli were excised and propagated on the same medium. After two weeks, the calli were multiplied or propagated by subculture on the same medium for another 2 weeks. Embryogenic callus pieces were sub-cultured on fresh medium 3 days before co-cultivation (to boost cell division activity).

Agrobacterium strain LBA4404 containing the expression vector was used for co-cultivation. 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 (OD_{600}) of about 1. The suspension was then transferred to a Petri dish and the calli immersed in the suspension for 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. Co-cultivated calli were grown on 2,4-D-containing medium for 4 weeks in the dark at 28°C in the presence of a selection agent. During this period, rapidly growing resistant callus islands developed. After transfer of this material to a regeneration medium and incubation in the light, the embryogenic potential was released and shoots developed in the next four to five 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.

Approximately 35 independent TO rice transformants were generated for one construct. The primary 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 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 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 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 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.

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 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 K2S04, 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.

Example 10: Phenotypic evaluation procedure

10.1 Evaluation setup

Approximately 35 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.

Drought screen

Plants from T2 seeds 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. Humidity 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.

Nitrogen use efficiency screen

Rice plants from T2 seeds 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

Plants are grown on a substrate made of coco fibers and 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. Seed-related parameters are then measured.

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

Biomass-related parameter measurement

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.

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 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. The early vigour is the plant (seedling) aboveground area three weeks post-germination. Increase in root biomass is expressed as an increase in total root biomass (measured as 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).

Early vigour was determined by counting the total number of pixels from aboveground plant parts 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. The results described below are for plants three weeks post-germination.

Seed-related parameter measurements

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 filled husks were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining fraction was counted again. The filled husks were weighed on an analytical balance. The number of filled seeds was determined by counting the number of filled husks that remained after the separation step. The total seed yield was measured by weighing all filled husks harvested from a plant. Total seed number per plant was measured by counting the number of husks harvested from a plant. Thousand Kernel Weight (TKW) is extrapolated from the number of filled seeds counted and their total weight. The Harvest Index (HI) in the present invention is defined as the ratio between the total seed yield and the above ground area (mm^2), multiplied by a factor 10^6. The total number of flowers per panicle as defined in the present invention is the ratio between the total number of seeds and the number of mature primary panicles. The seed fill rate as defined in the present invention is the proportion (expressed as a %) of the number of filled seeds over the total number of seeds (or florets).

Example 11: Phenotypic evaluation procedure

11.1 O-FUT-like polypeptides

The results of the evaluation of transgenic rice plants under non-stress conditions are presented below. An increase of more than 5 % was observed for aboveground biomass (AreaMax), emergence vigour (early vigour), total seed yield, number of filled seeds, fill rate, number of flowers per panicle, harvest index, and of (2.5-3)% for thousand kernel weight

Table C1: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for each parameter the p-value is <0.05.

Parameter	Overall
AreaMax	7.1
totalwgseeds	22.6
fillrate	8.3
harvestindex	13.8
nrfilledseed	19.0

11.2 By-Pass (BPS) polypeptides

The results of the evaluation of transgenic rice plants expressing a nucleic acid encoding the BPS polypeptide of SEQ ID NO: 267 under non-stress conditions are presented below in Table C2.

When grown under non-stress conditions, an increase of more than 5 % was observed for total seed yield, fill rate and harvest index.

Table C2: Results of the evaluation of transgenic rice plants expressing a nucleic acid encoding the BPS polypeptide of SEQ ID NO: 268 under non-stress conditions - for each parameter, the percentage overall is shown if it reaches $p \leq 0:05$ and above the 5% threshold.

Parameter	Overall increase
totalwgseeds	11.2
fillrate	13.9
harvestindex	12.7

11.3 SIZ1 polypeptides

The results of the evaluation of transgenic rice plants under non-stress conditions are presented below (Table D1). An increase of at least 5 % was observed for total seed yield (totalwgseeds), number of filled seeds, fill rate, number of flowers per panicle (flowerperpan), harvest index, centre of gravity of the canopy (GravityYMax), proportion of the thick root in the root system (RootThickMax) and of thousand kernel weight (TKW).

Table D1. Evaluation of transgenic rice plants under non-stress conditions

Parameter	Overall
totalwgseeds	25.7
nrfilledseed	18.2
fillrate	18.0
flowerperpan	15.4
harvestindex	16.5
TKW	6.3
GravityYMax	5.0
RootThickMax	5.9

For each parameter, the percentage overall is shown if it reaches $p \leq 0:05$ and above the 5% threshold.

11.4 bZIP-S polypeptides

The results of the evaluation of transgenic rice plants in the T1 generation and comprising and expressing a nucleic acid comprising the longest Open Reading Frame in SEQ ID NO: 421

encoding the polypeptide of SEQ ID NO: 422 under non-stress conditions are presented below. See previous Examples for details on the generations of the transgenic plants.

The results of the evaluation of transgenic rice plants under non-stress conditions are presented below (Table D2). An increase of at least 5 % was observed for the total seed yield (totalwgseeds), number of filled seeds (nrfilledseed), number of flowers per panicle (flowerperpan) and harvest index (harvestindex) in the transgenic compared to the control plants.

Table D2.

(Parameter) Yield Trait	% increase in transgenic compared to control plants
totalwgseeds	14.2
harvestindex	9.9
nrfilledseed	13.2
flowerperpan	7.9

In a similar experiment, rice plants transformed with a Populus trichocarpa bZIP-like coding sequence (SEQ ID NO: 465) under control of the GOS2 promoter (SEQ ID NO: 629) were evaluated in a drought screen as described above. One of the six tested lines showed an increase in total weight of seeds, fillrate, harvest index, TKW, number of filled seeds. A second line had increased fill rate and harvest index and a third line showed increased TKW.

11.5 SPA15-like polypeptides

The results of the evaluation of transgenic rice plants in the T1 and T2 generations and expressing a nucleic acid encoding the SPA15-like polypeptide of SEQ ID NO: 634 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 - fill rate, harvest index and thousand kernel weight (TKW).

In addition, plants expressing a SPA15-like nucleic acid showed higher total seed weight, number of filled seeds, flowers per panicle and the maximum gravity of the plants (Gravity YMax - height (in mm) of the gravity centre of the leafy biomass.

Table D3: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for T1 generation and the confirmation (T2 generation), for each parameter the p-value is <0.05.

Parameter	Overall increase	
	T1	T2
fill rate	9.1	9.1
harvestindex	17.7	8.2
TKW	8.2	8.9

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 O-FUT polypeptide, wherein said O-FUT polypeptide comprises a domain with a PFam accession number PF10250.

2. Method, according to claim 1, wherein said O-FUT polypeptide comprises one or more of the following motifs:
   (i) Motif 1: HYIALHLRYEKDM (SEQ ID NO: 261),
   (ii) Motif 2: IYIVAGEIYGGHSMD (SEQ ID NO: 262),
   (iii) Motif 3: ALDYNVAVQSDVFVYTYDGNMAKAVQGH (SEQ ID NO: 263)

3. Method, according to claim 1 or 2, wherein said O-FUT polypeptide may comprise a conserved Arginine residue in Motif 1.

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

5. Method according to any one of claims 1 to 4, wherein said nucleic acid encoding an O-FUT polypeptide encodes any one of the proteins listed in Table A or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

6. Method according to any one of claims 1 to 5, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A1.

7. Method according to any preceding claim, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

8. Method according to any one of claims 1 to 7, wherein said enhanced yield-related traits are obtained under non-stress conditions.

9. Method according to any one of claims 1 to 7, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

10. Method according to any one of claims 1 to 9, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

11. Method according to any one of claims 1 to 10, wherein said nucleic acid encoding a O-FUT polypeptide is of any origin, preferably of plant origin, more preferably from a monocotyledonous plant, further preferably from the family Poaceae, particularly preferably from the genus Oryza, most preferably from Oryza sativa.

12. Plant or part thereof, including seeds, obtainable by a method according to any one of claims 1 to 11, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a O-FUT polypeptide.

13. Construct comprising:
(i) nucleic acid encoding a O-FUT polypeptide as defined in any of the claims 1 to 3;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

14. Construct according to claim 13, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

15. Use of a construct according to claim 13 or 14 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

16. An isolated nucleic acid molecule selected from:
(i) a nucleic acid represented by SEQ ID NO: 1;
(ii) the complement of a nucleic acid represented by SEQ ID NO: 1;
(iii) a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 2, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be derived from a polypeptide sequence as represented by SEQ ID NO: 2 and further preferably confers enhanced yield-related traits relative 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%, 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 table A1 and further preferably conferring enhanced yield-related traits relative to control plants;

(v) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;

(vi) a nucleic acid encoding a O-FUT 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 SEQ ID NO: 2 and any of the other amino acid sequences in Table A1 and preferably conferring enhanced yield-related traits relative to control plants.

17. An isolated polypeptide selected from:

(i) an amino acid sequence represented by SEQ ID NO: 2;

(ii) 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 amino acid sequence represented by any one of SEQ ID NO: 2 or 22 and any of the other amino acid sequences in Table A1 and preferably conferring enhanced yield-related traits relative to control plants.

(iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.

18. Plant, plant part or plant cell transformed with a construct according to claim 13 or 14.

19. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:

(i) introducing and expressing in a plant a nucleic acid encoding an O-FUT polypeptide as defined in any of the claims 1 to 3; and

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

20. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding an O-FUT polypeptide as defined in any of the claims 1 to 3, or a transgenic plant cell derived from said transgenic plant.

21. Transgenic plant according to claim 12, 18 or 20, or a transgenic plant cell derived thereof, wherein said plant is a crop plant such as sugarbeet, 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.

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

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

24. Use of a nucleic acid encoding a O-FUT polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

25. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a BPS polypeptide.

26. A method, according with claim 25, wherein said BPS polypeptide further comprises at least one of the following motifs:
    (i) Motif 4: SWM[KT][LQ]A[MI]ESLC[EA][TI]H[TN]DIKTLIT[DE]LELP (SEQ ID NO: 341)
    (ii) Motif 5: D[IL]C[IN]AFSSE[LI][ST]RLNQGHL[LY]L[QK]C[AV]LHNL[DE][SG]SS (SEQ ID NO: 342)
    (iii) Motif 6: GKVLM[RQ]A[ML]YGV[KR]V[VQ]TV[FY][IV]CS[VI]FA[AV]AFSGS (SEQ ID NO: 343)

27. Method according to any of the claims 25 or 26, wherein said BPS polypeptide further comprises at least one or more of the following motifs:
    (i) Motif 7: SWM[KT][LQ]A[MI]ESLC[EA][TI]H[NT]D[IV]KTLIT[DE]LELPVSDW[DE][ED] KW[IV]DVYLD[IN]SVKL (SEQ ID NO: 344)
    (ii) Motif 8: SL[ND]LPK[VI]KNSAKGKVLM[RQ]A[ML]YGV[KR]V[QV]TV[FY][IV]CSVFA [AV]AFSGS (SEQ ID NO:345)
    (iii) Motif 9: PQ[ED]P[HP]R[PS]F[FL]PFGNPF (SEQ ID NO: 346)

28. Method according to any of the claims 25 to 27, wherein said BPS polypeptide further comprises one or more of the following motifs:
    (i) Motif 10: [VM]PK[EDN]K[SDN][DQ]ILT[LV]SWM[KS][QL]AM[EA]SLC[EQ]TH[KN] [NAS]I[KNR]TL[IV]TDL[EQ]LPVSD[WL]E[ED][KN][WF][VI][DY][IV]Y (SEQ ID NO: 347)
    (ii) Motif 11: LPK[VK]KNSAKGKVL[ML]RA[LF]YGVKV[KQ]T[LV]YI[CS][SG]VF[AT]AA [FW]S[GD]S[ST][NQK][ND]L[FL][YD][LV][TP][VI][SP][NE][EK] (SEQ ID NO: 348)

(iii) Motif 12: [PL]WA[KQP][SVA]F[MT][DE][MLV]Q[NS][TV][VM]N[AGPS]EI[KR][ND][IM][FL][LS]S[DG][GR][LFS]T[VI][LIM]K[ED]LE[AS]V[DE][AS][GS]V[KE][KQ]L[YA][PT][AM][IV]Q[DQE]G[SV] (SEQ ID NO: 349)

29. Method according to any of the claims 25 to 28, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a BPS polypeptide.

30. Method according to any one of claims 25 to 29, wherein said nucleic acid encoding a BPS polypeptide encodes any one of the proteins listed in Table A2 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

31. Method according to any one of claims 25 to 30, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A2.

32. Method according to any preceding claims, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

33. Method according to any one of claims 25 to 32, wherein said enhanced yield-related traits are obtained under non-stress conditions.

34. Method according to any one of claims 25 to 32, wherein said enhanced yield-related traits are obtained under conditions of a type of stress affecting the plant fertility.

35. Method according to any one of claims 25 to 34, wherein said nucleic acid is operably linked to a promoter active in roots.

36. Method according to any one of claims 25 to 34, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

37. Method according to any one of claims 25 to 36, wherein said nucleic acid encoding a BPS polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Brassicaceae, more preferably from the genus Arabidopsis, most preferably from Arabidopsis thaliana.

38. Plant or part thereof, including seeds, obtainable by a method according to any one of claims 25 to 37, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a BPS polypeptide.

39. Construct comprising:
    (i) nucleic acid encoding a BPS polypeptide as defined in any of the claims 25 to 27;
    (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
    (iii) a transcription termination sequence.

40. Construct according to claim 39, wherein one of said control sequences is a promoter active in roots.

41. Construct according to claim 39, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

42. Use of a construct according to any of the claims 39 to 41 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

43. Plant, plant part or plant cell transformed with a construct according to any of the claims 39 to 41.

44. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
    (i) introducing and expressing in a plant a nucleic acid encoding a BPS polypeptide as defined in any of the claims 25 to 28; and
    (ii) cultivating the plant cell under conditions promoting plant growth and development.

45. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding a BPS polypeptide as defined in any of the claims 25 to 28, or a transgenic plant cell derived from said transgenic plant.

46. Transgenic plant according to claim 38, 43 or 45, or a transgenic plant cell derived thereof, wherein said plant is a crop plant such as sugarbeet, 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.

47. Harvestable parts of a plant according to claim 46, wherein said harvestable parts are preferably shoot biomass and/or seeds.

48. Products derived from a plant according to claim 46 and/or from harvestable parts of a plant according to claim 47.

49. Use of a nucleic acid encoding a BPS polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

50. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a SIZ1 polypeptide, wherein said SIZ1 polypeptide comprises a DUF206 domain.

51. Method according to claim 50, wherein said SIZ1 polypeptide comprises one or more of the following motifs:
(i) Motif 13: MSCNGCRXLRKGCX (SEQ ID NO: 409),
(ii) Motif 14: QXXATXFXAKFXGR (SEQ ID NO: 410),
(iii) Motif 15: FXSLLXEAXG (SEQ ID NO: 411)

52. Method according to claim 50 or 51, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a SIZ1 polypeptide.

53. Method according to any one of claims 50 to 52, wherein said nucleic acid encoding a SIZ1 polypeptide encodes any one of the proteins 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.

54. Method according to any one of claims 50 to 53, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A3.

55. Method according to any preceding claim, wherein said enhanced yield-related traits comprise increased yield, preferably increased biomass and/or increased seed yield relative to control plants.

56. Method according to any one of claims 50 to 55, wherein said enhanced yield-related traits are obtained under non-stress conditions.

57. Method according to any one of claims 50 to 55, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

58. Method according to any one of claims 52 to 57, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

59. Method according to any one of claims 50 to 58, wherein said nucleic acid encoding a SIZ1 polypeptide is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Brassicaceae, more preferably from the genus Arabidopsis, most preferably from Arabidopsis thaliana.

60. Plant or part thereof, including seeds, obtainable by a method according to any one of claims 50 to 59, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a SIZ1 polypeptide.

61. Construct comprising:
(i) nucleic acid encoding a SIZ1 polypeptide as defined in claims 50 or 51;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

62. Construct according to claim 61, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

63. Use of a construct according to claim 61 or 62 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

64. Plant, plant part or plant cell transformed with a construct according to claim 61 or 62.

65. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding a SIZ1 polypeptide as defined in claim 50 or 51; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

66. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding a SIZ1 polypeptide as defined in claim 50 or 51, or a transgenic plant cell derived from said transgenic plant.

67. Transgenic plant according to claim 60, 64 or 66, or a transgenic plant cell derived thereof, wherein said plant is a crop plant such as sugarbeet, 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.

68. Harvestable parts of a plant according to claim 67, wherein said harvestable parts are preferably shoot biomass and/or seeds.

69. Products derived from a plant according to claim 67 and/or from harvestable parts of a plant according to claim 68.

70. Use of a nucleic acid encoding a SIZ1 polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

71. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a bZIP-S polypeptide.

72. Method according to claim 71, wherein said bZIP-S polypeptide comprises one or more of the following motifs:
    (i) Motif 19 as represented by SEQ ID NO: 522;
    (ii) Motif 20 as represented by SEQ ID NO: 587;
    (iii) Motif 21 as represented by SEQ ID NO: 600.

73. Method according to claim 71 or 72, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a bZIP-S polypeptide.

74. Method according to any one of claims 71 to 73, wherein said nucleic acid encoding a bZIP-S polypeptide encodes any one of the proteins 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.

75. Method according to any one of claims 71 to 74, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A4.

76. Method according to any preceding claim, wherein said enhanced yield-related traits comprise increased seed yield relative to control plants.

77. Method according to any one of claims 71 to 76, wherein said enhanced yield-related traits are obtained under non-stress conditions.

78. Method according to any one of claims 71 to 76, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

79. Method according to any one of claims 73 to 78, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

80. Method according to any one of claims 71 to 79, wherein said nucleic acid encoding a bZIP-S polypeptide 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.

81. Plant or part thereof, including seeds, obtainable by a method according to any one of claims 71 to 80, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a bZIP-S polypeptide.

82. Construct comprising:
(i) nucleic acid encoding a bZIP-S polypeptide as defined in claims 71 or 72;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

83. Construct according to claim 82, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

84. Use of a construct according to claim 82 or 83 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

85. Plant, plant part or plant cell transformed with a construct according to claim 82 or 83.

86. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding a bZIP-S polypeptide as defined in claim 71 or 72; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

87. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid

encoding a bZIP-S polypeptide as defined in claim 71 or 72, or a transgenic plant cell derived from said transgenic plant.

88. Transgenic plant according to claim 81, 85 or 87, or a transgenic plant cell derived thereof, wherein said plant is a crop plant, such as beet or sugarbeet, 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.

89. Harvestable parts of a plant according to claim 88, wherein said harvestable parts are preferably shoot biomass and/or seeds.

90. Products derived from a plant according to claim 88 and/or from harvestable parts of a plant according to claim 89.

91. Use of a nucleic acid encoding a bZIP-S polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

92. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a SPA15-like polypeptide, wherein said SPA15-like polypeptide comprises an Armadillo-type fold domain with an InterPro accession number IPR016024 and SuperFamily accession number SSF48371 and a "winged helix" DNA-binding domain with a SuperFamily accession number SSF46785.

93. Method according to claim 92, wherein said SPA15-like polypeptide comprises one or more of the following motifs:

(i) Motif 22: AAD[KR]HWSDGALEADLR[RL]ADF[RV][AV][KR][QR]RAMEDA[LF]MAL [EK]F[VI]K[ND][IV]HDMMV[SN][KR][ML][YQ][KE] (SEQ ID NO: 691);

(ii) Motif 23: RA[RC]QDVA[IV]LGS[GE]FLKLDARAR[EK]DTEKID[RHN] (SEQ ID NO: 692);

(iii) Motif 24: L[SA]EA[DC]GIDY[TN]D[PA]E[EF][LV] (SEQ ID NO: 693).

94. Method according to any of the previous claims, wherein said SPA15-like polypeptide comprises one or more of the following motifs:

(i) Motif 25: EADGIDYTDPEELELLV[AT]TLIDLDAMDGK[SG]S[VA]SLLAECSSSPD VNTR[KQ]AL (SEQ ID NO: 694);

(ii) Motif 26: APSMW[TI]LGNAGMGALQRLA[EQ]DSN[PY]A[IV]A[AR]A (SEQ ID NO: 695);

(iii) Motif 27: FPGEVS[TA]D[RQ]ITAI[QE]EAYW[SD]MA (SEQ ID NO: 696).

95. Method according to any of the previous claims, wherein said SPA15-like polypeptide comprises one or more of the following motifs:
    (i) Motif 28: DGIDYTDPEELELLV[AT]TLIDLDAMDGK[KSR]S[VA]SL[LI]AECSSSPD VNTRKALAN (SEQ ID NO: 697);
    (ii) Motif 29: PSMW[TI]LGNAGMGALQRLA[QE]D[SP]N[YP]A[VI]A[RA]AA[ST]RAI [ND][EA]L[KT]KQWE[LV]EEGDSLRF (SEQ ID NO: 698);
    (iii) Motif 30: [GL][SV][ST]S[PER][AT][NG][ST][TR][SDG][FR]I[TS]LEKNG[NKI][TA] [LF][EG][LF]FP[GH]EVS[TSA]D[QR]I[TSY]AIE[EQ]AY[WKQ]SMASA[LF]SEA (SEQ ID NO: 699).

96. Method according to any of the previous claims, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding a SPA15-like polypeptide.

97. Method according to any of the previous claims, wherein said nucleic acid encoding a SPA15-like polypeptide encodes any one of the proteins listed in Table A5 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.

98. Method according to any of the previous claims, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the proteins given in Table A5.

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

100. Method according to claim 99, wherein said enhanced yield-related traits are obtained under non-stress conditions.

101. Method according to claim 99, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.

102. Method according to any one of claims 92 to 98, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

103. Method according to claim 102, wherein said nucleic acid encoding a SPA15-like 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.

104. Plant or part thereof, including seeds, obtainable by a method according to any one of claims 92 to 101, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a SPA15-like polypeptide.

105. Construct comprising:
(i) nucleic acid encoding a SPA15-like polypeptide as defined in any of the claims 92 to 95;
(ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
(iii) a transcription termination sequence.

106. Construct according to claim 105, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

107. Use of a construct according to claim 105 or 106 in a method for making plants having increased yield, particularly increased biomass and/or increased seed yield relative to control plants.

108. Plant, plant part or plant cell transformed with a construct according to claim 105 or 106.

109. Method for the production of a transgenic plant having increased yield, particularly increased biomass and/or increased seed yield relative to control plants, comprising:
(i) introducing and expressing in a plant a nucleic acid encoding a SPA15-like polypeptide as defined in any of the claims 92 to 95; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

110. Transgenic plant having increased yield, particularly increased biomass and/or increased seed yield, relative to control plants, resulting from modulated expression of a nucleic acid encoding a SPA15-like polypeptide as defined in any of the claims 92 to 95, or a transgenic plant cell derived from said transgenic plant.

111. Transgenic plant according to any of the claims 104, 108 or 110, or a transgenic plant cell derived thereof, wherein said plant is a crop plant, such as beet or sugarbeet, 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.

112. Harvestable parts of a plant according to claim 111, wherein said harvestable parts are preferably shoot biomass and/or seeds.

113. Products derived from a plant according to claim 111 and/or from harvestable parts of a plant according to claim 112.

114. Use of a nucleic acid encoding a SPA15-like polypeptide in increasing yield, particularly in increasing seed yield and/or shoot biomass in plants, relative to control plants.

115. An isolated nucleic acid molecule selected from:

(i) a nucleic acid represented by SEQ ID NO: 633;

(ii) the complement of a nucleic acid represented by SEQ ID NO: 633;

(iii) a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 634, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be derived from a polypeptide sequence as represented by SEQ ID NO: 634, and further preferably confers enhanced yield-related traits relative 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%, 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 table A5 and further preferably conferring enhanced yield-related traits relative to control plants;

(v) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;

(vi) a nucleic acid encoding a SPA15-like 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 SEQ ID NO: 634, and any of the other amino acid sequences in Table A5 and preferably conferring enhanced yield-related traits relative to control plants.

116. An isolated polypeptide selected from:

(i) an amino acid sequence represented by SEQ ID NO: 634;

(ii) 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 amino acid sequence represented by SEQ ID NO: 634, and any of the other amino acid sequences in Table A5 and preferably conferring enhanced yield-related traits relative to control plants.

(iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.

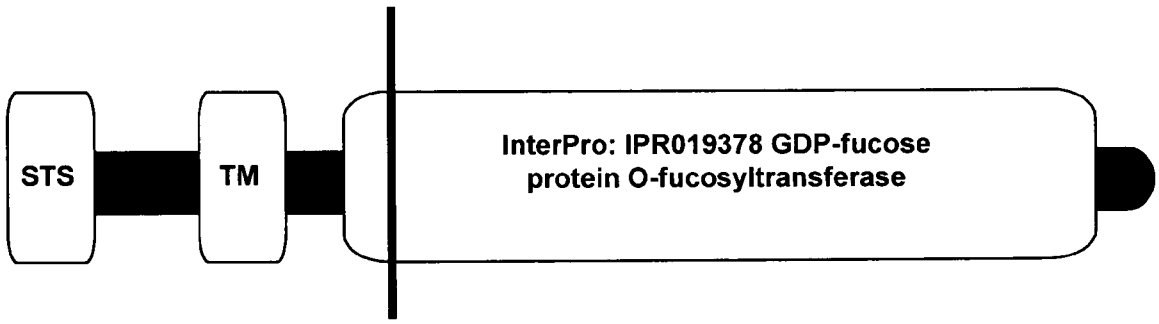
STS
TM
InterPro: IPR019378 GDP-fucose
protein O-fucosyltransferase

FIGURE 1

```
                      1                                                40
    AT2G37980     (1) --------MSAAGASPLAVAPITAPTTTTRRRVGDSLETT
    AT3G54100     (1) -------------MSVVGAINPVPIAPTTRRRVGDSLDVI
CP00056G00770     (1) -------MSSGGSNNPTGGSPLAAAGPTTTRRRVADSQDV
   VV08G10900     (1) -------MSSGGANSCNGSP--RLAGPTTTRRRVADMIDA
   PT06G07850     (1) ---------SLSGSSTTGGP-----------------HH
   PT16G09540     (1) ----------------------------------------
    AT5G01100     (1) -------MSVGVPVNPSSSSQLPAAPTTTTRRRVADSQED
   VV13G05630     (1) ----------------------------------------
    AT5G35570     (1) --MGHHHDGGDGVPQHHVNSP------RFSGPMTRRAQSF
CP00280G00020     (1) MGHHHHHHNTTDGVSQIVNSP------RFSGPMTRRAHSF
   PT00G00080     (1) -----NHNSASDGVSQRVNSP------RFSGPMTRRAHSF
   OS09G27080     (1) --MGYHQPVHGGAAAGAGGSG-----GGGGGDHLHQHHHP
  SB02G025960     (1) --MGYQPMHGNGGTAGAGTSGSGGGGGGGGGDHLH-HHHQ
PP00046G01420     (1) ----------------------------------------
PP00284G00340     (1) ----------------------------------------
PP00204G00120     (1) ----------------------------------------
PP00066G00650     (1) ----------------------------------------
PP00131G00330     (1) ----------------------------------------
  SB03G040020     (1) --------MVASSSSTASAPSPCSCSPVTSGAHRRRLTDA
   OS01G63230     (1) ----------MAAAPPSVASSSSSSSPVLSAAHRRRLNDV
 SEQ ID NO: 2     (1) ----------------------------------------
    Consensus     (1)

                      41                                               80
    AT2G37980    (33) SERPSISSDYCNTNTVNIAISPDIDDGETGLQGG--A---
    AT3G54100    (28) SERPSNSSDYCNTINQPIVGSTDLDGGDATGQDS--S---
CP00056G00770    (34) SERPSNSSDYHSVNTATNTVGNINGSTTASNRNSNVSLSF
   VV08G10900    (32) E-RSSNFSDVS----------------------------
   PT06G07850    (14) --FHHNYHCHN-----------------------------
   PT16G09540     (1) ----------------------------------------
    AT5G01100    (34) HSHVNTVGGGNAVVYVPDEEETATSCCG-----------
   VV13G05630     (1) ----------------------------------------
    AT5G35570    (33) KRGGSAGSSSNNNNTHVGVSGGDGN----NNNNTSSTLRV
CP00280G00020    (35) KRN-NNGSTHNNSNTANNTSTNNGNN--CSSTSCSSVLTT
   PT00G00080    (30) KRN--NTSSNNNSNAGNANSSNNGSNNVSNGNSNNSILSP
   OS09G27080    (34) RLHSPRISGGG-SMTRRANSFKR-------GEIELQIGSP
  SB02G025960    (38) RLHSPRMAAGGGSMTRRASSFKRGA----SGEIELQIGSP
PP00046G01420     (1) ----------------------------------------
PP00284G00340     (1) ----------------------------------------
PP00204G00120     (1) ----------------------------------------
PP00066G00650     (1) ----------------------------------------
PP00131G00330     (1) ----------------------------------------
  SB03G040020    (33) AAAVDYFVVPCDDDSDDGRG-G-----------------
   OS01G63230    (31) ERDAFDYGGPCDVDVDHDDDDGG-GGVR------------
 SEQ ID NO: 2     (1) ----------------------------------------
    Consensus    (41)
```

FIGURE 2

```
                      81                                                120
    AT2G37980   (68)  --CSSPSSIGSSSSGSHYHHDHHYHHHPTIRYFLLR----
    AT3G54100   (63)  --PSSESSASSTGSHYHHDHYHRFYSHPVIRYLLLR----
CP00056G00770   (74)  DLDEEDNPNGSYYSALGGSQLHHYQHHPVIRYLQFRRKTL
   VV08G10900   (42)  --EEEDNPNG------NNGDHHHHCHHPSIRYWLLRRRIF
   PT06G07850   (23)  ----------------------HSHRHHPLIKYLLLRRKFF
   PT16G09540    (1)  ----------------------------------------
    AT5G01100   (62)  ---------GGGGGGSLSCCPSGSHHNYLVGFLSLRKFR-
   VV13G05630    (1)  ----------------------------------------
    AT5G35570   (69)  HHEIDLPLNSPRSEIVSGSSGSDPSGGFDSALN-RKHQTY
CP00280G00020   (72)  HHEIDLHLNSPRSEITPSSASSD---GFDSILQMEKKYTY
   PT00G00080   (68)  HLEIDLPLNSPRSETVDGFERES-----HSRQN-LSQRVH
   OS09G27080   (66)  RSPRGDGVGSPLAESSSASEASAGGGG-AVHHHHHQSQQ-
  SB02G025960   (74)  RSPRCDGLGSPPS-----DSVEPSGGG-LLHHHHHQGQQN
PP00046G01420    (1)  ----------------------------------------
PP00284G00340    (1)  ----------------------------------------
PP00204G00120    (1)  ----------------------------------------
PP00066G00650    (1)  ----------------------------------------
PP00131G00330    (1)  ----------------------------------------
  SB03G040020   (54)  --HGAGAGAARVKALFFARRASGGKRVPVVDQAWVRN---
   OS01G63230   (58)  --RGHGAGVAGVRALFSSARRSK-RASVIIDQAWLRN---
 SEQ ID NO: 2    (1)  ----------------------------------------
    Consensus   (81)

                      121                                               160
    AT2G37980  (102)  ----KLRLPFLFDGVGSTAVVGQGWWLCSGRNVGRRILGL
    AT3G54100   (97)  ----KFWIPFYGG--ASTVVIGQGFR--SGRNVGRRILGL
CP00056G00770  (114)  LCVPEAWLFKIEDCCHWTAAMAQRLG--SGRNAGRRILAL
   VV08G10900   (74)  FFAPETWLLKIENGCSWAATAVQWLR--SGRHMGRRIFVG
   PT06G07850   (42)  FFVPESWLLGVED---LTATISRGLR--SGKNMGRRIFGV
   PT16G09540    (1)  ----------------------------------------
    AT5G01100   (92)  ----LVWMLMVENKSKWTAGIARNMR--STTNLGRFILTL
   VV13G05630    (1)  ------------------------------MYGEIFLG
    AT5G35570  (108)  GQLRERVVKGLL--RKPMGSVVSDFSLRERKKLGHWMFFA
CP00280G00020  (109)  HNLREREVKSLL--KKPIGSTVLELGIRERKKLGHWMFLV
   PT00G00080  (102)  GGVVRILTNK----KGSIGSVILDFGFKERKKLGHWMFFF
   OS09G27080  (104)  QQLRFRLFKRPG-SG--AGEVVLGLGIRERRRIGNLLFLA
  SB02G025960  (108)  QHLRFRLFKPPVGSGGSAVEVGLGLGIRERRKLGNMLFLA
PP00046G01420    (1)  ----------------------------------------
PP00284G00340    (1)  ----------------------------------------
PP00204G00120    (1)  ----------------------------------------
PP00066G00650    (1)  ----------------------------------------
PP00131G00330    (1)  ----------------------------------------
  SB03G040020   (89)  -------AVACLLGVAVVVGLVMSSRR--GATTGAEGRLV
   OS01G63230   (92)  -------VVACLLGLTVVAGLVLSSHRVSGAGGGRLVQRM
 SEQ ID NO: 2    (1)  ----------------------------------------
    Consensus  (121)                       V            LG  I
```

FIGURE 2 (continued)

```
                        161                                       200
    AT2G37980  (138)  LMIFVVVSLFLRVSLMSG-RVVDHAHRRDLNELVVMRALH
    AT3G54100  (129)  LMLLVVASVFLRVYLMGGVRVVDHAR---LKEFVVVRTLR
CP00056G00770  (152)  LMVLLVFSVFLRVSLLTSGVEVERKRK--EVINMFILQVQ
   VV08G10900  (112)  LLLLVFFSVFVKFSLMSSHVEANGKR---ENGLLIIQTFK
   PT06G07850   (77)  LMVMAVLSVFLKVSFWSQTERNIHEN----SNLVIFRHFK
   PT16G09540    (1)  ----------------------------------------
    AT5G01100  (126)  LSILVVTFFLIVALSGGVGRRRKHVEKHEFVVSIHPRPTI
   VV13G05630    (9)  LLTNAFMDFYCCVVFASH----------------FSG--
    AT5G35570  (146)  FCGVCLFLGVFKICATGWLGSAIDGAASDQDLSIP----R
CP00280G00020  (147)  FCGVCLFLGVFKICATGWFGSTIEMASSNQDLSDPSIS-Q
   PT00G00080  (138)  FCGLCLFLGVFKICLYGWFGSTLERAASNQVLHLIDVFGS
   OS09G27080  (141)  FCGVCLLLGVAKIWAGGWF--ALPGDDKDADLQDLSISFS
  SB02G025960  (148)  FCCVCLLLGVAKILAGGWF--ALPGNDKDADLKDLSVSFS
PP00046G01420    (1)  ----------------------------------------
PP00284G00340    (1)  ----------------------------------------
PP00204G00120    (1)  ----------------------------------------
PP00066G00650    (1)  ----------------------------------------
PP00131G00330    (1)  ----------------------------------MLQAV
  SB03G040020  (120)  RRVDAEVLRWREENLTAVARRPPDPPL--CMDLETIRTCR
   OS01G63230  (125)  DLGDGEVMGWTEENLTAVARQSPDTP--------------
 SEQ ID NO: 2    (1)  ----------------------------------------
    Consensus  (161)

                        201                                       240
    AT2G37980  (177)  EDWSMAQRAMTEN-------------------------VV
    AT3G54100  (166)  DDWSMAQREVAENQASSQPM-----------------RV
CP00056G00770  (190)  QDGSSVQHALAENHASSMPK-----------------RV
   VV08G10900  (149)  EDWAMAQKVVAEDEASDAVKP----------------LRV
   PT06G07850  (113)  DDWARAQRSIIEHHPSISTP-----------------LL
   PT16G09540    (1)  ----------------------------------------
    AT5G01100  (166)  EKIIREDESSN-S---------------------FQVLV
   VV13G05630   (30)  -DGDPTFQSVPAT---------------------------
    AT5G35570  (182)  VNLLDHSSHDYIYKDGGNDVDPTLVMVASD-----VVGDQ
CP00280G00020  (186)  LNLISQDSHKYGYREGGSDVERTLMMVASG-----VVGSQ
   PT00G00080  (178)  ITRQEQDSYRYMGSE--NDQKRMIIEVGSD-----VVDRL
   OS09G27080  (179)  SDGGYQFDSHFGHVGG-KESDRMLMTVGS----------E
  SB02G025960  (186)  GEKVHQVDHHFVYMGG-KESDRTLMTLESN-----IDGRE
PP00046G01420    (1)  ----------------------------------------
PP00284G00340    (1)  ----------------------------------------
PP00204G00120    (1)  ----------------------MPHTRH-----EDRRK
PP00066G00650    (1)  ----------------------------------------
PP00131G00330    (6)  KVESNEAKYGHESQILVDTSAVSIPSQEHT-----AVSKD
  SB03G040020  (158)  VKRIQAKSATGGNLRGLGLGGDIQTGCRRHFRGCPIQEGV
   OS01G63230  (151)  ----------------------------------------
 SEQ ID NO: 2    (1)  ----------------------------------------
    Consensus  (201)
```

FIGURE 2 (continued)

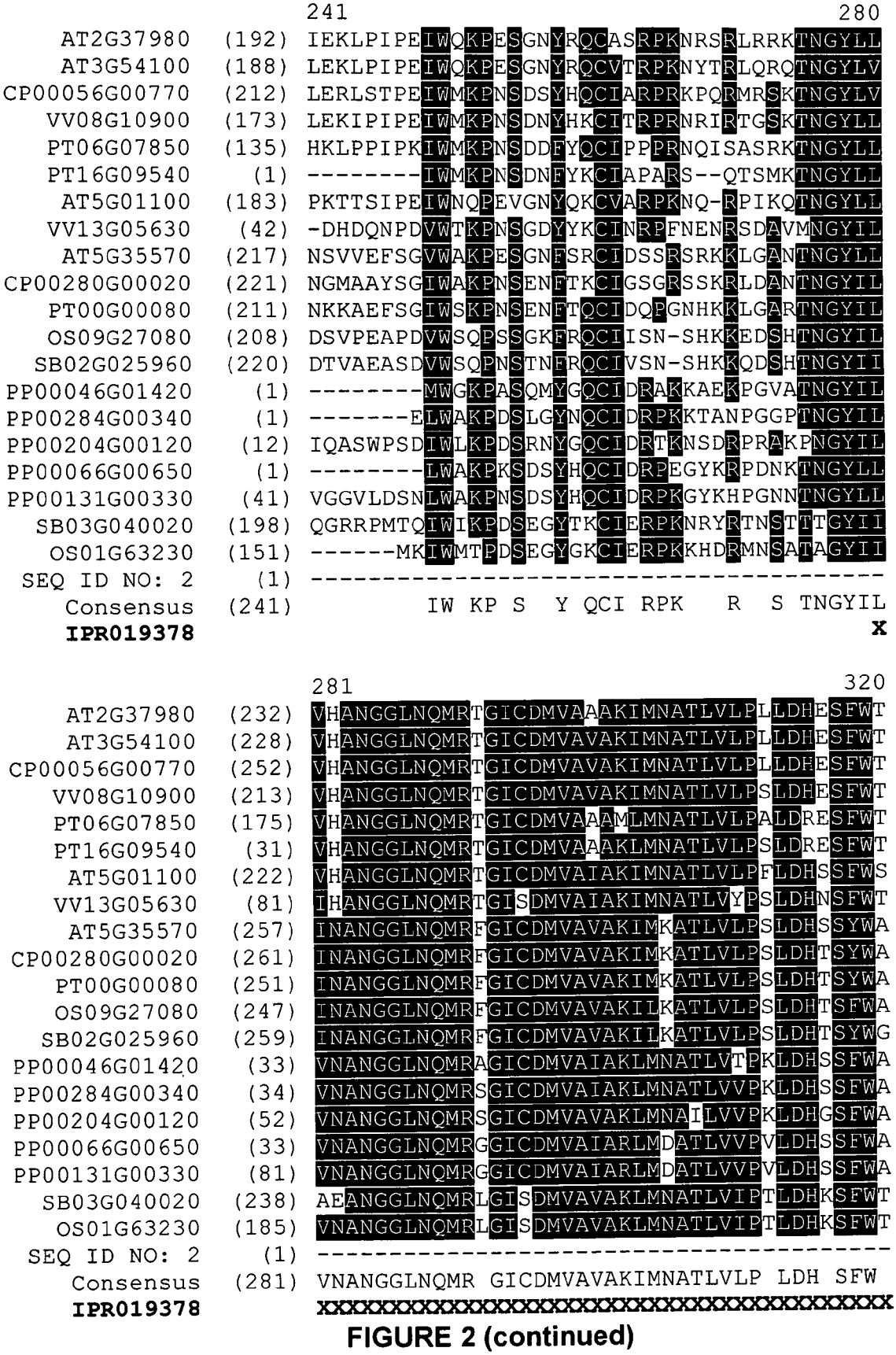

FIGURE 2 (continued)

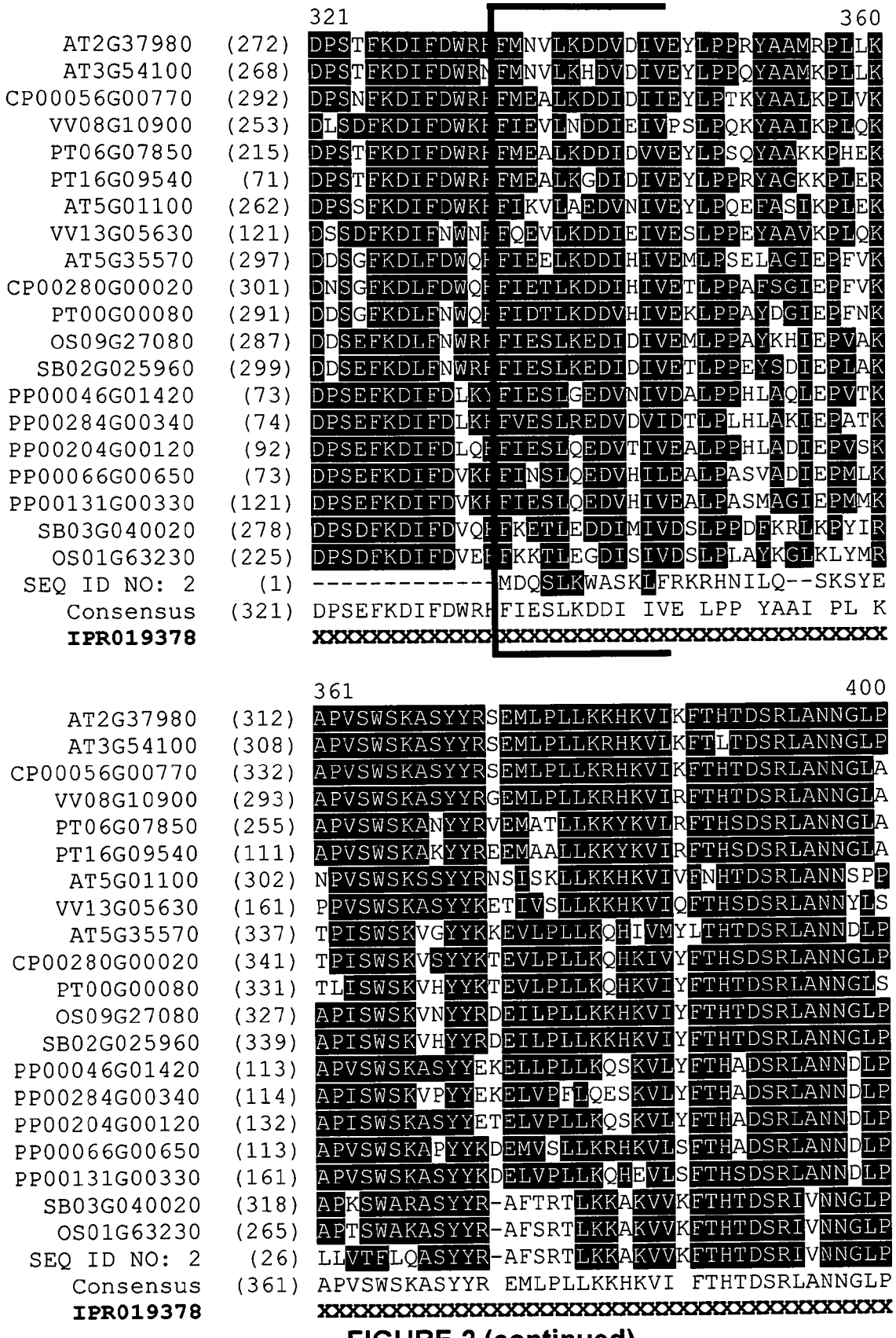

FIGURE 2 (continued)

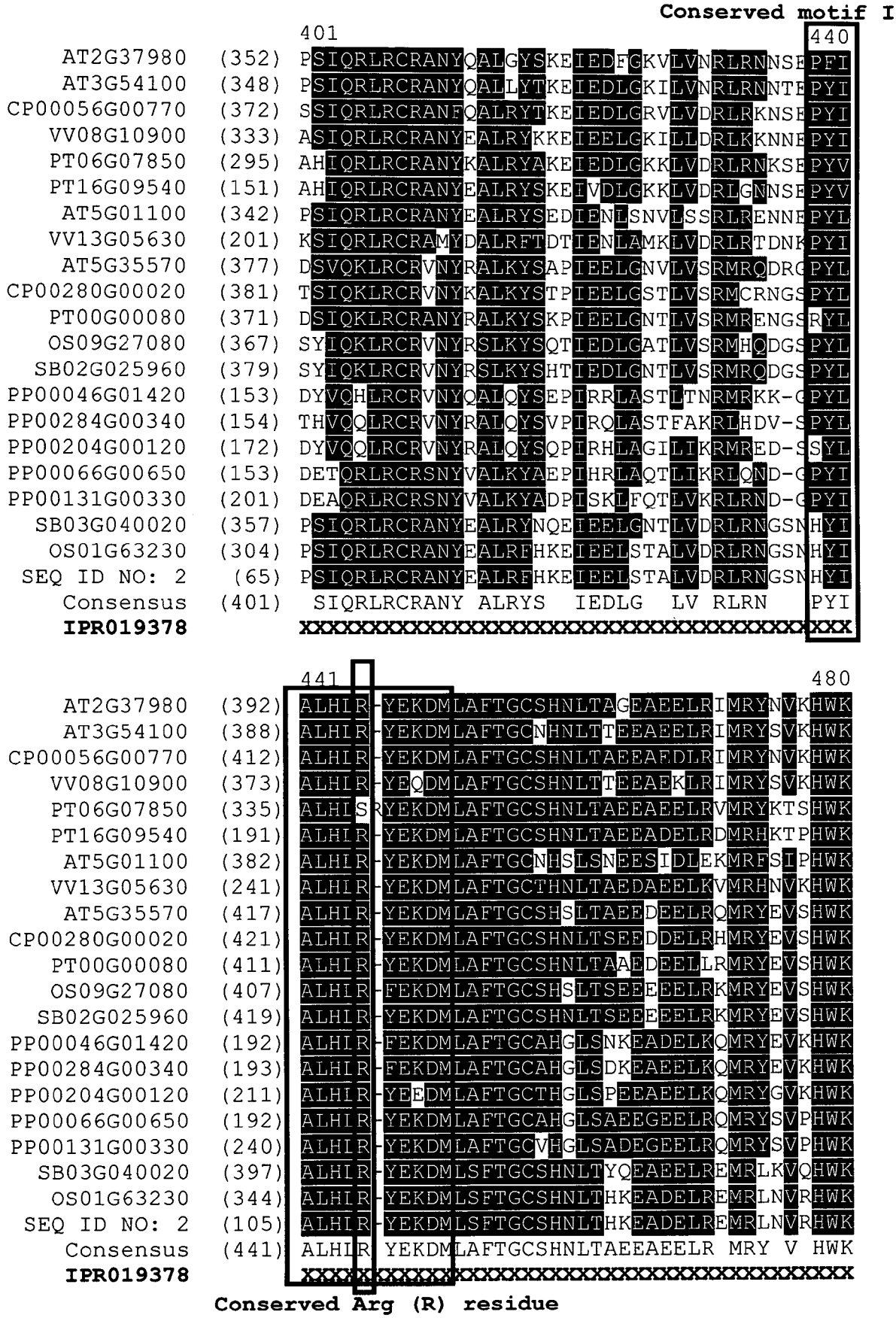

FIGURE 2 (continued)

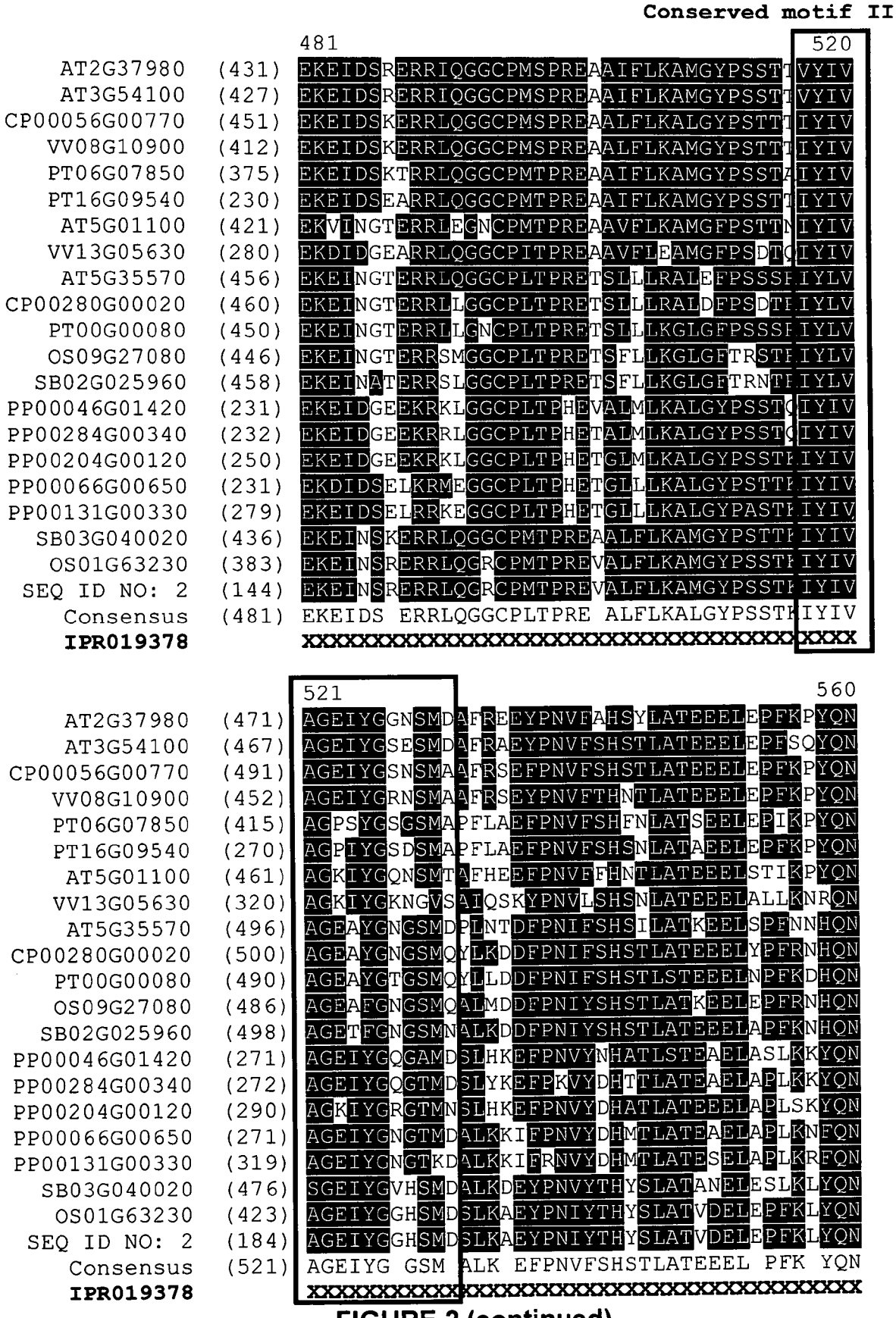

FIGURE 2 (continued)

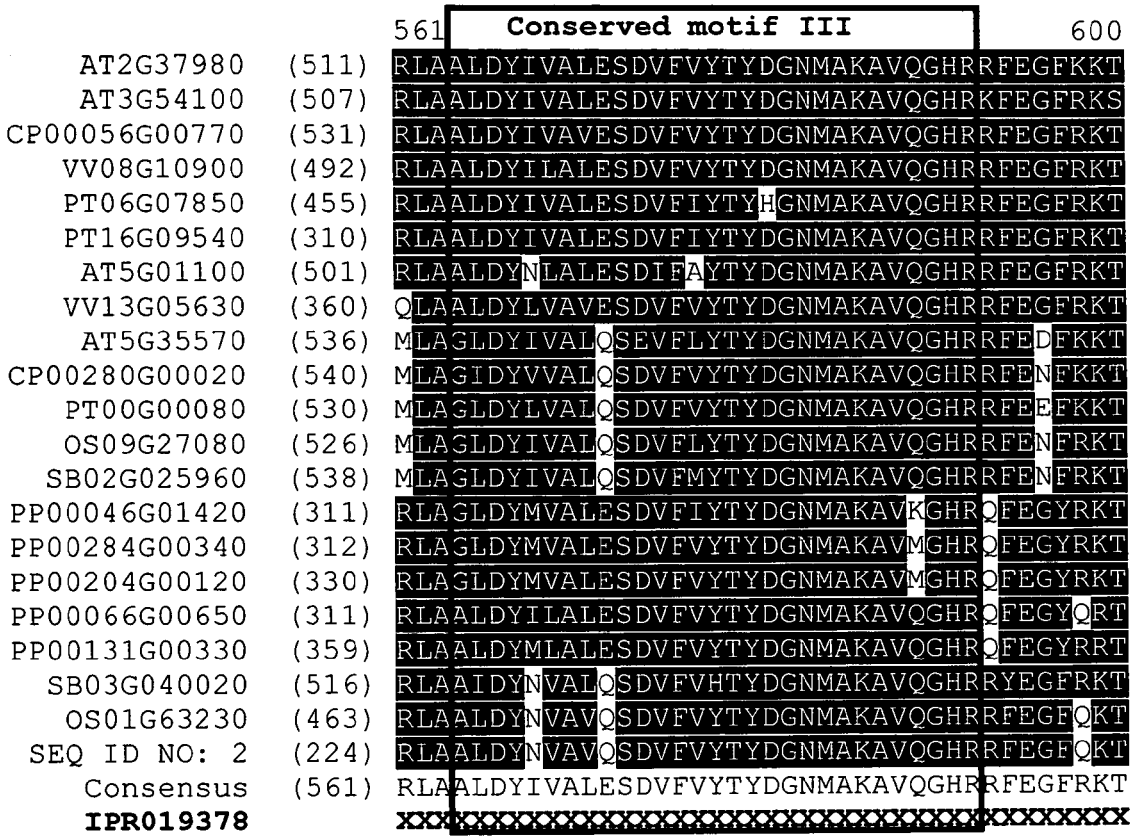

FIGURE 2 (continued)

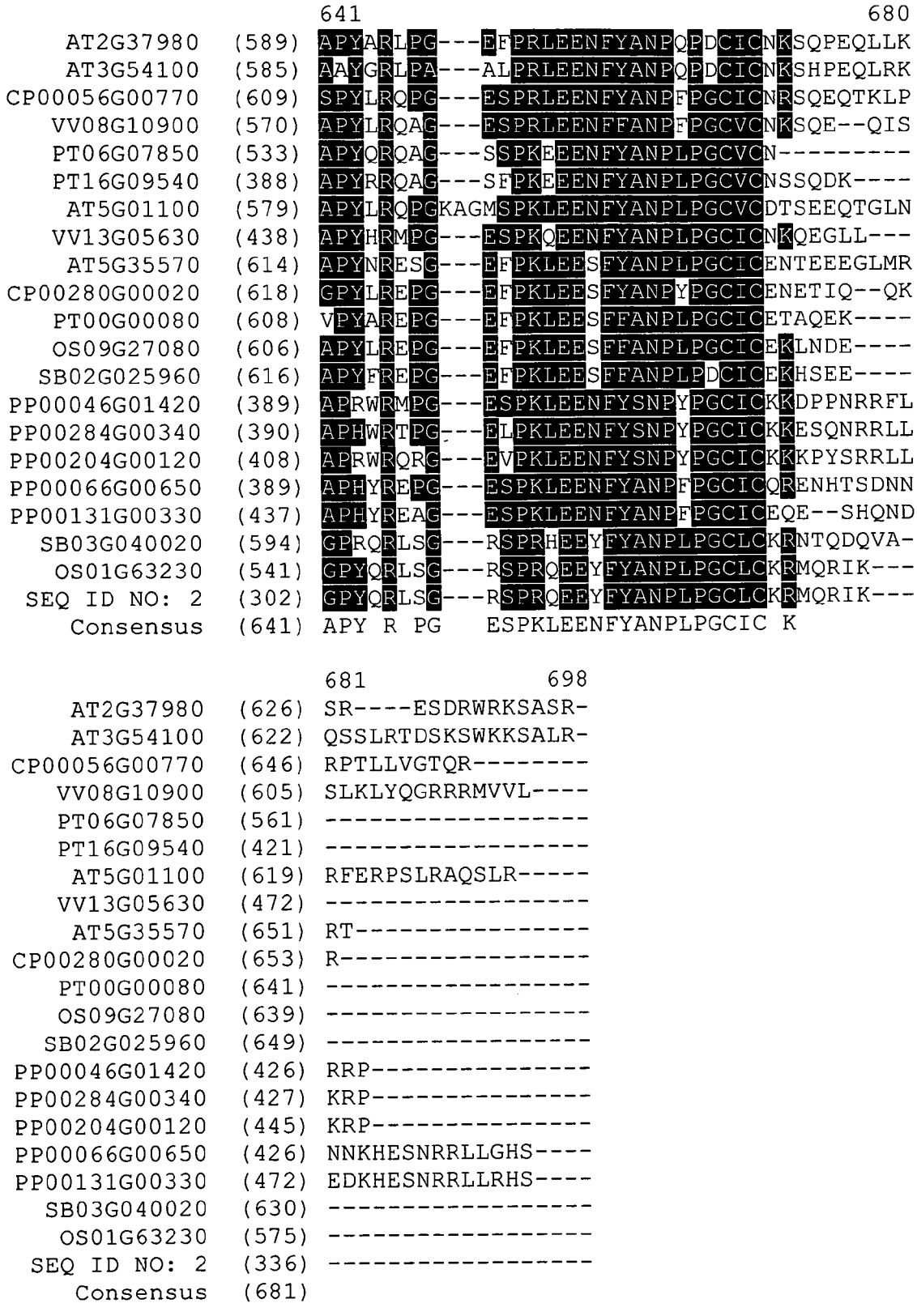

FIGURE 2 (continued)

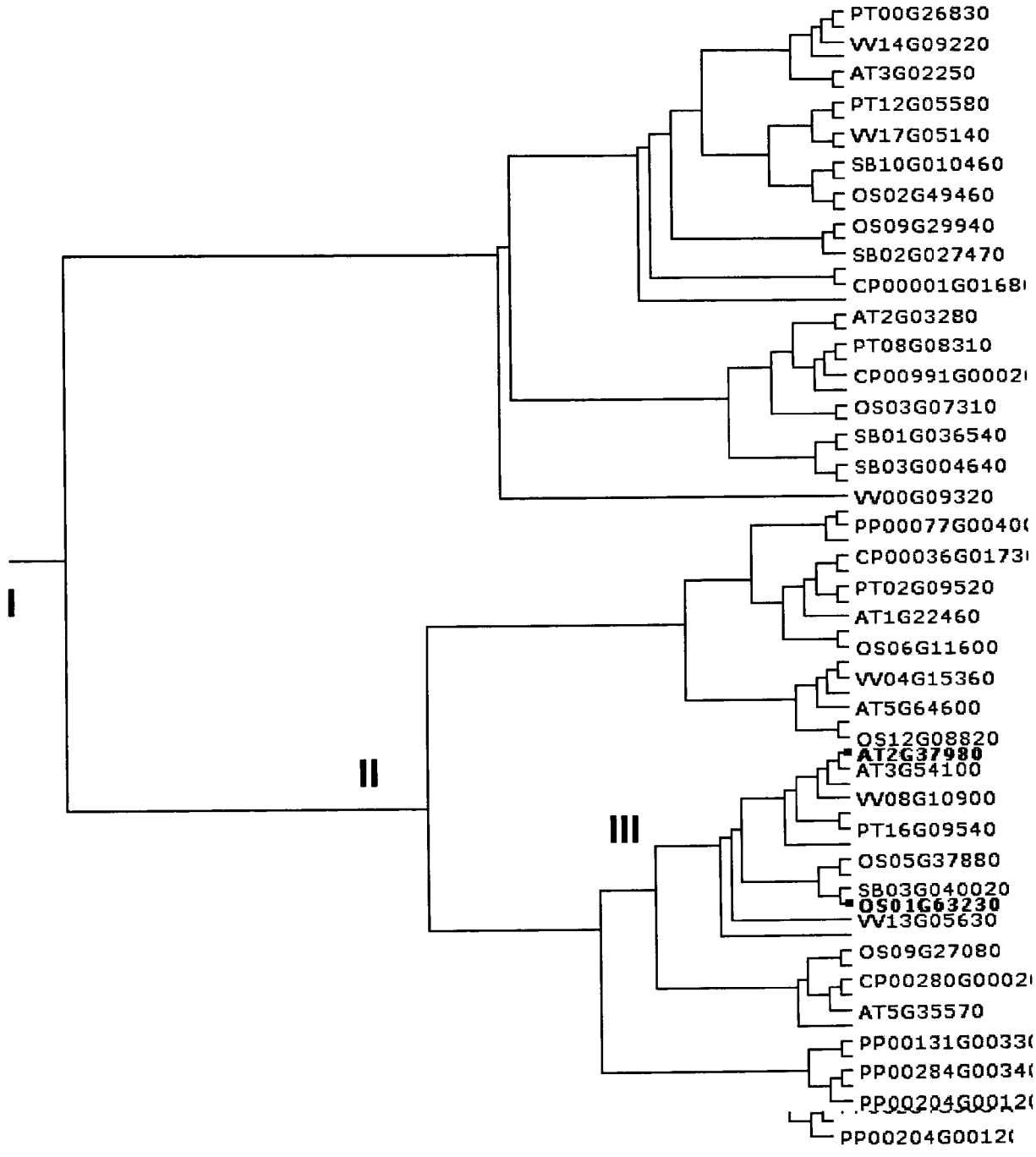
PT00G26830
VV14G09220
AT3G02250
PT12G05580
VV17G05140
SB10G010460
OS02G49460
OS09G29940
SB02G027470
CP00001G0168
AT2G03280
PT08G08310
CP00991G0002
OS03G07310
SB01G036540
SB03G004640
VV00G09320
PP00077G0040
CP00036G0173
PT02G09520
AT1G22460
OS06G11600
VV04G15360
AT5G64600
OS12G08820
AT2G37980
AT3G54100
VV08G10900
PT16G09540
OS05G37880
SB03G040020
OS01G63230
VV13G05630
OS09G27080
CP00280G0002
AT5G35570
PP00131G0033
PP00284G0034
PP00204G0012
PP00204G0012
I
II
III

FIGURE 3

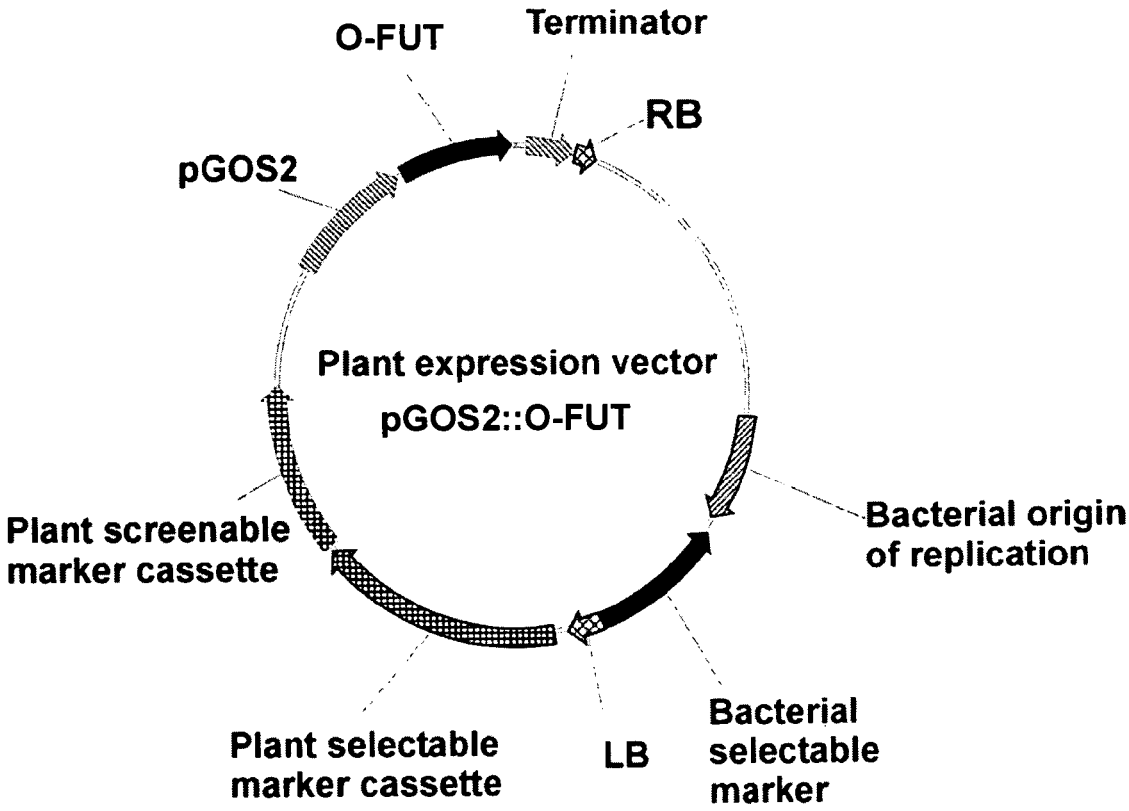
O-FUT
Terminator
RB
pGOS2
Plant expression vector
pGOS2::O-FUT
Plant screenable
marker cassette
Bacterial origin
of replication
Plant selectable
marker cassette
LB
Bacterial
selectable
marker

FIGURE 4

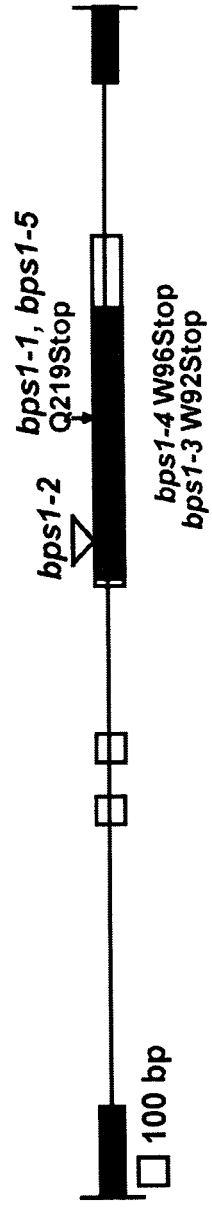
bps1-1, bps1-5
Q219Stop
bps1-2
bps1-4 W96Stop
bps1-3 W92Stop
100 bp

FIGURE 5

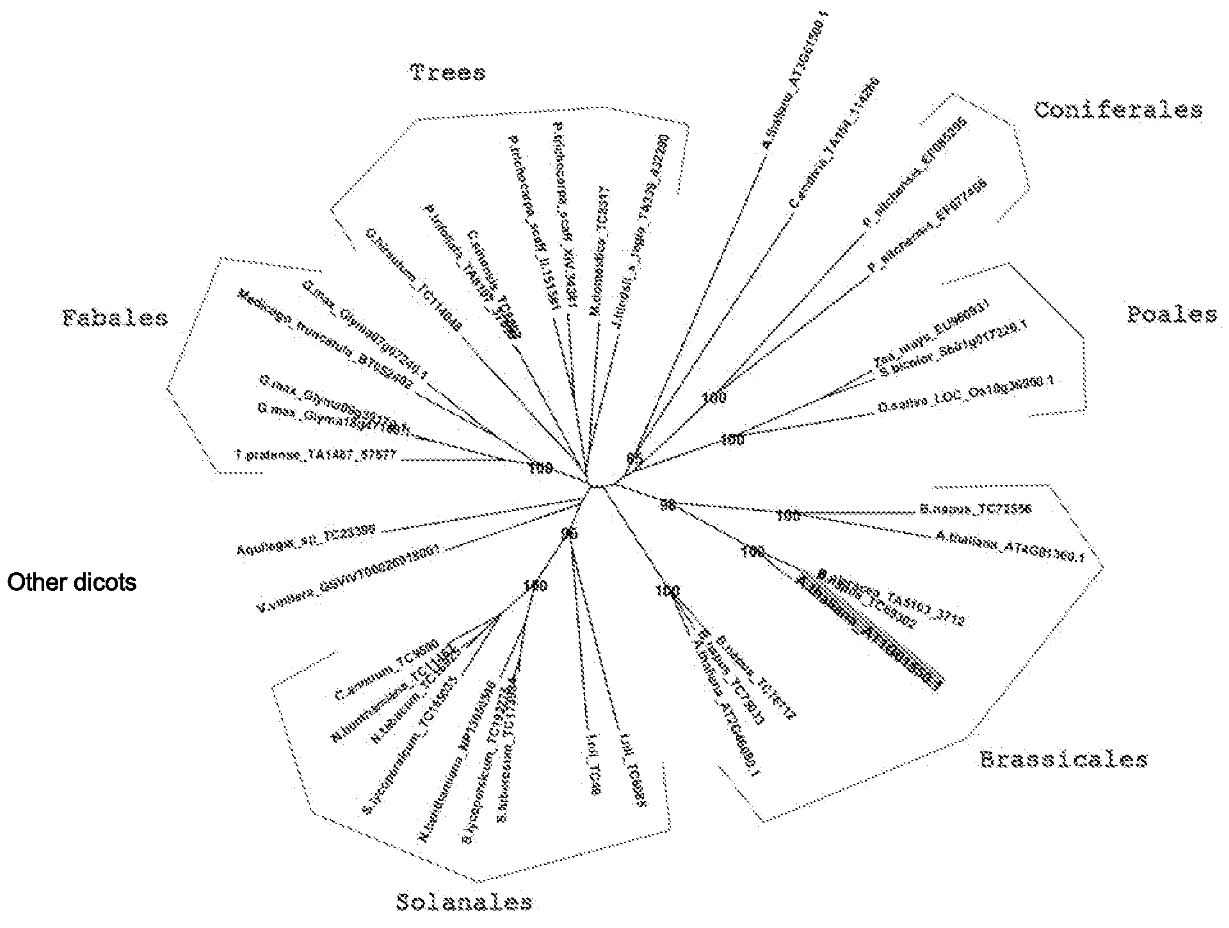
Trees
Coniferales
Fabales
Poales
Other dicots
Brassicales
Solanales

FIGURE 6

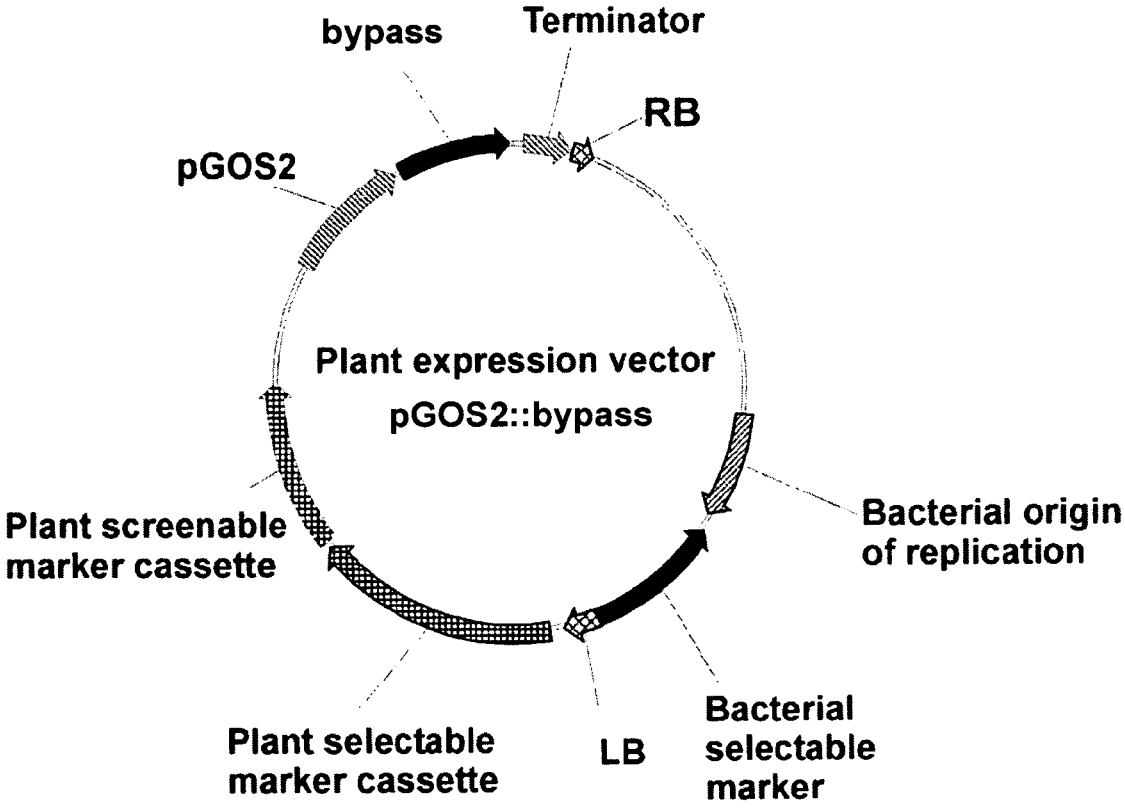
bypass
Terminator
RB
pGOS2
Plant expression vector
pGOS2::bypass
Plant screenable
marker cassette
Bacterial origin
of replication
Plant selectable
marker cassette
LB
Bacterial
selectable
marker

FIGURE 7

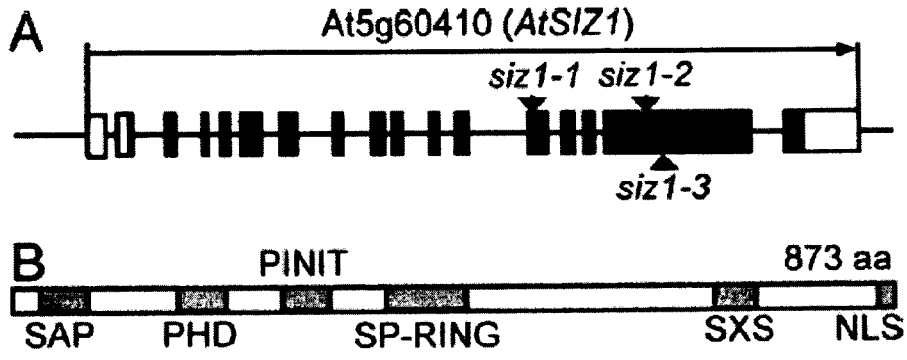
A
At5g60410 (AtSIZ1)
siz1-1 siz1-2
siz1-3
B
PINIT
873 aa
SAP
PHD
SP-RING
SXS
NLS

FIGURE 8

```
CLUSTAL 2.0.11 multiple sequence alignment

arabidospsisECsequence          ------------------------------------------------
A.thaliana_AT5G60410.5#1        ------------------------------------------------
Athaliana                       ------------------------------------------------
P.trichocarpa_scaff_66.246#1    ------------------------------------------------
P.trichocarpa_scaff_IX.1493#1   ------------------------------------------------
gi_255570825_ref_XP_002526365.  ------------------------------------------------
gi_225435251_ref_XP_002284945.  ------------------------------------------------
M.truncatula_AC152176_5.4#1     ------------------------------------------------
AK244805Soybean                 ------------------------------------------------
M.truncatula_AC150891_19.5#1    ------------------------------------------------
gi_223944535_gb_ACN26351.1_     ------------------------------------------------
sorghum                         ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0  ------------------------------------------------
gi_226493325_ref_NP_001140473.  ------------------------------------------------
H.vulgare_TA46195_4513#1        ------------------------------------------------
O.sativa_Os05g0125000#1         ------------------------------------------------
P.trichocarpa_scaff_VIII.362#1  ------------------------------------------------
P.trichocarpa_scaff_X.2133#1    ------------------------------------------------
S.bicolor_5288383#1             ------------------------------------------------
S.bicolor_5285414#1             ------------------------------------------------
O.sativa_Os03g0719100#1         ------------------------------------------------
P.patens_159214#1               ------------------------------------------------
P.patens_165698#1               MDTLKVTEIFSSIYQSIDSSKHTSIHSYIHSWGESEKCIDGGGGYSWMWG
P.patens_159935#1               ------------------------------------------------
AT5G41580"                      ------------------------------------------------
AT1G08910.1EMB3001protein       ------------------------------------------------
C.vulgaris_83729#1              ----------MPALLPLLRSVCLGLLPLGLSNICASFPTLCCELSGGSW


arabidospsisECsequence          ---------MDLEANCKEKLSYFRIKELKDVLTQLGLS------------
A.thaliana_AT5G60410.5#1        ---------MDLEANCKEKLSYFRIKELKDVLTQLGLS------------
Athaliana                       ---------MDLEANCKEKLSYFRIKELKDVLTQLGLS------------
P.trichocarpa_scaff_66.246#1    ---------MDLVASCKDKLAFFRIKELKDVLTQLGLS------------
P.trichocarpa_scaff_IX.1493#1   -------------------------------MFQLMVQ------------
gi_255570825_ref_XP_002526365.  ---------MDLVTSCKDKLAYFRIKELKDVLTQLGLS------------
gi_225435251_ref_XP_002284945.  ---------MDLVTSCKDKLAYFRIKELKDVLTQLGLS------------
M.truncatula_AC152176_5.4#1     ---------MDLVAGIKEKLTYFRIKELKDVLTQLGLS------------
AK244805Soybean                 ---------MDLVPSVKEKLNYFRIKELKDVLTQLHLS------------
M.truncatula_AC150891_19.5#1    --------MDDLVSSCKEKLQYFRVKDLKDVLTQIGIS------------
gi_223944535_gb_ACN26351.1_     ------------------------------------------------
sorghum                         ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0  ------------------------------------------------
gi_226493325_ref_NP_001140473.  ------------------------------------------------
H.vulgare_TA46195_4513#1        --------MADLASTCKDKLAYFRIKELKDILHQLGLP------------
O.sativa_Os05g0125000#1         --------MADLVSSCKDKLAYFRIKELKDILNQLGLP------------
P.trichocarpa_scaff_VIII.362#1  -------MEMDSMASCKEKLAYFRIKELKDILSLLGLS------------
P.trichocarpa_scaff_X.2133#1    -------MEMNLVASCKGKLAYFRIKELKDILSLLGLS------------
S.bicolor_5288383#1             ------------------------------------------------
S.bicolor_5285414#1             ------------------------------------------------
O.sativa_Os03g0719100#1         --MALDPADDPLLADCKYKLNHFRIKELKDVLHQLGLP------------
P.patens_159214#1               --------MADIATRCRIQLSSFQIKHLKDTLTRLGLS------------
P.patens_165698#1               TSHSHEGSFCNHHTNSRTQFMRWRRKDLVTNEPVLQLHPLLQPGTLSPCE
P.patens_159935#1               --------MADVASRCQSQLGSFRIRELKDVLARLGLP------------
AT5G41580"                      ------------------------------------------------
AT1G08910.1EMB3001protein       ------------------------------------------------
C.vulgaris_83729#1              TRAGSSSAQIMASATLRRYIFAFRVAELQLCLQELGLS------------
```

FIGURE 9

```
arabidospsisECsequence                ----------------------------------------------
A.thaliana_AT5G60410.5#1               ----------------------------------------------
Athaliana                              ----------------------------------------------
P.trichocarpa_scaff_66.246#1           ----------------------------------------------
P.trichocarpa_scaff_IX.1493#1          ----------------------------------------------
gi_255570825_ref_XP_002526365.         ----------------------------------------------
gi_225435251_ref_XP_002284945.         ----------------------------------------------
M.truncatula_AC152176_5.4#1            ----------------------------------------------
AK244805Soybean                        ----------------------------------------------
M.truncatula_AC150891_19.5#1           ----------------------------------------------
gi_223944535_gb_ACN26351.1_            ----------------------------------------------
sorghum                                ----------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0         ----------------------------------------------
gi_226493325_ref_NP_001140473.         ----------------------------------------------
H.vulgare_TA46195_4513#1               ----------------------------------------------
O.sativa_Os05g0125000#1                ----------------------------------------------
P.trichocarpa_scaff_VIII.362#1         ----------------------------------------------
P.trichocarpa_scaff_X.2133#1           ----------------------------------------------
S.bicolor_5288383#1                    ----------------------------------------------
S.bicolor_5285414#1                    ----------------------------------------------
O.sativa_Os03g0719100#1                ----------------------------------------------
P.patens_159214#1                      -KRGTKPLIM------------------------------------
P.patens_165698#1                      RQREHTKLIENVPGTPRAQWERRIMRTFEALHGMGRGLAISPRLMLVWIL
P.patens_159935#1                      -KQGKKQILIDKIMDLINPAEKESLAKG-----------------
AT5G41580"                             ----------------------------------------------
AT1G08910.1EMB3001protein              ----------------------------------------------
C.vulgaris_83729#1                     ----------------------------------------------

arabidospsisECsequence                ----------------------------------------------
A.thaliana_AT5G60410.5#1               ----------------------------------------------
Athaliana                              ----------------------------------------------
P.trichocarpa_scaff_66.246#1           ----------------------------------------------
P.trichocarpa_scaff_IX.1493#1          ----------------------------------------------
gi_255570825_ref_XP_002526365.         ----------------------------------------------
gi_225435251_ref_XP_002284945.         ----------------------------------------------
M.truncatula_AC152176_5.4#1            ----------------------------------------------
AK244805Soybean                        ----------------------------------------------
M.truncatula_AC150891_19.5#1           ----------------------------------------------
gi_223944535_gb_ACN26351.1_            ----------------------------------------------
sorghum                                ----------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0         ----------------------------------------------
gi_226493325_ref_NP_001140473.         ----------------------------------------------
H.vulgare_TA46195_4513#1               ----------------------------------------------
O.sativa_Os05g0125000#1                ----------------------------------------------
P.trichocarpa_scaff_VIII.362#1         ----------------------------------------------
P.trichocarpa_scaff_X.2133#1           ----------------------------------------------
S.bicolor_5288383#1                    ----------------------------------------------
S.bicolor_5285414#1                    ----------------------------------------------
O.sativa_Os03g0719100#1                ----------------------------------------------
P.patens_159214#1                      ----------------------------------------------
P.patens_165698#1                      FILWGPVSSGDCGSSGSGVHCGALAAGERSFWEYAPGLLDVLEGLVGSIR
P.patens_159935#1                      ----------------------------------------------
AT5G41580"                             ----------------------------------------------
AT1G08910.1EMB3001protein              ----------------------------------------------
C.vulgaris_83729#1                     ----------------------------------------------
```

FIGURE 9 (continued)

```
arabidospsisECsequence                ----------------------------------------------
A.thaliana_AT5G60410.5#1               ----------------------------------------------
Athaliana                              ----------------------------------------------
P.trichocarpa_scaff_66.246#1           ----------------------------------------------
P.trichocarpa_scaff_IX.1493#1          ----------------------------------------------
gi_255570825_ref_XP_002526365.         ----------------------------------------------
gi_225435251_ref_XP_002284945.         ----------------------------------------------
M.truncatula_AC152176_5.4#1            ----------------------------------------------
AK244805Soybean                        ----------------------------------------------
M.truncatula_AC150891_19.5#1           ----------------------------------------------
gi_223944535_gb_ACN26351.1_            ----------------------------------------------
sorghum                                ----------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0         ----------------------------------------------
gi_226493325_ref_NP_001140473.         ----------------------------------------------
H.vulgare_TA46195_4513#1               ----------------------------------------------
O.sativa_Os05g0125000#1                ----------------------------------------------
P.trichocarpa_scaff_VIII.362#1         ----------------------------------------------
P.trichocarpa_scaff_X.2133#1           ----------------------------------------------
S.bicolor_5288383#1                    ----------------------------------------------
S.bicolor_5285414#1                    ----------------------------------------------
O.sativa_Os03g0719100#1                ----------------------------------------------
P.patens_159214#1                      -------PHFCNVKWLKIHVVALRPLYSSLHLDVSKFD------------
P.patens_165698#1                      RHALCCDANFNEIEFAPIILLVTQRNRKSLPVCNQSFTNLHPHCRIAIFL
P.patens_159935#1                      -----IGLFILELELAIYVALAGSKSSKKAISGDEAIAIVDEQYS-----
AT5G41580"                             ----------------------------------------------
AT1G08910.1EMB3001protein              ----------------------------------------------
C.vulgaris_83729#1                     ----------------------------------------------

arabidospsisECsequence                ----------------------------------KQGKKQELVDRILTLLS
A.thaliana_AT5G60410.5#1               ----------------------------------KQGKKQELVDRILTLLS
Athaliana                              ----------------------------------KQGKKQELVDRILTLLS
P.trichocarpa_scaff_66.246#1           ----------------------------------KQGKKQDLVDRILAILS
P.trichocarpa_scaff_IX.1493#1          ----------------------------------AQRR------------
gi_255570825_ref_XP_002526365.         ----------------------------------KQGKKQDLVDRILAVLT
gi_225435251_ref_XP_002284945.         ----------------------------------KQGKKQDLVDRILAILS
M.truncatula_AC152176_5.4#1            ----------------------------------KQGKKQDLVDRILSILS
AK244805Soybean                        ----------------------------------KQGKKQDLVDRILSVLS
M.truncatula_AC150891_19.5#1           ----------------------------------KQGKKQDLIDRILSIIS
gi_223944535_gb_ACN26351.1_            ----------------------------------------------
sorghum                                ----------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0         ----------------------------------------------
gi_226493325_ref_NP_001140473.         ----------------------------------------------
H.vulgare_TA46195_4513#1               ----------------------------------KQGKKQDLIDRVLALLS
O.sativa_Os05g0125000#1                ----------------------------------KQGKKQDLIDRVLALLT
P.trichocarpa_scaff_VIII.362#1         ----------------------------------KQGKKQDLMDRVIGLLS
P.trichocarpa_scaff_X.2133#1           ----------------------------------KQGKKQDLMDRVLGLLS
S.bicolor_5288383#1                    ----------------------------------------------
S.bicolor_5285414#1                    ----------------------------------------------
O.sativa_Os03g0719100#1                ----------------------------------KQGRKQELVDKIIAVLS
P.patens_159214#1                      ------------------------------------GVMLMDKIMGLIN
P.patens_165698#1                      LLAMADVATRCRSQLGSFRIRELKDVLSRLGLPKQGKKQILMDKIMGLIN
P.patens_159935#1                      ----------------------------------VFRRFVMKVYIQQLF
AT5G41580"                             ----------------------------------------------
AT1G08910.1EMB3001protein              ----------------------------------------------
C.vulgaris_83729#1                     ----------------------------------KKGLKGELQSRLFAYFG
```

FIGURE 9 (continued)

```
arabidospsisECsequence                DEQA----ARLLSKKNTVAKEAVAKLVDDTYRKMQVSGASDLASKGQVSS
A.thaliana_AT5G60410.5#1              DEQA----ARLLSKKNTVAKEAVAKLVDDTYRKMQVSGASDLASKGQVSS
Athaliana                             DEQA----ARLLSKKNTVAKEAVAKLVDDTYRKMQVSGASDLASKGQVSS
P.trichocarpa_scaff_66.246#1          DEQV----SKLWAKKSAIGKEEVAKLVDDTYRKMQVSGATDLASRGQVAS
P.trichocarpa_scaff_IX.1493#1         ----------------------------VWRKHFIS-------------
gi_255570825_ref_XP_002526365.        DEQV----PKTSAKKSVVGKEEVAKLVDDIYRKMQVSGATDLASKGEGVL
gi_225435251_ref_XP_002284945.        DEQV----SRMWAKKNAVGKEEVAKLVEDTYRKMQVSGATDLASKGQVLS
M.truncatula_AC152176_5.4#1           DEQV----SKIWAKKNAVGKEQVAKLVDDTYRKMQISGATDLASKGQVVS
AK244805Soybean                       DEQV----SKMWAKKNAGGKEQVAKLVDDTYRKMQISGATDLASKGQGAS
M.truncatula_AC150891_19.5#1          DEQV----AKVRAKKNAVEKEQVVKLVEDTYRKLQVSGATDIASKGQVAS
gi_223944535_gb_ACN26351.1_           ------------------------------------------------
sorghum                               ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0        --------HHGWGRKNALTREAVAKVVDDTYSRKMQVCAPDLPSRSHSGS
gi_226493325_ref_NP_001140473.        ------------------------------------------------
H.vulgare_TA46195_4513#1              DEQGQ--RHHGWGRKNSFTKEAVAKIVDDIYRKMQIQSAPDLATRSHTGS
O.sativa_Os05g0125000#1               DEQGQ--RHHGWGRKNSLTKEAVAKIVDDTYRKMQIQCAPDLATRSHSGS
P.trichocarpa_scaff_VIII.362#1        DDEIC--SARRLARKKQIGKEAVVKIIDDAYRKMQIMDASDLAAKAPSGL
P.trichocarpa_scaff_X.2133#1          DDEIC--SARSFVRKQQIGKEAVVKIIDDAYRKMHITDASDLAVGAPSGF
S.bicolor_5288383#1                   ----------------------MVLKIVEDTFRKMQEATNTVTPSRSHGS-
S.bicolor_5285414#1                   ------------------------------------------------
O.sativa_Os03g0719100#1               DQQEQDSRLNGLPNKKMVGKETVAKIVDDTFAKMNGSTNAVPASRNQTDS
P.patens_159214#1                     PAEKQ-PTKGLKYSKQIVSREEAIAIIEEEYRKLRHSRSE-SKRKAAKSG
P.patens_165698#1                     PADKQSLTKGSKSSKKVVSREEAIAIIDEQYRKLRHSGTE-SKHKVAKSG
P.patens_159935#1                     LRESRVVTLSGNILELFQG--NIVTNTSQRSTKLRSIGTDLSRHKSAKSA
AT5G41580"                            ------------------MSTAAAARPVAGTGLREKTAASLVNSFRLASVT
AT1G08910.1EMB3001protein             ------------------MVIPATSR--FGFRAEFNTKEFQASCISLANFT
C.vulgaris_83729#1                    DYTG-VAARGVNPPKEQHRLDTAARLVTQIYHRMKGLPSPEALPARETIP


arabidospsisECsequence                DTSNLKVKGEPEDPFQPE------------------IKVRCVCGNSLETD
A.thaliana_AT5G60410.5#1              DTSNLKVKGEPEDPFQPE------------------IKVRCVCGNSLETD
Athaliana                             DTSNLKVKGEPEDPFQPE------------------IKVRCVCGNSLETD
P.trichocarpa_scaff_66.246#1          DCSNSKFNGEMDDPSHSD------------------TKVRCPCGSSLETE
P.trichocarpa_scaff_IX.1493#1         ------------------------------------------------
gi_255570825_ref_XP_002526365.        ESSKPVIKGEIDDSFHFD------------------TKVRCPCGSSLETE
gi_225435251_ref_XP_002284945.        DSSNVKFKEELEDSYN-D------------------MKIRCPCGSALPNE
M.truncatula_AC152176_5.4#1           DSSNVKVKAEVEDSFQIQTT--------------TTTKIRCLCGSTLETG
AK244805Soybean                       DSSSVKVKSEFDDAFQRD-------------------VKIRCLCGSRLETE
M.truncatula_AC150891_19.5#1          DSSNVKIKGEVEDSVQSA------------------TKVRCLCGSSLETD
gi_223944535_gb_ACN26351.1_           ------------------------------------------------
sorghum                               ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0        DFSHFRPKEEAPDFYHVD------------------TKVRCLCNSTLLND
gi_226493325_ref_NP_001140473.        ------------------------------------------------
H.vulgare_TA46195_4513#1              DL--FRPKDEVNDSFQPQP----------------VTKVRCICDSKLLND
O.sativa_Os05g0125000#1               DFS-FRPIEEAYDSFQP------------------EAKVRCICSSTMVND
P.trichocarpa_scaff_VIII.362#1        DI--TSVTEEVEDFITP------------------GKTIRCPCGSSLPTE
P.trichocarpa_scaff_X.2133#1          HT--MSVKEEVEDFISP------------------EKRIRCPCGSSLPTE
S.bicolor_5288383#1                   ---VKPKKKPE--SAQAVK-------------------VRCPCGDSKPND
S.bicolor_5285414#1                   ------------------------------------------------
O.sativa_Os03g0719100#1               GHIVKPKRKSDDSAQLDVK-------------------VRCPCGYSMAND
P.patens_159214#1                     FGSKHPSAGLEELTCVEEA------------------KTRCPCGNNTETG
P.patens_165698#1                     SSSGYPSAGPEDHEVVEET------------------RTRCPCGSNVETG
P.patens_159935#1                     SSSGYPSPGLD--------------------------ESRCPCGSSVEAG
AT5G41580"                            QRLRYHIQDGAKVDPKEFQI--------------CCISFAKGIDFAIANN
AT1G08910.1EMB3001protein             ENFTPGFGECSEIDA-----------------------------AIGRN
C.vulgaris_83729#1                    TAGYLQGSGDVPLAPAAPILPNGNAAAAVQAAARSNTQIRCICGSNYDRG
```

FIGURE 9 (continued)

```
arabidospsisECsequence                SMIQCEDPRCHVWQHVGCVILP--DKPMDGN-PPLPESFYCEICRLTRAD
A.thaliana_AT5G60410.5#1              SMIQCEDPRCHVWQHVGCVILP--DKPMDGN-PPLPESFYCEICRLTRAD
Athaliana                             SMIQCEDPRCHVWQHVGCVILP--DKPMDGN-PPLPESFYCEICRLTRAD
P.trichocarpa_scaff_66.246#1          SMIKCEDFKCHVWQHIGCVIIP--EKPMEGI-PQVPDVFYCEICRLSRAD
P.trichocarpa_scaff_IX.1493#1         ----------------------------------------LEQTR
gi_255570825_ref_XP_002526365.        SMIKCEDPRCRVWQHIGCVIIP--EKPMEAI-PQVPDLFYCEICRLCRAD
gi_225435251_ref_XP_002284945.        TMLKCDDLKCQVWQHIGCVIIP--EKTMEGI-PPTPDPFYCEICRLSRAD
M.truncatula_AC152176_5.4#1           DLIKCDDARCQVWQHISCVIIP--EKPMEGI-PPVPDKFYCELCRLSRAD
AK244805Soybean                       DLVKCDDPRCHVWQHISCVIIP--EKPTEGI-PPVPDKFYCELCRLTRAD
M.truncatula_AC150891_19.5#1          LLIKCEDRKCPVSQHLNCVIIP--DTPTEGL-PPIPDTFYCEICRLSRAD
gi_223944535_gb_ACN26351.1_           ------------------------------------------------
sorghum                               ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0        NMIKCEDGKCQVWQHITCVLIP--DKPTEGAGPDIPPHFYCELCRLKRAD
gi_226493325_ref_NP_001140473.        ------------------------------------------------
H.vulgare_TA46195_4513#1              NMIQCEDDQCHVWQHMSCVLVP--DKPTEGVGPEVPPHFYCELCRLSRAD
O.sativa_Os05g0125000#1               SMIQCEDQRCQVWQHLNCVLIP--DKPGE--SAEVPPVFYCELCRLSRAD
P.trichocarpa_scaff_VIII.362#1        FMIQCIDSKCQVQQHISCVIFT--EIPME-SEH--PPVFYCETCRIERAD
P.trichocarpa_scaff_X.2133#1          FMIQCIDSKCQVQQHISCVIFL--ENPVE-SDHPIPPVFYCETCRIDRAD
S.bicolor_5288383#1                   SMIKCIDPQCNMWQHVGCVIIPDAEKSADNISPELPSCFYCEVCRLSRAD
S.bicolor_5285414#1                   ------------------------------------------------
O.sativa_Os03g0719100#1               SMIKCEGPQCNTQQHVGCVIIS--EKPADSVPPELPPHFYCDMCRITRAD
P.patens_159214#1                     TMIQCVNLKCRVWQHMSCVVIP--EKSGDGTQTGIPSNFYCELCRISRSD
P.patens_165698#1                     TMIQCDNNKCKIWQHMDCVVIP--EKPSDGTQPEIPSSFYCELCRIARGD
P.patens_159935#1                     RMIQCDSHGCRIWQHRSCVDFP--KKPKDGVPVETPPNFYCELCRISQGD
AT5G41580"                            DIPKKVEEFPWLLKQLCRHGTDVY-TKTALMVLMISVKHACHLGWFSDSE
AT1G08910.1EMB3001protein             EVPGNIQELALILNNVCRRKCDDYQTRAVVMALMISVKSACQLGWFPERE
C.vulgaris_83729#1                    TMIQCEDEACGVWQHCDCVGVD---------LNVMPEHYLCELCRLARAD

arabidospsisECsequence                P----------------FWVTVAHPLSPVRLTATTIPNDGASTMQSVERT
A.thaliana_AT5G60410.5#1              P----------------FWVTVAHPLSPVRLTATTIPNDGASTMQSVERT
Athaliana                             P----------------FWVTVAHPLSPVRLTATTIPNDGASTMQSVERT
P.trichocarpa_scaff_66.246#1          P----------------FWVTVAHPLSPVKLVATNVPADGSRPVQGVEKT
P.trichocarpa_scaff_IX.1493#1         I----------------YWQ--------------------------------
gi_255570825_ref_XP_002526365.        P----------------FWVSVAHPLYPVKLTTN-IQADGSTPVQSAEKT
gi_225435251_ref_XP_002284945.        P----------------FWVTVAHPLLPVKLTTTSIPTDGTNPVQSVEKT
M.truncatula_AC152176_5.4#1           P----------------FWVSVSHPLLPVKLATTSIPTDGTNPVQCVERT
AK244805Soybean                       P----------------FWVSVAHPLHPVKLTTTSNPTDGNNPVQSVERT
M.truncatula_AC150891_19.5#1          P----------------FSVSMMHPLHPVKLSTTLVPTEGSNPMQSVEKT
gi_223944535_gb_ACN26351.1_           ------------------------------------------------
sorghum                               -------------------------------MSSGVGNDGASVPQIVEKT
Z.mays_ZM07MSbpsHQ_62031266.f0        P----------------FWVTTGNPLLPVKFMSSGVGNDGASVPQIVEKT
gi_226493325_ref_NP_001140473.        ------------------------------------------------
H.vulgare_TA46195_4513#1              P----------------FWVTTGNPLPPLKFMSSGVANDGTSVLQTVEKT
O.sativa_Os05g0125000#1               P----------------FWVTAGNPLLPVKFVSSGVTNDGTSVPQSVEKS
P.trichocarpa_scaff_VIII.362#1        P----------------FWVTVAHLLFPVKLPSSNISIDGNNTLQNVETT
P.trichocarpa_scaff_X.2133#1          PFANKGFYYINKLLVDSFWVTVAHLSLPVKLTSSNISMDGNITSQKVETT
S.bicolor_5288383#1                   P----------------FWVTMHHLLLPVLIGPSTVAADG-----------
S.bicolor_5285414#1                   ------------------------------------------------
O.sativa_Os03g0719100#1               P----------------FWVTVNHPVLPVSITPCKVASDGSYAVQYFEKT
P.patens_159214#1                     P----------------FCEAQLQTLMPSKLIPSGANTEGSNIVQTLEKS
P.patens_165698#1                     P----------------FCEAQTHTLMPTKLLSSTAKTEGSNTLQTIEKS
P.patens_159935#1                     P----------------FCEALFHPLLPVKFPSSTAKSERAITLQSIDEQ
AT5G41580"                            SQE-----------LIALADEIRTCFGSSGSTSPGIKSPGSTFSQIMERF
AT1G08910.1EMB3001protein             TQE-----------LLAIIDLMWNGFSCPENVTSCVNSPVTLISQVIERF
C.vulgaris_83729#1                    PFWRRVG------APVMSPVKLAPVQPPRSFPDGRTQEEDVVQVADRNFM
```

FIGURE 9 (continued)

```
arabidospsisECsequence                FQITRADKDLLAKPEYDVQAWCMLLNDKVLFRMQWPQYADLQVNGVPVRA
A.thaliana_AT5G60410.5#1               FQITRADKDLLAKPEYDVQAWCMLLNDKVLFRMQWPQYADLQVNGVPVRA
Athaliana                             FQITRADKDLLAKPEYDVQAWCMLLNDKVLFRMQWPQYADLQVNGVPVRA
P.trichocarpa_scaff_66.246#1          FQLTRADKDLLAKQEYDVQAWCMLLNDKVPFRMQWPQDTDLQVNGLAVRA
P.trichocarpa_scaff_IX.1493#1         ------------------------NKNMIFRMQWPQYADLQVNGIAVRA
gi_255570825_ref_XP_002526365.        FHLTRADKDLLAKQEYDVQAWCMLLNDKVPFRMQWPQYADLQVNGVPVRA
gi_225435251_ref_XP_002284945.        FHLTRADRDMVSKHEYDVQAWCILLNDKVSFRMQWPQYADLQVNGMAVRA
M.truncatula_AC152176_5.4#1           FQLTRADKDMVSKQEFDVEAWCMLLNDKVPFRIQWPQYTDLAVNGLPIRT
AK244805Soybean                       FQLTRADMDLVSKPEFDVEAWCMLLNDKVPFRMQWPQYTDLQVNGVPVRA
M.truncatula_AC150891_19.5#1          FQLARAHKDIVLKSEFDIQAWCMLLNDKVPFRMQWPQYADLVVNGYSVRA
gi_223944535_gb_ACN26351.1_           ----------------------------------------------------
sorghum                               FQLSRADRETVQRPEYDLQVWCILINDKVQFRMQWPQYAELQVNGIPVRV
Z.mays_ZM07MSbpsHQ_62031266.f0        FQLSRADRETVQRQEYDLQVWCILINDKVQFRMQWPQYAELQVNGIPVRV
gi_226493325_ref_NP_001140473.        ----------------------------------------------------
H.vulgare_TA46195_4513#1              FQLSRSGG-TVQRSEYDLQV------------------------------
O.sativa_Os05g0125000#1               FQLSRSDRETVQRQEYDLQVWCMLLNDKVQFRMQWPQYAELHVNGISVRV
P.trichocarpa_scaff_VIII.362#1        FQLTRSDQHLLKNCEYDAQAWCMLLNDKVLFRMQWPLYADLQVNGMPVKT
P.trichocarpa_scaff_X.2133#1          FQLTRSDQHLLQNCEYDVQAWCMLLNDNVLFRMQWPLFANLQVNDMSVRM
S.bicolor_5288383#1                   --------------------------------MQWPLHSDMQVNGIYVRV
S.bicolor_5285414#1                   ----------------------------------------------------
O.sativa_Os03g0719100#1               FPLSRANWEMLQKDEYDLQVWCILFNDSVPFRMQWPLHSDIQINGIPIRV
P.patens_159214#1                     FYLSRSDRELLQKPNHDLQVWCVLLSDNVSFRMHWPSFADLRVNGIGVRV
P.patens_165698#1                     FFLSRADRELLQKLNYDLQVWCVLLSDKVSFRMHWPSYADLRVNGISVRV
P.patens_159935#1                     FTLSLAHQELLQSPNYDLQVWCVLLSDKVSFRMHWPLSAVLRVNDANVRV
AT5G41580"                            YPFVKLGHVLVS---FEVKAGYTMLAHDFYISKNMPHSLQEKIRLFVAQT
AT1G08910.1EMB3001protein             YPCVKLGHILVS---FEAKPESKMMMKDFHISKKMPHSPKQKVGLFVVRT
C.vulgaris_83729#1                    LTHAQIDPARRQSHNFQLQVACIMMGDSVPMRIHWPRHADLRLNNMLYRP


arabidospsisECsequence                INRPGGQLLGVNGRDDGP-------------------IITSCIRDGVNRI
A.thaliana_AT5G60410.5#1              INRPGGQLLGVNGRDDGP-------------------IITSCIRDGVNRI
Athaliana                             INRPGGQLLGVNGRDDGP-------------------IITSCIRDGVNRI
P.trichocarpa_scaff_66.246#1          INRPGSQLLGANGRDDGP-------------------IVTPFVKDGINKI
P.trichocarpa_scaff_IX.1493#1         INRPGSQLLGANGRDDGP-------------------IITSCAKDGINKI
gi_255570825_ref_XP_002526365.        INRPGSQLLGINGRDDGP-------------------IITPCTKDGINKI
gi_225435251_ref_XP_002284945.        INRPGSQLLGANGRDDGP-------------------VITPCTKDGINKI
M.truncatula_AC152176_5.4#1           TTRPGSQLLGANGRDDGP-------------------IITPHTKDGINKI
AK244805Soybean                       TNRPGSQLLGANGRDDGP-------------------IITPYTKDGINKI
M.truncatula_AC150891_19.5#1          INRPGSQLLGANGRDDGP-------------------IITPYIKEGVNKI
gi_223944535_gb_ACN26351.1_           ----------------------------------------------------
sorghum                               MTRPGSQLLGINGRDDGP-------------------LVTTCSREGINKI
Z.mays_ZM07MSbpsHQ_62031266.f0        MTRPGSQLLGINGRDDGP-------------------LVTTCSREGINKI
gi_226493325_ref_NP_001140473.        ----------------------------------------------------
H.vulgare_TA46195_4513#1              ----------------------------------------------------
O.sativa_Os05g0125000#1               VTRPGSQLLGINGRDDGP-------------------LITTCSREGINKI
P.trichocarpa_scaff_VIII.362#1        LNRPGSQSLGASGRDDGA-------------------LIKSCIGEGINRV
P.trichocarpa_scaff_X.2133#1          LDRLVSQSLGANGRDDGAQVSPFPSLLSYEMINILLGIIKLCIREGINRI
S.bicolor_5288383#1                   VNRQPHQKLGANGRDDGP-------------------LLTDYLKEGPNKI
S.bicolor_5285414#1                   ----------------------------------------------------
O.sativa_Os03g0719100#1               VNRQPTQQLGVNGRDDGP-------------------VLTAYVREGSNKI
P.patens_159214#1                     TNRTGQQLLGANGRDEGT-------------------SVTVCAREGLNRL
P.patens_165698#1                     TNRPGQQLLGANGRDEGP-------------------GITVCAREGMNRL
P.patens_159935#1                     TNRPAEQPLGANSRDEGH-------------------SITSYTREGLNRL
AT5G41580"                            DNIDTSACISNPPE-----------------------VSFLLNGKGVEKR
AT1G08910.1EMB3001protein             EDISRSNCIVHPQG-----------------------VSFLLNGKGIDKR
C.vulgaris_83729#1                    YSRNSATKLGANARDEPAS------------------VGVMCSQGRNRL
```

FIGURE 9 (continued)

```
arabidospsisECsequence              SLSGGDVRIFCFGVRLVKRRTLQQVLNLIPEEGKGETFEDALARVRRCIG
A.thaliana_AT5G60410.5#1             SLSGGDVRIFCFGVRLVKRRTLQQVLNLIPEEGKGETFEDALARVRRCIG
Athaliana                           SLSGGDVRIFCFGVRLVKRRTLQQVLNLIPEEGKGETFEDALARVRRCIG
P.trichocarpa_scaff_66.246#1        LLSGCDARIFCLGVRIVKRRTVQ-------QDSEGERFEDALARVCRCVG
P.trichocarpa_scaff_IX.1493#1       SLTGCDARIFCLGVRIVKRRTVQ-------QVS-----------------
gi_255570825_ref_XP_002526365.      SLNGCDARIFCLGVRIVKRRTVQQILNMIPKESDGERFEDALARVCRCVG
gi_225435251_ref_XP_002284945.      SLTGCDARIFCLGVRIVKRRTVQQILSLIPKESDGERFEDALARVRRCIG
M.truncatula_AC152176_5.4#1         SLTVCDARIFCLGVRIVRRRSLQQILNLIPKESDGEPFEDALARVCRCVG
AK244805Soybean                     SLTGCDARIFCLGVRIVKRRSMQQILNSIPKESDGEKFEEALARVCRCVG
M.truncatula_AC150891_19.5#1        SLTGCDTRIFCLGVRIVRRRTLQQILNMIPKESDGERFEVALARVCCRVG
gi_223944535_gb_ACN26351.1_         ----------------------------------------------------
sorghum                             SLSRVDARTFCFGVRIVRRRTVPQVLNLIPKEGEGESFEDALARVRRCLG
Z.mays_ZM07MSbpsHQ_62031266.f0      SLSRVDARTFCFGVRIVRRRTVPQVLNLIPKEGEGESFEDALARVRRCLG
gi_226493325_ref_NP_001140473.      ----------------------------------------------------
H.vulgare_TA46195_4513#1            ----------------------------------------------------
O.sativa_Os05g0125000#1             CLSRVDARTFCFGVRIAKRRTVAQVLNLVPKEAEGESFEHALARVRRCLG
P.trichocarpa_scaff_VIII.362#1      SLSGCDSRAFCFGIRLVKRQTVEQVLNLIPKD--GEPFEDALARVCRCIG
P.trichocarpa_scaff_X.2133#1        SLSGCDSRVFCFGIRLVKRRTVEQVLNLIPKV--GESFEDALARVCRCIG
S.bicolor_5288383#1                 SLSRNDSRTFCLGIRIAKRRSLE-------QEQDGEKFDDALARVRRCVG
S.bicolor_5285414#1                 ----------------------------------------------------
O.sativa_Os03g0719100#1             VLSRSDSRTFCLGVRIAKRRSVEQVLSLVPKEQDGENFDNALARVRRCVG
P.patens_159214#1                   NISTYDARSFCLGVRIIRRLSLEQIMESIPNEKDGEKLEEAMARVRRCIN
P.patens_165698#1                   NMSAYDARPFCLGVRIIRRLTLEQVKDLIPNEKDGEPFEEAMARVRRCIN
P.patens_159935#1                   NMSCDDARPFCLGVRIIRRRSLEEVMDMIPNEKDGEPFDEAVARVRRCIN
AT5G41580"                          VNIAMDTGPQLPTNVTAQLKYGTNLLQVMGNFKGNYIIIIAFTGLVVPPE
AT1G08910.1EMB3001protein           VNISMESGPQLPTNVTALLNLGANLLQAIGCFGGSYLIAIAFMDVIPLPN
C.vulgaris_83729#1                  WVSVMESRSFCVMVQLAQRRTMDEVKALMAPPETEQAALKRVVRQTRGVK

arabidospsisECsequence              GGGGDDN-----------------------------------------
A.thaliana_AT5G60410.5#1             GGGGDDN-----------------------------------------
Athaliana                           GGGGDDN-----------------------------------------
P.trichocarpa_scaff_66.246#1        GGTATDNADSDSDLEVVADSFGVNLRCPDKLAFFRIKELKDVLTQLGLSK
P.trichocarpa_scaff_IX.1493#1       ----------------------------------------------------
gi_255570825_ref_XP_002526365.      GGAADN------------------------------------------
gi_225435251_ref_XP_002284945.      GGGATDN-----------------------------------------
M.truncatula_AC152176_5.4#1         GGNAADN-----------------------------------------
AK244805Soybean                     GGNAADD-----------------------------------------
M.truncatula_AC150891_19.5#1        GGNSADD-----------------------------------------
gi_223944535_gb_ACN26351.1_         ----------------------------------------------------
sorghum                             GGGATDN-----------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0      GGGATDN-----------------------------------------
gi_226493325_ref_NP_001140473.      ----------------------------------------------------
H.vulgare_TA46195_4513#1            ----------------------------------------------------
O.sativa_Os05g0125000#1             GGDTAEN-----------------------------------------
P.trichocarpa_scaff_VIII.362#1      GGMGASN-----------------------------------------
P.trichocarpa_scaff_X.2133#1        GGMSTTN-----------------------------------------
S.bicolor_5288383#1                 GGAEANN-----------------------------------------
S.bicolor_5285414#1                 ----------------------------------------------------
O.sativa_Os03g0719100#1             GGTEADN-----------------------------------------
P.patens_159214#1                   GGGSQGLGA---------------------------------------
P.patens_165698#1                   GGGGQGLGG---------------------------------------
P.patens_159935#1                   GGGGQGLGS---------------------------------------
AT5G41580"                          KPVLKDYLQS--------------------------------------
AT1G08910.1EMB3001protein           KPLLKDYVHP--------------------------------------
C.vulgaris_83729#1                  G-----------------------------------------------
```

FIGURE 9 (continued)

```
arabidospsisECsequence          ------------------------------------------------
A.thaliana_AT5G60410.5#1        ------------------------------------------------
Athaliana                       ------------------------------------------------
P.trichocarpa_scaff_66.246#1    QGKKQDLVDRILAILSDEQVSKLWAKKSAIGKEEVAKLVDDTYRKMQVSG
P.trichocarpa_scaff_IX.1493#1   ------------------------------------------------
gi_255570825_ref_XP_002526365.  ------------------------------------------------
gi_225435251_ref_XP_002284945.  ------------------------------------------------
M.truncatula_AC152176_5.4#1     ------------------------------------------------
AK244805Soybean                 ------------------------------------------------
M.truncatula_AC150891_19.5#1    ------------------------------------------------
gi_223944535_gb_ACN26351.1_     ------------------------------------------------
sorghum                         ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0  ------------------------------------------------
gi_226493325_ref_NP_001140473.  ------------------------------------------------
H.vulgare_TA46195_4513#1        ------------------------------------------------
O.sativa_Os05g0125000#1         ------------------------------------------------
P.trichocarpa_scaff_VIII.362#1  ------------------------------------------------
P.trichocarpa_scaff_X.2133#1    ------------------------------------------------
S.bicolor_5288383#1             ------------------------------------------------
S.bicolor_5285414#1             ------------------------------------------------
O.sativa_Os03g0719100#1         ------------------------------------------------
P.patens_159214#1               ------------------------------------------------
P.patens_165698#1               ------------------------------------------------
P.patens_159935#1               ------------------------------------------------
AT5G41580"                      ------------------------------------------------
AT1G08910.1EMB3001protein       ------------------------------------------------
C.vulgaris_83729#1              ------------------------------------------------

arabidospsisECsequence          ------------------------------------------------A
A.thaliana_AT5G60410.5#1        ------------------------------------------------A
Athaliana                       ------------------------------------------------A
P.trichocarpa_scaff_66.246#1    ATDLASRGQVASDCSNSKFNGEMDDPSHSDTKVRCPCGSSLETESMIKCE
P.trichocarpa_scaff_IX.1493#1   ------------------------------------------------
gi_255570825_ref_XP_002526365.  ------------------------------------------------A
gi_225435251_ref_XP_002284945.  ------------------------------------------------A
M.truncatula_AC152176_5.4#1     ------------------------------------------------A
AK244805Soybean                 ------------------------------------------------A
M.truncatula_AC150891_19.5#1    ------------------------------------------------A
gi_223944535_gb_ACN26351.1_     ------------------------------------------------
sorghum                         ------------------------------------------------A
Z.mays_ZM07MSbpsHQ_62031266.f0  ------------------------------------------------A
gi_226493325_ref_NP_001140473.  ------------------------------------------------
H.vulgare_TA46195_4513#1        ------------------------------------------------
O.sativa_Os05g0125000#1         ------------------------------------------------A
P.trichocarpa_scaff_VIII.362#1  ------------------------------------------------E
P.trichocarpa_scaff_X.2133#1    ------------------------------------------------E
S.bicolor_5288383#1             ------------------------------------------------A
S.bicolor_5285414#1             ------------------------------------------------
O.sativa_Os03g0719100#1         ------------------------------------------------A
P.patens_159214#1               ------------------------------------------------DDD
P.patens_165698#1               ------------------------------------------------NDD
P.patens_159935#1               ------------------------------------------------DDD
AT5G41580"                      ------------------------------------------------GVI
AT1G08910.1EMB3001protein       ------------------------------------------------EVV
C.vulgaris_83729#1              ------------------------------------------------E
```

FIGURE 9 (continued)

```
arabidospsisECsequence                DSDSDIEVVADFFGVNLRCPMS-----GSRI------------------
A.thaliana_AT5G60410.5#1              DSDSDIEVVADFFGVNLRCPMS-----GSRI------------------
Athaliana                             DSDSDIEVVADFFGVNLRCPMS-----GSRI------------------
P.trichocarpa_scaff_66.246#1          DFKCHVWQHIGCVIIPEKPMEG-----IPQVPDVFYCEICRLSRADPFWV
P.trichocarpa_scaff_IX.1493#1         VLKMHLLVFVAALVVELQQMMP-----IVTV------------------
gi_255570825_ref_XP_002526365.        DSDSDLEVVADSFAVNLRCPMS-----GSRM------------------
gi_225435251_ref_XP_002284945.        DSDSDLEVVADFFTVNLRCPMS-----GSRM------------------
M.truncatula_AC152176_5.4#1           DSDSDLEVVSDTFSISLRCPMS-----GSRM------------------
AK244805Soybean                       DSDSDLEVVSDTFTINLRCPMS-----GSRM------------------
M.truncatula_AC150891_19.5#1          GSDSDLEVVSDTFSISLRCPMS-----GSRM------------------
gi_223944535_gb_ACN26351.1_           ----------------------------------------------------
sorghum                               DSDSDLEVVTESVTVNLRCPNSGSRMR----------------------
Z.mays_ZM07MSbpsHQ_62031266.f0        DSDSDLEVVTESVTVNLRCPNSGSRMR----------------------
gi_226493325_ref_NP_001140473.        ----------------------------------------------------
H.vulgare_TA46195_4513#1              ----------------------------------------------------
O.sativa_Os05g0125000#1               DSDSDLEVVAESVTVNLRCPNSGSRM-----------------------
P.trichocarpa_scaff_VIII.362#1        DSDSDLEVIAEAIIVNLRCP------------------------------
P.trichocarpa_scaff_X.2133#1          DSDSDLEVIAEAITVNLRCPMSGSRM-----------------------
S.bicolor_5288383#1                   DSDSDIEVVADSVSVNLRCPMTASR-----------------------I
S.bicolor_5285414#1                   ---------------------MTASR-----------------------I
O.sativa_Os03g0719100#1               DSDSDIEVVADSVSVNLRCPMTGSR-----------------------I
P.patens_159214#1                     -SDSDLEVVADFITVNLRCPMSGSR-----------------------I
P.patens_165698#1                     -SDSDLEVVAESITVNLRCPMSGSR-----------------------I
P.patens_159935#1                     GADSDLEIVAESLTVNLRCPMSGSQ-----------------------I
AT5G41580"                            EASPDSDIIEGPSRVSLSCPISRKR-----------------------I
AT1G08910.1EMB3001protein             GSNSDCDIIEGPSRISLSCPISRTR-----------------------I
C.vulgaris_83729#1                    GDESDDEVEIGRTVVSLRCPMSGSR-----------------------M

arabidospsisECsequence                KVAGRFLPCVHMGCFDLDVFVELNQRSRK---------------------
A.thaliana_AT5G60410.5#1              KVAGRFLPCVHMGCFDLDVFVELNQRSRK---------------------
Athaliana                             KVAGRFLPCVHMGCFDLDVFVELNQRSRK---------------------
P.trichocarpa_scaff_66.246#1          TVAHPLSPVKLVATNVPADGSRPVQGVEKTFQLTRADKDLLAKQEYDVQD
P.trichocarpa_scaff_IX.1493#1         TVTWKLLQILLVSIFV-----------------------------------
gi_255570825_ref_XP_002526365.        KVAGRFKPCAHMGCFDLEVFLEMNQRSRK---------------------
gi_225435251_ref_XP_002284945.        KVAGRFKPCAHMGCFDLEIFVEMNQRSRK---------------------
M.truncatula_AC152176_5.4#1           KIAGRFKPCIHMGCFDLDVFVEMNQRSRK---------------------
AK244805Soybean                       KIAGRFKPCIHMGCFDLEVFVEMNQRSRK---------------------
M.truncatula_AC150891_19.5#1          KIAGRFKPCVHMGCFDLEVFVEMNQRSRK---------------------
gi_223944535_gb_ACN26351.1_           ----------------------------------------------------
sorghum                               -IAGRFKPCVHMGCFDLETFVELNQRSRK---------------------
Z.mays_ZM07MSbpsHQ_62031266.f0        -IAGRFKPCVHMGCFDLETFVELNQRSRK---------------------
gi_226493325_ref_NP_001140473.        ----------------------------------------------------
H.vulgare_TA46195_4513#1              ----------------------------------------------------
O.sativa_Os05g0125000#1               RIAGRFKPCIHMGCFDLETFVELNQRSRK---------------------
P.trichocarpa_scaff_VIII.362#1        ----------------------------------------------------
P.trichocarpa_scaff_X.2133#1          KIAGRFKPCAHMGCFDLETFVKLNQRSRK---------------------
S.bicolor_5288383#1                   QIAGRFKPCAHMGCFDLEAFIEINQRSRK---------------------
S.bicolor_5285414#1                   QIAGRFKPCAHMGCFDLEAFIEINQRSRK---------------------
O.sativa_Os03g0719100#1               KIAGRFKPCVHMGCFDLEAFVELNQRSRK---------------------
P.patens_159214#1                     KVAGRFKPCLHMGCFDLDTFVELNQQARK---------------------
P.patens_165698#1                     KVAGRFKPCLHMGCFDLDTCVELNQRARK---------------------
P.patens_159935#1                     KVAGRFKPCPHMGCFDLDTYVEMNQRTRK---------------------
AT5G41580"                            KLPVKGQLCKHLQCFDFSNYVHINMRNPT---------------------
AT1G08910.1EMB3001protein             KLPVKGHVCKHLQCFDFWNYVNMNTR-----------------------
C.vulgaris_83729#1                    RVPARFASVGGLNAFDLDTFLDVVQRSRK---------------------
```

FIGURE 9 (continued)

```
arabidospsisECsequence                  ------------------------------------------------
A.thaliana_AT5G60410.5#1                ------------------------------------------------
Athaliana                               ------------------------------------------------
P.trichocarpa_scaff_66.246#1            AMAAGYRSPGLAVRAINRPGSQLLGANGRDDGPIVTPFVKDGINKILLSG
P.trichocarpa_scaff_IX.1493#1           ------------------------------------------------
gi_255570825_ref_XP_002526365.          ------------------------------------------------
gi_225435251_ref_XP_002284945.          ------------------------------------------------
M.truncatula_AC152176_5.4#1             ------------------------------------------------
AK244805Soybean                         ------------------------------------------------
M.truncatula_AC150891_19.5#1            ------------------------------------------------
gi_223944535_gb_ACN26351.1_             ------------------------------------------------
sorghum                                 ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0          ------------------------------------------------
gi_226493325_ref_NP_001140473.          ------------------------------------------------
H.vulgare_TA46195_4513#1                ------------------------------------------------
O.sativa_Os05g0125000#1                 ------------------------------------------------
P.trichocarpa_scaff_VIII.362#1          ------------------------------------------------
P.trichocarpa_scaff_X.2133#1            ------------------------------------------------
S.bicolor_5288383#1                     ------------------------------------------------
S.bicolor_5285414#1                     ------------------------------------------------
O.sativa_Os03g0719100#1                 ------------------------------------------------
P.patens_159214#1                       ------------------------------------------------
P.patens_165698#1                       ------------------------------------------------
P.patens_159935#1                       ------------------------------------------------
AT5G41580"                              ------------------------------------------------
AT1G08910.1EMB3001protein               ------------------------------------------------
C.vulgaris_83729#1                      ------------------------------------------------

arabidospsisECsequence                  ------------------------------------------------
A.thaliana_AT5G60410.5#1                ------------------------------------------------
Athaliana                               ------------------------------------------------
P.trichocarpa_scaff_66.246#1            CDARIFCLGVRIVKRRTVQQILNLIPKDSEGERFEDALARVCRCVGGGTA
P.trichocarpa_scaff_IX.1493#1           ------------------------------------------------
gi_255570825_ref_XP_002526365.          ------------------------------------------------
gi_225435251_ref_XP_002284945.          ------------------------------------------------
M.truncatula_AC152176_5.4#1             ------------------------------------------------
AK244805Soybean                         ------------------------------------------------
M.truncatula_AC150891_19.5#1            ------------------------------------------------
gi_223944535_gb_ACN26351.1_             ------------------------------------------------
sorghum                                 ------------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0          ------------------------------------------------
gi_226493325_ref_NP_001140473.          ------------------------------------------------
H.vulgare_TA46195_4513#1                ------------------------------------------------
O.sativa_Os05g0125000#1                 ------------------------------------------------
P.trichocarpa_scaff_VIII.362#1          ------------------------------------------------
P.trichocarpa_scaff_X.2133#1            ------------------------------------------------
S.bicolor_5288383#1                     ------------------------------------------------
S.bicolor_5285414#1                     ------------------------------------------------
O.sativa_Os03g0719100#1                 ------------------------------------------------
P.patens_159214#1                       ------------------------------------------------
P.patens_165698#1                       ------------------------------------------------
P.patens_159935#1                       ------------------------------------------------
AT5G41580"                              ------------------------------------------------
AT1G08910.1EMB3001protein               ------------------------------------------------
C.vulgaris_83729#1                      ------------------------------------------------
```

FIGURE 9 (continued)

```
arabidospsisECsequence              ------------------------WQCPICLKNYSVEHVIVDPYFNRITS
A.thaliana_AT5G60410.5#1            ------------------------WQCPICLKNYSVEHVIVDPYFNRITS
Athaliana                           ------------------------WQCPICLKNYSVEHVIVDPYFNRITS
P.trichocarpa_scaff_66.246#1        TDNADSDSDLEVVADSFGVNLRCPWQCPICLKNYSLENIIIDPYFNRITS
P.trichocarpa_scaff_IX.1493#1       ------------------------------------------------
gi_255570825_ref_XP_002526365.      ------------------------WQCPVCLKNYSLENVIIDPYFNRVTS
gi_225435251_ref_XP_002284945.      ------------------------WQCPICLKNYSLENVIIDPYFNRITS
M.truncatula_AC152176_5.4#1         ------------------------WQCPICLKNYALENIIIDPYFNRITS
AK244805Soybean                     ------------------------WQCPICLKNYALENIIIDPYFNRITS
M.truncatula_AC150891_19.5#1        ------------------------WQCPICLKNYALENIIIDPYFNRITS
gi_223944535_gb_ACN26351.1_         ------------------------------------------------
sorghum                             ------------------------WQCPICLKNYSLENLMIDAYFNRITS
Z.mays_ZM07MSbpsHQ_62031266.f0      ------------------------WQCPICLKNYSLENLMIDPYFNRITS
gi_226493325_ref_NP_001140473.      ------------------------------------------------
H.vulgare_TA46195_4513#1            ------------------------------------------------
O.sativa_Os05g0125000#1             ------------------------WQCPICLKNYSLESLMIDPYFNRITS
P.trichocarpa_scaff_VIII.362#1      ------------------------WQCPICLKNYSLEDIVIDPYFNRITT
P.trichocarpa_scaff_X.2133#1        ------------------------WQCPICLKNYCLEDIVIDPYFNRITT
S.bicolor_5288383#1                 ------------------------WQCPICLKNYSLENIIIDPYFNRITS
S.bicolor_5285414#1                 ------------------------WQCPICLKNYSLENIIIDPYFNRITS
O.sativa_Os03g0719100#1             ------------------------WQCPICLKNYSLDNIIIDPYFNRITA
P.patens_159214#1                   ------------------------WQCPICLKNYCIDNLIIDPFFNRITN
P.patens_165698#1                   ------------------------WQCPICLKNYSIENLIIDPFFNQITN
P.patens_159935#1                   ------------------------WQCPICLKNYSIEHLIIDPFFNRITN
AT5G41580"                          --------------------------WRCPHCNQPVCYPDIRLDQNMAK
AT1G08910.1EMB3001protein           ------------------------------------------RHHGAARI
C.vulgaris_83729#1                  ------------------------WQCPHSMRNLPVQQLMVDAYLSHILA

arabidospsisECsequence              KMKHCDEEVTEIEVKPDGSWRVKFKRESERRELGELSQWHAPDGSLCPS-
A.thaliana_AT5G60410.5#1            KMKHCDEEVTEIEVKPDGSWRVKFKRESERRELGELSQWHAPDGSLCPS-
Athaliana                           KMKHCDEEVTEIEVKPDGSWRVKFKRESERRELGELSQWHAPDGSLCPS-
P.trichocarpa_scaff_66.246#1        KMTHCSEDITEIEVKPDGSWRVKTKTEAERRDVGGLAQWHNPDSTPCFPD
P.trichocarpa_scaff_IX.1493#1       -VLHCVEDITEIEVKPDGSWRVKTKTESDHRDAGELAQWHNPDSTLCVPY
gi_255570825_ref_XP_002526365.      KMQHCGEDITEIEVKPDGSWRAKTKSEAERRDVGELAQWHNPDGSLCVPI
gi_225435251_ref_XP_002284945.      SMQSCGEDVTEIQVKPDGCWRVKPENER-----GILAQWHNADGTLCPLA
M.truncatula_AC152176_5.4#1         MMINCGEDVTEVEVKPDGSWRVKAKSESERLDLGILGQWHLPNGSLCTST
AK244805Soybean                     MMMNCGEEIAEIEVKPDGSWRVKVKSESERLELGNLAQWRLPDGTLCVST
M.truncatula_AC150891_19.5#1        MMKNCGEEFTDVEVKPDGYWRVKAKSESECRELGNLAKWHCPDGSLPVST
gi_223944535_gb_ACN26351.1_         ------------------------------------------------
sorghum                             LLQNCSEDVNELDVKPDGSWRVKG-----DAATRDLSQWHMPDGTLCDSK
Z.mays_ZM07MSbpsHQ_62031266.f0      LLHNCSEDVNELDVKPDGSWRVKG-----DAATRDLSQWHMPDGTLCDSK
gi_226493325_ref_NP_001140473.      ------------------------------------------------
H.vulgare_TA46195_4513#1            ------------------------------------------------
O.sativa_Os05g0125000#1             LLRNCNEDVNEVDVKPDGSWRVKG-----DAASRELSQWHMPDGTLCNPK
P.trichocarpa_scaff_VIII.362#1      MMGHCEEDITDIEVKPDGSWTAKT-----KVEIGDLGQWHFPDGSLCAFM
P.trichocarpa_scaff_X.2133#1        LMGHCEEDITEIEVKPDGSWTVKT-----KVDIGDLRQWHFPDGSLCALT
S.bicolor_5288383#1                 LIKSCGDGTSEIDVKPDGSWRVKG-----RAELKDLVQWHQPDGTLSVAT
S.bicolor_5285414#1                 LIKSCRDDTSEIDVKPDGSWRVKG-----RAELKDLIQWHQPDGTLCVAT
O.sativa_Os03g0719100#1             LVQSCGDDVSEIDVKPDGSWRVKG-----GAELKGLAQWHLPDGTLCMPT
P.patens_159214#1                   AVNCLHEDIAAVELKSDGFWRPKLEG-----RVRSREPWRPSPVVTNV-V
P.patens_165698#1                   AVRTMDEDITEVELKADGSWRPKLEG-----HAKNGESWRPPPVAVGANI
P.patens_159935#1                   ALRTLDEDVTEVELKADGSWRPKLEG-----NVKNGEPWSPSPAAAVAIV
AT5G41580"                          ILKDVEHNAADVIIDAGGTWKVTKNTGETPEPVREIIHDLEDPMSLLNSG
AT1G08910.1EMB3001protein           ILEEVGRNAADVVISADGTWMVETENDEDVELVPETTHDHGDPNSFINLG
C.vulgaris_83729#1                  RLKAVSP------------------------VQLLQGPRYAAWFCYVTL
```

FIGURE 9 (continued)

```
arabidospsisECsequence                AVDIKRKMEMLP-VKQEGYSDGPAP--LKLGIRKNRNGIWEVSK--PNTN
A.thaliana_AT5G60410.5#1              AVDIKRKMEMLP-VKQEGYSDGPAP--LKLGIRKNRNGIWEVSK--PNTN
Athaliana                             AVDIKRKMEMLP-VKQEGYSDGPAP--LKLGIRKNRNGIWEVSK--PNTN
P.trichocarpa_scaff_66.246#1          GGEIKPKVEIVKQIRQEGISEGNAGTGLKLGIRKNRNGIWEVSK--PEDM
P.trichocarpa_scaff_IX.1493#1         TGELKSKVEM-KQIKQEGGSEGNAGASLKLGIRKNRNGFWEVSK--PDDM
gi_255570825_ref_XP_002526365.        SGEHKSKVEMEKQIKQEGNSEGYNGTGLKLGIRKNRNGFWEVSK--PEDV
gi_225435251_ref_XP_002284945.        EGEFKPKMDVLKQIKQEGISECHS--SLKLQIK-NRNGVWEVSK--PDEM
M.truncatula_AC152176_5.4#1           AGDIK-RVETLKQVKQEGFSDGPA--GLKLGIRRNRNGNWEVSK--PETT
AK244805Soybean                       DVDVK-RVDTLKQVKQEGVSDSPA--GLKLGIKKNCNGVWEVSK--PEGT
M.truncatula_AC150891_19.5#1          SGEDK-RVETLN-VKQEGVSDSPN--GLRLGIRKNCNGDWEVSK--PKDT
gi_223944535_gb_ACN26351.1_           ----------------------------------------------------
sorghum                               EDTNP-GVTSVNEFKREGTSDGHRT--LKI--KKNPNGSWQVSS--KADD
Z.mays_ZM07MSbpsHQ_62031266.f0        EDTNP-GVVSVNEFKREGTSDGHRT--LKLGIKKTPNGLWQVSS--KPDD
gi_226493325_ref_NP_001140473.        ----------------------------------------------------
H.vulgare_TA46195_4513#1              ----------------------------------------------------
O.sativa_Os05g0125000#1               EDVKPAMQNGNEQMMEGTSDGQKSL---KIGIKRNPNGIWEVSSKADDKK
P.trichocarpa_scaff_VIII.362#1        DEVTSCYEISRQIEKGDGLKAHGSP---EIGIKSNFGGMMQGR---KHQL
P.trichocarpa_scaff_X.2133#1          DEVTSCYKIPRQIEKGDGLKAHFSP---ETGIKNNLSGIVQGR---NPQL
S.bicolor_5288383#1                   DTAAKPEICIVKHEVKEEPLSEEVG-CLKLGLRKKSNGQWEISKIGDADL
S.bicolor_5285414#1                   DTAAKPEICIVKHEVKEEPLSEEVG-CLNLGLRKKSKGQWEISKLGDADL
O.sativa_Os03g0719100#1               DTRSKPNIRIVKQEIKEEPLSEETGGRLKLGIRRNNNGQWEINKR-----
P.patens_159214#1                     NETT-SVPVLFS-ENVRTEEKLSSHDNGSSCFNRNPCGVLVLHC------
P.patens_165698#1                     HKAT-AAPVLFF-KHVKTEEGMSSHDNGFSRFKVNPSGDWARNA------
P.patens_159935#1                     SNGNKSAPVLFSSHHVKIEAGRSSHDHGSLQLKWTPEVQRVVNG------
AT5G41580"                            PVVFDLTGDDDAELEVFGDNKVEDRKPCMSDAQGQSNNNNTNKHPSNDDY
AT1G08910.1EMB3001protein             PTVKNPARD---ENEMETSTQVEEHNPCLSEIQGPSN----DTHRPASDY
C.vulgaris_83729#1                    YGVHLIICQMCYFPGEVASAEEATDMAEVTEVEVSPGGEWRVGG------

arabidospsisECsequence                GLSSSN------------RQEKVGYQEKNIIPMSSSATGSGRDGDDASVN
A.thaliana_AT5G60410.5#1              GLSSSN------------RQEKVGYQEKNIIPMSSSATGSGRDGDDASVN
Athaliana                             GLSSSN------------RQEKVGYQEKNIIPMSSSATGSGRDGDDASVN
P.trichocarpa_scaff_66.246#1          NTFSSG------------RLQENFEHHEQKVIPMSSSATGSGRDGEDQSVN
P.trichocarpa_scaff_IX.1493#1         NTSSSG------------RLQENFELYEQKVIPMSSSATGSGRDGEDPSVN
gi_255570825_ref_XP_002526365.        NTSSSGN-----------RLPERFEIIEQKVIPMSSSATGSGRDGEDPSVN
gi_225435251_ref_XP_002284945.        NTLTCN------------RLQEKFEDPGQQVIPMSSSATGSGRDGEDPSVN
M.truncatula_AC152176_5.4#1           NTSSGH------------ILKEVFGNPEQVVIPMSSSGSESGRDGDDPSVN
AK244805Soybean                       NTSSGN------------NLKRVFGNPEQVVIPMSSSATGSGRDGDDPSVN
M.truncatula_AC150891_19.5#1          NISSDN------------RLNADLGNHEVVVIQMSSSGSESRLDGDDPSVN
gi_223944535_gb_ACN26351.1_           ----------------------------------------------------
sorghum                               KKPVVRH----------HIQNNNGFSTP-NMPIISSPTGSYRDGEDASVN
Z.mays_ZM07MSbpsHQ_62031266.f0        KKPVVRN----------HIQNNTGYSIPNIVPMISSPTGSCRDGEDVSVN
gi_226493325_ref_NP_001140473.        ----------------------------------------------------
H.vulgare_TA46195_4513#1              ----------------------------------------------------
O.sativa_Os05g0125000#1               PSVVGNR----------MQNNSGFRALNNIMHMSNSPTSSYRDGEDPSVN
P.trichocarpa_scaff_VIII.362#1        AFCSSKN----------QIEGNFVNQGQRTKTMSSSITGSSKYEEDPSIN
P.trichocarpa_scaff_X.2133#1          AFCSSKN----------QLEGSFVNHGQRTLTMSSSTTGSGGDEEDPSIN
S.bicolor_5288383#1                   VPSSGN-------------DHSRYNENKNCITLSSNIGDTNITDEGYNLE
S.bicolor_5285414#1                   VRSSGN-------------DHSRYNENKNDITLSSNIGDTNIANEGYNLE
O.sativa_Os03g0719100#1               -LDSNN-------------GQNGYIEDENCVVSASNTDDENSKNGIYNPE
P.patens_159214#1                     -TENLNG----------GHSQEVINVSRLSRSNSITDGNLRGDEDEHSAN
P.patens_165698#1                     -TKHLND----------GPSQKVVDAPRLSRSSSATDSNLKIDEDEHSVN
P.patens_159935#1                     -RNYRNG----------GPSLQVMPLPRLSRSSSATGSNLKVGEDENSVN
AT5G41580"                            SSIFDISDVIALDPEILSALGNTAPQPHQASNTGTGQQYSNLSQIPMSID
AT1G08910.1EMB3001protein             TMLNQS-----------HTSTNTLPQLPRTLNAFDGQQFVNLPQVINTRD
C.vulgaris_83729#1                    ---------------------TEGRWHSISEDPSLPIEDVKVKADPETAK
```

FIGURE 9 (continued)

```
arabidospsisECsequence              QDAIGTFDFVAN-GMELDSISMNVDS-GYNFPDRNQSGEGGNNEVIVLSD
A.thaliana_AT5G60410.5#1            QDAIGTFDFVAN-GMELDSISMNVDS-GYNFPDRNQSGEGGNNEVIVLSD
Athaliana                           QDAIGTFDFVAN-GMELDSISMNVDS-GYNFPDRNQSGEGGNNEVIVLSD
P.trichocarpa_scaff_66.246#1        QDAGGNYDFTNN-GMELDSLSLNVYT-TYGFTDQNLPVPLGNAEVIVLSD
P.trichocarpa_scaff_IX.1493#1       QDTGENFEFTNN-GMELDSLSLNVYS-TYGFTDQNLSAPVGNAEVIVLSD
gi_255570825_ref_XP_002526365.      QDGGGNFDFTNN-GIELDSLPLNVDS-TYGFPDRNFSAPVEDPEVIVLSD
gi_225435251_ref_XP_002284945.      QDGGGNYDFSTNPGIELDSISLNIDNNAYAFPERNTPAPMGDTELIVLSD
M.truncatula_AC152176_5.4#1         QGGGGHIDYSTTNGIEMDSQSRNNVDLARGYTVHNTSAQVGGAEIIVLSD
AK244805Soybean                     QGGGGHIDYSTTNGIEMDSLCLNNVDLAYEYTAPNTSAQVGGAEVIVLSD
M.truncatula_AC150891_19.5#1        QSGGGHTDYSPTNGIETNSVCHTNVDSTYGYTIPNTSAPMANAEVIVLSD
gi_223944535_gb_ACN26351.1_         ----------------------------------------------------
sorghum                             QEGGG-IQFDIALNQEFDSFAHNFGQ-TYNTEDRQQ-PQHNAADVIVLSD
Z.mays_ZM07MSbpsHQ_62031266.f0      QEGGG-IQFDISLNQEFDSFAHNFGQ-TYNTEDRQQEPQHNAADVIVLSD
gi_226493325_ref_NP_001140473.      ----------------------------------------------------
H.vulgare_TA46195_4513#1            ----------------------------------------------------
O.sativa_Os05g0125000#1             QESNRHVDLSLN-NGNNEFDSFSLNFGQACNTDDRPQQQHNATDVIVLSD
P.trichocarpa_scaff_VIII.362#1      QDYSGHVEISPS-NVN-EINSICHYFDPTLAINNGSFVPSRNADIIVLSD
P.trichocarpa_scaff_X.2133#1        QDYSGHVEISPS-SVN-EINSICHYFDPTLAINHGSSVAPGNADIIILSD
S.bicolor_5288383#1                 P-------------ATNGDPTTHVHDLDSSSSDQNGPPASTGQDIIVLSD
S.bicolor_5285414#1                 P-------------ATNGDPTTHVHDLGSSSSDENGPSASTGQDIIVLSD
O.sativa_Os03g0719100#1             P-------------GQFDQLTSNIYDLDSSPMDAHFPPAPTEQDVIVLSD
P.patens_159214#1                   QD---RGAVSTVDGDNVGLSSRSPQPTSKTLQTCAD-LANGAQVIVLSDS
P.patens_165698#1                   HDPSEKNAVSMDDNEEVEYLARPRDAPGGTWQTSGDDPANGADVIVLSDS
P.patens_159935#1                   QDASEKNTVSMDDTD-VEFLSILQEAARGTWQASGDDPANGGDIIVLSDS
AT5G41580"                          P---MPVPVPFSQTPSPRDRPATTSTVFTIPNPSPQYSQVHASPVTPTGT
AT1G08910.1EMB3001protein           SPASQALPMTFSPTPSPQDILATNAANFGTSMPAAQSSQFQGSHVTSLGN
C.vulgaris_83729#1                  QEAGSAGAFDSD---SDEEMSEGEELRRAAAAAKSTVRPKPPPDVIVISD

arabidospsisECsequence              SDDENDLVITPGPAYS-GCQT-DGGLTFPLNPPGIINSYNEDPHSIAGGS
A.thaliana_AT5G60410.5#1            SDDENDLVITPGPAYS-GCQT-DGGLTFPLNPPGIINSYNEDPHSIAGGS
Athaliana                           SDDENDLVITPGPAYS-GCQT-DGGLTFPLNPPGIINSYNEDPHSIAGGS
P.trichocarpa_scaff_66.246#1        SDDDNDILISPGSVYK-SNQND-GDATFSVPSPGIADPFPEDPTLVTGAN
P.trichocarpa_scaff_IX.1493#1       SDEENDILMSSGSVYK-SNQN--GGATISVPSPEIADHFLEDPTLGTGGN
gi_255570825_ref_XP_002526365.      SDDDNDILMTTGTVYK-NSQTDDGGAGFSMPPNGISNPYPEDPTVGNG--
gi_225435251_ref_XP_002284945.      SEEENDTLMSSGTLYN-NSRADAGGINFSIPT-GIPDSYAEDPTAGPGGS
M.truncatula_AC152176_5.4#1         SEEDNDILVSPPIANN-NHQN-DTADGYSMPPPGIVDPYVED--QNLGGS
AK244805Soybean                     SEEDNDLLASPAIAYK-NNRN-DATDGYSVPPPVIVDSYTED--HNLGGN
M.truncatula_AC150891_19.5#1        SEDD-EILISPTVGYG-NNQTGDAVDAYSVPPPGIMDPYAGD--HSIGGN
gi_223944535_gb_ACN26351.1_         ----------------------------------------------------
sorghum                             SDEENDPIVRPPAVYA-NATT--NGDSFPFVTDAAGTGYPERYQEDAGVG
Z.mays_ZM07MSbpsHQ_62031266.f0      SDEENDPIVRLPAVYA-NTPT--NGDSFPFVTDAAVSGYPEGYQEDAGVG
gi_226493325_ref_NP_001140473.      ----------------------------------------------------
H.vulgare_TA46195_4513#1            ----------------------------------------------------
O.sativa_Os05g0125000#1             SDEENDAMVCPPAVYDNTTTANGSGFPFTTNGIGYTERYQEDAGVGT---
P.trichocarpa_scaff_VIII.362#1      SDEENVNLVPPETVYD-TCPVDGSCSSLVAN-PGIADSYLEDLALDAGAD
P.trichocarpa_scaff_X.2133#1        SDEENVNLVPPETVYD-TYRVNGSGSSLAVN-PGITDSYLEDLALDAGAN
S.bicolor_5288383#1                 SDDDDVMVLSPGAVNCGSTHD--TGNMFPLNTPENLGVCSEQT-GVCPKE
S.bicolor_5285414#1                 SDDD-VMVLSPGAVNCGSTHD--TGNLFSINTPENLGVCSEQT-VECPKE
O.sativa_Os03g0719100#1             SDDDNVMVLSPGDVNFSSAHD--NGNAFPPNPPEASGICGEQPRGAGPDV
P.patens_159214#1                   EDEGEDEETIVVSSCASIYVESNVVGNWRMR-SRKSYQTNGDSSGLAVGL
P.patens_165698#1                   DDDGGEEEDVVASSAASMYAES-DVGNRRMR-SREVSHTNGDSSGLALGL
P.patens_159935#1                   DD-GEDEVTVVGSSHVSMYSDVIGNGGVRFSEGRVPNDINGDPSRFALGL
AT5G41580"                          YLGRTTS--PRWNQTY---QSQAP--PMTTPYTSRKVSVPVTSQS-----
AT1G08910.1EMB3001protein           CEGRTSDLMARWNHIYGRVQTQFPPAPLSHHHYSMQNQSPSPAQQRPVPS
C.vulgaris_83729#1                  SDEEDDAEMSRQGPAP----------APLRPPAPPQQHQQTGYQPAG--
```

FIGURE 9 (continued)

```
arabidospsisECsequence                SGLGLFNDDDE----------------------------FDTP--LWSF
A.thaliana_AT5G60410.5#1              SGLGLFNDDDE----------------------------FDTP--LWSF
Athaliana                             SGLGLFNDDDE----------------------------FDTP--LWSF
P.trichocarpa_scaff_66.246#1          SCLGLFNAN-DE---------------------------YGMP--LWSL
P.trichocarpa_scaff_IX.1493#1         SCLGLFNA--DE---------------------------YGMP--LWPL
gi_255570825_ref_XP_002526365.        --LGFLNPNDDE---------------------------FGIP--LWPL
gi_225435251_ref_XP_002284945.        SCLGLFSTADDD---------------------------FGMSGSLWPL
M.truncatula_AC152176_5.4#1           SCLGLFPNEDD----------------------------FGISS-LWSL
AK244805Soybean                       SCLGLFPNDDD----------------------------FGMSS-LWPL
M.truncatula_AC150891_19.5#1          PCLGVFDNPNESI--------------------------FGIPS-VWPL
gi_223944535_gb_ACN26351.1_           ------------------------------------------------M
sorghum                               TSGLGLLNNNTGD--------------------------FEINN--WQM
Z.mays_ZM07MSbpsHQ_62031266.f0        TSGLGLLSNNTGD--------------------------FEINN--WQM
gi_226493325_ref_NP_001140473.        ------------------------------------------------M
H.vulgare_TA46195_4513#1              -------------------------------------------------
O.sativa_Os05g0125000#1               SGLGLLSNNVDD---------------------------FEMNN--WQM
P.trichocarpa_scaff_VIII.362#1        SCFDLFDTGVND---------------------------VGMSS--WSF
P.trichocarpa_scaff_X.2133#1          SCFGLFDTGVND---------------------------VGMSN--WSY
S.bicolor_5288383#1                   SSFVALREGFGD-----------------------------LGLSFWEC
S.bicolor_5285414#1                   SSFLAVKEGFGD-----------------------------LGLSFWEC
O.sativa_Os03g0719100#1               TSFL---DGFDD-----------------------------LELPFWES
P.patens_159214#1                     EYADVTPASHSVCHPSDPATSRRIDIVIN-EQMTDDSGLGLHLDSGSLSA
P.patens_165698#1                     ESTDVTLASNSVCNPFDPPLSTRVGVGVSDEGTSNNSGFGLQSDRNSF-W
P.patens_159935#1                     G------------------------------------------------
AT5G41580"                            ----------------------------------------PANVSSFVQSQ
AT1G08910.1EMB3001protein             YIAHPQTFHVNYGENADQRWMPSSIAHPQ--------TLPVNYGGNTNQR
C.vulgaris_83729#1                    -------------------------------------------------

arabidospsisECsequence                PS-ETPEAPGFQLFRSDADVSGGLVGLHHHSPLNCSPEINGGYTMAPETS
A.thaliana_AT5G60410.5#1              PS-ETPEAPGFQLFRSDADVSGGLVGLHHHSPLNCSPEINGGYTMAPETS
Athaliana                             PS-ETPEAPGFQLFRSDADVSGGLVGLHHHSPLNCSPEINGGYTMAPETS
P.trichocarpa_scaff_66.246#1          PS-GNQAGPGFQLFN--SDVSDALVDLP-HGSVNCPLSMNG-YTLAPETV
P.trichocarpa_scaff_IX.1493#1         PP-GNQAGPGFQLFN--SDVSDALVDLP-HDPVNCPSSMNG-YTLAPETV
gi_255570825_ref_XP_002526365.        PP-GSQAGPGFQLFN--SDVPDALVDIQ-HGPISCPMTING-YTLAPETV
gi_225435251_ref_XP_002284945.        PP-GTQPGPGFQFFGTDTDVSDALADLQ-HNPINCPTSMNG-YTLGPEVV
M.truncatula_AC152176_5.4#1           PS-ASQAGPGFQLFGSDADASDALVHLQ-HVPINCTSSLNG-YALAPETA
AK244805Soybean                       PS-GSQAGPGFQLFGSDADVSDALVHLQ-HGPMNCSSSLNG-YALAPDTA
M.truncatula_AC150891_19.5#1          HS-GTQASSGFQLFSSDVDVSDALA----HGDINCSSSLNS-YTLAPDTA
gi_223944535_gb_ACN26351.1_           LS-YPQPEQGFQFFVTDTDVGNPFVAPH-----NSFTIAPEDYSLGCNVG
sorghum                               HS-YPQPEQGFQFFGTDTDVGNPFVGPH-----NSFNIAPEDYSLDCNVG
Z.mays_ZM07MSbpsHQ_62031266.f0        HS-YPQPEQGFQFFGTDTDVGNPFVGPH-----NSFSITPEDYSLDCNVG
gi_226493325_ref_NP_001140473.        YQ-CTVVHQTFQLSRADRET-----------------VQRQEYDL-----
H.vulgare_TA46195_4513#1              -------------------------------------------------
O.sativa_Os05g0125000#1               HSSYQQPEQGFQFFGNDTDVHNTFVG-SHN---SFGLAPND-YSLDCNVG
P.trichocarpa_scaff_VIII.362#1        SS-GIQAGAHFQLFNTDSDVSDAFID-LEHSSISCAAPMNG-STLASTPT
P.trichocarpa_scaff_X.2133#1          SS-CTQAGPHFQLFNTDPDVSDAFID-LDHPSISCAVPMNG-CTLASTPA
S.bicolor_5288383#1                   PR-SPRDDPTSQILDTSTKATDNPGE-------------------------
S.bicolor_5285414#1                   PG-SPRDDPTSQILDTSTKATDNSGE-------------------------
O.sativa_Os03g0719100#1               SS-SQD--------AAGTQVTDNQCE-------------------------
P.patens_159214#1                     VPAEIARNGECGVHESQTEVVNVLQGQP-VRQTPVRFVAGCGSSFRPLDK
P.patens_165698#1                     VSAEKTRNGQYSFYAPRTESVNVLQHPPPVRPTPVQALAGPGFLSVPLVR
P.patens_159935#1                     ---------QYEYYGSQTEAV-ALQHPP-LQPTPVEVVGSSDFLCTSLIG
AT5G41580"                            HVPRVLSQPN-------------------------NYG-----------
AT1G08910.1EMB3001protein             PIPSSIAHPQTLPVNYRGNTDHRSTPYSITHLQTLLNYGGNADQRPMPSS
C.vulgaris_83729#1                    -------------------------------------------------
```

FIGURE 9 (continued)

```
arabidospsisECsequence                MASVPVVPG-STGRS--EANDGLVDNP----------
A.thaliana_AT5G60410.5#1              MASVPVVPG-STGRS--EANDGLVDNP----------
Athaliana                             MASVPVVPG-STGRS--EANDGLVDNP----------
P.trichocarpa_scaff_66.246#1          MGSTCLIPDSSLGRSEMDVNDGLLDNP----------
P.trichocarpa_scaff_IX.1493#1         MRSTCLIPDSSIGRSDTDVNDGLVDNP----------
gi_255570825_ref_XP_002526365.        MGPSSLVADSSVGRSDTDTNDGLVNNP----------
gi_225435251_ref_XP_002284945.        MGSAALVPDPSIGRTDTDMNDGLVDNP----------
M.truncatula_AC152176_5.4#1           LGSGSLLQDSSAGRSDADLNGGLVDNP----------
AK244805Soybean                       LGSGGILQESSAGRSVADLNGGLVDNP----------
M.truncatula_AC150891_19.5#1          LGSSALIPNSSTDRSDTDLNGGLVDNP----------
gi_223944535_gb_ACN26351.1_           IEDPSAAHDVSICRNSNDVHGSLVDNP----------
sorghum                               IEDPSAAHDVSICRNSNDVHGSLVDNP----------
Z.mays_ZM07MSbpsHQ_62031266.f0        IEDPSAAHDVSICRNSNDVHGSLVDNP----------
gi_226493325_ref_NP_001140473.        ----------------QVWCILINDK----------
H.vulgare_TA46195_4513#1              ----------
O.sativa_Os05g0125000#1               VEEASVTPALSVCRNSNEMHGSLVDNP----------
P.trichocarpa_scaff_VIII.362#1        ITSGGEVPDSLACVANVDMDVGLVDNP----------
P.trichocarpa_scaff_X.2133#1          ITSGGEVLDSLACDANVDMDVGLVDNP----------
S.bicolor_5288383#1                   VENYPANDQPKQGPVSGAVEA--NPV-----------
S.bicolor_5285414#1                   VENYPANDQSMPGSVSGADLG--VPAN-----------
O.sativa_Os03g0719100#1               MQNFIVNHQFLHEPILGVNLGGTAASN-----------
P.patens_159214#1                     DDTTSPSGWESLCLPNGATD---TDGGS----------
P.patens_165698#1                     DDASLASGWKRYSRPYGATG---TDGGS----------
P.patens_159935#1                     NDPS-QLDWESFGQSHTPTLYEVIDSES----------
AT5G41580"                            ---VRGLTSSHASTSRQHPSGPTVQS-----------
AT1G08910.1EMB3001protein             ITNLQTLPATYGGYAHQRPMSSSITHPRTSPVNYGGTPDQRPMPSSITHP
C.vulgaris_83729#1                    --HAPQPEASSSGRGQLQHSPNSYGRQ----------

arabidospsisECsequence                --LAFGRDDPSLQIFLPTKP-DASAQSGFKNQADMSNGLRS-EDWISLRL
A.thaliana_AT5G60410.5#1              --LAFGRDDPSLQIFLPTKP-DASAQSGFKNQADMSNGLRS-EDWISLRL
Athaliana                             --LAFGRDDPSLQIFLPTKP-DASAQSGFKNQADMSNGLRS-EDWISLRL
P.trichocarpa_scaff_66.246#1          --LAFGGEDPSLQIFLPTGPSDASMHSDMRDQVDVSNGVRS-EDWISLRL
P.trichocarpa_scaff_IX.1493#1         --LAFGREDPSLQIFLPTGPSDASMQSDMRDQADVSNGVRT-DDWISLRL
gi_255570825_ref_XP_002526365.        --LAFGGEDPSLQIFLPTRPSDASGQSDLRDQADVSNGVRT-EDWISLRL
gi_225435251_ref_XP_002284945.        --LAFGGDDPSLQIFLPTRPSDASVPTDLRNQADVSNGSRP-DDWISLRL
M.truncatula_AC152176_5.4#1           --LAFAGDDPSLQIFLPTRPAESSMQNELRDQANVSNGVST-EDWTSLTL
AK244805Soybean                       --LAFGGDDPSFQIFLPTRPADSSMHNELRDQANVANGVCTEEDWISLRL
M.truncatula_AC150891_19.5#1          --LAFGGQDPSLQIFLPTRPAESSVQHELRNHTDVSNGVCT-EDWISLSL
gi_223944535_gb_ACN26351.1_           --LALTGDDPSLQIFLPSQPSTVPLQQELSERSDTPNGVHPNDWRISLTL
sorghum                               --LALAGDDPSLQIFLPSQPSTVPLQEELSERANTPNGVHPDDWRISLTL
Z.mays_ZM07MSbpsHQ_62031266.f0        --LALAGDDPSLQIFLPSQPSTVPLQEELSERANTPNGVHPDDWRISLTL
gi_226493325_ref_NP_001140473.        --VQFR-------------MQWPQYAELQVNG------------IPVRV
H.vulgare_TA46195_4513#1              --------------------------------------------------
O.sativa_Os05g0125000#1               --LALVGDDPSLQIFLPSQPSSVPLQEELSERANAPNGVQS-DDWISLTL
P.trichocarpa_scaff_VIII.362#1        --MRFVSEDPSLQTFLPTQP----VQPDLVIQPPVSNPVPT-EDWISLRL
P.trichocarpa_scaff_X.2133#1          --MRFVAEDPSLQTFLPVQP----VQPDLVHKPPVSNRAPT-EDWFPLSL
S.bicolor_5288383#1                   --PVEDEHDSALHPCPSNERDSAMGLANLGADTETCVDVHS-DKWTDVSI
S.bicolor_5285414#1                   --LVEDGDDGALQPCSLSERDSAMGLANLGADTQTCGDVHS-DKWTAVSI
O.sativa_Os03g0719100#1               --TLECEHDGALQACQSSDQD--------GDQNQTCHDGHS-GDLTNLSI
P.patens_159214#1                     --MDNEGTGLPLQKFLPSQPVRVEVLVAEKNHALEEAT------------
P.patens_165698#1                     --VDSDTDASPLQNFLPSQPARAEVQDNLRGPFLPEADDIDNSWFSLSLG
P.patens_159935#1                     --IDINKDGSPLQNLLPSQPARSEVQENLRDLTLPAEVMDN-SWFSLSLG
AT5G41580"                            -------------------VSRLSDLVDVDLTVPDTSNWRP--RMRGSL
AT1G08910.1EMB3001protein             QTLPVSYGGTTDQILNPGGAMGQFSSREFMNLTPANTENWRPQSRMRGSV
C.vulgaris_83729#1                    -----QGPAPTGRSYVGPYPGSLPAAGGHRAADRPGGGTVVRGRGITLRL
```

FIGURE 9 (continued)

```
arabidospsisECsequence                GD-SASGNHGDPAT-TNGINSSHQMSTREGSMDTTTE------------
A.thaliana_AT5G60410.5#1              GD-SASGNHGDPAT-TNGINSSHQMSTREGSMDTTTE------------
Athaliana                             GD-SASGNHGDPAT-TNGINSSHQMSTREGSMDTTTE------------
P.trichocarpa_scaff_66.246#1          GG-SATSNHGDLVPPTNGLNSRQQMPSSLDSLP----------------
P.trichocarpa_scaff_IX.1493#1         GGGGATGNHSEAVPSTNRLNSRQQMPSREDGMD----------------
gi_255570825_ref_XP_002526365.        GGGGATGSHGDSVS-ANGVNSRQQMPPRDGAMDSLAD------------
gi_225435251_ref_XP_002284945.        GG--SSGGHAESPA-ANGLNTRQQLPSKDGDMDSLAD------------
M.truncatula_AC152176_5.4#1           GG-GAGGSNGDAST-QNGLNSRHQVPSRDNGTNTLAD------------
AK244805Soybean                       GG-GAGGNNGDAPT-QNGLNSRHQIPTREGAKNTLDD------------
M.truncatula_AC150891_19.5#1          GG-GAGGSIGDAST-TNGLNSRPQIQSREDAPDSLTDSLNEA----DLLL
gi_223944535_gb_ACN26351.1_           AA---GGGGNEESTSVGGLKSQPKVSSKEAGVEPLLDAACQAFPGSSRLT
sorghum                               AA---GGGGNEESTSVDGLKSQPKVPSKEAGVEPLLDAA----------
Z.mays_ZM07MSbpsHQ_62031266.f0        AA---GGGGNEEPTSVDGLKSQPKVPSKEAGVEPLLDAA----------
gi_226493325_ref_NP_001140473.        MT---RPGS--QLLGINGRDDGPLVTT----------------------
H.vulgare_TA46195_4513#1              ----------------------------------------------------
O.sativa_Os05g0125000#1               AA----GGGGNEEPAPADVNSQPQIPSTETGIEPLTDA------------
P.trichocarpa_scaff_VIII.362#1        GS----TNESFGSHTMDRPHRAARDG-VDLRNQLGLNQ------------
P.trichocarpa_scaff_X.2133#1          SS----TSESFGNHTRNHDQGAAKIG-VDLRNQLGSNQ------------
S.bicolor_5288383#1                   SG-SGEDFTSAKIASKKRNNPGDRITALDGSLVGSANDD-----------
S.bicolor_5285414#1                   SV-SDEDLTSAKIASKKRSNPGDRITALNG--------------------
O.sativa_Os03g0719100#1               IS-TQDSLTNGKNASQKRTNCEDGTAGLDGSVVRSANG------------
P.patens_159214#1                     --------------------HRTLLLQLTCTS------------------
P.patens_165698#1                     GGENLAEPSPARSNLRTSLEQRSPIPQPRVTAHDDTHGRSGR--------
P.patens_159935#1                     GIGNVVKPTPERSLFTTSIEWCSPISQPGVNEYDPHG------------
AT5G41580"                            VPGSHSTALDHMIIRPSQQSQTSTRLNSSQPVQTPSVQ------------
AT1G08910.1EMB3001protein             APG---TGYDHMIIHPTRPVHPQAQT-PPAPLSTSYDG------------
C.vulgaris_83729#1                    PPRPERPPSGQSPFPMSGV--------------------------------

arabidospsisECsequence                --TASLLLGMNDSRQDKAKKQ-RSDNPFSFPRQKRSVRPRMYLSIDSDSE
A.thaliana_AT5G60410.5#1              --TASLLLGMNDSRQDKAKKQ-RSDNPFSFPRQKRSVRPRMYLSIDSDSE
Athaliana                             --TASLLLGMNDSRQDKAKKQ-RSDNPFSFPRQKRSNNEQDHQTRHRSLN
P.trichocarpa_scaff_66.246#1          --GTASSLGINDGRSEKASRQ-RSESTFSFPRQKRSVRPRPYLSIDSDSE
P.trichocarpa_scaff_IX.1493#1         --LAGTDL--------KRQVD-KDQTVLSHFLAKNVL-------------
gi_255570825_ref_XP_002526365.        --TASLLLGMNDGRSEKASRQ-RSDSPFQFPRQKRSIRPRLYLSIDSDSE
gi_225435251_ref_XP_002284945.        --TASLLLGMNDGRSDKTSSRQRSDSPFSFPRQRRSVRPRLYLSIDSDSE
M.truncatula_AC152176_5.4#1           --SASLLLGMNDVRSDRASRP-RSGSPFTFPRQKRSVRPRLYLSIDSESE
AK244805Soybean                       --TASLLLGMNDVRSDRARRQ-RSDSPFSFPRQKRSVRPRLYLSIDSDSE
M.truncatula_AC150891_19.5#1          AETASLLRSVDDAESDKASRK-RSDGPFSFPRQKRSVRPRLNLSIGSDSE
gi_223944535_gb_ACN26351.1_           FEASALGSMNNDRCTGSNLNPKRIENIFSHPRQPRSVRPRLCLSLDTDSE
sorghum                               ---SALPSMN--RCNGSNLNPRRIENIFSHPRQPRSVRPRLCLSLDTDSE
Z.mays_ZM07MSbpsHQ_62031266.f0        ---SALPSMNNDRCNGSNLNPRRIENIFSHPRQPRSVRPRLCLSLDSDSE
gi_226493325_ref_NP_001140473.        ---CSREGIN--KISLSRVDAR--TFCFGVRIVRRTVPQVLFI-------
H.vulgare_TA46195_4513#1              ----------------------------------------------------
O.sativa_Os05g0125000#1               --ASAFLSTNIERRSGADLNPRRIENIFSHPRQPRSVRPRLCLSIDTDSE
P.trichocarpa_scaff_VIII.362#1        --ATSVAALNDEARSNGKYNKKISDGPFSFPRQPRSVRQRVYSQ------
P.trichocarpa_scaff_X.2133#1          --GA----------------------------------------------
S.bicolor_5288383#1                   -------------------ESARERDGPFSPPQQRSVRPRVILDIASDSD
S.bicolor_5285414#1                   ----------------------------------MEEVFTPFAFM-------
O.sativa_Os03g0719100#1               ---------------------LRGEMPPLGQEQDRTVRQKLILTIESDSD
P.patens_159214#1                     ----------------------------------------------------
P.patens_165698#1                     --LGKHGSSQRLMDSRPPVRSHFTADGHPFPRHLSSFRRDGCLLSCSVRH
P.patens_159935#1                     --LGSNGSSQHLMDSRHPLRGHFTADGYPFPVRPCSGHIHARRVSR----
AT5G41580"                            ---TSQAQSPFTTAAYRT---ETVLGNRNHPVPAPP---GIVRPTGPTS-
AT1G08910.1EMB3001protein             ---ADEIQAFIGHPSYPVSNNETQAGTSSLPVAEGLGYSGSFWSMPPETW
C.vulgaris_83729#1                    ----------------------------------------------------
```

FIGURE 9 (continued)

```
arabidospsisECsequence          ----------------------------------------------
A.thaliana_AT5G60410.5#1        TMNRIIRQDTGV----------------------------------
Athaliana                       KICIILCAGKN-----------------------------------
P.trichocarpa_scaff_66.246#1    ----------------------------------------------
P.trichocarpa_scaff_IX.1493#1   ----------------------------------------------
gi_255570825_ref_XP_002526365.  ----------------------------------------------
gi_225435251_ref_XP_002284945.  ----------------------------------------------
M.truncatula_AC152176_5.4#1     ----------------------------------------------
AK244805Soybean                 ----------------------------------------------
M.truncatula_AC150891_19.5#1    ----------------------------------------------
gi_223944535_gb_ACN26351.1_     ----------------------------------------------
sorghum                         ----------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0  ----------------------------------------------
gi_226493325_ref_NP_001140473.  ----------------------------------------------
H.vulgare_TA46195_4513#1        ----------------------------------------------
O.sativa_Os05g0125000#1         ----------------------------------------------
P.trichocarpa_scaff_VIII.362#1  ----------------------------------------------
P.trichocarpa_scaff_X.2133#1    ----------------------------------------------
S.bicolor_5288383#1             ----------------------------------------------
S.bicolor_5285414#1             ----------------------------------------------
O.sativa_Os03g0719100#1         ----------------------------------------------
P.patens_159214#1               ----------------------------------------------
P.patens_165698#1               SYVTTQNGHPQLAASAFHNARVVLFCQDLMGLTRKNFNGARSRKRLQPVF
P.patens_159935#1               ----------------------------------------------
AT5G41580"                      ----------------------------------------------
AT1G08910.1EMB3001protein       ----------------------------------------------
C.vulgaris_83729#1              ----------------------------------------------

arabidospsisECsequence          ----------------------------------------------
A.thaliana_AT5G60410.5#1        ----------------------------------------------
Athaliana                       ----------------------------------------------
P.trichocarpa_scaff_66.246#1    ----------------------------------------------
P.trichocarpa_scaff_IX.1493#1   ----------------------------------------------
gi_255570825_ref_XP_002526365.  ----------------------------------------------
gi_225435251_ref_XP_002284945.  ----------------------------------------------
M.truncatula_AC152176_5.4#1     ----------------------------------------------
AK244805Soybean                 ----------------------------------------------
M.truncatula_AC150891_19.5#1    ----------------------------------------------
gi_223944535_gb_ACN26351.1_     ----------------------------------------------
sorghum                         ----------------------------------------------
Z.mays_ZM07MSbpsHQ_62031266.f0  ----------------------------------------------
gi_226493325_ref_NP_001140473.  ----------------------------------------------
H.vulgare_TA46195_4513#1        ----------------------------------------------
O.sativa_Os05g0125000#1         ----------------------------------------------
P.trichocarpa_scaff_VIII.362#1  ----------------------------------------------
P.trichocarpa_scaff_X.2133#1    ----------------------------------------------
S.bicolor_5288383#1             ----------------------------------------------
S.bicolor_5285414#1             ----------------------------------------------
O.sativa_Os03g0719100#1         ----------------------------------------------
P.patens_159214#1               ----------------------------------------------
P.patens_165698#1               RVVPLVLARTTIQKGAEIQTGDCMLNLGWAVFVKLHSPEGFGLSGD
P.patens_159935#1               ----------------------------------------------
AT5G41580"                      ----------------------------------------------
AT1G08910.1EMB3001protein       ----------------------------------------------
C.vulgaris_83729#1              ----------------------------------------------
```

FIGURE 9 (continued)

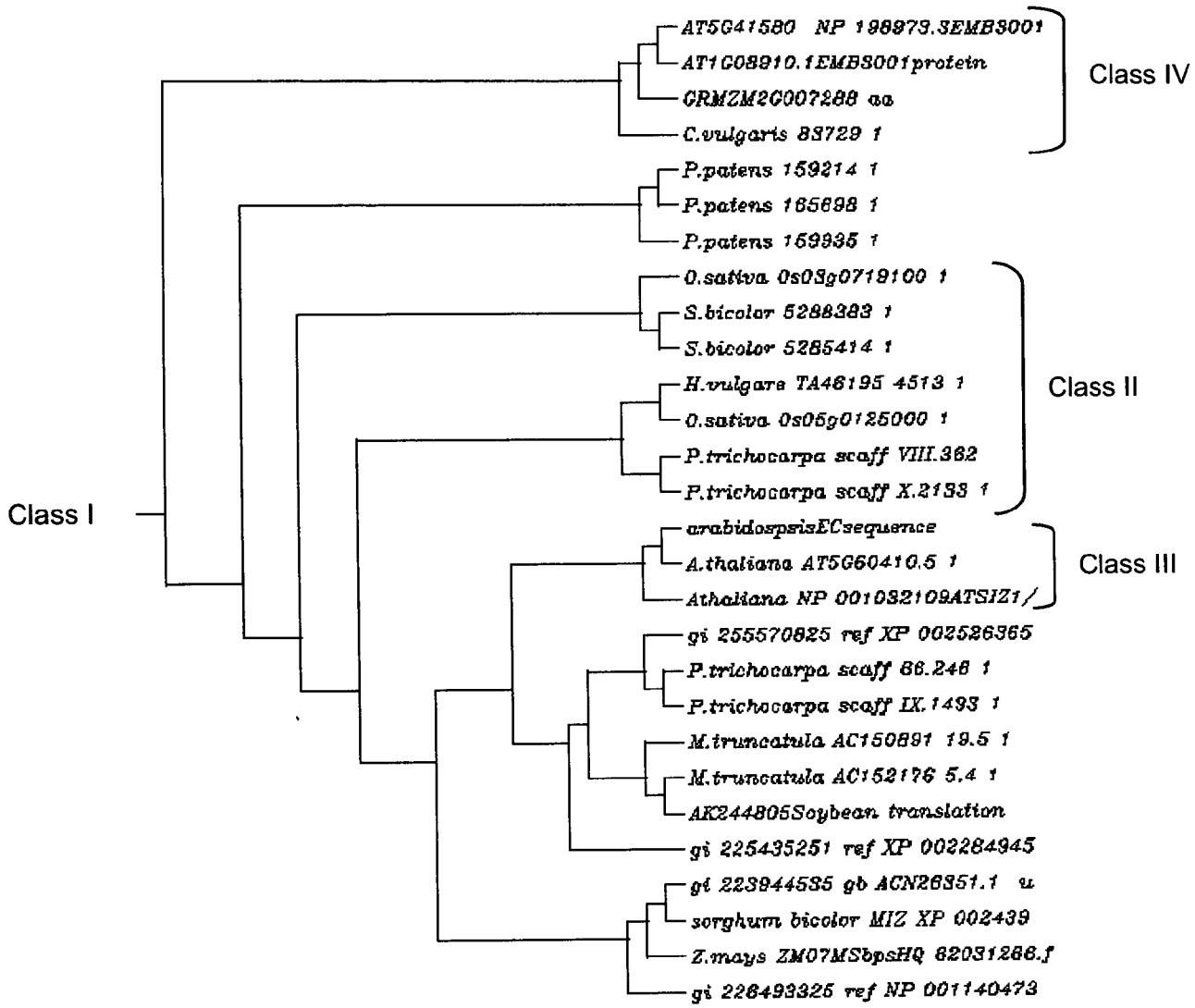
AT5G41580 NP 198973.3EMBS001
AT1G08910.1EMBS001protein
GRMZM2G007288 aa
C.vulgaris 83729 1
P.patens 159214 1
P.patens 165698 1
P.patens 159935 1
O.sativa Os03g0719100 1
S.bicolor 5288383 1
S.bicolor 5285414 1
H.vulgare TA46195 4513 1
O.sativa Os05g0125000 1
P.trichocarpa scaff VIII.362
P.trichocarpa scaff X.2133 1
arabidopsisECsequence
A.thaliana AT5G60410.5 1
Athaliana NP 001032109ATSIZ1/
gi 255570825 ref XP 002526365
P.trichocarpa scaff 66.246 1
P.trichocarpa scaff IX.1493 1
M.truncatula AC150891 19.5 1
M.truncatula AC152176 5.4 1
AK244805Soybean translation
gi 225435251 ref XP 002284945
gi 223944535 gb ACN26351.1 u
sorghum bicolor MIZ XP 002439
Z.mays ZM07MSbpsHQ 62031286.f
gi 226493325 ref NP 001140473
Class IV
Class II
Class III
Class I

FIGURE 10

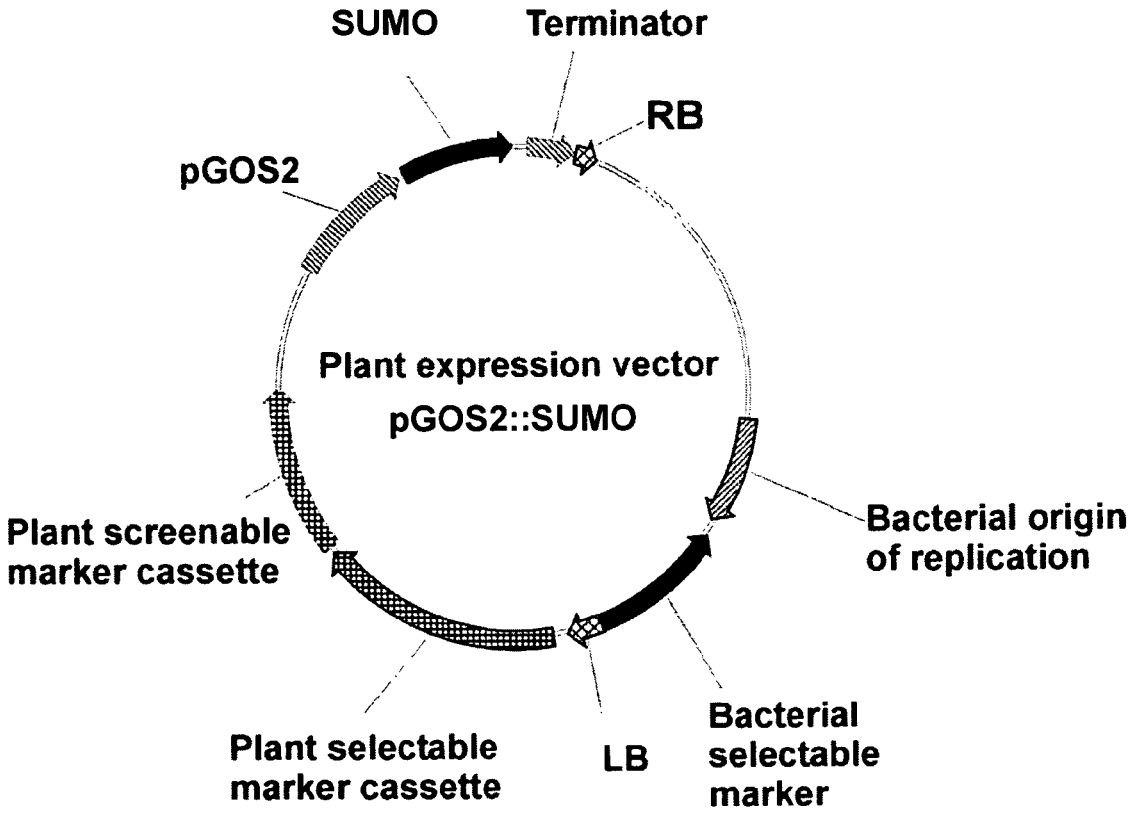
SUMO
Terminator
RB
pGOS2
Plant expression vector
pGOS2::SUMO
Plant screenable
marker cassette
Bacterial origin
of replication
Plant selectable
marker cassette
LB
Bacterial
selectable
marker

FIGURE 11

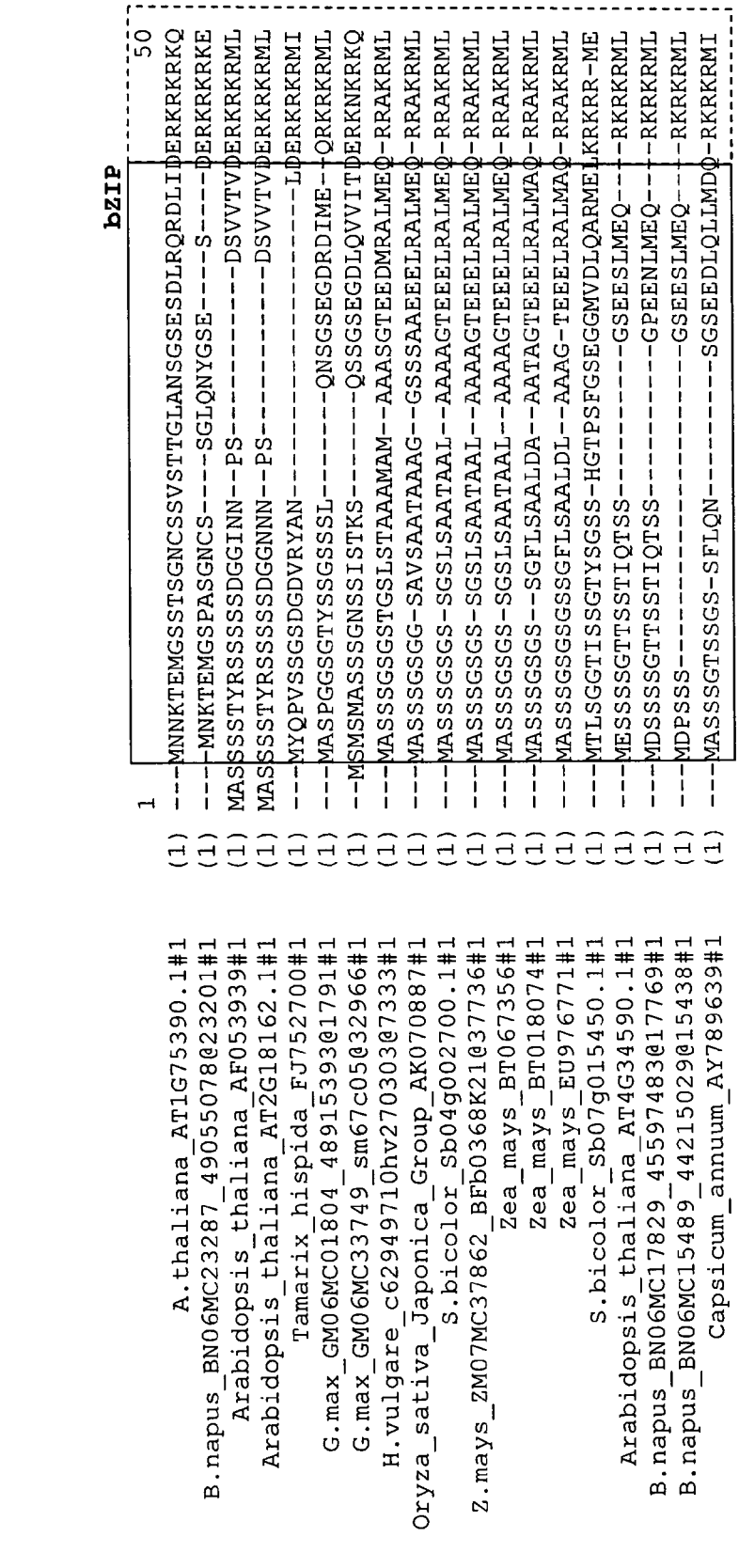
bZIP
1
50
A.thaliana_AT1G75390.1#1 (1) ---MNNKTEMGSSTSGNCSSVSTTGLANSGSESDLRQRDLIDERKRKRKQ
B.napus_BN06MC23287_490550780232O1#1 (1) ---MNKTEMGSPASGNCS----SGLQNYGSE----S----DERKRKRKE
Arabidopsis_thaliana_AF053939#1 (1) MASSSSTYRSSSSSSDGGINN--PS----------DSVVTVDERKRKRML
Arabidopsis_thaliana_AT2G18162.1#1 (1) MASSSTYRSSSSSSDGGNNN--PS----------DSVVTVDERKRKRML
Tamarix_hispida_FJ752700#1 (1) ---MYQPVSSGSDGDVRYAN----------LDERKRKRMI
G.max_GM06MC01804_48915393@1791#1 (1) ---MASPGGSGTYSSGSSSL----QNSGSEGDRDIME--QRKRKRML
G.max_GM06MC33749_sm67c05@32966#1 (1) --MSMSMASSSGNSSISTKS----QSSGSEGDLQVVITDERKNKRKQ
H.vulgare_c62949710hv270303@7333#1 (1) ---MASSSGSGSTGSLSTAAAMAM--AAASGTEEDMRALMEQ-RRAKRML
Oryza_sativa_Japonica_Group_AK070887#1 (1) ---MASSSGSGG-SAVSAATAAAG--GSSSAAEEELRALMEQ-RRAKRML
S.bicolor_Sb04g002700.1#1 (1) ---MASSSGSGS-SGSLSAATAAL--AAAAGTEEELRALMEQ-RRAKRML
Z.mays_ZM07MC37862_BFb0368K21@37736#1 (1) ---MASSSGSGS-SGSLSAATAAL--AAAAGTEEELRALMEQ-RRAKRML
Zea_mays_BT067356#1 (1) ---MASSSGSGS--SGFLSAALDA--AATAGTEELRALMAQ-RRAKRML
Zea_mays_BT018074#1 (1) ---MASSSGSGSGSGFLSAALDL--AAAG-TEELRALMAQ-RRAKRML
Zea_mays_EU976771#1 (1) ---MTLSGGTISSGTYSGSS-HGTPSFGSEGGMVDLQARMELKRKRR-ME
S.bicolor_Sb07g015450.1#1 (1) ---MESSSSGTTSSTIQTSS----------GSEESLMEQ----RKRKRML
Arabidopsis_thaliana_AT4G34590.1#1 (1) ---MDSSSSGTTSSTIQTSS----------GPEENLMEQ----RKRKRML
B.napus_BN06MC17829_45597483@17769#1 (1) ---MDPSSS----------GSEESLMEQ----RKRKRML
B.napus_BN06MC15489_44215029@15438#1 (1) ---MASSSGTSSGS-SFLQN----------SGSEEDLQLLMDQ-RKRKRMI
Capsicum_annuum_AY789639#1

FIGURE 12

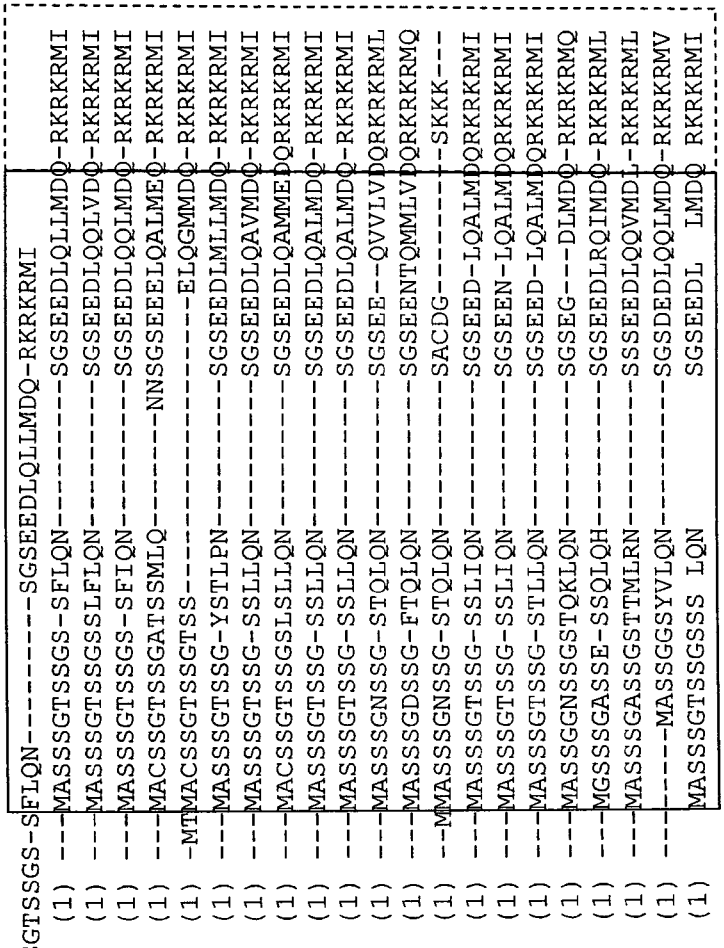

FIGURE 12 (continued)

```
                                              51                                              100
               A.thaliana_AT1G75390.1#1  (48) SNRESARRS--RMRKQKHLDDLTAQVTHLRKENAQIVAGIAVTTQHYVTI
   B.napus_BN06MC23287_49055078@23201#1  (35) SNRESARRS--RMRKQKHLDDLTAQIAHFLEENSHIVAGISVTTQQYVTI
        Arabidopsis_thaliana_AF053939#1  (38) SNRESARRS--RVRKQEHVDDLTAQINQLSNDNRQILNSLTVTSQLYMKI
     Arabidopsis_thaliana_AT2G18162.1#1  (38) SNRESARRS--RMRKQKHVDDLTAQINQLSNDNRQILNSLTVTSQLYMKI
             Tamarix_hispida_FJ752700#1  (28) SNRESARRS--RMRKQQHLGDLLNQVSKLQAENSQFVAKINSASQMYVKV
     G.max_GM06MC01804_48915393@1791#1  (39) SNRESARRS--RMRKQQHLEGLSAQLDQLKKENTQMNTNIGISTQLYLNV
     G.max_GM06MC33749_sm67c05@32966#1  (42) SNRESARRS--RMRKRNHLDQLTKQLSQLAKNNGEILATIDITTQHYLNV
   H.vulgare_c62949710hv270303@7333#1  (45) SNRESARRS--RMRKQRHLDDLAAQAAHLRRENAHVAAALGLTARGLLAV
 Oryza_sativa_Japonica_Group_AK070887#1  (44) SNRESARRS--RMRKQRHLDDLTAQVAHLRRENAHVATALGLTTQGLLAV
            S.bicolor_Sb04g002700.1#1  (44) SNRESARRS--RMRKQRHLDELTAQAAHLRRENAHVATALGLTTQGLLAV
  Z.mays_ZM07MC37862_BFb0368K21@37736#1  (44) SNRESARRS--RMCKQRHLDELTAQAAHLRRENAHVATALGLTAQGLLAV
                  Zea_mays_BT067356#1  (44) SNRESARRS--RMRKQRHLDELTAQAAHLRRENAHVATALGLTAQGLLAV
                  Zea_mays_BT018074#1  (43) SNRESARRS--RMRKQRHLDELTAQAAHLRRENAHVATALGLTAQGLLAV
                  Zea_mays_EU976771#1  (44) SNRESARRS--RMRKQRHLDELTAQAAHLRRENAHVATALGLTAQGLLAV
            S.bicolor_Sb07g015450.1#1  (46) SNRESAKRS--RQRKQQHLDDLNLQVDKLRTTKQQLMTALNITTQNYTAA
     Arabidopsis_thaliana_AT4G34590.1#1  (34) SNRESARRS--RMKKQKLLDDLTAQVNHLKKENTEIVTSVSITTQHYLTV
   B.napus_BN06MC17829_45597483@17769#1  (34) SNRESARRS--RKKKQKLLDDLTAQVNQLRKENSEIVTSVSITTQHYLTV
   B.napus_BN06MC15489_44215029@15438#1  (23) SNRESARRS--RMKKQKLLDDLTAQVNRLKEQNNEILTSVSITTQHYLTV
             Capsicum_annuum_AY789639#1  (37) SNRESARRS--RMRKQKHLNDLMAQVSTLRKENDQILTSMNVTTQHYLNV
```

FIGURE 12 (continued)

```
                          Capsicum_chinense_AF430372#1 (37) SNRESARRS--RMRKQKHLDDLMAQVSTLRKENDQILTSMNVTTQHYLNV
                          Capsicum_chinense_AF127797#1 (37) SNRESARRS--RMRKQKHLDDLMAQVSTLRKENDQILTSMNVTTQHYLNV
                          Nicotiana_tabacum_AY045570#1 (38) SNRESARRS--RMRKQKHLDDLMAQVATLRKENNQILTSMNVTTQHYLNV
                       Solanum_lycopersicum_FJ647190#1 (37) SNRESARRS--RMRKQKHLDDLMSQVTNLRKENNQILTSMNVTTQHYLNV
                  G.max_GM06MC17143_59654278@16848#1 (40) SNRESARRS--RMRKQKHLDDLASQVTQLRNENHQILTSVNLTTQKYLAV
                  G.max_GM06MC32046_sj86f07@31311#1 (31) SNRESARRS--RMRKQKHLDDLASQLTQLRSQNQQLLTSVNLTSHKYLAV
                                              Mt_bZIP2 (37) SNRESARRS--RMRKQKHLDDLAVQLSQLRNENQQILTSVNLTTQRFLAV
                  G.max_GM06MC32426_sk55g01@31681#1 (37) SNRESARRS--RMRKQKHLDDLVSQVAQLRKENQQILTSVNITTQQYLSV
                             Glycine_max_AF532621#1 (39) SNRESARRS--RMRKQKHLDDLVSQVAQLRKENQQILTSVNITTQQYLSV
                      Medicago_truncatula_BT053497#1 (37) SNRESARRS--RMRKQKHLDDLVSQVSKLRKENQEILTSVNITTQKYLSV
                                               Mt_bZIP (37) SNRESARRS--RMRKQKHLDDLVSQVSKLRKENQEILTSVNITTQKYLSV
                              P.trichocarpa_710131#1 (36) SNRESARRS--RMRKQKYLGDLMAQVAQLRTDNNQILTTINVTTQHFLNV
                              P.trichocarpa_715285#1 (38) SNRESARRS--RMKKQKHLDDLMAQVTQLRKDNNQILTTINVTTQHYLNV
                    V.vinifera_GSVIVT00014558001#1 (27) --EENAIKSGIRKAKQKHLDDLMAQVAQLRKENNEILSSINITNQRYLTV
                              P.trichocarpa_818112#1 (37) SNRESARRS--RMRKQKHLDDLMAQVSQLRKENHQIITGINITTQRYLSV
                       Populus_trichocarpa_EF147315#1 (37) SNRESARRS--RMRKQKHLDDLVAQVAQLKKENHQIITSINITTQHYLNV
                    V.vinifera_GSVIVT00036338001#1 (37) SNRESARRS--RMRKQKHLDDLMAQAAQLRKENSQIITSMNVTTQHYFNI
                    Petroselinum_crispum_AJ292745#1 (35) SNRESARRS--RQRKQNHLDELMAQAAQLRKENNQIITTTNLTTQQFVKV
                    V.vinifera_GSVIVT00036899001#1 (37) SNRESARRS--RMRKQKHLDDMMAQMVHLRKENNRILTTMNVTTQFHMNV
                              P.trichocarpa_719591#1 (38) SNRESARRS--RVKKQKHLDDLMGQLGQLSKENNEILKRMNVTSQLYMNI
T.erecta_SIN_01b-CS_Scarletade-12-L23.b1@917#1 (33) SNRESARRS--RMKKQKHLDDLMTQLSQLKKENNEIMAHVSITMQHYTNM
                                             Consensus (51) SNRESARRS  RMRKQKHLDDLMAQVAQLRKEN QILTSVNITTQ YL V
```

FIGURE 12 (continued)

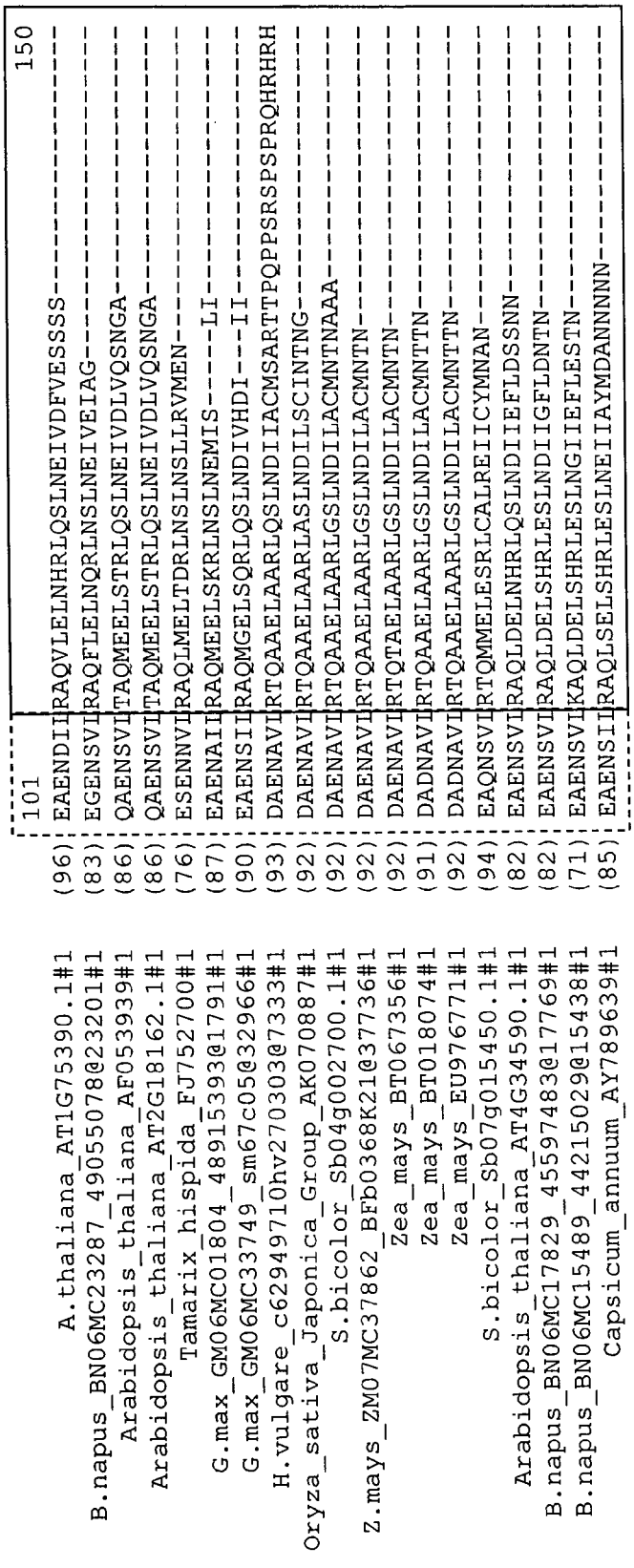

FIGURE 12 (continued)

```
Capsicum_chinense_AF430372#1      (85) EAENSILRAQLSELSHRLESLNEIIAYMDANNNNN----------
Capsicum_chinense_AF127797#1      (85) EAENSILRAQLSELSHRLESLNEIIAYMDANN------------
Nicotiana_tabacum_AY045570#1      (86) EAENSILRAQLAELNHRLESLNEIIAFLDANN------------
Solanum_lycopersicum_FJ647190#1   (85) EAENSILRAQLSELSRRLESLNEIIAVLDANS------------
G.max_GM06MC17143_596542?8@16848#1 (88) EAENSVLRAQVNELSHWLESLNEIIHFLNATDGG---------
G.max_GM06MC32046_sj86f07@31311#1 (79) EAENSVLRAQVNELSHRLDSLNQIIHLLNFFEP----------
Mt_bZIP2                          (85) ESENSVLRAQLNELNSRFESLNEIINFMNVAN------------
G.max_GM06MC32426_sk55g01@31681#1 (85) EAENSVLRAQVGELSHRLESLNEIVDVLNATTV----------
Glycine_max_AF532621#1            (87) EAENSVLRAQVGELSHRLESLNEIVDVLNATTT----------
Medicago_truncatula_BT053497#1    (85) EAENSVLRAQMGD-------------------------------
Mt_bZIP                           (85) EAENSVLRAQMGELSNRLESLNEIVGALNSSNG----------
P.trichocarpa_710131#1            (84) EAENSILRAQMMELNHRLDSLNEILNYINTSNG----------
P.trichocarpa_715285#1            (86) EAENSILRAQMMELNHRLDSLNEILNYINTSNG----------
V.vinifera_GSVIVT00014558001#1    (75) EADNSILRAQAMELSHRYQSLNDILNYMNTSNG----------
P.trichocarpa_818112#1            (85) EADNSILRVQISELSNRLESLNEIIGSLNSNNG----------
Populus_trichocarpa_EF147315#1    (85) EADNSILRAQVSELSHRLEFLNGIISLLNSSNG----------
V.vinifera_GSVIVT00036338001#1    (85) EAENSVLRAQFSELSNRLQYLVEIISFLNTSNG----------
Petroselinum_crispum_AJ292745#1   (83) EAENSVLRAQMDELTQRLQSLNDILHYINTTTTAA---------
V.vinifera_GSVIVT00036899001#1    (85) EAENAILRAQMAELTLRLQTLNEIMDYLNSSN-----------
P.trichocarpa_719591#1            (86) EAENSILRAQMAELSHRLNSLNEIIEYVNFCSG----------
T.erecta_SIN_01b-CS_Scarletade-12-L23.b1@917#1 (81) EAENLVLRAQVAELSHRLESLDQIKEFMSQPVD----------
Consensus                         (101) EAENSVLRAQM ELS RL SLNEII  LNT
```

FIGURE 12 (continued)

```
                                                   151                                              200
           A.thaliana_AT1G75390.1#1      (129) ------------GFGMETGQG-------LFDGGLFDGVMN--PMNLGFYN
  B.napus_BN06MC23287_49055078@23201#1   (112) ------------GFWMETGQGGGVIGYGGGGGGFYDGVMN--PLNLGFYN
      Arabidopsis_thaliana_AF053939#1    (119) ------------GFGVDQIDGCGFDDRTVGIDGYYDDMNM--MSNVNHWG
    Arabidopsis_thaliana_AT2G18162.1#1   (119) ------------GFGVDQIDGCGFDDRTVGIDGYYDDMNM--MSNVNHWG
            Tamarix_hispida_FJ752700#1   (106) ----------------VSGLN-------ADVEELPDALLE--PWQS---P
    G.max_GM06MC01804_48915393@1791#1    (115) ------------NSTTTTNNCLMFDEAQETTTQLFNDCGF--MDAWNYGI
    G.max_GM06MC33749_sm67c05@32966#1    (120) ------------NTTTTYERDCYLTSAQNFTTTTYTLCLN--QQAVFQW-
   H.vulgare_c62949710hv270303@7333#1    (143) RPVRLRRPVARRLPLQILPRDVHVPSPPKLAPTAMAWQTTAASTRSSERA
Oryza_sativa_Japonica_Group_AK070887#1   (124) ----------------AAAAAAAALTVAAATDPLLAFDSAA--FDDLFRS
          S.bicolor_Sb04g002700.1#1      (126) -----------AVGAVAVSVSLTAAAASSS-DPYLAFDGAT-ALDDLLRS
 Z.mays_ZM07MC37862_BFb0368K21@37736#1   (123) -----------AVGAVAVSLTAAAAAAAASSDPCLAFDGAT-ALDDLLRS
                  Zea_mays_BT067356#1    (123) -----------AVGAVAVSLTAAAAAAASS-DPCLAFDGAT-ALDDLLRS
                  Zea_mays_BT018074#1    (123) -----------------AAAAS--------SDPYLAFDGAATALDDLLKS
                  Zea_mays_EU976771#1    (124) -----------------AAAAAAF------SDPYLAFDGAATALDDLLRS
          S.bicolor_Sb07g015450.1#1      (125) ----------------HVANAATTMNAHPATIMSGAANYNTFGTSATAWN
    Arabidopsis_thaliana_AT4G34590.1#1   (115) ----------------NNNNNMGMCSNPLVGLECDDFFVNQMNMS-YIMN
  B.napus_BN06MC17829_45597483@17769#1   (114) ----------------------GICSNTLSGPESDDFLVSQFNMNMFYMN
  B.napus_BN06MC15489_44215029@15438#1   (103) ----------------------GTCPNPLTGPEPDDFLVNQMNMNMFCMN
           Capsicum_annuum_AY789639#1    (120) -SCSNGLAMDHN--EPYSFNFAQSDTVVDGFN-----MTN--SWNYFCSN
```

FIGURE 12 (continued)

```
Capsicum_chinense_AF430372#1  (120) -SCSNGLAMDHN--EPYSFNFAQSDTVVDGFN-----MTN--SWNYFCSN
Capsicum_chinense_AF127797#1  (117) -SCSNGLAMDHN--EPYSFNFAQSDTVVDGFN-----MTN--SWNYFCSN
Nicotiana_tabacum_AY045570#1  (118) -NCNGLANMDHNQEEPYSFNFAQNEPMVDGFN-----MTN--SWNYLCAN
Solanum_lycopersicum_FJ647190#1 (117) ----G-LVMDHN--EPYSFNFAQNDIMFDGFN-----VTN--SWNYLSAN
G.max_GM06MC17143_596542780@16848#1 (122) ----------PPPPPSSFF----------EPDATFFNKAYLS
G.max_GM06MC32046_sj86f07@31311#1 (112) ----------DAS----------TSTFFNNPFNFS
Mt_bZIP2                      (117) ----------GVFEPVDNNINENYFN----------N--PLNMGYLN
G.max_GM06MC32426_sk55g01@31681#1 (118) ----------AGFGAAATSSTFVEPINNNS----------FFN--PLNMGYLN
Glycine_max_AF532621#1        (120) ----------VAGFGAAASSTFVEPMNNNNNSFF-----NFN--PLNMGYLN
Medicago_truncatula_BT053497#1 (98) ----------
Mt_bZIP                       (118) ----------VFGASNAFVEQNNGFFFNSLN----------NM--SYMN
P.trichocarpa_710131#1        (117) ----------I-------FEIDHHEDLQTS-ADHG-----FMN--PLNLILLN
P.trichocarpa_715285#1        (119) ----------I-------FENDHHEDLP----DHS-----FMN--PSNLFYLN
V.vinifera_GSVIVT00014558001#1 (108) ----------V-------FETEDLPVTVDP----------FMN--PMNYLYLN
P.trichocarpa_818112#1        (118) ----------VFGDSITFNEP-AADS-----FLN--PWNMAYLN
Populus_trichocarpa_EF147315#1 (118) ----------LFGDSSIFNEP-AADS-----FLN--PFNMSYLN
V.vinifera_GSVIVT00036338001#1 (118) ----------G-------FESGEPWTLPEPTTPDS-----LMN--PLSLLYLS
Petroselinum_crispum_AJ292745#1 (118) -AAATAAATTGIDGVFEMDNLFGLDDHQSSYMNNNNNNNN--SWNMMYPN
V.vinifera_GSVIVT00036899001#1 (117) ----------VPFGTEYQGTQIADECF-----MNI--PWVPPFIN
P.trichocarpa_719591#1        (119) ----------TFERHEDAAAPTTGAFGHQLVDDFFMNPWNNADFHLN
T.erecta_SIN_01b-CS_Scarletade-12-L23.b1@917#1 (114) ----------YGGGFVDELYCG-G--GTELMDEFMNN--SLSCLCVN
Consensus                     (151)                                N     N     N
```

FIGURE 12 (continued)

```
                                  201                                 240
          A.thaliana_AT1G75390.1#1 (158) QPIMASASTAG-DVFNC--------------------------
B.napus_BN06MC23287_490550780@23201#1 (148) QTLMASASTVSADVFNCC-----------------------
     Arabidopsis_thaliana_AF053939#1 (155) GSVYTNQPIMANDINMY----------------------------
Arabidopsis_thaliana_AT2G18162.1#1 (155) GSVYTNQPIMANDINMY----------------------------
          Tamarix_hispida_FJ752700#1 (128) CPVQLPVMTNTADMFQF-------------------------
     G.max_GM06MC01804_489153930@1791#1 (151) PLNQQIMAYADNDMLMMY---------------------
     G.max_GM06MC33749_sm67c05@32966#1 (155) ----------------------------------------
     H.vulgare_c62949710hv270303@7333#1 (193) HEERSSIHLFSLMLLVFQVNYVVLKIWRAVEFLLHLTRKAS
Oryza_sativa_Japonica_Group_AK070887#1 (156) SPEL-------FQLC-----------------------
          S.bicolor_Sb04g002700.1#1 (163) CPEN----MYQLC-----------------------------
Z.mays_ZM07MC37862_BFb0368K21@37736#1 (161) CPEN----MFQYQLC-----------------------
               Zea_mays_BT067356#1 (160) CPEN----MFQYQLC-----------------------------
               Zea_mays_BT018074#1 (148) CPDSM--MFEYQLC-----------------------------
               Zea_mays_EU976771#1 (151) CPDSM--MFEYQLC-----------------------------
          S.bicolor_Sb07g015450.1#1 (159) SAMQMVQQPIDHLLYQCF-------------------------
     Arabidopsis_thaliana_AT4G34590.1#1 (148) QPLMASSDALMY---------------------------
B.napus_BN06MC17829_455974830@17769#1 (142) QPLMASSDALLY---------------------------
B.napus_BN06MC15489_442150290@15438#1 (131) QPLMASSDALLY---------------------------
          Capsicum_annuum_AY789639#1 (160) QPIMTAHVLQY----------------------------
```

FIGURE 12 (continued)

```
Capsicum_chinense_AF430372#1 (160) QPIMTAHVLQY----------
Capsicum_chinense_AF127797#1 (157) QPIMTAHVLQY----------
Nicotiana_tabacum_AY045570#1 (160) QPIMTADVLQY----------
Solanum_lycopersicum_FJ647190#1 (153) QPIMTADVLQY----------
G.max_GM06MC17143_596542780@16848#1 (144) QPIMASADMWQD----------
G.max_GM06MC32046_sj86f07@31311#1 (127) LPIMASADMLQY----------
Mt_bZIP2 (142) QPIMASADMNMIHY----------
G.max_GM06MC32426_sk55g01@31681#1 (149) HPIMASADILQY----------
Glycine_max_AF532621#1 (155) QPIMASADILQY----------
Medicago_truncatula_BT053497#1 (98) ----------
Mt_bZIP (145) QPIMASADILQY----------
P.trichocarpa_710131#1 (145) QPIMASPDLFQY----------
P.trichocarpa_715285#1 (144) QPIMASPDLFQY----------
V.vinifera_GSVIVT00014558001#1 (132) QPIIASVDMFPY----------
P.trichocarpa_818112#1 (144) QPIMASAEMFHY----------
Populus_trichocarpa_EF147315#1 (144) QPISASADMFQY----------
V.vinifera_GSVIVT00036338001#1 (147) QPIMAS-DIFQY----------
Petroselinum_crispum_AJ292745#1 (165) QPIMADMFMY----------
V.vinifera_GSVIVT00036899001#1 (145) QPIMASADGMFMY----------
P.trichocarpa_719591#1 (156) KPIMD---MIMY----------
T.erecta_SIN_01b-CS_Scarletade-12-L23.b1@917#1 (146) KPILASSDMIQY----------
Consensus (201) QPIMAS  M  Y
```

FIGURE 12 (continued)

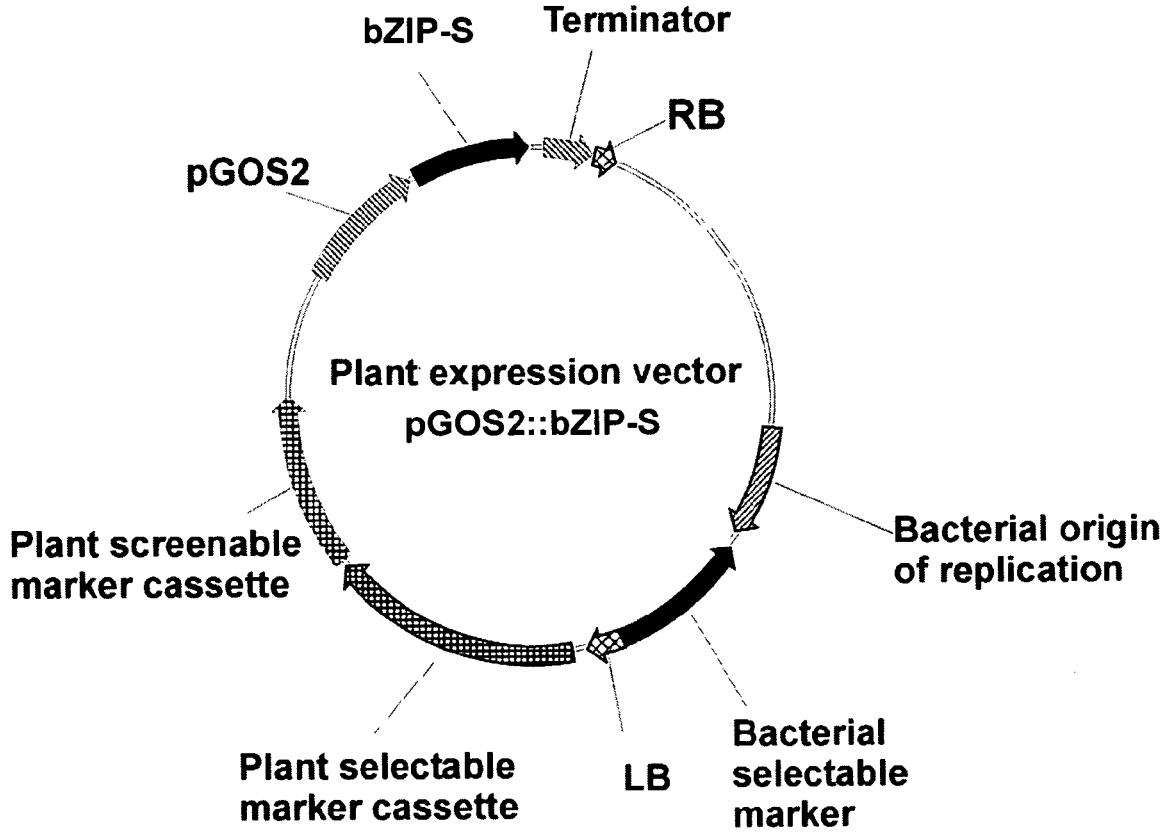
bZIP-S
Terminator
RB
pGOS2
Plant expression vector
pGOS2::bZIP-S
Plant screenable
marker cassette
Bacterial origin
of replication
Plant selectable
marker cassette
LB
Bacterial
selectable
marker

FIGURE 13

```
                          1                                                  50
LOC_Os05g05600.1     (1) MATRIPGTVAASGVYYNDQYRMPCKLKGIHCMALNCIPQKAKVRKCMNGY
Os_SPA15             (1) MATRIPGTVAASGVYYNDQYRMPCKLKGIHCMALNCIPQKAKVRKCMNGY
Consensus            (1) MATRIPGTVAASGVYYNDQYRMPCKLKGIHCMALNCIPQKAKVRKCMNGY

                         51                                                 100
LOC_Os05g05600.1    (51) QSTFRFCVNEKNGQTTGQSNGSLIQQGQNFRCHSYGSHNSSETKECSLED
Os_SPA15            (51) QSTFRFCVNEKNGQTTGQSNGSLIQQGQNFRCHSYGSHNSSETKECSLED
Consensus           (51) QSTFRFCVNEKNGQTTGQSNGSLIQQGQNFRCHSYGSHNSSETKECSLED

                        101                                                 150
LOC_Os05g05600.1   (101) GTDSYRDFEEHSRGASQFSDNQVAAKKKSVKSSQGLAEACKFVYNDAKFV
Os_SPA15           (101) GTDSYRDFEEHSRGASQFSDNQVAAKKKSVKSSQGLAEACKFVYNDAKFV
Consensus          (101) GTDSYRDFEEHSRGASQFSDNQVAAKKKSVKSSQGLAEACKFVYNDAKFV

                        151                                                 200
LOC_Os05g05600.1   (151) NERAQNDILLLSRGITRLNKRACQDVAVLGSGFLKLDARARKDTKKIDHS
Os_SPA15           (151) NERAQNDILLLSRGITRLNKRACQDVAVLGSGFLKLDARARKDTKKIDHS
Consensus          (151) NERAQNDILLLSRGITRLNKRACQDVAVLGSGFLKLDARARKDTKKIDHS

                        201                                                 250
LOC_Os05g05600.1   (201) VKERAARLTHFARILKEQAQSDLKKAADQHWSDGALEADLRRADSVVRRR
Os_SPA15           (201) VKERAARLTHFARILKEQAQSDLKKAADQHWSDGALEADLRRADSVVRRR
Consensus          (201) VKERAARLTHFARILKEQAQSDLKKAADQHWSDGALEADLRRADSVVRRR

                        251                                                 300
LOC_Os05g05600.1   (251) AMEDAFMALKFVRDIHDMMANRLQEQFAKDGSSSPANSRSFITLEKNGNT
Os_SPA15           (251) AMEDAFMALKFVRDIHDMMANRLQEQFAKDGSSSPANSRSFITLEKNGNT
Consensus          (251) AMEDAFMALKFVRDIHDMMANRLQEQFAKDGSSSPANSRSFITLEKNGNT

                        301                                                 350
LOC_Os05g05600.1   (301) FELFPHEVSTDQITAIEQAYWSMASALSEADGIDYTDPEELELLVATLID
Os_SPA15           (301) FELFPHEVSTDQITAIEQAYWSMASALSEADGIDYTDPEELELLVATLID
Consensus          (301) FELFPHEVSTDQITAIEQAYWSMASALSEADGIDYTDPEELELLVATLID

                        351                                                 400
LOC_Os05g05600.1   (351) LDAMDGKKSVSLLAECSSSPDVNTRKALANALAAAPSMWILGNAGMGALQ
Os_SPA15           (351) LDAMDGKKSVSLLAECSSSPDVNTRKALANALAVAPSMWILGNAGMGALQ
Consensus          (351) LDAMDGKKSVSLLAECSSSPDVNTRKALANALA APSMWILGNAGMGALQ

                        401                                                 450
LOC_Os05g05600.1   (401) RLAQDSNYAVARAATRAINELTKQWELEEGDSLRFVLNQNMVSKETADDS
Os_SPA15           (401) RLAQDSNYAVARAATRAINELTKQWELEEGDSLRFVLNQNMVSKETADDS
Consensus          (401) RLAQDSNYAVARAATRAINELTKQWELEEGDSLRFVLNQNMVSKETADDS

                        451
LOC_Os05g05600.1   (451) AAADDTR
Os_SPA15           (451) AAADDTR
Consensus          (451) AAADDTR
```

FIGURE 14

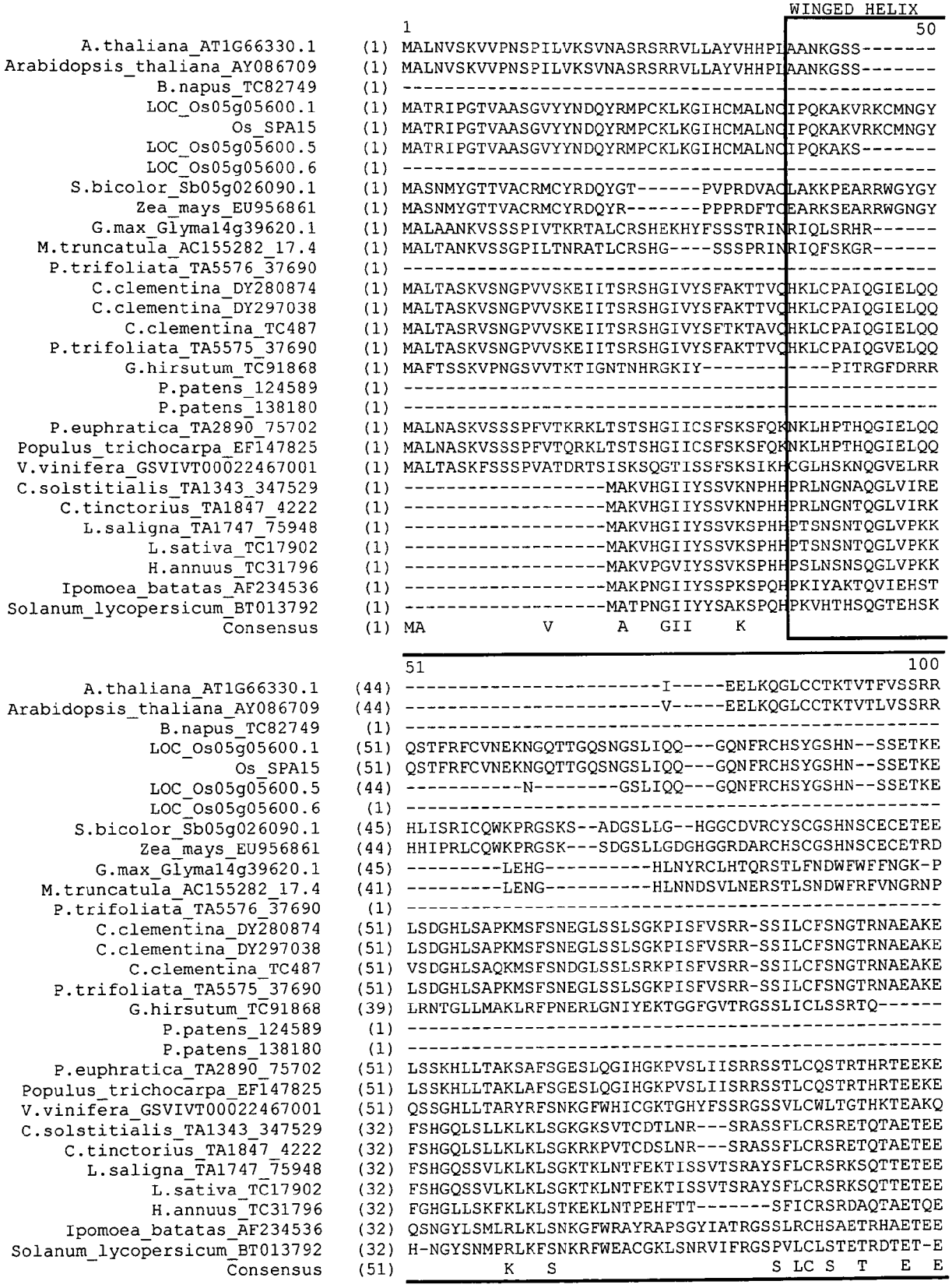
WINGED HELIX
1 50
A.thaliana_AT1G66330.1 (1) MALNVSKVVPNSPILVKSVNASRSRRVLLAYVHHPLAANKGSS-------
Arabidopsis_thaliana_AY086709 (1) MALNVSKVVPNSPILVKSVNASRSRRVLLAYVHHPLAANKGSS-------
B.napus_TC82749 (1) ----------
LOC_Os05g05600.1 (1) MATRIPGTVAASGVYYNDQYRMPCKLKGIHCMALNCIPQKAKVRKCMNGY
Os_SPA15 (1) MATRIPGTVAASGVYYNDQYRMPCKLKGIHCMALNCIPQKAKVRKCMNGY
LOC_Os05g05600.5 (1) MATRIPGTVAASGVYYNDQYRMPCKLKGIHCMALNCIPQKAKS-------
LOC_Os05g05600.6 (1) ----------
S.bicolor_Sb05g026090.1 (1) MASNMYGTTVACRMCYRDQYGT------PVPRDVACLAKKPEARRWGYGY
Zea_mays_EU956861 (1) MASNMYGTTVACRMCYRDQYR-------PPPRDFTCEARKSEARRWGNGY
G.max_Glyma14g39620.1 (1) MALAANKVSSSPIVTKRTALCRSHEKHYFSSSTRINRIQLSRHR------
M.truncatula_AC155282_17.4 (1) MALTANKVSSGPILTNRATLCRSHG----SSSPRINRIQFSKGR------
P.trifoliata_TA5576_37690 (1) ----------
C.clementina_DY280874 (1) MALTASKVSNGPVVSKEIITSRSHGIVYSFAKTTVQHKLCPAIQGIELQQ
C.clementina_DY297038 (1) MALTASKVSNGPVVSKEIITSRSHGIVYSFAKTTVQHKLCPAIQGIELQQ
C.clementina_TC487 (1) MALTASRVSNGPVVSKEIITSRSHGIVYSFTKTAVQHKLCPAIQGIELQQ
P.trifoliata_TA5575_37690 (1) MALTASKVSNGPVVSKEIITSRSHGIVYSFAKTTVQHKLCPAIQGVELQQ
G.hirsutum_TC91868 (1) MAFTSSKVPNGSVVTKTIGNTNHRGKIY----------PITRGFDRRR
P.patens_124589 (1) ----------
P.patens_138180 (1) ----------
P.euphratica_TA2890_75702 (1) MALNASKVSSSPFVTKRKLTSTSHGIICSFSKSFQKNKLHPTHQGIELQQ
Populus_trichocarpa_EF147825 (1) MALNASKVSSSPFVTQRKLTSTSHGIICSFSKSFQKNKLHPTHQGIELQQ
V.vinifera_GSVIVT00022467001 (1) MALTASKFSSSPVATDRTSISKSQGTISSFSKSIKHCGLHSKNQGVELRR
C.solstitialis_TA1343_347529 (1) ----------MAKVHGIIYSSVKNPHHPRLNGNAQGLVIRE
C.tinctorius_TA1847_4222 (1) ----------MAKVHGIIYSSVKNPHHPRLNGNTQGLVIRK
L.saligna_TA1747_75948 (1) ----------MAKVHGIIYSSVKSPHHPTSNSNTQGLVPKK
L.sativa_TC17902 (1) ----------MAKVHGIIYSSVKSPHHPTSNSNTQGLVPKK
H.annuus_TC31796 (1) ----------MAKVPGVIYSSVKSPHHPSLNSNSQGLVPKK
Ipomoea_batatas_AF234536 (1) ----------MAKPNGIIYSSPKSPQHPKIYAKTQVIEHST
Solanum_lycopersicum_BT013792 (1) ----------MATPNGIIYYSAKSPQHPKVHTHSQGTEHSK
Consensus (1) MA V A GII K
51 100
A.thaliana_AT1G66330.1 (44) ----------I-----EELKQGLCCTKTVTFVSSRR
Arabidopsis_thaliana_AY086709 (44) ----------V-----EELKQGLCCTKTVTLVSSRR
B.napus_TC82749 (1) ----------
LOC_Os05g05600.1 (51) QSTFRFCVNEKNGQTTGQSNGSLIQQ---GQNFRCHSYGSHN--SSETKE
Os_SPA15 (51) QSTFRFCVNEKNGQTTGQSNGSLIQQ---GQNFRCHSYGSHN--SSETKE
LOC_Os05g05600.5 (44) ----------N--------GSLIQQ---GQNFRCHSYGSHN--SSETKE
LOC_Os05g05600.6 (1) ----------
S.bicolor_Sb05g026090.1 (45) HLISRICQWKPRGSKS--ADGSLLG--HGGCDVRCYSCGSHNSCECETEE
Zea_mays_EU956861 (44) HHIPRLCQWKPRGSK---SDGSLLGDGHGGRDARCHSCGSHNSCECETRD
G.max_Glyma14g39620.1 (45) ---------LEHG----------HLNYRCLHTQRSTLFNDWFWFFNGK-P
M.truncatula_AC155282_17.4 (41) ---------LENG----------HLNNDSVLNERSTLSNDWFRFVNGRNP
P.trifoliata_TA5576_37690 (1) ----------
C.clementina_DY280874 (51) LSDGHLSAPKMSFSNEGLSSLSGKPISFVSRR-SSILCFSNGTRNAEAKE
C.clementina_DY297038 (51) LSDGHLSAPKMSFSNEGLSSLSGKPISFVSRR-SSILCFSNGTRNAEAKE
C.clementina_TC487 (51) VSDGHLSAQKMSFSNDGLSSLSRKPISFVSRR-SSILCFSNGTRNAEAKE
P.trifoliata_TA5575_37690 (51) LSDGHLSAPKMSFSNEGLSSLSGKPISFVSRR-SSILCFSNGTRNAEAKE
G.hirsutum_TC91868 (39) LRNTGLLMAKLRFPNERLGNIYEKTGGFGVTRGSSLICLSSRTQ------
P.patens_124589 (1) ----------
P.patens_138180 (1) ----------
P.euphratica_TA2890_75702 (51) LSSKHLLTAKSAFSGESLQGIHGKPVSLIISRRSSTLCQSTRTHRTEEKE
Populus_trichocarpa_EF147825 (51) LSSKHLLTAKLAFSGESLQGIHGKPVSLIISRRSSTLCQSTRTHRTEEKE
V.vinifera_GSVIVT00022467001 (51) QSSGHLLTARYRFSNKGFWHICGKTGHYFSSRGSSVLCWLTGTHKTEAKQ
C.solstitialis_TA1343_347529 (32) FSHGQLSLLKLKLSGKGKSVTCDTLNR---SRASSFLCRSRETQTAETEE
C.tinctorius_TA1847_4222 (32) FSHGQLSLLKLKLSGKRKPVTCDSLNR---SRASSFLCRSRETQTAETEE
L.saligna_TA1747_75948 (32) FSHGQSSVLKLKLSGKTKLNTFEKTISSVTSRAYSFLCRSRKSQTTETEE
L.sativa_TC17902 (32) FSHGQSSVLKLKLSGKTKLNTFEKTISSVTSRAYSFLCRSRKSQTTETEE
H.annuus_TC31796 (32) FGHGLLSKFKLKLSTKEKLNTPEHFTT-------SFICRSRDAQTAETQE
Ipomoea_batatas_AF234536 (32) QSNGYLSMLRLKLSNKGFWRAYRAPSGYIATRGSSLRCHSAETRHAETEE
Solanum_lycopersicum_BT013792 (32) H-NGYSNMPRLKFSNKRFWEACGKLSNRVIFRGSPVLCLSTETRDTET-E
Consensus (51) K S S LC S T E E

FIGURE 15

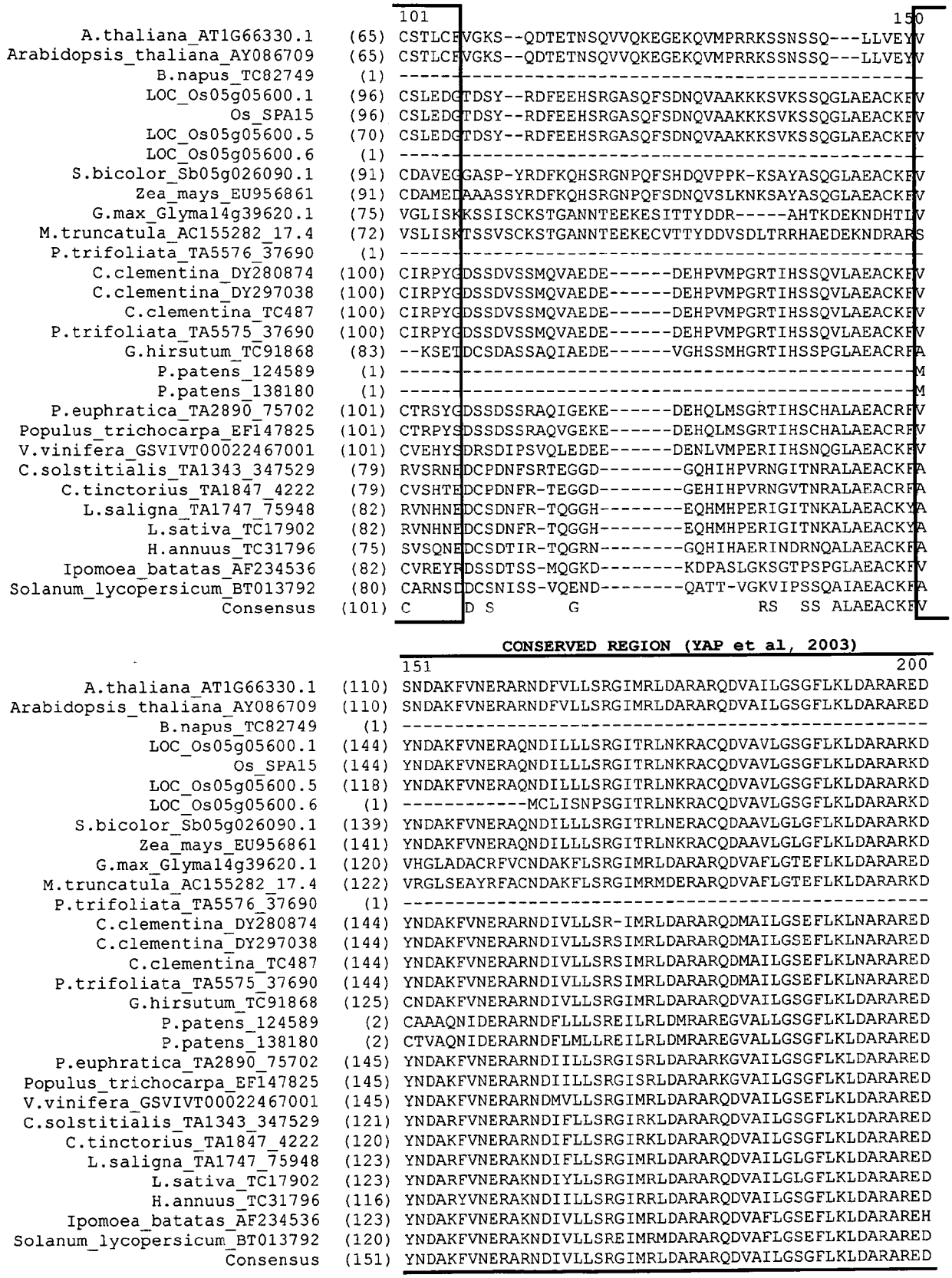

FIGURE 15 (continued)

```
                                        201                                              250
         A.thaliana_AT1G66330.1  (160)  TEKIDRDVKRKAERLHHIATIFKNIAESKLKNAADKHWSDGALEADLRRA
 Arabidopsis_thaliana_AY086709  (160)  TEKIDRDVKRKAERLHHIATILKNIAESKLKNAADKHWSDGALEADLRRA
                B.napus_TC82749    (1)  --------------------------------------------------
             LOC_Os05g05600.1  (194)  TKKIDHSVKERAARLTHFARILKEQAQSDLKKAADQHWSDGALEADLRRA
                      Os_SPA15  (194)  TKKIDHSVKERAARLTHFARILKEQAQSDLKKAADQHWSDGALEADLRRA
             LOC_Os05g05600.5  (168)  TKKIDHSVKERAARLTHFARILKEQAQSDLKKAADQHWSDGALEADLRRA
             LOC_Os05g05600.6   (39)  TKKIDHSVKERAARLTHFARILKEQAQSDLKKAADQHWSDGALEADLRRA
       S.bicolor_Sb05g026090.1  (189)  TQKIDHTVKERAARLNHFARAFKERAQSDLKKAADKHWSDGALEADLRRA
              Zea_mays_EU956861  (191)  TQKIDHTVKERAALLNHFARAFKERAQSDLKKAADKHWSDGALEADLRRA
         G.max_Glyma14g39620.1  (170)  TEKIDRDVKEKASRLSHIATILKDKAQSRLKNAADEHWSDGALEADLRLA
     M.truncatula_AC155282_17.4  (172)  TEKIDRGVKEKAKRLNRIATILKDIAQTRLKSAADEHWSDGALEADLRLA
     P.trifoliata_TA5576_37690    (1)  --------------------------------------------------
       C.clementina_DY280874  (193)  TEKIDRDVKQKAECLHHIATILKDKAESRLKHAADKHWSDGALEADLRRA
       C.clementina_DY297038  (194)  TEKIDRDVKQKAECLHHIATILKDKAESRLKHAADKHWSDGALEADLRRA
           C.clementina_TC487  (194)  TEKIDRDVKQKAECLHHIATILKDKAESRLKHAADKHWSDGALEADLRRA
     P.trifoliata_TA5575_37690  (194)  TEKIDRDVKQKAECLHHIATILKDKAESRLKHAADKHWSDGALEADLRRA
          G.hirsutum_TC91868  (175)  TEKIDRGVKKKAERLHHIATILKAKAESRLKKAADKHWSDGALEADLRRA
            P.patens_124589   (52)  VLKLDTQARRDMLRLRHIALGLTETASMELSSAAEEHWSDGALDADLRLA
            P.patens_138180   (52)  VEKLDTQARKDMMRFRHIALGLTETASMELSSAAEEHWNDGALDADLRLA
     P.euphratica_TA2890_75702  (195)  TEKIDRDVKEKAERLHHIATIIKDRAQTKLKTAADKHWSDGALEADLRLA
  Populus_trichocarpa_EF147825  (195)  TEKIDRDVKEKAERLHHIATIIKDRAQTKLKTAADKHWSDGALEADLRLA
  V.vinifera_GSVIVT00022467001  (195)  TEKIDHGVKKKAERLHHIATILKDKAQSRLKSAADKHWSDGALEADLRRA
  C.solstitialis_TA1343_347529  (171)  TEKIDNDVKKRAEHLHHIALILKNKAQSKLKSVADRHWSDGALEADLRRA
     C.tinctorius_TA1847_4222  (170)  TEKIDNDVKKRAEHLHHIALILKNKAQSKLKSVADRHWSDGALEADLRRA
       L.saligna_TA1747_75948  (173)  TEKIDNNVKKRAERLHHIALILKNKAESKLKSVADRHWSDGALEADLRRA
             L.sativa_TC17902  (173)  TEKIDNNVKKRAERLHHIALILKNKAESKLKSVADRHWSDGALEADLRRA
             H.annuus_TC31796  (166)  TEKIDNDVKKRAEHLHHIALVLQNKAHSKLKSVADRHWSDGALEADLRRA
     Ipomoea_batatas_AF234536  (173)  TEKIDNDVKRKAERLHHVATILKNKAQSRLKNAADRHWSDGALEADLRRA
Solanum_lycopersicum_BT013792  (170)  TEKVDHNVKKRAERIQHVATILKNIAQTRLKKAADRHWSDGALEADLRRA
                     Consensus  (201)  TEKID  VK KAERLHHIA ILKDKAQSRLK AADKHWSDGALEADLRRA
```

```
                                        251                                              300
         A.thaliana_AT1G66330.1  (210)  DFRAKQRAMEDALMALEFIKNIHDMMVNKMVDSLVTS-------ETGTTD
 Arabidopsis_thaliana_AY086709  (210)  DFRAKQRAMEDAFMALEFIKNIHDMMVNKMLDSLLTS-------ETGTTD
                B.napus_TC82749    (1)  -----------------------------MVDSLVTS-------EIGTTD
             LOC_Os05g05600.1  (244)  DSVVRRRAMEDAFMALKFVRDIHDMMANRLQEQFAKDG---SSSPANSRS
                      Os_SPA15  (244)  DSVVRRRAMEDAFMALKFVRDIHDMMANRLQEQFAKDG---SSSPANSRS
             LOC_Os05g05600.5  (218)  DSVVRRRAMEDAFMALKFVRDIHDMMANRLQEQFAKDG---SSSPANSRS
             LOC_Os05g05600.6   (89)  DSVVRRRAMEDAFMALKFVRDIHDMMANRLQEQFAKDG---SSSPANSRS
       S.bicolor_Sb05g026090.1  (239)  DLVVKRRAMEDAFMALKFVQDIHDMMVNRLYEQLPKDG---SSSRTNSTG
              Zea_mays_EU956861  (241)  DLVVKRRAMEDAFMALKFVQDIHDMMVNRLYEQLPKDG---SSSRTNSTG
         G.max_Glyma14g39620.1  (220)  DLRAKQRAMEDPLMALELIKNIHNRMVSKMYNFPLRRDKG-SLSENNVRG
     M.truncatula_AC155282_17.4  (222)  DFRAKQRAMEDALMSLELIKNIHDMMVSKTYNFPIFRDKG-SLSENNVRG
     P.trifoliata_TA5576_37690    (1)  -----------------------------MEFSPLRREKI-SLSDPEMMG
       C.clementina_DY280874  (243)  DFRAKQRAMEDALMALEFLKNIHDMMVRKMYKFPLRR-EKILYQILR---
       C.clementina_DY297038  (244)  DFRAKQRAMENALKALEFLKNIHDMMVRKMNKFPLPKGKDFSNQILK---
           C.clementina_TC487  (244)  DFRAKQRAMEDALMALEFLKNIHDMMVRKMYKFPLRR-EKISLSDPEMMG
     P.trifoliata_TA5575_37690  (244)  DFRAKQRAMEDALMALEFLKNIHAMMVRKMYKFPLRR-EKISLSDPEMMG
          G.hirsutum_TC91868  (225)  DFRAKQRAMEDALMALEFVKNIHDMMVSKGYKFHCDG-KKVHYQLMT---
            P.patens_124589  (102)  DLRARRRAMEDLYAAVLVVKNVHDALVSTLKTRSVDASK------DNAER
            P.patens_138180  (102)  DLRARRRAMEDLYAALQAVKSVHDALVRTLKIRSVLSAK------QGSNV
     P.euphratica_TA2890_75702  (245)  DFRAKQRAMEDALMALEFVKNIHDLMVSKMYKFPLRK-EEGSLTANGILG
  Populus_trichocarpa_EF147825  (245)  DFRAKQRAMEDALMALEFVKNIHELMVSKMYKFPLRK-EEGSLTANGILG
  V.vinifera_GSVIVT00022467001  (245)  DFCAKQRAMEDALMALEFVKNIHDMMVSKMYKLKKGT-----PSSNDTMG
  C.solstitialis_TA1343_347529  (221)  DYIAKQRAMEDALMALEFIKDVHDMMISKMYELKNGALSS-----DDMAG
     C.tinctorius_TA1847_4222  (220)  DYIAKQRAMEDALMALEFIKDVHDMMMSKMYELKNGALSS-----DDMAG
       L.saligna_TA1747_75948  (223)  DYIAKQRAMEDALMALEFIKDVSDMMMSKMYKLKNGALSS-----DDMAG
             L.sativa_TC17902  (223)  DYIAKKRAMEDALMALEFIKDVQDMMMSKMYKLKNGALSS-----DDMAG
             H.annuus_TC31796  (216)  DYIAKQRAMEDALMALEFMKDVHDMMISKMYKLRNGALSF-----DDMAG
     Ipomoea_batatas_AF234536  (223)  DFAAKQRAMEDALMALEFVKNIHDMMVSKMCNLKRSSLNP-----NKMTE
Solanum_lycopersicum_BT013792  (220)  DFVAKQRAMEDSLMALEFVKNIHDRMVSKMCKLKRSSID------TEKTR
                     Consensus  (251)  DF AKQRAMEDALMALEFVKNIHDMMV KMY                  G
```

FIGURE 15 (continued)

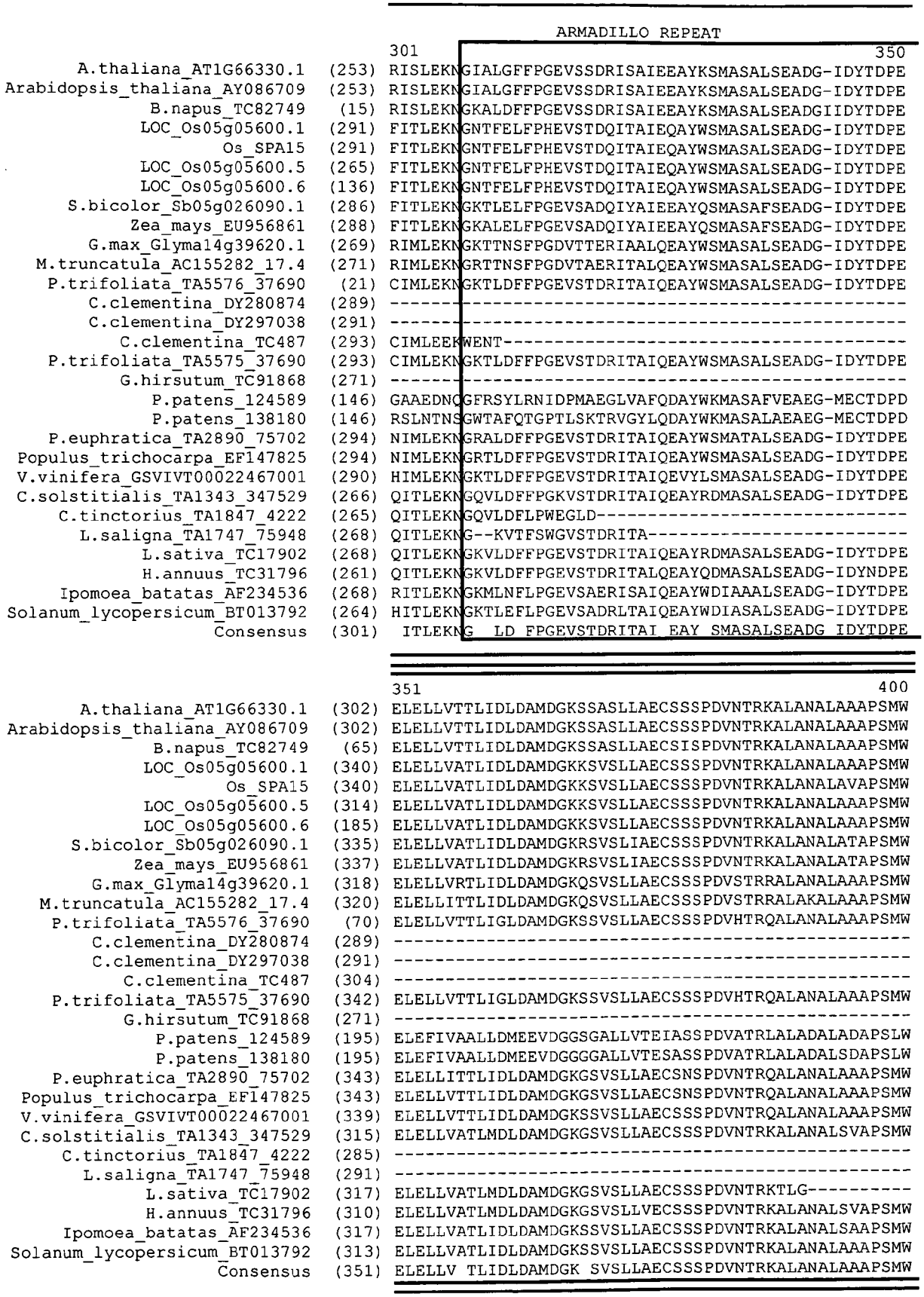

FIGURE 15 (continued)

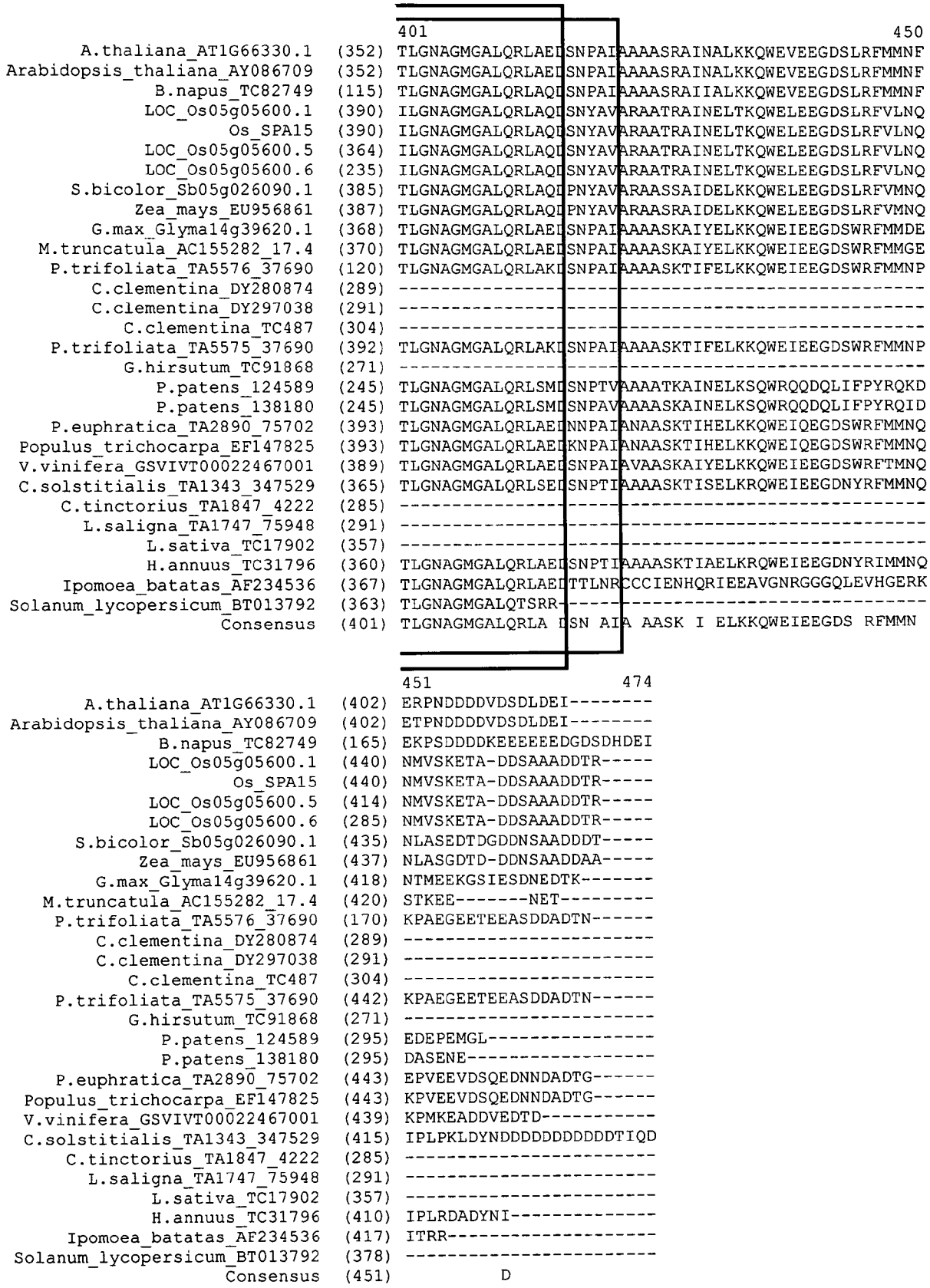

FIGURE 15 (continued)

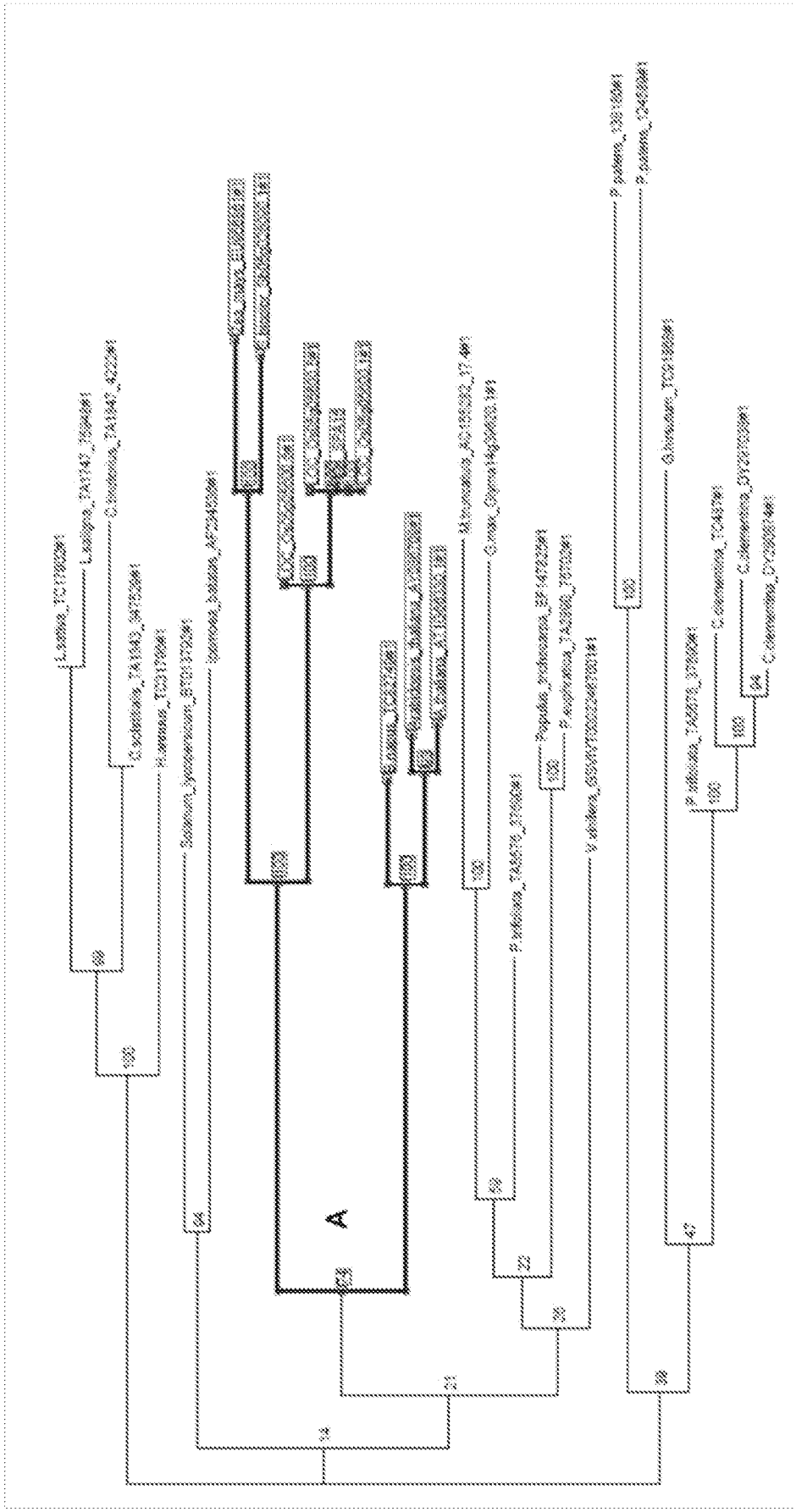

FIGURE 16

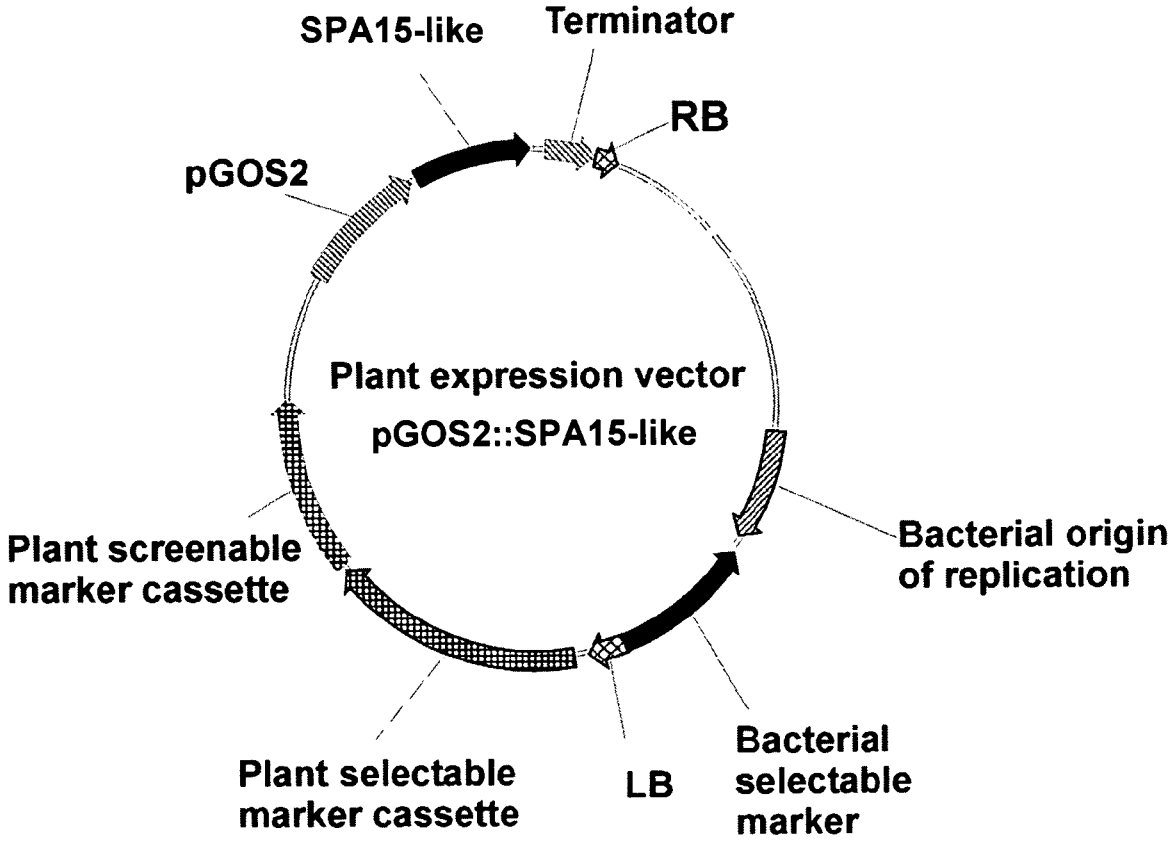
SPA15-like
Terminator
RB
pGOS2
Plant expression vector
pGOS2::SPA15-like
Plant screenable
marker cassette
Bacterial origin
of replication
Plant selectable
marker cassette
LB
Bacterial
selectable
marker

FIGURE 17

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2010/067164

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82 C12N15/29 C07K14/415 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal , Sequence Search , WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE Uni Prot [Online] 3 October 2006 (2006-10-03) , "SubName: Full =0s01g0851100 protein;" XP002621566 retrieved from EBI accession no. UNIPROT:Q0JHP5 Database accession no. Q0JHP5 the whole document -----	12 , 13 , 16-18, 20-23
X	DATABASE EMBL [Online] 19 July 2003 (2003-07-19) , "Oryza sativa Japonica Group cDNA clone: J023081E01 , full insert sequence. " XP002621567 retrieved from EBI accession no. EMBL:AK100337 Database accession no. AK100337 compound -----	12 , 13 , 16-18, 20-23

-/- ·

X Further documents are listed in the continuation of Box C.

X See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 February 2011	29/03/2011

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Maddox, Andrew

3

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

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2006/123505 A1 (KIKUCHI SHOSHI [JP] ET AL) 8 June 2006 (2006-06-08) paragraph [0124] - paragraph [0126] ; claims 1-26; sequences 24895,53239	1-24
X	US 2004/123343 A1 (LA ROSA THOMAS J [US] ET AL LA ROSA THOMAS [US] ET AL) 24 June 2004 (2004-06-24) claims 1-3; sequences 152505, 152516,108726,6243,50033,50022	1-24
X	WO 2009/091518 A2 (MONSANTO TECHNOLOGY LLC [US]; KOVALIC DAVID [US]; TABASKA JACK [US]; Q) 23 July 2009 (2009-07-23) claims 1-16; sequences 34042,75821	1-24

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2010/067164

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006123505 A1	08-06-2006	JP 2005185101 A	14-07-2005
US 2004123343 A1	24-06-2004	NONE	
WO 2009091518 A2	23-07-2009	NONE	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2010/067164

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

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

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

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

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

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

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

see additional sheet

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

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

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

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

1-24(partially)

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

Invention: 1; Claims: l-24(partially)

Method, plant, construct, use, isolated nucleic acid molecule, isolated polypeptide, transgenic plant, harvestable part, products and and use of a nucleic acid as defined in the claims and relating to SEQ ID NOS: 1 and 2

Inventions: 2-130; Claims: l-24(partially)

Method, plant, construct, use, isolated nucleic acid molecule, isolated polypeptide, transgenic plant, harvestable part, products and and use of a nucleic acid as defined in the claims and relating to each independent pair of odd and even SEQ ID NOS of table Al concerning separate and unlinked claimed inventions i.e. 3/4 through to 259/260

Inventions: 131-167; Claims: 25-49 (partially)

Method, plant, construct, use, isolated nucleic acid molecule, isolated polypeptide, transgenic plant, harvestable part, products and and use of a nucleic acid as defined in the claims and relating to each independent pair of odd and even SEQ ID NOS of table A2 concerning separate and unlinked claimed inventions i.e. 267/268 through to 339/340

Inventions: 168-195; Claims: 50-70(partially)

Method, plant, construct, use, isolated nucleic acid molecule, isolated polypeptide, transgenic plant, harvestable part, products and and use of a nucleic acid as defined in the claims and relating to each independent pair of odd and even SEQ ID NOS of table A3 concerning separate and unlinked claimed inventions i.e. 353/354 through to 407/408

Inventions: 196-235; Claims: 71-91 (partially)

Method, plant, construct, use, isolated nucleic acid molecule, isolated polypeptide, transgenic plant, harvestable part, products and and use of a nucleic acid as defined in the claims and relating to each independent pair of odd and even SEQ ID NOS of table A4 concerning separate and unlinked claimed inventions i.e. 421/422 through to 499/500

Inventions: 236-264; Claims: 92- 116 (partially)

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Method, plant, construct, use, isolated nucleic acid molecule, isolated polypeptide, transgenic plant, harvestable part, products and and use of a nucleic acid as defined in the claims and relating to each independent pair of odd and even SEQ ID NOS of table A5 concerning separate and unlinked claimed inventions i.e. 633/634 through to 689/690
