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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 February 2009 (05.02.2009)

PCT

(10) International Publication Number
WO 2009/016212 A2

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

(21) International Application Number: PCT/EP2008/060030

(22) International Filing Date: 31 July 2008 (31.07.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
07113569.3 31 July 2007 (31.07.2007) EP
07113568.5 31 July 2007 (31.07.2007) EP

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

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **HATZFELD, Yves** [FR/FR]; 18C, rue des Dondaines, F-59000 Lille (FR).

(74) Agent: **MISTRY, Meeta**; BASF SE, GVX - C 006, Carl-Bosch-Str., 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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(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 a NITR (Nitrite Reductase) polypeptide or an ASNS (Asparagine Synthase) polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid encoding a NITR polypeptide or an ASNS polypeptide, which plants have enhanced yield-related traits relative to control plants. The invention also provides constructs comprising NITR-encoding nucleic acids or ASNS-encoding nucleic acids, useful in performing the methods of the invention.

WO 2009/016212 A2

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 improving various plant growth characteristics by modulating expression in a plant of a nucleic acid encoding a NITR (Nitrite Reductase). The present invention also concerns plants having modulated expression of a nucleic acid encoding a NITR, which plants have improved growth characteristics relative to corresponding wild type plants or other control plants. The present invention furthermore concerns a method for improving various plant growth characteristics by modulating expression in a plant of a nucleic acid encoding an ASNS (Asparagine Synthase). The present invention also concerns plants having modulated expression of a nucleic acid encoding an ASNS, which plants have improved growth characteristics 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 vigor 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 (2003) 218: 1-14). 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.

Surprisingly, it has now been found that modulating expression of a nucleic acid encoding a NITR polypeptide or of an ASNS 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 of a plant relative to control plants, comprising modulating expression of a nucleic acid encoding a

NITR polypeptide in a plant. The improved yield related traits comprised one or more of increased biomass, increased early vigour, and increased seed yield.

According another embodiment, there is provided a method for improving yield related traits of a plant relative to control plants, comprising modulating expression of a nucleic acid encoding an ASNS polypeptide in a plant. The improved yield related traits comprised one or more of increased biomass, increased early vigour, and increased seed yield.

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.

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.

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 α-helical structures or β-sheet structures). Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide; 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

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.

Motif/Consensus sequence/Signature

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).

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 (T_m) for the specific sequence at a defined ionic strength and pH. Medium stringency conditions are when the temperature is 20°C below T_m, and high stringency conditions are when the temperature is 10°C below T_m. High stringency hybridisation conditions are typically used for isolating hybridising sequences that have high sequence similarity to the target nucleic acid sequence. However, nucleic acids may deviate in sequence and still encode a substantially identical polypeptide, due to the degeneracy of the genetic code. Therefore medium stringency hybridisation conditions may sometimes be needed to identify such nucleic acid molecules.

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):

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

2) DNA-RNA or RNA-RNA hybrids:

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

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 = 2×(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. 1×SSC is 0.15M NaCl 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.

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).

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
SbPRP1	Suzuki et al., Plant Mol. Biol. 21: 109-119, 1993.
LRX1	Baumberger et al. 2001, Genes & Dev. 15:1128
BTG-26 Brassica napus	US 20050044585
LeAMT1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
The LeNRT1-1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
class I patatin gene (potato)	Liu et al., Plant Mol. Biol. 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;1Np (N. plumbaginifolia)	Quesada et al. (1997, Plant Mol. Biol. 34:265)

A seed-specific promoter is transcriptionally active predominantly in seed tissue, but not necessarily exclusively in seed tissue (in cases of leaky expression). The seed-specific promoter may be active during seed development and/or during germination. Examples of seed-specific promoters 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 Itr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	Mena et al, The Plant Journal, 116(1): 53-62, 1998
blz2	EP99106056.7
synthetic promoter	Vicente-Carbajosa et al., Plant J. 13: 629-640, 1998.
rice prolamin NRP33	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice a-globulin Glb-1	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
rice α-globulin REB/OHP-1	Nakase et al. Plant Mol. Biol. 33: 513-522, 1997
rice ADP-glucose pyrophos-phorylase	Trans Res 6:157-68, 1997
maize ESR gene family	Plant J 12:235-46, 1997
sorghum α-kafirin	DeRose et al., Plant Mol. Biol 32:1029-35, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
rice oleosin	Wu et al, J. Biochem. 123:386, 1998
sunflower oleosin	Cummins et al., Plant Mol. Biol. 19: 873-876, 1992
PRO0117, putative rice 40S ribosomal protein	WO 2004/070039
PRO0136, rice alanine aminotransferase	unpublished
PRO0147, trypsin inhibitor ITR1 (barley)	unpublished
PRO0151, rice WSI18	WO 2004/070039
PRO0175, rice RAB21	WO 2004/070039

PRO005	WO 2004/070039
PRO0095	WO 2004/070039
α-amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β-like gene	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Barley Ltp2	Kalla et al., Plant J. 6:849-60, 1994
Chi26	Leah et al., Plant J. 4:579-89, 1994
Maize B-Peru	Selinger et al., Genetics 149;1125-38,1998

Table 2d: examples of endosperm-specific promoters

Gene source	Reference
glutelin (rice)	Takaiwa et al. (1986) Mol Gen Genet 208:15-22; Takaiwa et al. (1987) FEBS Letts. 221:43-47
zein	Matzke et al., (1990) Plant Mol Biol 14(3): 323-32
wheat LMW and HMW glutenin-1	Colot et al. (1989) Mol Gen Genet 216:81-90, Anderson et al. (1989) NAR 17:461-2
wheat SPA	Albani et al. (1997) Plant Cell 9:171-184
wheat gliadins	Rafalski et al. (1984) EMBO 3:1409-15
barley 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(14):9175-82
synthetic promoter	Vicente-Carbajosa et al. (1998) Plant J 13:629-640
rice prolamin NRP33	Wu et al, (1998) Plant Cell Physiol 39(8) 885-889
rice globulin Glb-1	Wu et al. (1998) Plant Cell Physiol 39(8) 885-889
rice globulin REB/OHP-1	Nakase et al. (1997) Plant Molec Biol 33: 513-522
rice ADP-glucose pyrophosphorylase	Russell et al. (1997) Trans Res 6:157-68
maize ESR gene family	Opsahl-Ferstad et al. (1997) Plant J 12:235-46
sorghum kafirin	DeRose et al. (1996) Plant Mol Biol 32:1029-35

Table 2e: Examples of embryo specific promoters:

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

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	Wagner & Kohorn (2001) Plant Cell

	meristems, and in expanding leaves and sepals	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.

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., WO9322443), or isolated promoters may be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene.

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

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

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.

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. 5,116,742). Alternatively, mRNA transcripts corresponding to a nucleic acid sequence can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (Bartel and Szostak (1993) Science 261, 1411-1418). The use of ribozymes for gene silencing in plants is known in the art (e.g., Atkins et al. (1994) WO 94/00012; Lenne et al. (1995) WO 95/03404; Lutziger et al. (2000) WO 00/00619; Prinsen et al. (1997) WO 97/13865 and Scott et al. (1997) WO 97/38116).

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

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, 1121-1133, 2006).

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

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

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 nptII that phosphorylates neomycin and kanamycin, or hpt, phosphorylating hygromycin, or genes conferring resistance to, for example, bleomycin, streptomycin, tetracyclin, chloramphenicol, ampicillin, gentamycin, geneticin (G418), spectinomycin or blasticidin), to herbicides (for example bar which provides resistance to Basta®; aroA or gox providing resistance against glyphosate, or the genes conferring resistance to, for example, imidazolinone, phosphinothricin or sulfonylurea), or genes that provide a metabolic trait (such as manA that allows plants to use mannose as sole carbon source or xylose isomerase for the utilisation of xylose, or 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.

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, 1099-1102); microinjection into plant material (Crossway A et al., (1986) Mol. Gen Genet 202: 179-185); DNA or RNA-coated particle bombardment (Klein TM et al., (1987) Nature 327: 70) infection with (non-integrative) viruses and the like. Transgenic plants, including transgenic crop plants, are preferably produced via Agrobacterium-mediated transformation. An advantageous transformation method is the transformation in planta. To this end, it is possible, for example, to allow the agrobacteria to act on plant seeds or to inoculate the plant meristem with agrobacteria. It has proved particularly expedient in accordance with the invention to allow a suspension of transformed agrobacteria to act on the intact plant or at least on the flower primordia. The plant is subsequently grown on until the seeds of the treated plant are obtained (Clough and Bent, Plant J. (1998) 16, 735–743). Methods for Agrobacterium-mediated transformation of rice include well known methods for rice transformation, such as those described in any of the following: European patent application EP 1198985 A1, Aldemita and Hodges (Planta 199: 612-617, 1996); Chan et al. (Plant Mol Biol 22 (3): 491-506, 1993), Hiei et al. (Plant J 6 (2): 271-282, 1994), which disclosures are incorporated by reference herein as if fully set forth. In the case of corn transformation, the preferred method is as described in either Ishida et al. (Nat. Biotechnol 14(6): 745-50, 1996) or Frame et al. (Plant Physiol 129(1): 13-22, 2002), which disclosures are incorporated by reference herein as if fully set forth. Said methods are further described by way of example in B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press (1993) 128-143 and in Potrykus Annu. Rev. Plant Physiol. Plant Molec. Biol. 42 (1991) 205-225). The nucleic acids or the construct to be expressed is preferably cloned into a vector, which is suitable for transforming Agrobacterium tumefaciens, for example pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984) 8711). Agrobacteria transformed by such a vector can then be used in known manner for the transformation of plants, such as plants used as a model, like Arabidopsis (Arabidopsis thaliana is within the scope of the present invention not considered as a crop plant), or crop plants such as, by way of example, tobacco plants, for example by immersing bruised leaves or chopped leaves in an agrobacterial solution and then culturing them in suitable media. The transformation of plants by means of Agrobacterium tumefaciens is described, for example, by Höfgen and Willmitzer in Nucl. Acid Res. (1988) 16, 9877 or is known inter alia from F.F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press, 1993, pp. 15-38.

In addition to the transformation of somatic cells, which then have to be regenerated into intact plants, it is also possible to transform the cells of plant meristems and in particular those cells which develop into gametes. In this case, the transformed gametes follow the natural plant development, giving rise to transgenic plants. Thus, for example, seeds of Arabidopsis are treated with agrobacteria and seeds are obtained from the developing plants of which a certain proportion is transformed and thus transgenic [Feldman, KA and Marks MD (1987). Mol Gen Genet 208:274-289; 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: 1194-1199], while in the case of the ”floral dip” method the developing floral tissue is incubated briefly with a surfactant-treated agrobacterial suspension [Clough, SJ and Bent AF (1998) The Plant J. 16, 735-743]. A certain proportion of transgenic seeds are harvested in both cases, and these seeds can be distinguished from non-transgenic seeds by growing under the above-described selective conditions. In addition the stable transformation of plastids is of advantages because plastids are inherited maternally 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).

<u>T-DNA activation tagging</u>

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.

<u>TILLING</u>

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

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.

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.

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 color) is calculated. The greenness index is expressed as the percentage of pixels for which the green-to-red ratio exceeds a given threshold. Under normal growth conditions, under salt stress growth conditions, and under reduced nutrient availability growth conditions, the greenness index of plants is measured in the last imaging before flowering. In contrast, under drought stress growth conditions, the greenness index of plants is measured in the first imaging after drought.

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.

Detailed description of the invention

<u>I NITR</u>

Surprisingly, it has now been found that modulating expression in a plant of a nucleic acid encoding a NITR 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 a NITR polypeptide.

A preferred method for modulating (preferably, increasing) expression of a nucleic acid encoding a NITR polypeptide is by introducing and expressing in a plant a nucleic acid encoding a NITR polypeptide.

Any reference hereinafter to a “protein useful in the methods of the invention” is taken to mean a NITR 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 NITR 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 “NITR nucleic acid” or “NITR gene”.

A “NITR polypeptide” as defined herein refers to the nitrite reductase protein represented by SEQ ID NO: 2 and to homologues (orthologues and paralogues) thereof. Nitrite reductases belong to the enzyme class EC 1.7.7.1 and catalyse the reduction of nitrite to ammonium.
Preferably, the homologues of SEQ ID NO: 2 have a NIR_SIR domain. NIR_SIR domains (Pfam entry PF01077, Nitrite and sulphite reductase 4Fe-4S region) are well known in the art and may readily be identified by persons skilled in the art. Preferably, the NITR polypeptides also comprise one or more of the following domains:

- InterPro: IPR005117 (Nitrite/sulphite reductase, hemoprotein beta-component, ferrodoxin-like)
- PFAM: PF03460 (NIR_SIR_ferr)
- InterPro: IPR006066 (Nitrite and sulphite reductase iron-sulphur/siroheme-binding site)
- PRINTS: PR00397 (SIROHAEM)
- PROSITE: PS00365 (NIR_SIR)
- InterPro: IPR006067 (Nitrite and sulphite reductase 4Fe-4S region)
- GENE3D: G3DSA:3.30.413.10 (G3DSA:3.30.413.10)

Alternatively, the homologue of a NITR 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 the conserved domains 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 polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 3, clusters with the group of NITR polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with Sulfite Reductases or any other group.

The term “domain” and “motif” is defined in the “definitions” section herein. 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.

Furthermore, NITR polypeptides (at least in their native form), as far as SEQ ID NO: 2 and its homologues are concerned, typically have oxidoreductase activity. Tools and techniques for measuring oxidoreductase activity are well known in the art, see for example Ferrari and Varner, Plant Physiol., 47(6), 790–794 (1971).

Nitrite reductases group together with Sulfite Reductases (EC 1.8.1.2, Hilz et al., Biochem. Z. 332, 151-166, 1959), which catalyse the reaction:
hydrogen sulfide + 3 NADP+ + 3 H2O = sulfite + 3 NADPH + 3 H+
However, it should be noted that the group of Sulfite Reductases are not encompassed by the term NITR polypeptides as used in the present invention.

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 NITR-encoding nucleic acid or NITR polypeptide as defined herein (thereby excluding the Sulfite Reductases).

Examples of nucleic acids encoding NITR polypeptides (such as those provided in Figure 2 or in the sequence listing) may be found in databases known in the art. Such nucleic acids are useful in performing the methods of the invention. Orthologues and paralogues, the terms “orthologues” and “paralogues” being as defined herein, may readily be identified by performing a so-called reciprocal blast search. Typically, this involves a first BLAST involving BLASTing a query sequence (for example using SEQ ID NO: 2) 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 (where the query sequence is SEQ ID NO: 1 or SEQ ID NO: 2, the second BLAST would therefore be against Arabidopsis thaliana sequences). 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.

Nucleic acid variants encoding homologues and derivatives of SEQ ID NO: 2 may also be useful in practising the methods of the invention, 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 SEQ ID NO: 2. 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 nucleic acid variants useful in practising the methods of the invention include portions of nucleic acids encoding NITR polypeptides, nucleic acids hybridising to nucleic acids encoding NITR polypeptides, splice variants of nucleic acids encoding NITR polypeptides, allelic variants of nucleic acids encoding NITR polypeptides and variants of nucleic acids encoding NITR polypeptides obtained by gene shuffling. The terms hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

Nucleic acids encoding NITR 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 SEQ ID NO: 1, or a portion of a nucleic acid encoding an orthologue, paralogue or homologue of SEQ ID NO: 2.

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.

Portions useful in the methods of the invention, encode a NITR polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in SEQ ID NO: 2. Preferably, the portion is a portion of any one of the nucleic acids given in SEQ ID NO: 1, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in SEQ ID NO: 1. Preferably the portion is at least 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750 consecutive nucleotides in length, the consecutive nucleotides being of SEQ ID NO: 1, or of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 1.

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 a NITR 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 SEQ ID NO: 1, 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 SEQ ID NO: 1.

Hybridising sequences useful in the methods of the invention encode a NITR polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in SEQ ID NO: 2. Preferably, the hybridising sequence is capable of hybridising to SEQ ID NO: 1, or to a portion of any of these sequences, a portion being as defined above, or the hybridising sequence is capable of hybridising to a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2.

Another nucleic acid variant useful in the methods of the invention is a splice variant encoding a NITR 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 SEQ ID NO: 1, or a splice variant of a nucleic acid encoding an orthologue, paralogue or homologue of SEQ ID NO: 2.

Another nucleic acid variant useful in performing the methods of the invention is an allelic variant of a nucleic acid encoding a NITR 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 SEQ ID NO: 1, or comprising introducing and expressing in a plant an allelic variant of a nucleic acid encoding an orthologue, paralogue or homologue of the amino acid sequences represented by SEQ ID NO: 2.

The allelic variants useful in the methods of the present invention have substantially the same biological activity as the NITR polypeptide of SEQ ID NO: 2. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Gene shuffling or directed evolution may also be used to generate variants of nucleic acids encoding NITR 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 SEQ ID NO: 1, or comprising introducing and expressing in a plant a variant of a nucleic acid encoding an orthologue, paralogue or homologue of SEQ ID NO: 2, which variant nucleic acid is obtained by gene shuffling.

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.).

Nucleic acids encoding NITR 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 NITR polypeptide-encoding nucleic acid is from a plant. In the case of SEQ ID NO: 1, the NITR polypeptide encoding nucleic acid is preferably from a monocotyledonous plant, more preferably from the family Brassicaceae, most preferably the nucleic acid is from Arabidopsis thaliana.

Performance of the methods of the invention gives plants having enhanced yield-related traits. In particular performance of the methods of the invention gives plants having increased early vigour and increased yield, especially increased biomass and 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 in 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 biomass and/or seeds, and performance of the methods of the invention results in plants having increased early vigour, biomass and/or seed yield relative to the early vigour, biomass or seed yield of control plants.

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, number of spikelets per panicle, number of flowers (florets) per panicle (which is expressed as a ratio of the number of filled seeds over the number of primary panicles), 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.

The present invention provides a method for increasing yield, especially biomass and/or seed yield of plants, relative to control plants, which method comprises modulating expression, preferably increasing expression, in a plant of a nucleic acid encoding a NITR 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.

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 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.

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, preferably increasing expression, in a plant of a nucleic acid encoding a NITR polypeptide as defined herein. In a particular embodiment, performance of the methods of the present invention gives plants with increased early vigour.

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% or 30%, preferably less than 25%, 20% or 15%, more preferably less than 14%, 13%, 12%, 11% or 10% or less 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 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.

Performance of the methods of the invention gives plants grown under non-stress conditions or under mild drought conditions increased yield and/or increased early vigour, relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield and/or early vigour in plants grown under non-stress conditions or under mild drought conditions, which method comprises increasing expression in a plant of a nucleic acid encoding a NITR 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 increasing expression in a plant of a nucleic acid encoding a NITR polypeptide. Nutrient deficiency may result from a lack of nutrients such as nitrogen, phosphates and other phosphorous-containing compounds, potassium, calcium, cadmium, magnesium, manganese, iron and boron, amongst others.

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 a POI polypeptide. 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.

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 a NITR polypeptide as defined above.

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding NITR 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 a NITR 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 a NITR 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).

Advantageously, any type of promoter, whether natural or synthetic, may be used to drive expression of the nucleic acid sequence. A constitutive promoter is particularly useful in the methods of the invention. Preferably the constitutive promoter is also a ubiquitous promoter. See the “Definitions” section herein for definitions of the various promoter types.

It should be clear that the applicability of the present invention is not restricted to the NITR polypeptide-encoding nucleic acid represented by SEQ ID NO: 1, nor is the applicability of the invention restricted to expression of a NITR polypeptide-encoding nucleic acid when driven by a constitutive specific promoter.

The constitutive promoter is preferably a medium strength promoter of plant origin, preferably a GOS2 promoter, more preferably a GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 3, most preferably the constitutive promoter is as represented by SEQ ID NO: 3. See Table 2 in 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. 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.

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 a NITR polypeptide as defined hereinabove.

More specifically, the present invention provides a method for the production of transgenic plants having increased enhanced yield-related traits, particularly increased early vigour and/or increased yield, which method comprises:

(i) introducing and expressing in a plant or plant cell a NITR 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 a NITR 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 genetically modified plant cells can be regenerated via all methods with which the skilled worker is familiar. Suitable methods can be found in the abovementioned publications by S.D. Kung and R. Wu, Potrykus or Höfgen and Willmitzer.

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).

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 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 a NITR 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, sunflower, canola, alfalfa, rapeseed, 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 a NITR 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.

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 (preferably, increasing) expression of a nucleic acid encoding a NITR polypeptide is by introducing and expressing in a plant a nucleic acid encoding a NITR 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 present invention also encompasses use of nucleic acids encoding NITR polypeptides as described herein and use of these NITR polypeptides in enhancing any of the aforementioned yield-related traits in plants.

Nucleic acids encoding NITR polypeptide described herein, or the NITR polypeptides themselves, may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to a NITR polypeptide-encoding gene. The nucleic acids/genes, or the NITR 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.

Allelic variants of a NITR polypeptide-encoding nucleic acid/gene may also find use in marker-assisted breeding programmes. 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.

Nucleic acids encoding NITR 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. Such use of NITR polypeptide-encoding nucleic acids requires only a nucleic acid sequence of at least 15 nucleotides in length. The NITR polypeptide-encoding 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 NITR-encoding nucleic acids. 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 NITR polypeptide-encoding nucleic acid 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.

The methods according to the present invention result in plants having enhanced yield-related traits, as described hereinbefore. These traits may also be combined with other economically advantageous traits, such as further yield-enhancing traits, tolerance to other abiotic and biotic stresses, traits modifying various architectural features and/or biochemical and/or physiological features.

II ASNS

Surprisingly, it has now been found that modulating expression in a plant of a nucleic acid encoding an ASNS 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 ASNS polypeptide.

A preferred method for modulating (preferably, increasing) expression of a nucleic acid encoding an ASNS polypeptide is by introducing and expressing in a plant a nucleic acid encoding an ASNS polypeptide.

Any reference hereinafter to a “protein useful in the methods of the invention” is taken to mean an ASNS 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 ASNS 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 “ASNS nucleic acid” or “ASNS gene”.

An “ASNS polypeptide” as defined herein refers to the Asparagine synthetase represented by SEQ ID NO: 63 and to homologues (orthologues and paralogues) thereof. SEQ ID NO: 63 comprises, compared to the wild type sequence (Os06g0265000, SEQ ID NO: 67), two point mutations: R382G and S165G (Fig. 6). Arginine on position 382 in SEQ ID NO: 67 is highly conserved among Asparagine synthetases and may be part of a large alpha-helix which delimits the molecular tunnel between the 2 active sites. It is also close to the AMP binding site. Serine on position 165 may be located in a distorted a-helix region, on the external side of the glutamine binding side according to the structure derived from E. coli. It is postulated that the S165G mutation will probably have little impact on the structure of this region.

Therefore, ASNS polypeptides useful in the methods of the present invention preferably have a substitution of the Arginine residue that corresponds to R382 in SEQ ID NO: 67, into an amino acid that distorts the alpha-helix, preferably into a Glycine. Optionally ASNS polypeptides useful in the methods of the present invention additionally have a substitution of the Serine residue that corresponds to S165 in SEQ ID NO: 67, into another amino acid, preferably into a Glycine. Arg residues corresponding to R382 in SEQ ID NO: 67 or Ser residues corresponding to S165 can be identified by aligning the amino acid sequence to the one of SEQ ID NO: 67, see for example the multiple alignment in Fig.4. Such alignment methods are well known in the art.

Preferably, the homologues of SEQ ID NO: 63 have a Asn_synthase domain. Asn_synthase domains (Pfam entry PF00733) are well known in the art and may readily be identified by persons skilled in the art. Besides the Asn_synthase domain, ASNS polypeptides preferably also have a Glutamine amidotransferase, class-II domain (InterPro IPR000583; GATase_2 (HMMPfam entry PF00310), GATase_Type_II (PROSITE entry PS00443)) and/or a Asparagine synthase, glutamine-hydrolyzing domain (asn_synth_AEB: asparagine synthase (glutami (TIGRFAMs entry TIGR01536))

Alternatively, the homologue of a ASNS 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: 63, provided that the homologous protein comprises the conserved motifs as outlined above and the substitution of the Arg residue that corresponds to R382 in SEQ ID NO: 67. 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. Compared to overall sequence identity, the sequence identity will generally be higher when only conserved domains or motifs are considered.

The term “domain” and “motif” is defined in the “definitions” section herein. 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).

Furthermore, ASNS polypeptides (at least in their native form), as far as SEQ ID NO: 2 and its homologues are concerned, typically have asparagine synthetase activity (Patterson and Orr, J. Biol. Chem. 243, 376-380, 1968; Enzyme Catalogue 6.3.5.4, reaction scheme:
ATP + L-aspartate + L-glutamine + H2O = AMP + diphosphate + L-asparagine + L-glutamate).
Tools and techniques for measuring asparagine synthetase activity are well known in the art.

The present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 62, encoding the polypeptide sequence of SEQ ID NO: 63. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any ASNS-encoding nucleic acid or ASNS polypeptide as defined herein.

Examples of nucleic acids encoding ASNS polypeptides may be found in databases known in the art, and some of them are listed in Figure 5. Such nucleic acids are useful in performing the methods of the invention. Orthologues and paralogues, the terms “orthologues” and “paralogues” being as defined herein, may readily be identified by performing a so-called reciprocal blast search. Typically, this involves a first BLAST involving BLASTing a query sequence (for example using SEQ ID NO: 63) 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 (where the query sequence is SEQ ID NO: 62 or SEQ ID

NO: 63, the second BLAST would therefore be against Oryza sativa sequences). 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. Examples of orthologues and paralogues of SEQ ID NO: 63 or SEQ ID NO: 67 are listed in Figure 5.

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

Nucleic acid variants encoding homologues and derivatives of SEQ ID NO: 63 may also be useful in practising the methods of the invention, 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 SEQ ID NO: 63. 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 nucleic acid variants useful in practising the methods of the invention include portions of nucleic acids encoding ASNS polypeptides, nucleic acids hybridising to nucleic acids encoding ASNS polypeptides, splice variants of nucleic acids encoding ASNS polypeptides, allelic variants of nucleic acids encoding ASNS polypeptides and variants of nucleic acids encoding ASNS polypeptides obtained by gene shuffling. The terms hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

Nucleic acids encoding ASNS 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 SEQ ID NO: 62, or a portion of a nucleic acid encoding an orthologue, paralogue or homologue of SEQ ID NO: 63.

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.

Portions useful in the methods of the invention, encode an ASNS polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in SEQ ID NO: 63. Preferably, the portion is a portion of any one of the nucleic acids given in SEQ ID NO: 62, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in SEQ ID NO: 62. Preferably the portion is at least 800, 900, 1000, 1100, 1200, 1300, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750 consecutive nucleotides in length, the consecutive nucleotides being of SEQ ID NO: 62, or of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 63. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 62.

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 ASNS 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 SEQ ID NO: 62, 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 SEQ ID NO: 62.

Hybridising sequences useful in the methods of the invention encode an ASNS polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in SEQ ID NO: 63. Preferably, the hybridising sequence is capable of hybridising to SEQ ID NO: 62, or to a portion of any of these sequences, a portion being as defined above, or the hybridising sequence is capable of hybridising to a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 63.

Another nucleic acid variant useful in the methods of the invention is a splice variant encoding an ASNS 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 SEQ ID NO: 62, or a splice variant of a nucleic acid encoding an orthologue, paralogue or homologue of SEQ ID NO: 63.

Another nucleic acid variant useful in performing the methods of the invention is an allelic variant of a nucleic acid encoding an ASNS 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 SEQ ID NO: 62, or comprising introducing and expressing in a plant an allelic variant of a nucleic acid encoding an orthologue, paralogue or homologue of the amino acid sequences represented by SEQ ID NO: 63.

The allelic variants useful in the methods of the present invention have substantially the same biological activity as the ASNS polypeptide of SEQ ID NO: 63. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Gene shuffling or directed evolution may also be used to generate variants of nucleic acids encoding ASNS 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 SEQ ID NO: 62, or comprising introducing and expressing in a plant a variant of a nucleic acid encoding an orthologue, paralogue or homologue of SEQ ID NO: 63, which variant nucleic acid is obtained by gene shuffling.

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.).

Nucleic acids encoding ASNS 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 ASNS polypeptide-encoding nucleic acid is from a plant. In the case of SEQ ID NO: 62, the ASNS polypeptide encoding nucleic acid is 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 early vigour and increased yield, especially increased biomass and 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 in 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 biomass and/or seeds, and performance of the methods of the invention results in plants having increased early vigour, biomass and/or seed yield relative to the early vigour, biomass or seed yield of control plants.

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, number of spikelets per panicle, number of flowers (florets) per panicle (which is expressed as a ratio of the number of filled seeds over the number of primary panicles), 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.

The present invention provides a method for increasing yield, especially biomass and/or seed yield of plants, relative to control plants, which method comprises modulating expression, preferably increasing expression, in a plant of a nucleic acid encoding an ASNS 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.

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 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.

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, preferably increasing expression, in a plant of a nucleic acid encoding an ASNS polypeptide as defined herein. In a particular embodiment, performance of the methods of the present invention gives plants with increased early vigour.

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% or 30%, preferably less than 25%, 20% or 15%, more preferably less than 14%, 13%, 12%, 11% or 10% or less 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 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.

Performance of the methods of the invention gives plants grown under non-stress conditions or under mild drought conditions increased yield and/or increased early vigour, relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield and/or early vigour in plants grown under non-stress conditions or under mild drought conditions, which method comprises increasing expression in a plant of a nucleic acid encoding an ASNS 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 increasing expression in a plant of a nucleic acid encoding an ASNS polypeptide. Nutrient deficiency may result from a lack of nutrients such as nitrogen, phosphates and other phosphorous-containing compounds, potassium, calcium, cadmium, magnesium, manganese, iron and boron, amongst others.

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 a ASNS polypeptide. 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.

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 ASNS polypeptide as defined above.

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding ASNS 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 ASNS 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 ASNS 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).

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 of the invention. Preferably the constitutive promoter is also a ubiquitous promoter of medium strength. See the “Definitions” section herein for definitions of the various promoter types.

It should be clear that the applicability of the present invention is not restricted to the ASNS polypeptide-encoding nucleic acid represented by SEQ ID NO: 62, nor is the applicability of the invention restricted to expression of an ASNS polypeptide-encoding nucleic acid when driven by a constitutive specific promoter.

The constitutive promoter is preferably a GOS2 promoter, preferably a GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 64, most preferably the constitutive promoter is as represented by SEQ ID NO: 64. See Table 2 in 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. 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.

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 ASNS polypeptide as defined hereinabove.

More specifically, the present invention provides a method for the production of transgenic plants having increased enhanced yield-related traits, particularly increased early vigour and/or increased yield, which method comprises:

(i) introducing and expressing in a plant or plant cell an ASNS 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 ASNS 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 genetically modified plant cells can be regenerated via all methods with which the skilled worker is familiar. Suitable methods can be found in the abovementioned publications by S.D. Kung and R. Wu, Potrykus or Höfgen and Willmitzer.

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).

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 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 ASNS 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, sunflower, canola, alfalfa, rapeseed, 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. 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.

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 (preferably, increasing) expression of a nucleic acid encoding an ASNS polypeptide is by introducing and expressing in a plant a nucleic acid encoding an ASNS 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 present invention also encompasses use of nucleic acids encoding ASNS polypeptides as described herein and use of these ASNS polypeptides in enhancing any of the aforementioned yield-related traits in plants.

Nucleic acids encoding ASNS polypeptide described herein, or the ASNS polypeptides themselves, may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to an ASNS polypeptide-encoding gene. The nucleic acids/genes, or the ASNS 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.

Allelic variants of an ASNS polypeptide-encoding nucleic acid/gene may also find use in marker-assisted breeding programmes. 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.

Nucleic acids encoding ASNS 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. Such use of ASNS polypeptide-encoding nucleic acids requires only a nucleic acid sequence of at least 15 nucleotides in length. The ASNS polypeptide-encoding 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 ASNS-encoding nucleic acids. 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 ASNS polypeptide-encoding nucleic acid 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.

The methods according to the present invention result in plants having enhanced yield-related traits, as described hereinbefore. These traits may also be combined with other economically advantageous traits, such as further yield-enhancing traits, tolerance to other abiotic and biotic stresses, traits modifying various architectural features and/or biochemical and/or physiological features.

Description of figures

The present invention will now be described with reference to the following figures in which:
Figure 1 represents the binary vector for increased expression in Oryza sativa of a NITR-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2::NITR)
Figure 2 details examples of sequences useful in performing the methods according to the present invention.
Figure 3 gives a phylogenetic tree of the NITR protein sequences listed in Fig. 2, in which tree the outgroup is represented by Sulfite Reductases, exemplified by SEQ ID NO: 9 (C. reinhardtii 59303), SEQ ID NO: 11 (C. reinhardtii 192232) and SEQ ID NO: 33 (A. thaliana At5g04590) .
Figure 4 represents the binary vector for increased expression in Oryza sativa of an ASNS-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2::ASNS)
Figure 5 details examples of sequences useful in performing the methods according to the present invention.
Figure 6 shows an alignment between SEQ ID NO: 63 and SEQ ID NO: 67. The S165G and R382G mutations are indicated.
Figure 7 is a multiple alignment of examples of ASNS polypeptides. The asterisks represent amino acids that are identical in all sequences, the colons indicate highly conserved residues, the dots represent conserved residues. The Arg residues corresponding to R382 in SEQ ID NO: 67 is shown in bold.

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 the invention

Sequences (full length cDNA, ESTs or genomic) related to the nucleic acid sequence used in the methods of the present invention are 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 acids used in the present invention are used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis is viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflects 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 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 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.

In some instances, related sequences may tentatively be assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR). 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 or polypeptide sequence of interest.

Example 2: Alignment of NITR polypeptide sequences

Alignment of polypeptide sequences is performed using the AlignX programme from the Vector NTI package (Invitrogen) which is based on the popular Clustal W algorithm of progressive alignment (Thompson et al. (1997) Nucleic Acids Res 25:4876-4882; Chenna et al. (2003). Nucleic Acids Res 31:3497-3500). Default values are for the gap open penalty of 10, for the gap extension penalty of 0,1 and the selected weight matrix is Blosum 62 (if polypeptides are aligned). Minor manual editing may be done to further optimise the alignment.

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

For the construction of the phylogenetic tree of Figure 3, the proteins of Figure 2 were aligned using MUSCLE (Edgar (2004), Nucleic Acids Research 32(5): 1792-97). A Neighbour-Joining tree was calculated using QuickTree (Howe et al. (2002), Bioinformatics 18(11): 1546-7). Support of the major branching is indicated for 100 bootstrap repetitions. A circular phylogram was drawn using 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

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. 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

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

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

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

The protein sequences representing the NITR are used as query to search the InterPro database.

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

TargetP 1.1 predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal pre-sequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP). Scores on which the final prediction is based are not really probabilities, and they do not necessarily add to one. However, the location with the highest score is the most likely according to TargetP, and the relationship between the scores (the reliability class) may be an indication of how certain the prediction is. The reliability class (RC) ranges from 1 to 5, where 1 indicates the strongest prediction. 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 were 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 protein sequence represented by SEQ ID NO: 2 was used to query TargetP 1.1. The “plant” organism group is selected, no cutoffs defined, and the predicted length of the transit peptide requested. The protein has a predicted location in the chloroplast (probability 0.793, reliability class 3).

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

Example 6: Cloning of the nucleic acid sequence used in the methods of the invention

Cloning of SEQ ID NO: 1:

The NITR encoding nucleic acid sequence SEQ ID NO: 1 used in the methods of the invention 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 prm07073 (SEQ ID NO: 4; sense, start codon in bold): 5'-ggggacaagtttgt acaaaaaagcaggcttaaacaatgacttctttctctctcactt-3' and prm07074 (SEQ ID NO: 5; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggtcaatagct tttgaatcaatct-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 recombines in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone". 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: 3) for seed specific expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::NITR (Figure 1) was transformed into Agrobacterium strain LBA4044 according to methods well known in the art.

Example 7: 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 T0 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).

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 K2SO4, and 100 µm acetosyringinone. The explants are washed in half-strength Murashige-Skoog medium (Murashige and Skoog, 1962) and plated on the same SH induction medium without acetosyringinone but with a suitable selection agent and suitable antibiotic to inhibit Agrobacterium growth. After several weeks, somatic embryos are transferred to BOi2Y development medium containing no growth regulators, no antibiotics, and 50 g/ L sucrose. Somatic embryos are subsequently germinated on half-strength Murashige-Skoog medium. Rooted seedlings were transplanted into pots and grown in a greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Example 8: Phenotypic evaluation procedure

<u>8.1 Evaluation setup</u>

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. Five 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%.

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 were grown in potting soil under normal conditions except for the nutrient solution. The pots were 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) was the same as for plants not grown under abiotic stress. Growth and yield parameters were 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.

8.2 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.

Because two experiments with overlapping events are carried out, a combined analysis is performed. This is useful to check consistency of the effects over the two experiments, and if this is the case, to accumulate evidence from both experiments in order to increase confidence in the conclusion. The method used is a mixed-model approach that takes into account the multilevel structure of the data (i.e. experiment - event - segregants). P values are obtained by comparing likelihood ratio test to chi square distributions.

8.3 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. 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.

Example 9: Results of the phenotypic evaluation of the transgenic plants

The transgenic rice plants expressing the NITR nucleic acid represented by SEQ ID NO: 1 under control of the GOS2 promoter showed an increase of more than 5% for biomass (root and shoot), early vigour, total weight of seeds, number of filled seeds, harvest index, total number of seeds and number of flowers per panicle when grown under nitrogen deficiency-stress conditions. When evaluated over two generations (T1 and T2) the following data were obtained (Table 3):

Table 3: Yield increase for transgenic plants expressing the NITR nucleic acid compared to the control plants. For each parameter the p value is ≤ 0.05.

Parameter	Overall increase (%)
Early vigour	17.5
Root/Shoot index	9.0
Total weight of seeds	6.1
Number of filled seeds	5.8
Total number of seeds	5.0

Example 10: Identification of sequences related to the nucleic acid sequence used in the methods of the invention

Sequences (full length cDNA, ESTs or genomic) related to the nucleic acid sequence used in the methods of the present invention are 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 acids used in the present invention are used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis is viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflects 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 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 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.

In some instances, related sequences may tentatively be assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR). 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 or polypeptide sequence of interest.

Example 11: Alignment of ASNS polypeptide sequences

Alignment of polypeptide sequences is performed using the AlignX programme from the Vector NTI package (Invitrogen) which is based on the popular Clustal W algorithm of progressive alignment (Thompson et al. (1997) Nucleic Acids Res 25:4876-4882; Chenna et al. (2003). Nucleic Acids Res 31:3497-3500). Default values are for the gap open penalty of 10, for the gap extension penalty of 0,1 and the selected weight matrix is Blosum 62 (if polypeptides are aligned). Minor manual editing may be done to further optimise the alignment. For the alignment of Figure 4, the ClustalW 2.0 algorithm was used with default parameters (Matrix: Gonnet, Gap-opening penalty: 10, Gap-extension penalty: 0.1).

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

Example 12: Calculation of global percentage identity between polypeptide sequences useful in performing the methods of the invention

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. 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

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

Example 13: Identification of domains comprised in polypeptide sequences useful in performing the methods of the invention

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

The protein sequences represented by SEQ ID NO: 63 was used as query to search the InterPro database.

Example 14: Topology prediction of the polypeptide sequences useful in performing the methods of the invention

TargetP 1.1 predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal pre-sequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP). Scores on which the final prediction is based are not really probabilities, and they do not necessarily add to one. However, the location with the highest score is the most likely according to TargetP, and the relationship between the scores (the reliability class) may be an indication of how certain the prediction is. The reliability class (RC) ranges from 1 to 5, where 1 indicates the strongest prediction. 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 were 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 protein sequence of SEQ ID NO: 63 was used to query TargetP 1.1. The “plant” organism group is selected, no cutoffs defined, and the predicted length of the transit peptide requested. No clear subcellular location was predicted by TargetP, but SubLoc (Hua and Sun, Bioinformatics) predicted a cytoplasmic localisation.

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

Example 15: Cloning of the nucleic acid sequence used in the methods of the invention

<u>Cloning of SEQ ID NO: 62:</u>

The nucleic acid sequence SEQ ID NO: 64 used in the methods of the invention 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 prm06049

(SEQ ID NO: 65; sense, start codon in bold): 5'-ggggacaagtttgtacaa aaaagcaggcttaaacaatgtgtggcatcctcgccgtgctcg-3' and prm06050 (SEQ ID NO: 66; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggtgcgacgatagaa agttaaacggcag-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 recombines in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone". Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 62 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: 64) for seed specific expression was located upstream of this Gateway cassette.

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

Example 16: 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 T0 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).

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 K2SO4, and 100 µm acetosyringinone. The explants are washed in half-strength Murashige-Skoog medium (Murashige and Skoog, 1962) and plated on the same SH induction medium without acetosyringinone but with a suitable selection agent and suitable antibiotic to inhibit Agrobacterium growth. After several weeks, somatic embryos are transferred to BOi2Y development medium containing no growth regulators, no antibiotics, and 50 g/ L sucrose. Somatic embryos are subsequently germinated on half-strength Murashige-Skoog medium. Rooted seedlings were transplanted into pots and grown in a greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Example 17: Phenotypic evaluation procedure

17.1 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%.

Four T1 events were further evaluated in the T2 generation following the same evaluation procedure as for the T1 generation but with more individuals per event.

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 were grown in potting soil under normal conditions except for the nutrient solution. The pots were 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) was the same as for plants not grown under abiotic stress. Growth and yield parameters were 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.

17.2 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.

Because two experiments with overlapping events are carried out, a combined analysis was performed. This is useful to check consistency of the effects over the two experiments, and if this is the case, to accumulate evidence from both experiments in order to increase confidence in the conclusion. The method used was a mixed-model approach that takes into account the multilevel structure of the data (i.e. experiment - event - segregants). P values were obtained by comparing likelihood ratio test to chi square distributions.

17.3 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.

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). Increase in root biomass is expressed as root thickness, which is the maximum biomass of roots above a certain thickness threshold observed during the lifespan of a plant (obtained by a root-imaging system).

Example 18: Results of the phenotypic evaluation of the transgenic plants

The transgenic rice plants expressing the ASNS nucleic acid represented by SEQ ID NO: 62 under control of the GOS2 promoter and grown whether under non-stress conditions or under conditions of reduced nitrogen availability, showed an increase of more than 5% for at least one of the following parameters: early vigour, total weight of seeds, number of filled seeds, fill rate, number of flowers per panicle, Harvest Index, total number of seeds and root thickness. For Thousand Kernel Weight the observed increase was at least 3%.

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 a NITR, wherein said NITR is represented by SEQ ID NO: 2 or an orthologue or paralogue thereof and wherein said plant is grown under conditions of nitrogen deficiency.

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

3. Method according to claim 1 or 2, wherein said nucleic acid encoding said NITR polypeptide is a portion of SEQ ID NO: 1, or a nucleic acid capable of hybridising with such a nucleic acid.

4. Method according to any preceding claim, wherein said nucleic acid sequence encodes SEQ ID NO: 2.

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

6. Method according to any one of claims 2 to 5, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

7. Method according to any one of claims 1 to 6, wherein said nucleic acid encoding a NITR polypeptide is of plant origin, preferably from a monocotyledonous plant, further preferably from the family Brassicaceae, more preferably from the genus Arabidopsis, most preferably from Arabidopsis thaliana.

8. Plant or part thereof, including seeds, obtainable by a method according to any preceding claim, wherein said plant or part thereof comprises a recombinant nucleic acid encoding a NITR polypeptide.

9. Construct comprising:
   (a) nucleic acid encoding a NITR polypeptide as defined in claim 1;
   (b) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
   (c) a transcription termination sequence.

10. Construct according to claim 9, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

11. Use of a construct according to claim 9 or 10 in a method for making plants having increased yield-related traits, particularly enhanced early vigour and/or increased yield relative to control plants.

12. Plant, plant part or plant cell transformed with a construct according to claim 9 or 10.

13. Method for the production of a transgenic plant having enhanced early vigour and/or 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 NITR polypeptide as defined in claim 1; and
    (ii) cultivating the plant cell under conditions promoting plant growth and development.

14. Transgenic plant having increased yield-related traits, particularly increased early vigour, increased biomass and/or increased seed yield, relative to control plants, resulting from increased expression of a nucleic acid encoding a NITR polypeptide as defined in claim 1, or a transgenic plant cell derived from said transgenic plant.

15. Transgenic plant according to claim 8, 12 or 14, or a transgenic plant cell derived thereof, wherein said plant is a crop plant or a monocot or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, secale, einkorn, teff, milo and oats.

16. Harvestable parts of a plant according to claim 15, wherein said harvestable parts are preferably shoot biomass, root biomass and/or seeds.

17. Products derived from a plant according to claim 15 and/or from harvestable parts of a plant according to claim 16.

18. Use of a nucleic acid encoding a NITR polypeptide in increasing yield-related traits, particularly in increasing one or more of: early vigour, seed yield, root biomass and shoot biomass in plants, relative to control plants.

19. 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 ASNS, wherein said ASNS is represented by SEQ ID NO: 63 or an orthologue or paralogue thereof.

20. Method according to claim 19, wherein said modulated expression is effected by introducing and expressing in a plant a nucleic acid encoding said ASNS polypeptide.

21. Method according to claim 19 or 20, wherein said nucleic acid encoding said ASNS polypeptide is a portion of SEQ ID NO: 62, or a nucleic acid capable of hybridising with such a nucleic acid.

22. Method according to any one of claims 19 to 21, wherein said nucleic acid sequence encodes SEQ ID NO: 63.

23. Method according to any one of claims 19 to 22, wherein said enhanced yield-related traits comprise increased early vigour and/or increased yield, preferably root thickness and/or increased seed yield, relative to control plants.

24. Method according to any one of claims 19 to 23, wherein said enhanced yield-related traits are obtained under non-stress conditions or under conditions of reduced nutrient availability.

25. Method according to any one of claims 20 to 24, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.

26. Method according to any one of claims 19 to 25, wherein said nucleic acid encoding an ASNS polypeptide is of plant origin, preferably from a monocotyledonous plant, further preferably from the family Poaceae, more preferably from the genus Oryza, most preferably from Oryza sativa.

27. Plant or part thereof, including seeds, obtainable by a method according to any preceding claim, wherein said plant or part thereof comprises a recombinant nucleic acid encoding an ASNS polypeptide.

28. Construct comprising:
    (i) nucleic acid encoding an ASNS polypeptide as defined in claim 1;
    (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
    (iii) a transcription termination sequence.

29. Construct according to claim 28, wherein one of said control sequences is a constitutive promoter, preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.

30. Use of a construct according to claim 28 or 29 in a method for making plants having increased yield-related traits, particularly increased early vigour and increased yield relative to control plants.

31. Plant, plant part or plant cell transformed with a construct according to claim 28 or 29.

32. 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 ASNS polypeptide as defined in claim 1; and
(ii) cultivating the plant cell under conditions promoting plant growth and development.

33. Transgenic plant having increased yield-related traits, particularly increased early vigour, increased root thickness and/or increased seed yield, relative to control plants, resulting from increased expression of a nucleic acid encoding an ASNS polypeptide as defined in claim 19, or a transgenic plant cell derived from said transgenic plant.

34. Transgenic plant according to claim 27, 31 or 33, or a transgenic plant cell derived thereof, wherein said plant is a crop plant or a monocot or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum and oats.

35. Harvestable parts of a plant according to claim 34, wherein said harvestable parts are preferably root biomass and/or seeds.

36. Products derived from a plant according to claim 34 and/or from harvestable parts of a plant according to claim 35.

37. Use of a nucleic acid encoding an ASNS polypeptide in increasing yield-related traits, particularly in increasing one or more of early vigour, seed yield and root biomass in plants, relative to control plants.

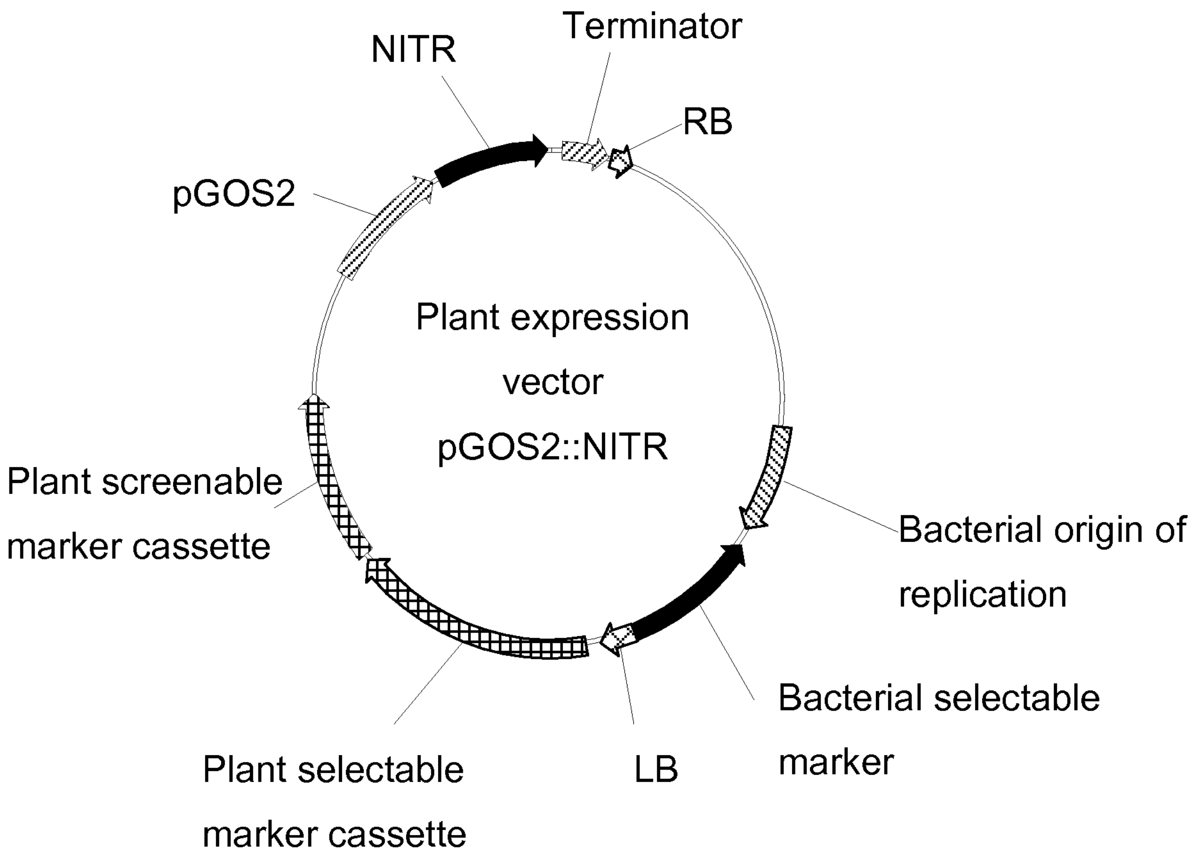
Terminator
NITR
RB
pGOS2
Plant expression
vector
pGOS2::NITR
Plant screenable
marker cassette
Bacterial origin of
replication
Bacterial selectable
marker
LB
Plant selectable
marker cassette

FIGURE 1

SEQ ID NO: 1, DNA – *Arabidopsis thaliana*, ORF 11 - 1768
GGCTTAAACAATGACTTCTTTCTCTCTCACTTTCACATCTCCTCTCCTCCCTTCCTCCTCCACCAA
ACCCAAAAGATCCGTCCTTGTCGCCGCCGCTCAGACCACAGCTCCGGCCGAATCCACCGCCTCTGT
TGACGCAGATCGTCTCGAGCCAAGAGTTGAGTTGAAAGATGGTTTTTTATTCTCAAGGAGAAGTT
TCGAAAAGGGATCAATCCTCAGGAGAAGGTTAAGATCGAGAGAGAGCCCATGAAGTTGTTTATGGA
GAATGGTATTGAAGAGCTTGCTAAGAAATCTATGGAAGAGCTTGATAGTGAAAAGTCTTCTAAAGA
TGATATTGATGTTAGACTCAAGTGGCTTGGTCTCTTTCACCGTAGAAAGCATCAGTATGGGAAGTT
TATGATGAGGTTGAAGTTACCAAATGGTGTGACTACAAGTGCACAGACTCGGTATTTAGCGAGTGT
GATTAGGAAGTATGGTGAAGATGGGTGTGCTGATGTGACTACTAGACAGAATTGGCAGATCCGTGG
TGTTGTGTTGCCTGATGTGCCTGAGATCTTGAAAGGTCTTGCTTCTGTTGGTTTAACGAGTCTTCA
AAGTGGTATGGATAACGTGAGGAACCCGGTTGGGAATCCTATAGCTGGGATTGATCCGGAGGAGAT
TGTTGACACGAGGCCTTACACGAATCTCCTTTCGCAGTTTATCACCGCTAATTCACAAGGAAACCC
CGATTTCACCAACTTGCCAAGAAAGTGGAATGTGTGTGTGGTGGGGACTCATGATCTCTATGAGCA
TCCACATATCAATGATTTGGCCTACATGCCTGCTAATAAAGATGGACGGTTTGGATTCAATTTGCT
TGTGGGAGGATTCTTTAGTCCCAAAAGATGTGAAGAAGCGATTCCTCTTGATGCTTGGGTCCCTGC
TGATGACGTTCTTCCACTCTGCAAAGCTGTTCTAGAGGCTTACAGAGATCTTGGAACTCGAGGAAA
CCGACAGAAGACAAGAATGATGTGGCTTATCGACGAACTTGGTGTTGAAGGATTTAGAACTGAGGT
AGAGAAGAGAATGCCAAATGGGAAACTCGAGAGAGGATCTTCAGAGGATCTTGTGAACAAACAGTG
GGAGAGGAGAGACTATTTCGGAGTCAACCCTCAGAAACAAGAAGGTCTTAGCTTCGTGGGGCTTCA
CGTTCCGGTTGGTAGGCTACAAGCTGATGACATGGATGAGCTTGCTCGGTTAGCTGATACCTACGG
GTCAGGTGAGCTAAGACTCACAGTAGAGCAAAACATCATCATCCCAAATGTAGAAACCTCGAAAAC
CGAAGCTTTGCTTCAAGAGCCGTTTCTCAAGAACCGTTTCTCCCCTGAACCATCTATCCTAATGAA
AGGCTTAGTTGCTTGTACCGGTAGCCAGTTCTGCGGACAAGCGATAATCGAGACTAAGCTAAGAGC
TTTAAAAGTGACAGAAGAAGTAGAGAGACTTGTATCTGTGCCAAGACCGATAAGGATGCATTGGAC
AGGATGTCCCAATACTTGCGGACAAGTCCAAGTAGCAGATATCGGATTCATGGGATGCTTAACACG
AGGCGAGGAAGGAAAGCCAGTCGAGGGTGCTGACGTGTACGTCGGGGGACGAATAGGAAGTGACTC
GCATATCGGAGAGATCTATAAGAAAGGTGTTCGTGTCACGGAGTTGGTTCCATTGGTGGCTGAGAT
TCTGATCAAAGAATTTGGTGCTGTGCCTAGAGAAAGAGAAGAGAATGAAGATTGATTCAAAAGCTA
TTGACCCAGCTTTCTTGTACAAAGT

SEQ ID NO: 2, protein – *Arabidopsis thaliana*
MTSFSLTFTSPLLPSSSTKPKRSVLVAAAQTTAPAESTASVDADRLEPRVELKDGFFILKEKFRKG
INPQEKVKIEREPMKLFMENGIEELAKKSMEELDSEKSSKDDIDVRLKWLGLFHRRKHQYGKFMMR
LKLPNGVTTSAQTRYLASVIRKYGEDGCADVTTRQNWQIRGVVLPDVPEILKGLASVGLTSLQSGM
DNVRNPVGNPIAGIDPEEIVDTRPYTNLLSQFITANSQGNPDFTNLPRKWNVCVVGTHDLYEHPHI
NDLAYMPANKDGRFGFNLLVGGFFSPKRCEEAIPLDAWVPADDVLPLCKAVLEAYRDLGTRGNRQK
TRMMWLIDELGVEGFRTEVEKRMPNGKLERGSSEDLVNKQWERRDYFGVNPQKQEGLSFVGLHVPV
GRLQADDMDELARLADTYGSGELRLTVEQNIIIPNVETSKTEALLQEPFLKNRFSPEPSILMKGLV
ACTGSQFCGQAIIETKLRALKVTEEVERLVSVPRPIRMHWTGCPNTCGQVQVADIGFMGCLTRGEE
GKPVEGADVYVGGRIGSDSHIGEIYKKGVRVTELVPLVAEILIKEFGAVPREREENED

SEQ ID NO: 3, DNA - *Oryza sativa* - GOS2 promoter sequence
AATCCGAAAAGTTTCTGCACCGTTTTCACCCCCTAACTAACAATATAGGGAACGTGTGCTAAATAT
AAAATGAGACCTTATATATGTAGCGCTGATAACTAGAACTATGCAAGAAAAACTCATCCACCTACT
TTAGTGGCAATCGGGCTAAATAAAAAAGAGTCGCTACACTAGTTTCGTTTTCCTTAGTAATTAAGT
GGGAAAATGAAATCATTATTGCTTAGAATATACGTTCACATCTCTGTCATGAAGTTAAATTATTCG
AGGTAGCCATAATTGTCATCAAACTCTTCTTGAATAAAAAAATCTTTCTAGCTGAACTCAATGGGT
AAAGAGAGAGATTTTTTTTAAAAAAATAGAATGAAGATATTCTGAACGTATTGGCAAAGATTTAAA

FIGURE 2

CATATAATTATATAATTTTATAGTTTGTGCATTCGTCATATCGCACATCATTAAGGACATGTCTTA
CTCCATCCCAATTTTTATTTAGTAATTAAAGACAATTGACTTATTTTTATTATTTATCTTTTTTCG
ATTAGATGCAAGGTACTTACGCACACACTTTGTGCTCATGTGCATGTGTGAGTGCACCTCCTCAAT
ACACGTTCAACTAGCAACACATCTCTAATATCACTCGCCTATTTAATACATTTAGGTAGCAATATC
TGAATTCAAGCACTCCACCATCACCAGACCACTTTTAATAATATCTAAAATACAAAAAATAATTTT
ACAGAATAGCATGAAAAGTATGAAACGAACTATTTAGGTTTTTCACATACAAAAAAAAAAAGAATT
TTGCTCGTGCGCGAGCGCCAATCTCCCATATTGGGCACACAGGCAACAACAGAGTGGCTGCCCACA
GAACAACCCACAAAAAACGATGATCTAACGGAGGACAGCAAGTCCGCAACAACCTTTTAACAGCAG
GCTTTGCGGCCAGGAGAGAGGAGGAGAGGCAAAGAAAACCAAGCATCCTCCTTCTCCCATCTATAA
ATTCCTCCCCCCTTTTCCCCTCTCTATATAGGAGGCATCCAAGCCAAGAAGAGGGAGAGCACCAAG
GACACGCGACTAGCAGAAGCCGAGCGACCGCCTTCTCGATCCATATCTTCCGGTCGAGTTCTTGGT
CGATCTCTTCCCTCCTCCACCTCCTCCTCACAGGGTATGTGCCTCCCTTCGGTTGTTCTTGGATTT
ATTGTTCTAGGTTGTGTAGTACGGGCGTTGATGTTAGGAAAGGGGATCTGTATCTGTGATGATTCC
TGTTCTTGGATTTGGGATAGAGGGGTTCTTGATGTTGCATGTTATCGGTTCGGTTTGATTAGTAGT
ATGGTTTTCAATCGTCTGGAGAGCTCTATGGAAATGAAATGGTTTAGGGATCGGAATCTTGCGATT
TTGTGAGTACCTTTTGTTTGAGGTAAAATCAGAGCACCGGTGATTTTGCTTGGTGTAATAAAGTAC
GGTTGTTTGGTCCTCGATTCTGGTAGTGATGCTTCTCGATTTGACGAAGCTATCCTTTGTTTATTC
CCTATTGAACAAAAATAATCCAACTTTGAAGACGGTCCCGTTGATGAGATTGAATGATTGATTCTT
AAGCCTGTCCAAAATTTCGCAGCTGGCTTGTTTAGATACAGTAGTCCCCATCACGAAATTCATGGA
AACAGTTATAATCCTCAGGAACAGGGGATTCCCTGTTCTTCCGATTTGCTTTAGTCCCAGAATTTT
TTTTCCCAAATATCTTAAAAAGTCACTTTCTGGTTCAGTTCAATGAATTGATTGCTACAAATAATG
GTGCAAATCAGGTCTATATGATTGATTTTGGGCTGGCCAAGAAGTATAGAGACTCATCAACTCATC
AGCATATTCCGTATAGAGAAAACAAAAATTTGACAGGAACTGCTAGATACGCAAGCATGAATACTC
ATCTTGGCATTGAACAAAGTCGAAGGGATGATTTGGAATCGCTGGGTTATGTTTTAATGTACTTCT
TAAGAGGAAGTCTCCCTTGGCAGGGGCTGAAAGCAGGCACTAAGAAACAGAAGTATGAGAAGATCA
GTGAGAAGAAAGTATCAACATCAATAGAGACCTTGTGTAGGGGATATCCTGCAGAGTTTGCATCAT
ATTTTCATTACTGTCGATCACTAAGATTTGATGATAAACCAGATTATGCTTATCTGAAGAGAATTT
TCCGTGATCTTTTCATTCGTGAAGGGTTTCAATTTGATTATATATTTGACTGGACCATTTTGAAAT
ATCAGCAATCACAGCTTGCCAATCCTCCATCTCGTGCTCTTGGTGGTACTGCTGGGCCAAGCTCAG
GGATGCCTCATGCTCTTGTTAATGTTGAGAGGCAATCAGGTGGAGATGAAGGTCGACCAACTGGTT
GGTCTTCATCAAATCTTACACGTAATAAGAGCACGGGGCTGCATTTCAATTCTGGAAGCTTATTGA
AGCAAAAAGGCACAGTTGCTAATGATTTATCCATGGGTAAAGAGTTATCCAGTTCTAATTTTTTCC
GGTCAAGTGGACCATTGAGGCGTCCAGTTGTCTCTAGCATCCGAGACCCAGTGATTGCAGGGGGTG
AACCTGACCCCTCCGGCACTCTGACAAAAGATGCAAGCCCGGGACCATTGCGTAAAGTATCCAGTG
CTGCACGGAGGAGTTCACCAGTTGTGTCCTCAGATCACAAGCGCAGCTCCTCTATCAAAAATGCCA
ACATAAAGAATTTAGAGTCCACCGTCAAGGGAATAGAGGGTTTAAGTTTTCGATGATGAGGGACTG
CATTAGTAGCTGTGCTTTGTCTCAGTTCTCCGTTCACTGTAAATTTTGGCACACCAACTTGGGGAG
TAAGAGTTCTGATATTAGTTGCTGTCAGGAAGTACCATAAAGCTGAATTATACAATTAAAATTTGG
GATCCAATCGCAAAAGCACATTAAGGATATGATGGGGTTGCAGATCCAAACTCACAGATTCCAGTT
TATGCTCGTCCATACAGTTATAGGCACTTTCCATATTCTTTTCTTTAATCTCTGTCTCTTGCTTGT
TATTGTTATGTCGTGGTATTCTTGTTGAGGTCATGTTTGTGAATTGCGAAGATGGTCATGTATAAT
TGCCGAGAAATCATGTACTAGTTTGTTTTAAACATGAGCAAACTGTTATTTTGTTCAAGCTACTTT
AATATCAAAAAAAAAAAAAAAAAGGGCGGCCGCTCTAGAGTATCCCTCGAGGGGCCCAAGCTTACGC
GTACCCAGCTTT

SEQ ID NO: 4, Artificial sequence - prm07073
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAACAATGACTTCTTTCTCTCTCACTT

FIGURE 2 (continued)

SEQ ID NO: 5, Artificial sequence - prm07074
GGGGACCACTTTGTACAAGAAAGCTGGGTCAATAGCTTTTGAATCAATCT

SEQ ID NO: 6, Aquilegia formosa - TA9067
SKNELCRLSSTFLSTMASLQFLAPSSSPLQSNRLMVRATSSTSPSVNQTMVAPDLSRLEPRVEERE
GGYWVLKEKYREKINPQEKIKIEKEPMKFVTEGGIHELAKTPFEELEKAKLTKDDIDVRLKWLGLF
HRRKNHYGRFMMRLKLPNGVTTSEQTRYLASVIRRYGKDGCADVTTRQNWQIRGVELPHVPEIMKG
LNQVGLTSLQSGMDNVRNPVGNPLAGIDPLEIVDTRPYNDQLSRFITGNFKGNLAFTNLPRKWNVC
VVGSHDLFEHPHINDLAYMPATKNGRFGFNLLVGGFFSPKRCAEAIPLDAWVSGEDVIPVCKAILE
AYRDLGTRGNRQKTRMMWLIDELGVEGFRSEVVKRMPEQELERSSTEELVQKQWERRDLIGVHAQK
QAGYSFVGLHIPVGRLQADDMDELARIADEYGSGELRLTVEQNIIIPNVENSRVEALLKEALLRDR
FSPTPPLLMKGLVACTGNQFCGQAIIETKARALKVTEEVERLVAVTKPVRMHWTGCPNTCAQVQVA
DIGFMGCMARDENGKPCEGADVYLGGRIGSDSHLGDIYKKSVPCKDLVPLVVDILIERFGAVPRER
EEDGED

SEQ ID NO: 7, Betula pendula - X60093
MSSLSVRFLSPPLFSSTPAWPRTGLAATQAVPPVVAEVDAGRLEPRVEEREGYWVLKEKFREGINP
QEKLKLEREPMKLFMEGGIEDLAKMSLEEIDKDKISKSDIDVRLKWLGLFHRRKHHYGRFMMRLKL
PNGVTTSAQTRYLASVIRKYGKDGCADVTTRQNWQIRGVVLSDVPEILKGLDEVGLTSLQSGMDNV
RNPVGNPLAGIDIHEIVATRPYNNLLSQFITANSRGNLAFTNLPRKWNVCVVGSHDLFEHPHINDL
AYMPAIKDGRFGFNLLVGGFFSPRRCAEAVPLDAWVSADDIILVCKAILEAYRDLGTRGNRQKTRM
MWLIDELGIEGFRSEVVKRMPNQELERAAPEDLIEKQWERRELIGVHPQKQEGLSYVGLHIPVGRV
QADDMDELARLADTYGCGELRLTVEQNIIIPNIENSKLEALLGEPLLKDRFSPEPPILMKGLVACT
GNQFCGQAIIETKARALKVTEEVQRQVAVTRPVRMHWTGCPNSCGQVQVADIGFMGCMARDENGKP
CEGAAVFLGGRIGSDSHLGNLYKKGVPCKNLVPLVVDILVKHFGAVPREREESED

SEQ ID NO: 8, Capsicum annuum - TA5054
MTATIITTLNNQESTKFLNSKFGEMASFSVKFSATSSLTSSKRFSKLHATPPQTVAVPPSGAVEVA
AERLEPRLEERDGYWVLKEKFRKGINPAEKAKIEKEPMKLFTENGIEDIAKISLEEIEKSKLAKDD
IDVRLKWLGLFHRRKHQYGRFMMRLKLPNGITTSAQTRYLASVIRKYGKDGCADVTTRQNWQIRGV
VLPDVPEILKGLDEVGLTSLQSGMDNVRNPVGNPLAGIDPQEIVDTRPYANLLSNLLSQYVTANFR
GNLSVHNLPRKWNVCVIGSHDLYEHPHINDLAYMPATKDGRFGFNLLVGGFFSPKRCAEAIPLDAW
VPADDVVPVCKTILEAYRDLGTRGNRQKTRMMWLIDELGVEGFRAEVVKRMPQKKLERESTEDLVQ
KQWERREYLGVNPQKQEGYSFVGLHIPVGRVQADDMDELARLAEEYGSGELRLTVEQNIIIPNIEN
SKIDALLNEPLLKQISPDPPILMRNLVACTGNQFCGQAIIETKARSMKITEEVQRLVSVTQPVRMH
WTGCPNSCGQVQVADIGFMGCLTRKEGKTVEGADVFLGGRIGTDSHLGDIYKKSVPCEDLVPIIVD
LLVNNFGAVPREREEAED

SEQ ID NO: 9, Chlamydomonas reinhardtii - 59303
MLLHAPHVKPLGQRSSIRRGNLVVANVACTAGKNPTSRPAKRSKVEFIKENSDHLRHPLMEELVND
ETFITEDSVQLMKFHGSYQQDNREKRAFGQGKAYSFLMRTRQPAGVVPNRLYLVMDDLADQFGNGT
LRLTTRQAYQLHGVLKKDLKTVFSSVIKNMGSTLAACGDVNRNVMGPAAPFTNRPDYLAAQKAALD
LADLLTPQSGAYYDVWLDGEKFMSSYKEDPAVTEARAFNGFGTNFDNSPEPIYGSQYLPRKFKIAT
TVPGDNSVDLFTQDLGVVVQGYNLYVGGGQGRSHRDADTFPRLADPLGYVAAADLFAAAKAVVAVF
RDYGRRDNRKQARTRHMLAEWGVDKFRSVAEQYLGKRFQEPVPLPPWQYKDYLGWGEQGDGRLYCG
VYVQNGRIKGEAKRALRAAIERYSLPVVLTPHQNLVLRDVRPEDREDIEQLLRAGGVKELVEWDGL
DRLSMACPALPLCGLAVTEAERALPDVNTRIRAMLTRAGLPPSQPLHVRMTGCPNGCVRPYMAELG
LVGDGPNSYQLWLGGGPAQTRLAQPYAERVKVKDLESTLEPLFGAWRAGRQPDEAFGDWVARLGFD
AVRQQAAAAAAAAPVGTA

FIGURE 2 (continued)

SEQ ID NO: 10, Chlamydomonas reinhardtii - 192085
MQSRQCLNRKASGARPCANSRSLTARVLATAAPVAPSATPASAPLPLPDGVGEHSGLKHLPEAART
RALDKKANKFEKVKVEKCGSRAWNDVFELSSLLKEGKTKWEDLNLDDVDIRLKWAGLFHRGKRTPG
KFMMRLKVPNGELTAAQLRFLASSIAPYGADGCADITTRANIQLRGVTMEDSETVIKGLWDVGLTS
FQSGMDSVRNLTGNPIAGVDPHELVDTRPLLRDMEAMLFNNGKGREEFANLPRKLNICISSTRDDF
PHTHINDVGYEAVAKPNGEVVYNVVVGGYFSIKRNIMSIPLGCSITQDQLMPFTEALLRVFRDHGP
RGDRQQTRLMWLVEAVGVDKFRQLLSEYMGGATFGEPVHVHHDQPWERRNLLGVHRQRQAGLNWVG
ACVPAGRLHAADFEEIAAVAEKYGDGTVRITCEENVIFTNVPDAKLEAMKAEPLFQRFPIFPGVLL
SGMVSCTGNQFCGFGLAETKAKAVKVVEALDAQLELSRPVRIHFTGCPNSCGQAQVGDIGLMGAPA
KHEGKAVEGYKIFLGGKIGENPALATEFAQGVPAIESVLVPRLKEILISEFGAKERATATA

SEQ ID NO: 11, Chlamydomonas reinhardtii - 192232
MLLKGITTPMLGQQRPTRGQLHVVNVATPSKNPSSRLAKRSKVEIIKEKSDYLRHPLMEELVNDAT
FITEDSVQLMKFHGSYQQDHREKRAFGQGKAYCFMMRTRQPAGVVPNRLYLVMDDLADQYGNGTLR
LTTRQAYQLHGVLKKDLKTVFSSVIKNMGSTLAACGDVNRNVMGPSAPFTNRPDYVAAQKAANDIA
DLLTPQSGAYYDVWLDGEKFMSAYKEDPKVTADRAYNGFGTNFENSPEPIYGAQFLPRKFKVATTV
PGDNSVDLFTQDLGVVVIMDESGKEVKGYNLTVGGGMGRTHRDDETFPRLADPLGYVDKDDLFHAV
KAVVAVQRDYGRRDNRKQARLKYLVGLPADQELHVRMTGCPNGCARPYMAELGFVGDGPNSYQLYF
GGNVNQTRLAQLFADRVKVKDLESTLEPIFAAWKASRRPKESFGDWVSRPSQDPKNLSSVQQGTQH
ESAVVAH

SEQ ID NO: 12, Gossypium hirsutum - TA24262
MSSLSVRFFAPQQPLLPSTASSFKPKTWVMAAPTTAPATSVDVDGGRLEPRVEEREGYFVLKEKFR
DGINPQEKIKIEKDPLKLFMEAGIDELAKMSFEDLDKAKATKDDIDVRLKWLGLFHRRKHQYGRFM
MRLKLPNGVTTSAQTRYLASVIRKYGKEGCADVTTRQNWQIRGAVLPDVPEILKGLDEVGLTSLQS
GMDNVRNPVGNPLAGIDPEEIVDTRPYTNLLSQFITANSRGNPAVANLPRKWNVCVVGSHDLYEHP
HINDLAYMPATKNGRFGFNLLVGGFFSAKRCDEAIPLDAWVSADDVIPLCKAVLEAYRDLGYRGNR
QKTRMMWLIDELGIEVFRSEVAKRMPQKELERASDEDLVQKQWERRDYLGVHPQKQEGFSYIGIHI
PVGRVQADDMDELARLADTYGSGEFRLTVEQNIIIPNVENSKLEALLNEPLLKDRFSPQPSILMKG
LVACTGNQFCGQAIIETKARALKVTEEVERLVSVSRPVRMHWTGCPNTCGQVQVADIGFMGCMARD
ENGKPCEGADIFLGGRIGSDSHLGELYKKGVPCKNLVPVVADILVEPFGAVPRQREEGED

SEQ ID NO: 13, Hordeum vulgare - TA43088
MASSASLQSFLPPSAHAATSSSRLRPSRARPVQCAAVSAPSSSSSSASPSASAVPSERLEPRVEQR
EGGYWVLKEKYRTSLNPQEKVKLGKEPMALFTEGGINDLAKLPMEQIDADKLTKEDVDVRLKWLGL
FHRRKQQYGRFMMRLKLPNGVTTSEQTRYLASVIDKYGEEGCADVTTRQNWQIRGVTLPDVPEILD
GLRSVGLTSLQSGMDNVRNPVGSPLAGIDPLEIVDTRPYTNLLSSYITNNSEGNLAITNLPRKWNV
CVIGTHDLYEHPHINDLAYMPAEKDGKFGFNLLVGGFISPKRWGEALPLDAWVPGDDIIPVCKAVL
EAFRDLGTRGNRQKTRMMWLIDELGMEAFRSEIEKRMPNGVLERAAPEDLIDKKWERRDYLGVHPQ
KQEGLSFVGLHVPVGRLQAADMFELARLADEYGSGELRLTVEQNIVLPNVKNEKVEALLAEPLLHK
FSAHPSLLMKX

SEQ ID NO: 14, Lotus japonicus - TA2640
MSSSFSIRFLAPPFPSTSRPKSCLSAATPAVAPTDAAVSRLEPRVEERNGYWVLKEEHRGGINPQE
KVKLEKEPMALFMEGGIDELAKVSIEELDSSKLTKDDVDVRLKWLGLFHRRKHQYGRFMMRLKLPN
GVTTSAQTRYLASVIRKYGKDGCADVTTRHNWQIRGVVLPDVPEILKGLAEVGLTSLQSGMDNVRN
PVGNPLAGIDPDEIVDTRPYTNLLSHFITANSRGNPTVSNLPRKWNVCVVGSHDLFEHPHINDLAY
MPANKDGRFGFNLLVGGFFSPKRCAEAIPLDAWVSAEDVIPVCKAILEMYRDLGTRGNRQKTRMMW

FIGURE 2 (continued)

LIDELGIEVFRSEVVKRMPLGQQLERASQEDLVQKQWERRDYFGANPQKQEGLSYVGIHIPVGRIQ
ADEMDELARLADEYGTGELRLTVEQNIIIPNVENSKLSALLNEPLLKEKFSPEPSLLMKTLVACTG
SQFCGQAIIETKARALKVTEEVERLVAVTRPVRMHWTGCPNTCGQVQVADIGFMGCMARDENGKPG
EGVDIFLGGRIGSDSHLAEVYKKAVPCKDLVPIVADILVKHFGAVQRNREEGDD

SEQ ID NO: 15, Nicotiana tabacum - TA16376

MASFSVKFSATSLPNPNRFSRTAKLHATPPQTVAVPPSGEAEIASERLEPRVEEKDGYWVLKEKFR
QGINPAEKAKIEKEPMKLFMENGIEDLAKISLEEIEGSKLTKDDIDVRLKWLGLFHRRKHHYGRFM
MRLKLPNGVTTSAQTRYLASVIRKYGKDGCGDVTTRQNWQIRGVVLPDVPEILKGLDEVGLTSLQS
GMDNVRNPVGNPLAGIDPHEIVDTRPYTNLLSQYVTANFRGNPAVTNLPRKWNVCVIGSHDLYEHP
HINDLAYMPASKDGRFGFNLLVGGFFSPKRCAEAVPLDAWVPADDVVPVCKAILEAYRDLGTRGNR
QKTRMMWLVDELGVEGFRAEVVKRMPQQKLDRESTEDLVQKQWERREYLGVHPQKQEGYSFVGLHI
PVGRVQADDMDELARLADNYGSGELRLTVEQNIIIPNVENSKIESLLNEPLLKNRFSTNPPILMKN
LVACTGNQFCGQAIIETKARSMKITEEVQRLVSVTKPVRMHWTGCPNSCGQVQVADIGFMGCLTRK
EGKTVEGADVYLGGRIGSDSHLGDVYKKSVPCEDLVPIIVDLLVNNFGAVPREREEAED

SEQ ID NO: 16, Nicotiana tabacum - TA13596

MASFSIKFLAPSLPNPARFSKNAVKLHATPPSVAAPPTGAPEVAAERLEPRVEEKDGYWILKEQFR
KGINPQEKVKIEKQPMKLFMENGIEELAKIPIEEIDQSKLTKDDIDVRLKWLGLFHRRKNQYGRFM
MRLKLPNGVTTSAQTRYLASVIRKYGKEGCADITTRQNWQIRGVVLPDVPEILKGLAEVGLTSLQS
GMDNVRNPVGNPLAGIDPEEIVDTRPYTNLLSQFITGNSRGNPAVSNLPRKWNPCVVGSHDLYEHP
HINDLAYMPATKDGRFGFNLLVGGFFSAKRCDEAIPLDAWVPADDVVPVCKAILEAFRDLGFRGNR
QKCRMMWLIDELGVEGFRAEVEKRMPQQQLERASPEDLVQKQWERRDYLGVHPQKQEGYSFIGLHI
PVGRVQADDMDELARLADEYGSGEIRLTVEQNIIIPNIENSKIEALLKEPVLSTFSPDPPILMKGL
VACTGNQFCGQAIIETKARSLMITEEVQRQVSLTRPVRMHWTGCPNTCAQVQVADIGFMGCLTRDK
NGKTVEGADVFLGGRIGSDSHLGEVYKKAVPCDDLVPLVVDLLVNNFGAVPREREETED

SEQ ID NO: 17, Nicotiana tabacum - AB093534

MASFSVKFSATSLPNHKRFSKLHATPPQTVAVAPSGAAEIASERLEPRVEEKDGYWVLKEKFRQGI
NPAEKAKIEKEPMKLFMENGIEDLAKISLEEIEGSKLTKDDIDVRLKWLGLFHRRKHHYGRFMMRL
KLPNGVTTSSQTRYLASVIRKYGKDGCADVTTRQNWQIRGVVLPDVPEILKGLDEVGLTSLQSGMD
NVRNPVGNPLAGIDPHEIVDTRPYTNLLSQYVTANFRGNPAVTNLPRKWNVCVIGSHDLYEHPQIN
DLAYMPATKDGRFGFNLLVGGFFSPKRCAEAVPLDAWVPADDVVPVCKAILEAYRDLGTRGNRQKT
RMMWLVDELGVEGFRAEVVKRMPQQKLDRESTEDLVQKQWERREYLGVHPQKQEGYSFVGLHIPVG
RVQADDMDELARLADEYGSGELRLTVEQNIIIPNVKNSKIEALLNEPLLKNRFSTDPPILMKNLVA
CTGNQFCGKAIIETKARSMKITEEVQLLVSITQPVRMHWTGCPNSCAQVQVADIGFMGCLTRKEGK
TVEGADVYLGGRIGSDSHLGDVYKKSVPCEDLVPIIVDLLVDNFGAVPREREEAED

SEQ ID NO: 18, Oryza sativa - Os01g0357100

MASSASLQRFLPPYPHAAASRCRPPGVRARPVQSSTVSAPSSSTPAADEAVSAERLEPRVEQREGR
YWVLKEKYRTGLNPQEKVKLGKEPMSLFMEGGIKELAKMPMEEIEADKLSKEDIDVRLKWLGLFHR
RKHQYGRFMMRLKLPNGVTTSEQTRYLASVIEAYGKEGCADVTTRRQIRGVTLPDVPAILDGLNAV
GLTSLQSGMDNVRNPVGNPLAGIDPDEIVDTRSYTNLLSSYITSNFQGNPTITNLPRKWNVCVIGS
HDLYEHPHINDLAYMPAVKGGKFGFNLLVGGFISPKRWEEALPLDAWVPGDDIIPVCKAVLEAYRD
LGTRGNRQKTRMMWLIDELGMEAFRSEVEKRMPNGVLERAAPEDLIDKKWQRRDYLGVHPQKQEGM
SYVGLHVPVGRVQAADMFELARLADEYGSGELRLTVEQNIVIPNVKNEKVEALLSEPLLQKFSPQP
SLLLKGLVACTGNQFCGQAIIETKQRALLVTSQVEKLVSVPRAVRMHWTGCPNSCGQVQVADIGFM
GCLTKDSAGKIVEAADIFVGGRVGSDSHLAGAYKKSVPCDELAPIVADILVERFGAVRREREEDEE

FIGURE 2 (continued)

SEQ ID NO: 19, Physcomitrella patens - 193361
MQGAMQTKMWRGELISTSTHFIGGTRLQPKLNQDARKPTKSENCIVRVSMEREVKAKAAVSPPAVA
ADRLTPRVQERDGYYVLKEEFRQGINPQEKIKLGKEPMKFFIENEIEELAKTPFAELDSSKPGKDD
IDVRLKWLGLFHRRKHQYGRFMMRFKLPNGITNSTQTRFLAETISKYGKEGCADLTTRQNWQIRGI
MLEDVPSLLKGLESVGLSSLQSGMDNVRNAVGNPLAGIDPDEIVDTIPICQALNDYIINRGKGNTE
ITNLPRKWNVCVVGTHDLFEHPHINDLAYVPATKNGVFGFNILVGGFFSSKRCAEAIPMDAWVPTD
DVVPLCKAILETYRDLGTRGNRQKTRMMWLIDEMGVEEFRAEVERRMPSGTIRRAGQDLIDPSWKR
RSFFGVNPQKQAGLNYVGLHVPVGRLHAPEMFELARIADEYGNGEIRITVEQNLILPNIPTEKIDK
LMQEPLLQKYSPNPTPLLANLVACTGSQFCGQAIAETKALSLQLTQQLEDTMETTRPIRLHFTGCP
NTCAQIQVADIGFMGTMARDENRKPVEGFDIYLGGRIGSDSHLGELVVPGVPATKLLPVVQELMIQ
HFGAKRKP

SEQ ID NO: 20, Physcomitrella patens - 144369
MQGTMQSQMWRGQVSGASLHFTGATRVQGNSHQDLVYPTQFHKHGVRASAEREVKAKAVAAPPTIA
ADRLVPRVEERDGYYVLKEEFRQGINPSEKIKIAKEPMKFFMENEIEELAKTPFAELDSSKAGKDD
IDVRLKWLGLFHRRKHQYGRFMMRFKLPNGITNSSQTRFLAETISKYGEYGCADLTTRQNWQIRGI
VLEDVPALLKGLESVGLSSLQSGMDNVRNPVGNPLAGIDPDEIVDTAPFCKVLSDYIINRGQGNPQ
ITNLPRKWNVCVVGTHDLFEHPHINDLAYMPATKNGVFGFNILVGGFFSPKRCAEAIPMDAWVPAD
DVVPLCKAILETYRDLGTRGNRQKTRMMWLIDEMGIEEFRAEVERRMPGGSILRAGKDLVDPSWTR
RSFYGVNPQKQPGLNYVGLHIPVGRLHAPEMFELARIADEYGNGEIRISVEQNLILPNVPTEKIEK
LLKEPLLEKYSPNPTPLLANLVACTGSQFCGQAIAETKARSLQLTQELEATMETTRPIRLHFTGCP
NTCAQIQVADIGFMGTMARDENRKPVEGFDIYLGGRIGSDSHLGELVVPGVPATKLLPVVQDLMIQ
HFGAKRKT

SEQ ID NO: 21, Pinus taeda - TA3139
MNLSSPVRFDEIRPLAHVVYNPVCCGHKPNRLRLMTAIQVRAVNHGGRNSEISTDGNSKGTTAKAV
ASPAGSHVAVDASRLEARVEERDGYWVLKEEFRAGINPQEKIKLQREPMKLFMENEIEELAKKPFA
EIESEKVNKDDIDVRLKWLGLFHRRKHHYGRFMMRLKLPNGVTTSLQTRYLASVIQQYGPEGCADI
TTRQNWQIRGVVLDDVPAILKGLKEVGLSSLQSGMDNVRNPVGNPLAGIDADEIIDTRPYTKVLTD
YIVNNGKGNPSITNLPRKWNVCVVGTHDLFEHPHINDLAYIPAMNSGRFGFNLLVGGFFSPKRCEE
AVPLDAWVAGEDVVPVCRAILEVYRDLGTRGNRQKTRMMWLIDELGIEGFRSEVVKRMPGEKLERA
ATEDMLDKSWERRSYLGVHPQKQEGLNFVGLHVPVGRLQAEDMLELARLAEQYGTQELRLTVEQNA
IIPNVPTDKIEALLQEPLLQKFSPSPPLLVSTLVACTGNQFCGQAIIETKARALKITEELDRTMEV
PKPVRMHWTGCPNTCGQVQVADIGFMGCMTRDENKKVVEGVDIFIGGRVGADSHLGDLIHKGVPCK
DVVPVVQELLIKHFGAIRKTDM

SEQ ID NO: 22, Populus trichocarpa - 130.69
MSSLSVRFLTPQLSPTVPSSSARPRTRLFAGPPTVAQPAETGVDAGRLEPRVEKKDGYYVLKEKFR
QGINPQEKVKIEKEPMKLFMENGIEELAKLSMEEIDKEKSTKDDIDVRLKWLGLFHRRKHQYGRFM
MRLKLPNGVTTSAQTRYLASVIRKYGKDGCADVTTRQNWQIRGVVLPDVPEILRGLAEVGLTSLQS
GMDNVRNPVGNPLAGIDPDEIVDTRPYTNLLSQFITANSRGNPEFTNLPRKWNVCVVGSHDLYEHP
HINDLAYMPAMKDGRFGFNLLVGGFFSPKRCAEAIPLDAWVSADDVLPSCKAVLEAYRDLGTRGNR
QKTRMMWLIDELGIEGFRSEVVKRMPRQELERESSEDLVQKQWERRDYFGVHPQKQEGLSYAGLHI
PVGRVQADDMDELARLADIYGTGELRLTVEQNIIIPNIEDSKIEALLKEPLLKDRFSPEPPLLMQG
LVACTGKEFCGQAIIETKARAMKVTEEVQRLVSVSKPVRMHWTGCPNTCGQVQVADIGFMGCMARD
ENGKICEGADVYVGGRVGSDSHLGELYKKSVPCKDLVPLVVDILVKQFGAVPREREEVDD

FIGURE 2 (continued)

SEQ ID NO: 23, Solanum lycopersicum - TA37687
MASFSIKFLAPSLPNPTRFSKSSIVKLNATPPQTVAAAGPPEVAAERLEPRVEEKDGYWILKEQFR
QGINPQEKVKIEKEPMKLFMENGIEELAKIPIEEIDQSKLTKDDIDVRLKWLGLFHRRKNQYGRFM
MRLKLPNGVTTSAQTRYLASVIRKYGEEGCADITTRQNWQIRGVVLPDVPEILKGLEEVGLTSLQS
GMDNVRNPVGNPLAGIDPEEIVDTRPYTNLLSQFITGNSRGNPAVSNLPRKWNPCVVGSHDLYEHP
HINDLAYMPAIKDGRFGFNLLVGGFFSAKRCDEAIPLDAWVPADDVVPVCKAILEAFRDLGFRGNR
QKCRMMWLIDELGVEGFRAEVVKRMPQQELERASPEDLVQKQWERRDYLGVHPQKQEGYSFIGLHI
PVGRVQADDMDDLARLADEYGSGELRLTVEQNIIIPNIENSKIDALLKEPILSKFSPDPPILMKGL
VACTGNQFCGQAIIETKARSLKITEEVQRQVSLTRPVRMHWTGCPNTCAQVQVADIGFMGCLTRDK
DKKTVEGADVFLGGRIGSDSHLGEVYKKAVPCDELVPLIVDLLIKNFGAVPREREETED

SEQ ID NO: 24, Solanum lycopersicum - TA37689
MTSFSVKFSATSLPNSNRFSKLHATPPQTVAVPSYGAAEIAAERLEPRVEQRDGYWVVKDKFRQGI
NPAEKAKIEKEPMKLFTENGIEDLAKISLEEIEKSKLTKEDIDIRLKWLGLFHRRKHHYGRFMMRL
KLPNGVTTSDQTRYLGSVIRKYGKDGCGDVTTRQNWQIRGVVLPDVPEILKGLDEVGLTSLQSGMD
NVRNPVGNPLAGIDLHEIVDTRPYTNLLSQYVTANFRGNVDVTNLPRKWNVCVIGSHDLYEHPHIN
DLAYMPATKDGRFGFNLLVGGFFSPKRCAEAIPLDAWVPADDVVPVCKAILEAYRDLGTRGNRQKT
RMMWLIDELGVEGFRAEVVKRMPQKKLDRESSEDLVLKQWERREYLGVHPQKQEGYSFVGLHIPVG
RVQADDMDELARLADEYGSGELRLTVEQNIIIPNIENSKIDALLNEPLLKNRFSPDPPILMRNLVA
CTGNQFCGQAIIETKARSMKITEEVQRLVSVTQPVRMHWTGCPNTCGQVQVADIGFMGCLTRKEGK
TVEGADVFLGGRIGSDSHLGEVYKKSVPCEDLVPIIVDLLINNFGAVPREREETEE

SEQ ID NO: 25, Arabidopsis thaliana - AT2G15620
MTSFSLTFTSPLLPSSSTKPKRSVLVAAAQTTAPAESTASVDADRLEPRVELKDGFFILKEKFRKG
INPQEKVKIEREPMKLFMENGIEELAKKSMEELDSEKSSKDDIDVRLKWLGLFHRRKHQYGKFMMR
LKLPNGVTTSAQTRYLASVIRKYGEDGCADVTTRQNWQIRGVVLPDVPEILKGLASVGLTSLQSGM
DNVRNPVGNPIAGIDPEEIVDTRPYTNLLSQFITANSQGNPDFTNLPRKWNVCVVGTHDLYEHPHI
NDLAYMPANKDGRFGFNLLVGGFFSPKRCEEAIPLDAWVPADDVLPLCKAVLEAYRDLGTRGNRQK
TRMMWLIDELGVEGFRTEVEKRMPNGKLERGSSEDLVNKQWERRDYFGVNPQKQEGLSFVGLHVPV
GRLQADDMDELARLADTYGSGELRLTVEQNIIIPNVETSKTEALLQEPFLKNRFSPEPSILMKGLV
ACTGSQFCGQAIIETKLRALKVTEEVERLVSVPRPIRMHWTGCPNTCGQVQVADIGFMGCLTRGEE
GKPVEGADVYVGGRIGSDSHIGEIYKKGVRVTELVPLVAEILIKEFGAVPREREENED

SEQ ID NO: 26, Vitis vinifera - GSVIVT00036600001
MASISVPFLSQAPTHLSNSTSLRLKTRISATPTPTPTPTTVAPSSTAAVDASRMEPRVEERGGYWV
LKEKFREGINPQEKVKIEKDPMKLFIEDGFNELASMSFEEIEKSKHTKDDIDVRLKWLGLFHRRKH
QYGRFMMRLKLPNGVTSSAQTRYLASAIRQYGKEGCADVTTRQNWQIRGVVLPDVPEILKGLSEVG
LTSLQSGMDNVRNPVGNPLAGIDPHEIVDTRPYTNLLSQFITANARGNTAFTNLPRKWNVCVVGSH
DLYEHPHINDLAYMPATKKGRFGFNLLVGGFFSPKRCADAIPLDAWIPADDVLPVCQAVLEAYRDL
GTRGNRQKTRMMWLIDELGIEQFRAEVVKRMPQQELERSSSEDLVQKQWERRDYLGVHPQKQEGFS
FVGIHIPVGRVQADDMDELARLADEYGSGELRLTVEQNIIIPNVENSRLEALLKEPLLRDRFSPEP
PILMKGLVACTGNQFCGQAIIETKARALKVTEDVGRLVSVTQPVRMHWTGCPNSCGQVQVADIGFM
GCMTRDENGNVCEGADVFLGGRIGSDCHLGEVYKKRVPCKDLVPLVAEILVNHFGGVPREREEEAE
D

SEQ ID NO: 27, Volvox sp. 83067
MQSQSLSRRTCTRTLGRGLVTPVLATAAPASAAQAADGINAHSGLKHLPEAARVRALDRKANKFEK
VKVEKCGSRAWTDVFELSRLLKEGNTKWEDLDLDDIDIRMKWAGLFHRGKRTPGKFMMRLKVPNGE

FIGURE 2 (continued)

LDARQLRFLASAIAPYGADGCADITTRANIQLRGVTLADADAIIRGLWDVGLTSFQSGMDSVRNLT
GNPIAGVDPHELIDTRPLLREMEAMLFNNGKGREEFANLPRKLNICISSTRDDFPHTHINDVGFEA
VRRPDDGEVVFNVVVGGFFSIKRNVMSIPLGCSVTQDQLMPFTEALLRVFRDHGPRGDRQQTRLMW
MVDAIGVEKFRQLLSEYMGGAELAPPVHVHHEGPWERRDVLGVHPQKQPGLNWVGACVPAGRLQAA
DFDEFARIAETYGDGTVRITCEENVIFTNVPDAKLPDMLAEPLFQRFKVNPGLLLRGLVSCTGNQF
CGFGLAETKARAVKVVEMLEEQLELTRPVRIHFTGCPNSCGQAQQVGDIGLMGAPAKLDGKAVEGY
KIFLGGKIGENPQLATEFAQGIPAVESHLVPKLKEILIKEFGAKEKETAVVV

SEQ ID NO: 28, Spinacia oleracea - X07568
MASLPVNKIIPSSTTLLSSSNNNRRRNNSSIRCQKAVSPAAETAAVSPSVDAARLEPRVEERDGFW
VLKEEFRSGINPAEKVKIEKDPMKLFIEDGISDLATLSMEEVDKSKHNKDDIDVRLKWLGLFHRRK
HHYGRFMMRLKLPNGVTTSEQTRYLASVIKKYGKDGCADVTTRQNWQIRGVVLPDVPEIIKGLESV
GLTSLQSGMDNVRNPVGNPLAGIDPHEIVDTRPFTNLISQFVTANSRGNLSITNLPRKWNPCVIGS
HDLYEHPHINDLAYMPATKNGKFGFNLLVGGFFSIKRCEEAIPLDAWVSAEDVVPVCKAMLEAFRD
LGFRGNRQKCRMMWLIDELGMEAFRGEVEKRMPEQVLERASSEELVQKDWERREYLGVHPQKQQGL
SFVGLHIPVGRLQADEMEELARIADVYGSGELRLTVEQNIIIPNVENSKIDSLLNEPLLKERYSPE
PPILMKGLVACTGSQFCGQAIIETKARALKVTEEVQRLVSVTRPVRMHWTGCPNSCGQVQVADIGF
MGCMTRDENGKPCEGADVFVGGRIGSDSHLGDIYKKAVPCKDLVPVVAEILINQFGAVPREREEAE

SEQ ID NO: 29, Nostoc sp. NC - 003272
MTDTVTTPKASLNKFEKFKAEKDGLAIKSEIEKIASLGWEAMDATDRDHRLKWVGVFFRPVTPGKF
MMRMRMPNGILTSDQMRVLAEVVQRYGDDGNADITTRQNIQLRGIRIEDLPHIFNKFHAVGLTSVQ
SGMDNIRNITGDPIAGLDADELYDTRELVQQIQDMLTNKGEGNREFSNLPRKFNIAIAGGRDNSVH
AEINDLAFVPAFKEGIGDWVLGNGEESSTYQKVFGFNVLVGGFFSAKRCEAAIPLNAWVTPEEVLP
LCRAILEVYRDNGLRANRLKSRLMWLIDEWGIDKFRAEVEQRLGKSLLPAAPKDEIDWEKRDHIGV
YKQKQEGLNYVGLHIPVGRLYAEDMFELARIADVYGSGEIRMTVEQNIIIPNITDSRLRTLLTDPL
LERFSLDPGALTRSLVSCTGAQFCNFALIETKNRALEMIKGLEAELTFTRPVRIHWTGCPNSCGQP
QVADIGLMGTKARKNGKAVEGVDIYMGGKVGKDAHLGSCVQKGIPCEDLHLVLRDLLITNFGAKPR
QEALVTSQ

SEQ ID NO: 30, Plectonema boryanum - D31732
MTDTLAAPTLNKFEKLKAEKDGLAVKAELEHFARLGWEAMDETDRDHRLKWLGVFFRPVTPGKFML
RMRVPNGIITSGQTRVLGEILQRYGDDGNADITTRQNFQLRGIRIEDLPEIFRKFDQAGLTSIQSG
MDNVRNITGSPVAGIDADELIDTRGLVRKVQDMITNNGRGNSSFSNLPRKFNIAIAGCRDNSVHAE
INDIAFVPAFKDGTLGFNILVGGFFSGKRCEAAIPLNAWVDPRDVVAVCEAILTVYRNLGLRANRQ
KARLMWLIDEMGLEPFREAVEKQLGYAFTPAAAKDEILWDKRDHIGIHAQKQPGLNYVGLHVPVGR
LYAQDLFDLARIAEVYGSGEIRLTVEQNVIIPNVPDSRVSALLREPIVKRFSIEPQNLSRALVSCT
GAQFCNFALIETKNRAVALMQELEQDLYCPRPVRIHWTGCPNSCGQPQVADIGLMGTKVRKDGKTV
EGVDLYMGGKVGKHAELGTCVRKSIPCEDLKPILQEILIEQFGARLWSDLPESARPNPTALITLDR
PTVETPNGKSTTVQELNAQEFDYVLSAPPVVKAPTEIAAPATIRFAQSGKEITCTQDDLILDIADQ
AEVAIESSCRSGTCGSCKCTLLEGEVSYDSEPDVLDEHDRASGQILTCIARPVGRILLDA

SEQ ID NO: 31, Anabaena variabilis CP000117
MTDTATTPKASLNKFEKFKAEKDGLAIKSEIEKIASLGWEAMDETDRDHRLKWVGVFFRPVTPGKF
MMRMRMPNGILTSDQMRVLAEVVQRYGDDGNADITTRQNIQLRGIRIEDLPHIFNKFHAVGLTSVQ
SGMDNIRNITGDPIAGLDADELYDTRELVQQIQDMLTNKGEGNREFSNLPRKFNIAIAGGRDNSVH
AEINDLAFVPAFKEGIGDWVLGGGEESSTHQKVFGFNVLVGGFFSAKRCEAAIPLNAWVTAEEVVA
LCRAVLEVYRDNGLRANRLKSRLMWLIDEWGIDKFRAEVEQRLGKSLLYAAPKDEIDWEKRDHIGV

FIGURE 2 (continued)

YKQKQEGLNYVGLHIPVGRLYAEDMFELARIADVYGSGEIRMTVEQNIIIPNITDSRLKTLLTDPL
LERFSLDPGALTRSLVSCTGAQFCNFALIETKNRALEMIKGLEAELTFTRPVRIHWTGCPNSCGQP
QVADIGLMGTKARKNGKAVEGVDIYMGGKVGKDAHLGSCVQKGIPCEDLHLVLRDLLITNFGAKPR
QEALVSSQ

SEQ ID NO: 32, Synechococcus sp. CP000239
MANQFERLKSEKDGLAVKAELEAFARMGWENIPEDDRDHRLKWLGIFFRKRTPGQFMLRLRLPNGI
LTSGQMRMLGAIIHPYGEQGVADITTRQNLQLRGIPIEEMPQILGYLKEVGLTSIQSGMDNVRNIT
GSPLAGIDPDELIDVRGLTRKVQDMVTNNGEGNPSFSNLPRKFNIAICGCRDNSVHAEINDLAFVP
AFKNGRLGFNVLVGGFFSARRCAEAIGLDVWVDPRDVVPLCEAVLLVYRDHGLRANRQKARLMWLI
DEWGLEKFRAAVERQIGHPLPRAAEKDEVVWHKRDLLGVHAQKQPGLNFVGLHVPVGRLNALEMME
LARLAEVYGSGELRLTVEQNVLIPNVPDSRVAPLLKEPLLKKFSPNPGPLQRGLVSCTGNQFCNFA
LIETKNRAVALMEELEAELEIPQTVRIHWTGCPNSCGQPQVADIGLMGTTARKDGRVVEAVDIYMG
GEVGKDAKLGECVRKGIPCEDLKPVLVELLIEHFGAKPRQHPSAAQASVLVTR

SEQ ID NO: 33, Arabidopsis thaliana - AT5G04590
MSSTFRAPAGAATVFTADQKIRLGRLDALRSSHSVFLGRYGRGGVPVPPSASSSSSSPIQAVSTPA
KPETATKRSKVEIIKEKSNFIRYPLNEELLTEAPNVNESAVQLIKFHGSYQQYNREERGGRSYSFM
LRTKNPSGKVPNQLYLTMDDLADEFGIGTLRLTTRQTFQLHGVLKQNLKTVMSSIIKNMGSTLGAC
GDLNRNVLAPAAPYVKKDYLFAQETADNIAALLSPQSGFYYDMWVDGEQFMTAEPPEVVKARNDNS
HGTNFVDSPEPIYGTQFLPRKFKVAVTVPTDNSVDLLTNDIGVVVVSDENGEPQGFNIYVGGGMGR
THRMESTFARLAEPIGYVPKEDILYAVKAIVVTQREHGRRDDRKYSRMKYLISSWGIEKFRDVVEQ
YYGKKFEPSRELPEWEFKSYLGWHEQGDGAWFCGLHVDSGRVGGIMKKTLREVIEKYKIDVRITPN
QNIVLCDIKTEWKRPITTVLAQAGLLQPEFVDPLNQTAMACPAFPLCPLAITEAERGIPSILKRVR
AMFEKVGLDYDESVVIRVTGCPNGCARPYMAELGLVGDGPNSYQVWLGGTPNLTQIARSFMDKVKV
HDLEKVCEPLFYHWKLERQTKESFGEYTTRMGFEKLKELIDTYKGVSQ

SEQ ID NO: 34, Aquilegia formosa - TA9067
CTGATCCAAGAATGAACTCTGCAGACTTTCTTCTACCTTTCTTTCAACAATGGCTTCATTACAGTT
TCTTGCACCTTCATCATCACCTTTGCAATCCAACCGACTCATGGTTCGAGCCACTAGTAGTACTAG
TCCATCAGTCAACCAGACCATGGTTGCACCAGACTTATCAAGATTGGAACCAAGAGTTGAAGAAAG
AGAAGGTGGTTATTGGGTTTTGAAAGAGAAATATAGAGAGAAAATAAATCCACAAGAGAAAATCAA
AATAGAGAAAGAACCAATGAAGTTTGTTACTGAAGGTGGTATACATGAATTAGCAAAAACTCCATT
TGAAGAACTTGAGAAAGCTAAACTTACTAAAGATGATATTGATGTTAGACTCAAGTGGCTTGGTCT
TTTTCATAGAAGAAAAAATCATTATGGTAGATTTATGATGAGATTGAAGTTGCCTAATGGAGTTAC
AACTAGTGAACAAACGCGATATCTTGCGAGTGTTATTAGAAGGTATGGAAAGGATGGATGTGCTGA
TGTTACAACTAGACAGAACTGGCAAATTCGCGGTGTTGAGTTACCTCATGTGCCTGAGATAATGAA
AGGATTAAATCAAGTTGGATTAACTAGTCTTCAGAGTGGTATGGATAATGTGCGTAATCCTGTTGG
TAATCCACTTGCTGGTATTGACCCACTAGAGATTGTCGATACTAGACCCTACAATGATCAGCTATC
TCGATTTATTACTGGCAATTTTAAAGGGAACCTGGCTTTTACTAATCTGCCGAGGAAATGGAATGT
ATGTGTGGTGGGCTCTCATGATCTTTTTGAGCATCCCCACATCAATGATCTTGCTTACATGCCAGC
CACAAAGAATGGCCGTTTTGGGTTTAATCTGTTAGTAGGTGGTTTCTTCAGTCCAAAAAGATGTGC
AGAGGCAATTCCTCTCGATGCCTGGGTTTCAGGAGAAGACGTGATCCCAGTTTGCAAAGCTATACT
TGAGGCATACAGAGATCTTGGCACCAGAGGAAACCGACAGAAAACACGAATGATGTGGTTGATTGA
TGAACTTGGGGTAGAAGGATTTAGGTCAGAAGTGGTGAAAAGGATGCCTGAACAAGAGCTGGAGAG
ATCTTCCACTGAAGAGTTGGTTCAAAAGCAATGGGAGAGGAGAGATCTAATCGGTGTCCATGCGCA
AAAGCAGGCAGGCTACAGTTTTGTTGGTCTCCACATACCAGTAGGCAGGCTTCAGGCTGATGACAT
GGATGAACTAGCCCGGATAGCTGATGAGTATGGCTCAGGGGAGCTCCGTCTCACTGTGGAACAAAA

FIGURE 2 (continued)

TATCATAATTCCTAATGTTGAGAACTCAAGAGTTGAAGCTTTGCTGAAGGAAGCCCTATTGAGGGA
CAGGTTTTCACCCACTCCACCTCTTCTAATGAAAGGACTTGTGGCCTGCACAGGCAACCAGTTCTG
TGGACAAGCCATCATTGAGACAAAGGCACGAGCACTGAAGGTGACAGAAGAGGTTGAAAGACTGGT
GGCAGTGACTAAACCAGTAAGAATGCATTGGACAGGATGCCCAAACACCTGCGCGCAGGTGCAAGT
AGCTGATATTGGGTTCATGGGGTGCATGGCAAGAGATGAAAACGGGAAACCGTGTGAAGGAGCAGA
TGTTTACTTAGGTGGGAGGATTGGTAGTGATTCTCATTTGGGAGATATATATAAGAAATCTGTGCC
TTGTAAGGACTTGGTTCCTCTGGTAGTTGACATCTTGATTGAGCGCTTTGGAGCTGTCCCTAGGGA
GAGAGAAGAAGATGGCGAAGACTAGATTATCAAATTCCTAACCGAAAGCCCTTTCTGATTTTAATA
AACTAATTTGGAAGGTGAATGCACATAGACAATTTGGATGAATAAAAGCCATGCAGAAGTGGTTCT
TTTTGGACTTGAGTTGAGGAAGCAACTTTATTGTTGTATCAGAAGACAGGTTATTTTAAATTTCAA
TTCGTTCTTATGTACTCAGAATACTTGGATCATATCTCTAGACATTCTTAATCACCGTTTT

SEQ ID NO: 35, Betula pendula - X60093
AAAAGCTGCTAGAGTATGGAAACATGCTTGTCCAGGAGCAGGACAATGTGAAGAGAGTTCAACTGG
CAGACACGTACTTGAGCCAAGCAGCTCTTGGAGATGCAAACGAGGATTCGATCAAGCGGGGAACTT
TCTATGGCAAGGCAGGCCAACAAGTTAATGTACCCGTTCCTGAAGGTTGCACCGATCCATCTGCTA
GTAACTTTGATCCAACAGCTAGGAGCGATAATGGTAGCTGCCAGTATTGAGGCTAAGCCATTTCTA
GCCTTCTACCTGCTAGGCTATATAAATGCTGTATGAGGTTGGGAGAACTATTCATTTCCACTATTG
CTTGCTTTCTCGATACGGAGAAGTATTCCTAATTTTGTTGTAATGAACGTATAATTTTATCTTAAT
CACAACCACGACTAAAATTACCATTACAAGCTTCAGTTTATTACCATGTCGTCGCTCTCAGTGCGC
TTTCTTTCACCTCCCCTTTTTTCTTCCACCCCTGCATGGCCAAGAACAGGGCTTGCCGCCACTCAG
GCGGTGCCACCGGTTGTGGCGGAGGTGGACGCGGGGAGGCTGGAGCCGAGAGTGGAGGAGAGAGAA
GGGTACTGGGTGTTGAAGGAGAAGTTCAGAGAAGGCATAAATCCTCAGGAGAAATTGAAGCTCGAG
AGAGAGCCTATGAAGCTTTTCATGGAAGGTGGGATAGAAGATTTGGCCAAGATGTCGCTCGAGGAA
ATTGACAAGGATAAGATTTCAAAGAGTGATATTGATGTAAGGCTCAAGTGGCTTGGTCTCTTCCAT
AGGAGAAAGCATCATTATGGTAGATTTATGATGAGACTGAAGCTACCTAATGGGGTAACAACAAGT
GCACAAACTCGATACTTAGCGAGTGTGATTAGGAAATATGGAAAGGACGGGTGCGCAGATGTGACC
ACCAGGCAAAATTGGCAAATTCGTGGTGTGGTACTGTCTGATGTGCCAGAAATACTTAAAGGTCTT
GATGAAGTTGGCTTGACAAGCCTGCAGAGTGGAATGGATAATGTGAGAAACCCTGTTGGGAACCCC
CTTGCAGGCATTGACATACATGAGATTGTTGCTACACGGCCTTACAACAACTTGTTATCACAATTT
ATCACTGCTAATTCGCGCGGTAATCTGGCCTTCACTAACTTGCCAAGGAAGTGGAATGTGTGTGTA
GTGGGTTCTCATGATCTCTTTGAGCATCCTCACATCAATGATCTTGCTTACATGCCTGCTATAAAG
GATGGAAGGTTTGGTTTCAATCTGCTGGTTGGTGGCTTCTTTAGTCCCAGGCGATGTGCAGAAGCA
GTCCCTCTCGATGCCTGGGTCTCAGCGGATGACATAATCCTCGTGTGCAAAGCCATACTGGAGGCT
TATAGGGATCTTGGCACCAGAGGGAACAGACAGAAAACAAGAATGATGTGGTTGATTGATGAACTT
GGAATAGAAGGATTCAGGTCTGAGGTAGTGAAAAGAATGCCCAACCAAGAGCTGGAGAGAGCTGCT
CCTGAAGATCTAATTGAGAAGCAATGGGAAAGGAGAGAGTTAATTGGTGTCCATCCACAGAAACAA
GAAGGCCTTAGTTACGTGGGTCTTCACATTCCGGTGGGTCGAGTCCAAGCAGATGACATGGATGAA
CTTGCTCGTTTAGCCGACACATATGGCTGTGGCGAACTTCGGCTCACTGTGGAGCAAAACATCATA
ATTCCCAACATTGAGAACTCAAAGCTCGAAGCCTTACTCGGAGAGCCTCTATTGAAAGACAGATTT
TCACCAGAACCGCCTATTCTCATGAAAGGGTTGGTGGCTTGCACTGGCAATCAGTTCTGTGGGCAA
GCCATTATAGAGACAAAGGCCAGGGCCTTGAAGGTGACTGAGGAAGTTCAACGGCAAGTGGCAGTG
ACTCGGCCGGTTAGGATGCACTGGACAGGCTGTCCAAATAGCTGTGGGCAGGTTCAAGTGGCTGAT
ATTGGTTTCATGGGGTGTATGGCAAGGGATGAGAATGGAAGCCTTGTGAAGGTGCTGCTGTTTTT
CTGGGAGGCAGAATTGGGAGCGACTCACATTTGGGAAATCTTTACAAAAAGGGTGTTCCTTGCAAG
AACTTGGTGCCATTGGTAGTGGACATTCTTGTTAAACATTTTGGAGCTGTACCAAGGGAGAGGGAA
GAGAGCGAGGATTGATTCAAACAGCAAGATTACTTCTTCTTTTACCATTTTGGATGACTCCCTGCA

FIGURE 2 (continued)

AAGCATTTGTTCTGGGAGAGGGAACGTGATGCATCAAAGAAATCCTTATGGGACTAAAATTTGTGA
GAGGGAGGCACATTTTAGTGCTATACCCAGCTTTTAACATGTTGGTTTTATAGGTTTGGTACGCTA
TAAGTACTCTGTTTGAATTAACTTATGTATTAAAACAGCTAAGAGTTGAATTGTAATATGAAAGTA
ATAAAATAGGAGGCTTTTGGTGCAAAAAAA

SEQ ID NO: 36, Capsicum annuum - TA5054
CCCACCTCACCCCACCTTACGACTACAAAAATGATCTTATTTCGCCATTTTAACCATGACCGCCAC
GATCATCACCACCCTCAATAATCAAGAATCAACTAAATTCCTCAATTCCAAATTTGGCGAAATGGC
ATCTTTTTCTGTTAAATTTTCAGCAACTTCTTCGCTGACAAGTTCTAAGAGATTTTCCAAGCTTCA
TGCCACTCCACCGCAGACAGTGGCAGTACCTCCATCTGGGGCAGTGGAGGTAGCTGCAGAGAGACT
AGAGCCTAGACTGGAGGAAAGAGATGGGTATTGGGTACTTAAGGAAAAGTTCAGAAAAGGCATAAA
TCCTGCTGAAAAGGCCAAGATTGAAAAGGAACCTATGAAATTGTTCACTGAAAATGGTATTGAAGA
TATTGCTAAGATCTCACTTGAAGAGATCGAAAAATCTAAGCTTGCTAAGGATGATATTGATGTTAG
GCTCAAGTGGCTTGGCCTCTTCCATAGGAGAAAGCATCAATATGGACGATTCATGATGCGACTGAA
GCTTCCAAATGGGATAACGACGAGTGCCCAAACTCGATATTTAGCAAGTGTGATTAGGAAATATGG
GAAAGATGGATGTGCAGATGTGACTACAAGGCAAAATTGGCAGATTCGTGGGGTTGTGCTACCTGA
TGTGCCTGAGATTCTAAAGGGACTGGATGAAGTTGGCTTGACCAGTCTGCAAAGTGGCATGGACAA
TGTTAGAAATCCCGTGGGGAACCCTCTGGCGGGGATTGATCCACAAGAAATTGTGGACACAAGGCC
TTACGCTAATTTGCTATCCAATTTGCTATCCCAATATGTCACTGCCAATTTTCGTGGCAATCTGTC
CGTGCATAACTTGCCAAGGAAGTGGAATGTATGTGTAATAGGGTCACACGATCTTTATGAGCATCC
CCATATCAATGATCTTGCCTATATGCCTGCAACGAAAGATGGACGATTTGGATTCAACCTGCTTGT
GGGTGGATTCTTCAGTCCGAAGCGATGTGCAGAGGCAATTCCTCTTGATGCATGGGTTCCAGCTGA
TGATGTAGTCCCTGTTTGCAAAACAATATTAGAAGCTTATAGAGATCTTGGTACCAGAGGGAACAG
GCAGAAAACAAGAATGATGTGGTTAATTGACGAACTGGGTGTTGAAGGATTCAGGGCAGAAGTTGT
GAAGAGAATGCCTCAAAAGAAGCTAGAGAGAGAATCCACAGAGGATTTGGTGCAGAAACAATGGGA
AAGGAGAGAGTATCTTGGGGTTAATCCACAGAAACAGGAAGGTTACAGCTTTGTTGGTCTTCACAT
TCCAGTGGGTCGTGTCCAAGCAGATGACATGGATGAGCTTGCTCGTTTAGCAGAAGAGTATGGTTC
AGGAGAGCTCCGGCTGACTGTTGAGCAAAACATCATTATTCCGAACATTGAGAACTCAAAGATTGA
TGCATTGCTCAATGAACCTCTTCTGAAACAGATTTCACCCGATCCACCTATTCTCATGAGAAATTT
GGTGGCTTGTACTGGTAACCAATTCTGTGGGCAAGCCATAATCGAGACTAAAGCACGTTCAATGAA
GATAACTGAGGAGGTTCAACGGCTAGTCTCTGTGACTCAGCCCGTGAGGATGCACTGGACTGGTTG
CCCAAATTCATGTGGACAAGTTCAAGTTGCAGATATCGGATTTATGGGATGCCTGACAAGAAAGGA
AGGAAAGACAGTGGAAGGCGCTGATGTTTTCTTGGGTGGCAGAATAGGGACTGACTCACACTTGGG
AGATATTTATAAGAAGTCTGTCCCCTGTGAAGATTTGGTACCAATAATTGTGGACTTACTAGTTAA
CAACTTTGGTGCTGTTCCAAGAGAGAGAGAAGAAGCAGAAGATTAATCTCAACATTTCAGAATCAG
CTCGTGGCTTTACTCAACATAGTAAATTGGACGTTGATGGAATGTGCTTACCATATTAAGATATTT
CCAAGGTACAGAACTGGTGGAGCTGTTGTTGGAAGTTAGTAGAATAATCAGAACATGAGCTGTTCT
TGACATGCTATGTGTGACATTCCACGATGCAAATACTTGTACTTGTTTCAGAATATTCACCCGGTG
TATTGTTTTGGAAAAGAGCTGATCCAAACTAAAAGGTTTTTGAATTGTGGGATTCCTAATAATAGA
TTTTTTAAAAATGTAATTTAATAATCATACATTTCAATTTTTACCTATTATTATATTCTTTGTT

SEQ ID NO: 37, Chlamydomonas reinhardtii - 59303
TTGCATCGTTATCTCCTTCGACCACCTTGAATTGCCTGCGGGCCCCTTGACCTCATCCGACGCAGC
CATGCTTCTGCACGCGCCGCATGTTAAGCCCCTGGGGCAGCGTAGTTCGATACGGCGTGGAAATTT
GGTGGTTGCGAACGTAGCGTGCACGGCGGGCAAGAACCCGACGTCGCGGCCAGCGAAACGCTCCAA
GGTGGAGTTCATCAAGGAGAACAGCGACCACCTGCGCCACCCGCTCATGGAAGAGCTGGTGAATGA
CGAGACATTCATCACCGAGGACTCGGTGCAGCTGATGAAATTTCACGGCTCCTACCAACAAGACAA
CCGTGAGAAACGCGCCTTCGGCCAAGGCAAAGCTTACTCATTCCTGATGCGGACTCGGCAGCCCGC

FIGURE 2 (continued)

TGGCGTTGTGCCCAACCGGCTCTACCTGGTGATGGACGACCTCGCCGACCAGTTCGGCAACGGCAC
GCTGCGCCTGACCACGCGCCAGGCCTACCAGCTGCACGGCGTGCTGAAGAAGGACCTCAAGACGGT
GTTCAGCTCCGTCATCAAGAACATGGGATCCACACTGGCCGCATGCGGCGACGTCAACCGCAACGT
GATGGGGCCCGCAGCGCCCTTCACCAACCGCCCCGACTACCTGGCCGCCCAGAAGGCGGCGCTGGA
CCTGGCGGATCTGCTAACGCCGCAGTCGGGCGCCTACTACGACGTGTGGCTGGACGGCGAGAAGTT
CATGAGCAGCTACAAGGAGGACCCCGCTGTGACCGAGGCCCGTGCCTTCAACGGCTTCGGAACCAA
TTTCGACAACAGCCCCGAGCCCATCTACGGCTCCCAGTACCTCCCCCGCAAGTTCAAGATCGCCAC
CACGGTGCCTGGTGACAACAGTGTGGACCTGTTCACTCAGGACCTGGGCGTGGTGGTTCAGGGCTA
CAACCTGTATGTGGGCGGTGGGCAGGGCCGCAGCCACAGAGACGCAGACACCTTCCCGCGCCTGGC
GGACCCGCTGGGCTACGTGGCCGCCGCCGACCTGTTCGCCGCGGCCAAGGCGGTGGTGGCGGTGTT
CCGCGACTACGGCCGCCGTGACAACCGCAAGCAGGCGCGAACACGGCACATGCTGGCGGAGTGGGG
CGTGGACAAGTTCCGCTCGGTGGCGGAGCAGTACCTGGGCAAGCGCTTCCAGGAGCCGGTGCCGCT
GCCGCCCTGGCAGTACAAGGACTACCTGGGCTGGGGCGAGCAGGGCGACGGGCGGCTGTACTGCGG
CGTGTATGTGCAGAACGGGCGCATCAAGGGCGAGGCCAAGCGGGCGCTGCGTGCGGCCATTGAGCG
CTACAGCCTGCCGGTGGTACTCACGCCGCACCAGAACCTGGTCCTGCGGGACGTGCGGCCCGAGGA
CCGGGAGGACATTGAGCAGCTGCTGCGGGCCGGCGGCGTCAAGGAGCTGGTGGAGTGGGACGGGCT
GGACCGGCTGTCCATGGCCTGCCCCGCGCTGCCGCTGTGCGGCCTGGCGGTCACGGAGGCGGAGCG
GGCGCTGCCGGACGTCAACACGCGCATCCGGGCCATGTTGACACGGGCGGGCCTGCCTCCCTCCCA
GCCGCTGCACGTGCGCATGACGGGCTGCCCCAACGGCTGCGTGCGGCCCTACATGGCCGAGTTGGG
GCTGGTGGGCGACGGACCCAACAGCTACCAGCTGTGGCTGGGCGGCGGGCCGGCGCAGACACGCCT
GGCGCAGCCGTACGCGGAGAGGGTCAAGGTGAAGGACTTGGAGTCCACGCTGGAGCCCCTGTTTGG
CGCCTGGAGGGCCGGGCGCCAGCCGGACGAGGCCTTTGGAGATTGGGTGGCGCGGCTCGGATTTGA
CGCCGTGCGGCAGCAGGCGGCGGCGGCGGCGGCGGCTCCTGTCGGCACCGCGTGAGGCGGCGG
CTCGGGGCTTTCCCGGTGCAAACGTACGTGCGTGCGTATGCGTGTTTACGTGTGTGTAAGTATGTA
TCTGTGTATGTGTACCGTATGTGTACGAGAAGCGAAAATGGTGGACGACGACTGCACAGTCGCAGC
ACCGGCGGCTTGTGGGGTAGGCTGTGGCTACCTCTCGCAATGCGGCCACGTAATGGTATTGCAAAA
TGCCCCTGCGTCAATGATAAGAGATTGCGTATTCATGCACGTGACTGAGGAGAAACGGTTCACAAC
GAAACCCTGCAGCCCGGCAATGCCATGTTCTAGATAGGTCACGCACGCAATCCGCATGCAGCGCGG
TCTTCGTATGTACTATGTAGCACTACCCTGTGCGCAGTGCACCATTTATATGCTTTGCTAGCAGCA
AGCGGTTTTGCTTGAGGTTCCTTTTGCCTGGATTCGCCTGCCAGCCCTCCGGGAGCTAGGGGTGCT
CTGTAGCGATCATGCAAAAGTAAGATGAGTTCTGTTTGGGTTGCGCGGAAGTGCTGAGGCGCTCTT
GTGCAATACGAGTACGG

SEQ ID NO: 38, Chlamydomonas reinhardtii - 192085

CTTGTAACTTGACAACCAAGGACAACCAAGGACCAGCCGCTTATAATCACTAGGGTTGCGCTCCAG
TCGGTGTCTTGTGAGCGTTGATTCCTCGCTGAAAGCTTTATCTTGAGCACCATACTAGTTGAGTCG
TGATTGCATTCGCAAGGGCAAAATAACCCGAGGCTTGTGACTACAATCAACAAACGGCAATGCAGT
CGCGCCAGTGCTTGAACCGCAAGGCCAGCGGCGCGCGGCCCTGCGCTAACTCGCGCAGCCTCACAG
CTCGCGTACTCGCTACGGCCGCGCCTGTCGCGCCGTCCGCCACACCCGCCTCCGCCCCCCTGCCCC
TCCCCGATGGCGTTGGCGAGCACAGCGGCCTGAAGCACCTGCCCGAGGCCGCCCGCACTCGTGCGC
TCGACAAGAAGGCCAACAAGTTTGAGAAGGTTAAGGTCGAGAAGTGCGGCTCGCGCGCCTGGAACG
ACGTGTTTGAGCTGTCTTCCCTGCTGAAGGAGGGCAAGACCAAGTGGGAGGACCTTAACCTCGATG
ATGTCGACATCCGTCTCAAGTGGGCCGGCCTGTTCCACCGCGGCAAGCGCACCCCCGGCAAGTTCA
TGATGCGTCTCAAGGTGCCCAACGGCGAGCTCACCGCCGCGCAGCTGCGCTTCCTGGCCTCCTCCA
TCGCGCCCTACGGCGCTGACGGCTGCGCCGACATCACCACCCGCGCCAACATCCAGCTGCGCGGCG
TCACCATGGAGGACTCGGAGACGGTCATCAAGGGGCTGTGGGATGTGGGCCTGACGTCCTTCCAGT
CGGGCATGGACTCCGTGCGCAACCTCACCGGCAACCCCATCGCCGGAGTCGACCCACACGAGCTGG
TGGACACGCGGCCGCTGCTGCGCGACATGGAGGCGATGCTGTTCAACAACGGCAAGGGCCGCGAGG

FIGURE 2 (continued)

AGTTTGCCAACCTGCCGCGCAAGCTGAACATCTGCATCTCCTCCACCCGCGACGACTTCCCGCACA
CCCACATCAACGACGTTGGCTACGAGGCCGTGGCCAAGCCCAACGGCGAGGTGGTGTACAATGTGG
TGGTGGGCGGCTACTTCTCCATCAAGCGCAACATCATGTCCATCCCGCTGGGCTGCTCCATCACCC
AGGACCAGCTGATGCCCTTCACTGAGGCCCTGCTGCGCGTGTTCCGGGATCACGGCCCGCGCGGCG
ACCGGCAGCAGACGCGGCTGATGTGGCTGGTGGAGGCGGTGGGCGTGGACAAGTTCCGCCAGCTGC
TGTCGGAGTACATGGGCGGCGCCACCTTCGGCGAGCCCGTGCACGTTCACCACGACCAGCCCTGGG
AGCGGCGCAACCTGCTGGGCGTGCACCGGCAGAGGCAGGCCGGCCTGAACTGGGTCGGCGCCTGCG
TGCCCGCGGGCCGCCTGCACGCCGCCGACTTTGAGGAGATCGCGGCTGTGGCTGAGAAGTACGGCG
ACGGCACGGTGCGCATCACGTGCGAGGAGAACGTGATCTTCACCAACGTGCCCGACGCCAAGCTGG
AGGCGATGAAGGCGGAGCCGCTGTTCCAGCGCTTCCCCATCTTCCCCGGCGTGCTGCTGTCGGGCA
TGGTGTCCTGCACCGGCAACCAGTTCTGCGGCTTCGGTCTGGCTGAGACCAAGGCGAAGGCCGTGA
AGGTGGTGGAGGCGCTGGACGCGCAGCTGGAGCTGAGCCGGCCCGTGCGCATCCACTTCACCGGCT
GCCCCAACTCATGCGGCCAGGCGCAGGTGGGCGACATCGGGCTGATGGGCGCGCCCGCCAAGCACG
AGGGCAAGGCCGTGGAGGGCTACAAGATCTTCCTGGGCGGCAAGATCGGCGAGAACCCCGCGCTCG
CCACCGAGTTCGCGCAGGGTGTGCCGGCCATTGAGAGCGTGCTGGTGCCTCGGCTAAAGGAGATTC
TGATCTCCGAGTTCGGTGCCAAGGAGCGCGCCACCGCCACCGCCTAAGAGCGTGGTGTCACGAGCG
TGGCGGCAGTGGAACGTGCTTGCAGCGTTGGTGTTTGGAGCGAGCTCCTCAGAGCGTGAGTGCCTT
GTTGAACACGCCGGCGTTGCGTGATGGGAAGGTGGGATTGGTGGTCGCCCTGAGGTGCATGAAGCA
TGCAGGGCAGGGGAGTGGGATTGGTTGGAGAGGAAAATGAGTAGGAGTGATGCGCACCTGCGGCTG
CCTATATAACATAAGGAAGTAAGCGTGATGGATGCACGGGCTGTGTTTTGCTTGAAGCGGCAGAGC
CCTGCAGGAGCCAGACGGCCGACATGTACTGCTAAGGCAGGAGCCAGTTCCTGCGTTGAGAAGAAG
CGTGCTTGCTTGCCGGCGGAGGCCGTCTTGCGTGCCATACCAGGGCACGGCAGCGCTGGAAGACTG
CATGCGACGCAGCGATCGGAGCACGCTGTGGTTCTTTACCCTCGTTTTACATATGCGTTGTCGTGT
TCCTTGTGTGTATGTACGTGTGTGTGTGTGTGTGTGTACGGTGTGTATACGGCGTGCGGGGCAGGC
AGGCGGAGGCTGCAAAGGGAGCGCAGATGCGCATCCTTAGGGAAAGGTACGTAGGAGCCGCCGCTG
CGTGTATGTATGTACTAGCAGCTAGATATGCACGTGGTGACCTGCAGCGCTGTGCTCAATGCGTGC
TGTGGCACCAGCGCAGGGGCAAGAAGCGTAGGCATTTCGGTAGTACGGTATTGTGTGCGCGTGCTG
GCGCTGGGAGGCGGTGCAGTGGTGCAGGTTTGTTGGCGCCGGCCGCTGCACCTGCTGCGCTTGCGA
CTAGGCAGGCGCCGTACGGTAATAGGGGTTGAGGCACATTGCGCATGCATTTGTCTAGATAATGGT
ATGCGGCGCCGACAAGTGGCAACTAGCGTTAGGGTGGCTTGTCTGTACTAAACCACGGCCCATACC
GCAGTGCGGCGTGTGGCTGCAACACCCGTGCCGGCGTGTAGGAGGAGCTGACGTGTGATCTAGAGT
GAATACCAATGGTACTGGAAGAGGTAACAGACTTTGCGACGAGCGTTGCAATGCGAGGCGCCCGCC
GGGGCAGGCGTGCACACAACCACCTAGATGGCTGCATCCCGGGCGAATGTAACAACACCGGAAGGA

SEQ ID NO: 39, Chlamydomonas reinhardtii - 192232
CTGTTTCGTCACGTCGTTATTGAATTCTATTAAGTGGTTTAACCGTAGGTAGCAGCCATGCTTCTC
AAGGGCATTACAACCCCGATGCTGGGGCAGCAGCGCCCCACTCGCGGCCAGCTGCACGTCGTGAAC
GTGGCTACGCCCTCCAAGAATCCCTCCTCTCGCCTGGCGAAGCGCAGCAAGGTGGAGATTATTAAG
GAGAAGAGCGACTACCTGCGGCACCCACTCATGGAGGAGCTGGTTAACGACGCCACCTTCATCACC
GAGGACTCGGTGCAGCTCATGAAGTTCCACGGCTCGTACCAGCAGGACCACCGCGAGAAGCGCGCG
TTTGGTCAGGGCAAGGCTTACTGCTTTATGATGCGCACGCGTCAGCCCGCTGGTGTCGTGCCCAAC
CGCCTGTACCTGGTGATGGACGACCTGGCCGATCAGTACGGCAACGGCACGCTGCGCCTGACTACG
CGCCAGGCCTACCAGCTGCACGGCGTGCTGAAGAAGGACCTCAAGACGGTGTTCAGCTCCGTCATC
AAGAACATGGGATCCACCCTGGCCGCCTGCGGCGACGTCAACCGCAACGTTATGGGCCCCTCCGCG
CCCTTCACCAACCGCCCCGACTACGTGGCCGCCCAGAAGGCCGCCAACGACATCGCCGACCTGCTG
ACGCCGCAGTCGGGCGCCTACTACGACGTGTGGCTGGACGGCGAGAAGTTCATGTCGGCTTACAAG
GAGGACCCCAAGGTGACCGCCGACCGTGCCTACAACGGCTTCGGCACCAACTTTGAGAACAGCCCC
GAGCCTATCTACGGCGCGCAGTTCCTGCCCCGCAAGTTCAAGGTGGCCACCACGGTGCCGGGCGAC

FIGURE 2 (continued)

AACAGCGTGGACCTGTTCACCCAGGACCTGGGCGTGGTGGTCATCATGGACGAGAGCGGCAAGGAG
GTCAAGGGCTACAACCTGACGGTGGGCGGCGGCATGGGCCGCACACACCGCGACGATGAGACCTTC
CCGCGTCTGGCTGACCCGCTGGGCTACGTGGACAAGGACGACCTGTTCCACGCCGTCAAGGCGGTT
GTTGCGGTTCAGCGCGACTACGGCCGCCGCGACAACCGCAAGCAGGCGCGCCTCAAGTACCTGGTG
GGCCTGCCCGCCGACCAGGAGCTGCACGTGCGCATGACGGGCTGCCCCAACGGCTGCGCGCGGCCC
TACATGGCCGAGCTGGGCTTCGTGGGCGACGGCCCCAACAGCTACCAGCTCTACTTCGGCGGCAAC
GTCAACCAGACGCGCCTGGCGCAGCTGTTCGCGGACAGGGTCAAGGTGAAGGACCTGGAGTCCACG
CTGGAGCCCATCTTCGCCGCCTGGAAGGCCAGCCGCCGGCCAAAGGAGTCGTTCGGCGACTGGGTG
TCGCGGCCGTCCCAAGATCCCAAGAATCTCAGTTCTGTACAACAGGGCACGCAGCACGAGAGCGCC
GTCGTCGCGCACTAA

SEQ ID NO: 40, Gossypium hirsutum - TA24262
TATCCCTTCACTTATCTTTCCACCACCACAATTCCACCAGTTCCAAGCTTCTTTTCAAACAACAAA
ACCCCACATGTCTTCCTTGTCGGTCCGTTTCTTTGCTCCACAACAGCCGTTACTGCCGTCCACAGC
TTCCTCTTTCAAGCCCAAAACATGGGTTATGGCAGCTCCCACGACGGCGCCGGCGACTTCGGTGGA
TGTCGACGGGGGGAGGTTGGAACCCCGAGTTGAAGAACGAGAGGGGTACTTCGTGTTGAAAGAGAA
GTTCAGAGATGGCATCAACCCTCAGGAGAAAATAAAGATCGAGAAAGACCCTTTGAAGCTTTTCAT
GGAAGCTGGGATTGATGAACTCGCTAAGATGTCGTTCGAGGATCTTGATAAAGCTAAGGCTACAAA
GGACGACATTGATGTTAGACTTAAATGGCTCGGCTTGTTCCATAGGAGAAAACATCAATATGGGAG
ATTTATGATGAGACTAAAACTACCAAATGGTGTAACAACAAGTGCACAAACACGGTACTTAGCCAG
TGTGATAAGGAAATACGGCAAAGAAGGGTGTGCCGATGTTACGACAAGGCAAAACTGGCAAATCCG
TGGAGCGGTGTTGCCTGATGTGCCTGAAATACTTAAGGGTCTCGACGAAGTAGGCTTGACGAGCCT
ACAGAGTGGCATGGACAATGTGAGGAACCCTGTCGGTAATCCTCTTGCCGGCATCGACCCCGAAGA
GATTGTCGATACTCGACCTTATACCAACTTGTTATCTCAGTTCATCACCGCCAATTCCCGCGGCAA
TCCGGCTGTTGCCAACTTGCCTAGGAAATGGAATGTCTGTGTCGTGGGGTCTCATGATCTTTACGA
ACATCCCCATATCAATGATCTCGCTTATATGCCGGCGACGAAAAACGGACGATTTGGGTTTAATTT
GCTGGTTGGTGGGTTCTTTAGTGCCAAGAGATGTGATGAGGCCATTCCTCTTGATGCTTGGGTCTC
AGCTGATGATGTGATTCCATTGTGCAAAGCTGTGTTAGAAGCCTATAGGGATCTTGGATACAGGGG
CAATAGGCAAAAGACTAGAATGATGTGGCTGATTGATGAACTGGGTATTGAAGTGTTCAGATCAGA
AGTAGCCAAAAGAATGCCTCAGAAAGAGTTGGAGAGAGCATCTGATGAAGATTTGGTTCAAAAGCA
ATGGGAAAGGAGAGACTACCTTGGTGTCCATCCGCAAAAGCAAGAAGGTTTCAGCTACATCGGCAT
TCACATCCCAGTCGGTCGAGTCCAAGCCGACGACATGGACGAACTAGCCCGGTTAGCCGACACGTA
TGGCTCGGGCGAATTCAGACTCACTGTGGAGCAAAACATCATAATCCCCAACGTTGAGAACTCGAA
ACTAGAAGCATTACTAAACGAGCCTCTATTGAAAGACCGGTTTTCACCCCAACCAAGTATTCTCAT
GAAAGGGCTAGTAGCTTGTACTGGTAACCAGTTTTGCGGACAAGCCATTATTGAAACAAAAGCTAG
AGCCTTGAAGGTGACGGAAGAGGTTGAAAGGCTAGTGTCGGTGAGCCGGCCGGTGAGGATGCATTG
GACCGGTTGCCCCAACACGTGTGGTCAAGTCCAAGTGGCGGATATAGGTTTCATGGGGTGCATGGC
AAGGGATGAGAATGGAAACCATGTGAAGGGGCAGACATATTCTTGGGAGGGAGAATTGGGAGTGA
CTCACATTTAGGAGAGCTTTATAAGAAGGGTGTCCCTTGTAAGAACTTGGTACCTGTAGTTGCTGA
CATTTTGGTGGAACCCTTTGGAGCTGTCCCTAGGCAAAGGGAAGAAGGGGAAGATTGATTCAAAAT
CAACTTCATTTCATTCCATTACTTTTATATTTGTTTTATTTTTTTTTTTTAATAACCAAGAAAAAT
GAAGGGTTTGAAAGATACTGGGGAGGATTAAATTTGGAGAATATTGATCAATGGCATGATGATGAA
GGGCTTTGTATTATAAAATATGTAACATTTTCAGCATATGTATTAGAATAAAGTTACTGGTAATAT
ATTTTCAGTTAAAATTTAGAGATGATCATGTTTG

FIGURE 2 (continued)

SEQ ID NO: 41, Hordeum vulgare - TA43088
ACCACCATCACCGCCACAGAGCAGCAGCAGCGGCACCACCACCACCGCAACCACAAGCAGCATCCA
TGGCGTCCTCGGCCTCCCTGCAGAGCTTCCTCCCGCCCTCGGCCCACGCGGCGACGTCGTCGTCCC
GGCTCCGGCCCAGCCGCGCCCGCCCCGTCCAGTGCGCTGCCGTCTCCGCGCCGTCGTCGTCGTCGT
CGTCCGCATCGCCGTCGGCCTCGGCCGTCCCGTCGGAGCGGCTGGAGCCGCGGGTGGAGCAGCGGG
AGGGCGGCTACTGGGTGCTCAAGGAGAAGTACCGCACCAGCCTGAACCCGCAGGAGAAGGTGAAGC
TGGGCAAGGAGCCCATGGCGCTCTTCACCGAGGGCGGCATCAACGACCTCGCCAAGCTGCCCATGG
AGCAGATCGACGCCGACAAGCTCACCAAGGAGGACGTCGACGTGCGCCTCAAGTGGCTCGGCCTCT
TCCACCGCCGCAAGCAGCAGTATGGGCGGTTCATGATGCGGCTGAAGCTGCCCAACGGCGTGACGA
CGAGCGAGCAGACGAGGTACCTGGCGAGCGTGATCGACAAGTACGGCGAGGAGGGGTGCGCCGACG
TGACGACCCGGCAGAACTGGCAGATCCGCGGCGTGACGCTGCCGGACGTGCCGGAGATCCTGGACG
GGCTCCGCTCCGTCGGCCTCACCAGCCTGCAGAGCGGCATGGACAACGTGCGCAACCCCGTCGGCA
GCCCGCTCGCCGGCATCGACCCCCTCGAGATCGTCGACACGCGCCCCTACACCAACCTCCTCTCCT
CCTACATCACCAACAACTCCGAGGGCAACCTCGCCATCACCAACCTTCCTAGGAAGTGGAACGTGT
GCGTGATCGGCACACATGATCTGTACGAGCACCCGCACATCAACGACCTGGCGTACATGCCGGCCG
AGAAGGACGGCAAGTTCGGGTTCAACCTGCTCGTGGGCGGGTTCATCAGCCCCAAGAGGTGGGGTG
AGGCCCTGCCGCTCGACGCCTGGGTCCCCGGCGACGACATCATCCCGGTCTGCAAGGCCGTCCTCG
AGGCGTTCCGCGACCTCGGCACCAGGGGCAACCGCCAGAAGACGCGCATGATGTGGCTCATCGACG
AGCTCGGGATGGAGGCGTTCCGGTCGGAGATCGAGAAGAGGATGCCCAACGGCGTGCTGGAGCGCG
CGGCGCCGGAGGACCTGATCGACAAGAAGTGGGAGAGGCGCGACTACCTCGGCGTGCACCCGCAGA
AGCAGGAGGGGCTCTCCTTCGTCGGCCTTCACGTGCCCGTCGGCCGGCTGCAGGCCGCGGACATGT
TCGAGCTGGCCCGCCTCGCCGACGAGTACGGCTCCGGCGAGCTCCGCCTCACGGTGGAGCAGAACA
TCGTGCTGCCCAACGTGAAGAACGAGAAGGTGGAGGCGCTGCTGGCGGAGCCGCTGCTGCACAAGT
TCTCGGCGCACCCGTCGCTGCTGATGAAGG

SEQ ID NO: 42, Lotus japonicus - TA2640
TCACCATGTCTTCTTCCTTCTCCATTCGCTTCCTCGCTCCTCCATTTCCCTCCACCTCTCGCCCCA
AGTCATGTCTCTCCGCCGCCACGCCGGCTGTGGCTCCAACCGATGCGGCGGTGTCGAGGTTGGAGC
CCAGAGTGGAGGAGAGAAATGGGTACTGGGTTTTGAAGGAAGAGCACAGGGGTGGCATTAATCCGC
AGGAAAAGGTGAAGCTGGAGAAAGAGCCTATGGCCCTTTTTATGGAAGGTGGGATTGATGAGTTGG
CTAAGGTTTCTATTGAAGAGCTTGATAGCTCTAAGCTTACTAAGGATGATGTTGATGTTAGGCTCA
AATGGCTTGGTCTTTTTCATAGGAGAAAGCATCAGTATGGTAGATTTATGATGAGGCTGAAACTTC
CAAATGGGGTGACAACGAGTGCGCAGACACGATACTTGGCGAGTGTGATCAGGAAGTACGGGAAAG
ATGGGTGTGCTGATGTGACCACAAGGCATAATTGGCAAATTCGTGGTGTAGTGCTACCTGATGTTC
CTGAAATTCTTAAGGGCCTTGCAGAGGTTGGCTTGACTAGTCTGCAGAGTGGTATGGACAATGTAA
GAAACCCTGTGGGTAACCCTCTTGCAGGCATTGACCCTGATGAGATTGTTGATACCCGACCTTACA
CGAACTTGTTGTCCCATTTCATCACTGCCAATTCACGTGGCAACCCAACCGTCTCAAACTTGCCAA
GGAAGTGGAATGTATGCGTTGTGGGTTCTCATGATCTCTTTGAGCATCCCCACATAAATGATCTTG
CTTACATGCCTGCTAACAAAGATGGTCGTTTTGGATTCAACTTATTGGTGGGGGGTTTCTTTAGTC
CCAAGCGATGTGCAGAGGCAATTCCACTTGATGCATGGGTCTCTGCAGAAGATGTAATCCCAGTTT
GTAAAGCAATCCTCGAGATGTACAGGGATCTTGGCACCAGAGGAAACAGACAGAAAACAAGAATGA
TGTGGTTGATTGACGAACTGGGGATAGAAGTATTCAGGTCAGAGGTGGTAAAAAGAATGCCATTAG
GGCAGCAGCTGGAGAGAGCATCCCAGGAAGATCTGGTTCAGAAACAATGGGAAAGAAGAGATTACT
TTGGTGCCAATCCACAGAAACAAGAGGGCTTAAGCTATGTTGGGATTCACATTCCAGTTGGTAGGA
TCCAAGCAGATGAGATGGACGAGCTGGCCCGTCTGGCCGATGAATACGGCACTGGTGAACTGAGGC
TCACTGTAGAGCAAAACATAATAATCCCAAATGTGGAAAACTCAAAACTCAGTGCCCTGCTCAATG
AGCCTCTCTTGAAAGAAAAGTTCTCACCTGAACCTTCCCTTCTAATGAAAACACTGGTGGCATGCA
CTGGTAGCCAATTTTGTGGGCAAGCCATAATTGAGACAAAGGCGAGGGCATTGAAGGTGACTGAAG

FIGURE 2 (continued)

AAGTGGAGAGACTAGTGGCAGTGACTAGGCCTGTGAGAATGCACTGGACTGGGTGTCCCAACACCT
GCGGGCAAGTGCAGGTTGCTGATATTGGTTTCATGGGGTGCATGGCCAGAGATGAGAATGGTAAGC
CTGGTGAAGGTGTGGATATTTTCCTGGGAGGGAGGATAGGAAGTGATTCACACTTAGCTGAGGTTT
ATAAGAAGGCTGTTCCTTGCAAGGACTTGGTGCCCATAGTGGCAGACATACTAGTAAAACATTTTG
GAGCTGTCCAGAGGAATAGAGAAGAAGGAGATGATTAAGTTATTTAGGTTTAACTTTTGAAATTAA
ACCTTCTGTTGTATCTATGACAAAATATCATTTTCTTGTCCAAAATTTATAATAGTAGTAAGGGTG
ATCAAGTGAGATATACCACATGTGCCAATGGGGAAAAAAAGTCGGATATGAAAGTTGTAATCTTAC
ATGAGTGGTTTTGAAATTACATGACACATTTTTATTGATCGGACGGAAAAGAAGATCCAAACAAAT
GTGTAAGAAATTTTTCTTAGTTTCTAATTTCCACTTTCTATTCATAAATAAATGTGTAAGCTATGG
TTCTTACTTTGTGACATTTGTTAAAATAAATATTTTCACTTTTTTT

SEQ ID NO: 43, Nicotiana tabacum - TA16376

ATGGCATCTTTTTCTGTTAAATTCTCAGCAACTTCATTGCCAAATCCTAACAGATTTTCCAGGACT
GCTAAGCTTCATGCAACACCGCCGCAGACGGTGGCAGTACCACCATCTGGGGAGGCGGAGATAGCT
TCCGAGAGGCTAGAGCCTAGAGTAGAGGAAAAAGATGGGTATTGGGTACTCAAGGAAAAATTCAGA
CAAGGGATAAATCCAGCTGAAAAGGCCAAGATTGAGAAAGAACCAATGAAATTATTTATGGAAAAT
GGTATTGAAGATCTTGCTAAGATCTCACTTGAAGAGATCGAAGGGTCTAAGCTTACTAAAGATGAT
ATTGATGTTAGGCTCAAGTGGCTTGGCCTTTTCCATAGGAGAAAGCATCATTATGGCCGATTCATG
ATGCGATTGAAGCTTCCAAATGGGGTAACAACGAGTGCCCAAACTCGATACTTAGCCAGTGTGATA
AGGAAATATGGAAAAGATGGATGTGGTGATGTGACTACAAGGCAAAATTGGCAGATTCGCGGGGTT
GTACTACCTGATGTACCCGAGATTCTAAAGGGACTGGATGAAGTTGGCTTGACCAGTCTGCAAAGT
GGCATGGACAACGTTCGAAATCCGGTGGGAAATCCTCTGGCGGGGATTGATCCACATGAAATTGTA
GACACAAGGCCTTACACTAATTTGCTCTCCCAATATGTTACTGCCAATTTTCGTGGCAATCCGGCT
GTTACTAACTTGCCAAGGAAGTGGAATGTATGTGTAATAGGGTCACATGATCTTTATGAGCATCCC
CATATCAATGATCTTGCCTATATGCCGGCATCAAAAGATGGACGATTTGGATTCAACCTGCTTGTG
GGTGGATTCTTCAGTCCGAAGCGATGTGCAGAGGCAGTTCCTCTAGATGCATGGGTTCCAGCTGAT
GACGTGGTCCCTGTTTGCAAAGCAATATTAGAAGCTTATAGAGATCTTGGTACCAGAGGGAACAGG
CAAAAAACAAGAATGATGTGGTTAGTTGATGAACTGGGCGTTGAAGGATTCAGGGCAGAGGTCGTA
AAGAGAATGCCTCAACAAAAGCTAGATAGAGAATCAACAGAGGACTTGGTTCAAAAACAATGGGAA
AGGAGAGAATACCTTGGCGTGCATCCGCAGAAACAAGAAGGATACAGCTTTGTTGGCCTTCACATT
CCGGTAGGTCGTGTCCAAGCAGATGACATGGACGAGCTAGCTCGTTTAGCGGATAACTATGGTTCA
GGAGAGCTCCGGTTGACTGTTGAACAGAACATCATTATTCCCAACGTTGAGAACTCAAAGATCGAG
TCATTGCTCAATGAGCCTCTCTTAAAGAACAGATTTTCGACCAATCCACCTATTCTCATGAAAAAT
CTGGTGGCTTGTACTGGTAACCAATTTTGCGGGCAAGCCATAATTGAGACTAAAGCGCGTTCCATG
AAGATAACTGAGGAGGTACAACGACTAGTTTCTGTGACAAAGCCGGTGAGGATGCATTGGACTGGT
TGCCCGAATTCATGTGGACAAGTTCAAGTCGCGGATATTGGATTTATGGGATGCTTGACAAGAAAA
GAAGGAAAAACTGTAGAAGGTGCTGATGTTTATTTGGGAGGCAGAATAGGGAGTGACTCACATTTG
GGAGATGTTTATAAGAAATCAGTACCTTGTGAGGATTTGGTGCCAATAATTGTGGACTTACTAGTT
AACAACTTTGGTGCTGTTCCAAGAGAAAGAGAAGAAGCAGAAGATTAATTTCAAGATTTCATAACA
GCTCGCGGATCGCGCTGCAGAATTGGACATTAATGGAATGTGCACACCATATCAAGTTATTTCGAA
GGTACAGAAATGGTGACACTGATCCTGAAAACCAAGGTTTTCTTTATTGAAAGTTAGTTGAATAAT
TGGTATATGTGCCGTTATTAACATGCTCATGTGTGATATAGCACGACAGAAATATTTGTACTTGTT
TCAGAATAATTATATTGTGTATTCTTTTGGAAAAACTGATACAAACCAAAAGGCTTTTAAACCACC
CTTCAGTTGGGATTCTAATAATCCATCTTTACATACCAATTAATCATGTTGTTGTATTCTTAATCA
TATTGTTATATTATAATAATCCATTCGGTTTGATGCC

FIGURE 2 (continued)

SEQ ID NO: 44, Nicotiana tabacum - TA13596
ATGGCATCTTTTTCTATTAAATTTCTGGCACCTTCATTGCCAAATCCAGCTAGATTTTCCAAGAAT
GCTGTCAAGCTCCACGCAACACCGCCGTCTGTGGCAGCGCCGCCAACTGGTGCTCCAGAGGTTGCT
GCTGAGAGGCTAGAACCCAGAGTTGAGGAAAAAGATGGTTATTGGATACTCAAAGAGCAGTTTAGA
AAAGGCATAAATCCTCAAGAAAAGGTCAAGATTGAGAAGCAACCTATGAAGTTGTTCATGGAAAAT
GGTATTGAAGAGCTTGCTAAGATACCCATTGAAGAGATAGATCAGTCCAAGCTTACTAAGGATGAT
ATTGATGTTAGGCTTAAGTGGCTTGGCCTCTTCCATAGGAGAAAGAACCAATATGGGCGGTTCATG
ATGAGATTGAAGCTTCCAAATGGAGTAACAACGAGTGCACAGACTCGATACTTAGCGAGTGTGATA
AGGAAATACGGGAAGGAAGGATGTGCTGATATTACGACAAGGCAAAATTGGCAGATTCGTGGAGTT
GTACTGCCTGATGTGCCGGAGATACTAAAGGGACTAGCAGAAGTTGGGTTGACCAGTTTGCAGAGT
GGCATGGACAATGTCAGGAATCCAGTAGGAAATCCTCTGGCTGGAATTGATCCAGAAGAAATAGTA
GACACAAGGCCTTACACTAATTTGCTCTCCCAATTTATCACTGGCAATTCACGAGGCAATCCCGCA
GTTTCTAACTTGCCAAGGAAGTGGAATCCGTGTGTAGTAGGCTCTCATGATCTTTATGAGCATCCC
CATATCAACGATCTCGCGTACATGCCTGCCACGAAAGACGGGCGATTTGGATTCAACCTGCTTGTG
GGAGGGTTCTTCAGTGCAAAAAGATGTGATGAGGCAATTCCTCTTGATGCATGGGTTCCAGCCGAT
GATGTTGTTCCGGTTTGCAAAGCAATACTGGAAGCTTTTAGAGATCTTGGTTTCAGAGGGAACAGA
CAGAAATGTAGAATGATGTGGTTAATCGATGAACTGGGTGTAGAAGGATTCAGGGCAGAGGTCGAG
AAGAGAATGCCACAGCAACAACTAGAGAGAGCATCTCCAGAGGACTTGGTTCAGAAACAATGGGAA
AGAAGAGATTATCTTGGTGTACATCCACAAAAACAAGAAGGCTACAGCTTTATTGGTCTTCACATT
CCAGTGGGTCGTGTTCAAGCAGACGATATGGATGAGCTAGCTCGTTTAGCTGATGAGTATGGTTCA
GGAGAGATCCGGCTTACTGTGGAACAAAACATTATTATTCCCAACATTGAGAACTCAAAGATTGAG
GCACTGCTCAAAGAGCCTGTTCTGAGCACATTTTCACCTGATCCACCTATTCTCATGAAAGGTTTA
GTGGCTTGTACTGGTAACCAGTTTTGTGGACAAGCCATAATCGAGACTAAAGCTCGTTCCCTGATG
ATAACTGAAGAGGTTCAACGGCAAGTTTCTTTGACACGGCCAGTGAGGATGCACTGGACAGGCTGC
CCGAATACGTGTGCACAAGTTCAAGTTGCGGACATTGGATTCATGGGATGCCTGACTAGAGATAAG
AATGGAAAGACTGTGGAAGGCGCCGATGTTTTCTTAGGAGGCAGAATAGGGAGTGATTCACATTTG
GGAGAAGTATATAAGAAGGCTGTTCCTTGTGATGATTTGGTACCACTTGTTGTGGACTTACTAGTT
AACAACTTTGGTGCAGTTCCACGAGAAAGAGAAGAAACAGAAGACTAATAAAATTTAGAATAGTTG
GTGATTTTGCTGTGTTCATAACATGTAATGTATGATAAATCAATGCAAACATTTCTACCTACGTGA
GAATTATTACATGCTACATATATTCTTTTGAAGAAAATTACATGCGTACTCCTC

SEQ ID NO: 45, Nicotiana tabacum - AB093534
ATGGCATCTTTTTCTGTTAAATTCTCAGCTACTTCATTACCAAATCATAAAAGATTTTCAAAGCTA
CATGCAACACCGCCGCAGACGGTGGCTGTAGCCCCATCTGGGGCGGCGGAGATAGCATCGGAGAGG
TTAGAGCCTAGAGTAGAAGAAAAAGATGGGTATTGGGTACTTAAGGAAAAATTCAGACAAGGGATA
AATCCAGCTGAAAAAGCTAAGATTGAGAAGGAACCAATGAAATTGTTTATGGAAAATGGTATTGAA
GATCTAGCTAAGATCTCACTTGAAGAGATCGAAGGGTCTAAGCTTACTAAAGATGATATTGATGTT
AGGCTCAAGTGGCTTGGCCTTTTCCATAGGAGAAAGCATCACTATGGCCGATTCATGATGAGATTG
AAGCTTCCAAATGGGGTAACAACGAGTTCCCAAACTCGATACTTAGCCAGTGTGATAAGGAAATAT
GGGAAAGATGGATGTGCTGATGTGACGACAAGGCAAAATTGGCAGATTCGTGGGGTTGTACTACCT
GATGTACCCGAGATTCTAAAGGGACTGGATGAAGTTGGCTTAACCAGTCTGCAGAGTGGCATGGAC
AATGTTAGAAATCCGGTGGGAAATCCTCTGGCGGGGATTGATCCACATGAAATTGTAGACACAAGG
CCTTACACTAATTTGCTCTCCCAATATGTTACTGCCAATTTTCGTGGCAATCCGGCTGTGACTAAC
TTGCCAAGGAAGTGGAATGTATGTGTAATAGGGTCACACGATCTTTATGAGCATCCCCAGATCAAC
GATCTTGCCTATATGCCGGCAACAAAAGATGGACGATTTGGATTCAACCTGCTTGTGGGTGGATTC
TTCAGTCCGAAGCGATGTGCAGAGGCAGTTCCTCTTGATGCATGGGTTCCAGCTGATGACGTAGTC
CCTGTTTGCAAAGCAATATTAGAAGCTTATAGAGATCTTGGCACCAGAGGGAACAGGCAGAAAACA
AGAATGATGTGGTTAGTTGATGAACTGGGCGTTGAAGGATTCAGGGCAGAGGTTGTAAAGAGAATG

FIGURE 2 (continued)

CCTCAACAAAAGCTAGATAGAGAATCAACAGAGGACTTGGTTCAAAAACAATGGGAAAGGAGAGAA
TACCTTGGCGTGCATCCACAGAAACAAGAAGGGTACAGCTTTGTTGGTCTTCACATTCCAGTGGGT
CGTGTCCAAGCAGATGACATGGACGAGCTAGCTCGTTTGGCCGATGAGTATGGTTCCGGAGAGCTC
CGGCTGACTGTTGAACAAAACATCATTATTCCCAATGTTAAGAACTCAAAGATCGAGGCATTGCTC
AATGAACCTCTCTTAAAGAACAGATTTTCAACCGATCCACCTATTCTCATGAAAAATTTGGTCGCT
TGTACTGGTAACCAATTTTGCGGGAAAGCCATAATTGAGACTAAGGCACGATCCATGAAAATAACT
GAGGAGGTTCAACTACTAGTTTCTATAACGCAGCCTGTGAGGATGCATTGGACTGGTTGCCCGAAT
TCATGTGCACAAGTTCAGGTCGCGGATATTGGATTTATGGGATGCTTGACAAGAAAAGAAGGAAAA
ACTGTAGAAGGTGCTGATGTTTATTTGGGAGGCAGAATAGGGAGTGACTCACATTTGGGAGATGTT
TATAAGAAATCAGTACCTTGTGAGGATTTGGTGCCAATAATTGTGGACTTACTAGTTGACAACTTT
GGTGCTGTTCCAAGAGAAAGAGAAGAAGCAGAAGATTAA

SEQ ID NO: 46, Oryza sativa - Os01g0357100

GAACCTTATCTCCTTCTCTCTCGTCGCTTTCTGCGTCTCCCCGTCTCTCCTTCGCCAACAGCCGAG
AAGAGGCAGAGAGAGCGCCGCCCCCCGTCCCTCTCTCTCCCTCTCGTCCTCGCCCCCATCCCTCTC
GTCTTTCCCTTGCCGGCAGCAGAGGAGGCGGCAGCGACGGCTTCAGCTGCTCCCACGGGCCGGATC
GGGCAGTGGCGGTGGCGTCGGCGGCTTCCGCTGGCGAATCCGGCGGGTGGATACAAATCAGTGTTC
CGATAGGTAAAACCCTGCTCTCAGCATCTGCCCTTTTGAATTCGCCAAGAGCCAGCATCTGCCCTT
TTGAATTCGCCAAGGGCCAGCATCTGCCCATTTGATTTTGAATTCGCCAAGAGCCAGCAACAGCGC
CCCCGCGCCCCCTCCCTCCTCCGCAATAAACAGCCACACGCGCCGCCCCCATGTCCACCCTCATCG
CCACAGCGCACCACCACCACCACCACCACCACCACCACCACCGTCTCCAGCCATGGCCTCCTCCGC
CTCCCTGCAGCGCTTCCTCCCCCCGTACCCCCACGCGGCAGCATCCCGCTGCCGCCCTCCCGGCGT
CCGCGCCCGCCCCGTGCAGTCGTCGACGGTGTCCGCACCGTCCTCCTCGACTCCGGCGGCGGACGA
GGCCGTGTCGGCGGAGCGGCTGGAGCCGCGGGTGGAGCAGCGGGAGGGCCGGTACTGGGTGCTCAA
GGAGAAGTACCGGACGGGGCTGAACCCGCAGGAGAAGGTGAAGCTGGGGAAGGAGCCCATGTCATT
GTTCATGGAGGGCGGCATCAAGGAGCTCGCCAAGATGCCCATGGAGGAGATCGAGGCCGACAAGCT
CTCCAAGGAGGACATCGACGTGCGGCTCAAGTGGCTCGGCCTCTTCCACCGCCGCAAGCATCAGTA
TGGGCGGTTCATGATGCGGCTGAAGCTGCCAAACGGTGTGACGACGAGCGAGCAGACGAGGTACCT
GGCGAGCGTGATCGAGGCGTACGGCAAGGAGGGCTGCGCCGACGTGACAACCCGCCGGCAGATCCG
CGGCGTCACGCTCCCCGACGTGCCGGCCATCCTCGACGGGCTCAACGCCGTCGGCCTCACCAGCCT
CCAGAGCGGCATGGACAACGTCCGCAACCCCGTCGGCAACCCGCTCGCCGGCATCGACCCCGACGA
GATCGTCGACACGCGATCCTACACCAACCTCCTCTCCTCCTACATCACCAGCAACTTCCAGGGCAA
CCCCACCATCACCAACCTGCCGAGGAAGTGGAACGTGTGCGTGATCGGGTCGCACGATCTGTACGA
GCACCCACACATCAACGACCTCGCGTACATGCCGGCGGTGAAGGGCGGCAAGTTCGGGTTCAACCT
CCTCGTCGGCGGGTTCATAAGCCCCAAGAGGTGGGAGGAGGCGCTGCCGCTCGACGCCTGGGTCCC
CGGCGACGACATCATCCCGGTGTGCAAGGCCGTTCTCGAGGCGTACCGCGACCTCGGCACCAGGGG
CAACCGCCAGAAGACCCGCATGATGTGGCTCATCGACGAACTTGGAATGGAGGCTTTTCGGTCGGA
GGTGGAGAAGAGGATGCCGAACGGCGTGCTGGAGCGCGCGGCGCCGGAGGACCTCATCGACAAGAA
ATGGCAGAGGAGGGACTACCTCGGCGTGCACCCGCAGAAGCAGGAAGGGATGTCCTACGTCGGCCT
GCACGTGCCCGTCGGCCGGGTGCAGGCGGCGGACATGTTCGAGCTCGCACGCCTCGCCGACGAGTA
CGGCTCCGGCGAGCTCCGCCTCACCGTGGAGCAGAACATCGTGATCCCGAACGTCAAGAACGAGAA
GGTGGAGGCGCTGCTCTCCGAGCCGCTGCTTCAGAAGTTCTCCCCGCAGCCGTCGCTGCTGCTCAA
GGGCCTCGTCGCGTGCACCGGCAACCAGTTCTGCGGCCAGGCCATCATCGAGACGAAGCAGCGGGC
GCTGCTGGTGACGTCGCAGGTGGAGAAGCTCGTGTCGGTGCCCCGGGCGGTGCGGATGCACTGGAC
CGGCTGCCCCAACAGCTGCGGCCAGGTGCAGGTCGCCGACATCGGCTTCATGGGCTGCCTCACCAA
GGACAGCGCCGGCAAGATCGTTGAGGCGGCCGACATCTTCGTCGGCGGCCGCGTCGGCAGCGACTC
GCACCTCGCCGGCGCGTACAAGAAGTCCGTGCCGTGCGACGAGCTGGCGCCGATCGTCGCCGACAT

FIGURE 2 (continued)

CCTGGTCGAGCGGTTCGGGGCCGTGCGGAGGGAGAGGGAGGAGGACGAGGAGTAGGAACACAGACT
GGGGTGTTTTGCTTGCTCCGGTGATCTCTCGCCGTCCTTGTAAAGTAGACGACAATATGCCTTCGC
CCATGGCACGCTTGTACTGTCACGTTTTGGTTTGATCTTGTAGCCCAAAAGTTGTGTTCATTCTCG
TTACAGTCTTACAGAGGATGATTGATTGATAAATAAAGAAGAAACAGATTCTGC

SEQ ID NO: 47, Physcomitrella patens - 193361

ATTAGAGAGTTGATGGACATCGTTTGATCGTTAACTGCAGCGAAATAAGTCCATGGGGTTTTTAGG
AAGTGGAGTGATACATCGTCGCATAGTTACTGGGAAAATTGTAATTGCTCGTGCTCAGGCTGGAAT
TTCAAGCAAGTTGAGGATTGCAGGCGAAATTTACTGAAGTAAAATTCGCCAGGCGCAATGCAAGGT
GCAATGCAGACAAAGATGTGGAGGGGAGAGCTGATCAGCACATCGACCCACTTTATAGGCGGCACT
CGACTGCAGCCCAAACTAAACCAGGATGCAAGGAAACCCACGAAAAGTGAAAATTGTATCGTTCGA
GTCTCCATGGAGCGTGAGGTCAAGGCTAAGGCCGCGGTTTCTCCACCCGCTGTTGCTGCAGACCGT
CTCACTCCACGAGTGCAAGAAAGAGATGGCTACTACGTTCTCAAAGAGGAATTCCGACAAGGAATT
AACCCCCAAGAGAAGATCAAACTTGGGAAAGAGCCGATGAAATTCTTCATAGAGAACGAGATAGAG
GAGCTTGCAAAGACGCCGTTCGCGGAGCTAGACAGCTCGAAGCCTGGGAAGGACGATATCGATGTT
AGACTCAAGTGGTTGGGTCTCTTCCACCGCCGCAAACATCAATATGGAAGGTTCATGATGCGGTTC
AAGCTTCCGAATGGAATCACGAACAGTACACAGACGAGGTTTTTGGCCGAGACCATCTCAAAATAC
GGAAAGGAAGGGTGTGCAGATTTGACGACAAGACAGAACTGGCAAATTCGTGGGATTATGCTCGAA
GATGTGCCCTCCCTTCTGAAAGGACTGGAATCCGTGGGCCTATCGTCTCTGCAGAGCGGGATGGAC
AATGTAAGAAATGCGGTCGGTAACCCTCTTGCTGGAATCGACCCCGACGAAATCGTCGACACCATT
CCTATCTGTCAGGCGCTGAACGACTACATCATCAACAGAGGGAAAGGAAATACTGAGATCACCAAC
TTACCTCGGAAGTGGAACGTGTGCGTGGTCGGGACGCACGACTTATTTGAACATCCGCACATCAAC
GATCTTGCGTACGTTCCCGCAACCAAGAACGGCGTCTTCGGTTTCAACATTCTTGTTGGAGGATTC
TTCAGCTCAAAGCGGTGCGCCGAAGCTATTCCGATGGACGCTTGGGTGCCGACAGACGACGTCGTC
CCGTTGTGCAAAGCAATTCTGGAGACTTATCGAGACCTCGGGACTCGCGGCAACCGACAGAAGACT
CGCATGATGTGGTTGATCGATGAGATGGGAGTCGAGGAGTTCAGAGCCGAGGTGGAAAGGCGCATG
CCCAGCGGCACTATCCGGCGAGCCGGACAGGATCTGATAGACCCGTCGTGGAAGCGCCGGAGCTTC
TTCGGAGTAAACCCCCAGAAGCAAGCAGGGCTGAACTACGTTGGTCTTCACGTCCCGGTCGGGCGT
TTGCACGCTCCAGAGATGTTCGAGCTGGCTCGCATTGCCGATGAGTACGGCAACGGCGAGATCCGG
ATCACTGTGGAGCAGAACCTGATTCTGCCCAACATCCCGACGGAGAAAATTGACAAGTTGATGCAG
GAGCCCCTCTTGCAGAAATACTCTCCGAATCCCACCCCCTTGTTGGCGAACTTGGTGGCCTGCACT
GGCAGCCAGTTCTGCGGCCAAGCGATCGCGGAGACGAAGGCCCTGTCCCTGCAACTCACGCAGCAG
CTCGAAGACACCATGGAAACGACTCGCCCGATCCGATTGCACTTCACGGGATGCCCCAACACATGC
GCTCAAATCCAGGTTGCGGATATCGGATTCATGGGCACCATGGCTCGAGATGAAAACCGAAAGCCC
GTTGAAGGGTTCGACATCTACCTCGGAGGCCGCATCGGCTCCGACTCTCACTTGGGAGAGCTTGTC
GTGCCTGGTGTGCCTGCCACCAAGCTGCTTCCGGTGGTGCAAGAGCTGATGATCCAGCATTTCGGC
GCTAAAAGGAAACCTTGAGATGCAAATCTGGGTATAGTAACAAAAAATCACTACTCGTCACACACA
CACACACACCGCTGATGTATAATTTACGTAAAACCAATCTATCGAATAGCACGATTCACAGTTACG
AAACTCTGGGTAAAACCCGGTTATAAATTGATGACCATTCATTCGTCTTGTGCAGCCTTCCAGTGA
CATTGTCAGTGTCGGTGGGCATGAGCTCTGTCGCTAATCCCCACTTCTCCAATAAAGTTTCGGCAA
ATCTGTGCCCACATGAATCAT

SEQ ID NO: 48, Physcomitrella patens – 144369

ATGCAAGGCACTATGCAGTCACAAATGTGGAGGGGACAGGTGAGCGGCGCATCGCTCCACTTCACA
GGCGCAACCCGAGTGCAGGGTAACAGCCACCAGGATTTAGTATATCCCACGCAATTTCACAAACAT
GGCGTTCGGGCCTCTGCGGAGCGCGAGGTCAAGGCCAAGGCTGTAGCTGCCCCACCTACCATCGCT
GCAGACCGCCTCGTGCCACGCGTGGAAGAACGAGATGGTTATTACGTTCTTAAGGAGGAATTTCGA
CAGGGCATCAACCCGTCGGAGAAGATAAAAATCGCCAAAGAACCCATGAAATTCTTCATGGAGAAC

FIGURE 2 (continued)

GAGATAGAAGAGCTGGCGAAAACGCCGTTCGCCGAGCTCGATAGTTCGAAGGCAGGAAAGGACGAC
ATTGATGTGAGATTGAAGTGGTTGGGCCTCTTCCACCGTCGCAAACATCAATATGGGAGATTCATG
ATGCGGTTCAAGCTTCCAAATGGGATCACGAATAGCTCGCAGACGCGGTTCTTGGCTGAGACAATC
TCCAAGTACGGAGAGTATGGGTGCGCTGATTTGACGACACGTCAAAACTGGCAAATCAGGGGGATT
GTTCTCGAAGACGTGCCTGCTCTTCTGAAGGGATTGGAATCAGTAGGCCTGTCATCTTTGCAGAGC
GGCATGGACAACGTTAGGAACCCAGTTGGTAACCCTCTTGCAGGAATCGACCCTGACGAAATTGTC
GACACTGCCCCGTTCTGCAAGGTACTCAGCGATTACATCATCAACCGAGGGCAAGGAAATCCTCAG
ATCACCAATTTACCTCGGAAATGGAACGTGTGCGTGGTTGGAACACATGACTTGTTCGAGCACCCG
CACATCAACGACCTGGCGTACATGCCAGCCACAAAGAACGGTGTCTTCGGTTTCAACATCCTGGTG
GGAGGATTCTTTAGCCCTAAGCGGTGTGCGGAAGCAATTCCCATGGATGCTTGGGTGCCAGCAGAT
GATGTCGTTCCCTTGTGCAAGGCAATTCTGGAAACCTACCGAGACCTTGGAACCCGAGGCAACCGA
CAGAAGACCCGCATGATGTGGTTGATCGACGAGATGGGAATTGAGGAATTCAGAGCCGAGGTAGAG
AGGCGCATGCCCGGTGGGTCCATTCTTAGAGCCGGGAAGGACCTGGTCGATCCATCCTGGACGCGC
CGGAGCTTCTATGGAGTGAACCCGCAGAAGCAACCGGGCTTAAACTACGTAGGCCTCCACATTCCC
GTCGGCCGGCTGCATGCTCCAGAGATGTTCGAGCTTGCGCGCATTGCAGACGAGTACGGCAACGGG
GAGATTCGGATCTCGGTGGAGCAGAACCTGATCCTGCCCAACGTCCCCACGGAGAAAATCGAGAAG
CTATTGAAGGAGCCCCTCCTGGAGAAATACTCCCCGAATCCCACCCCTCTGCTCGCCAACTTGGTG
GCCTGCACAGGCAGCCAGTTCTGTGGCCAGGCCATCGCGGAGACCAAGGCCCGGTCGTTGCAGCTC
ACGCAAGAGCTGGAAGCCACCATGGAAACCACTCGTCCTATTCGGTTGCACTTCACCGGATGCCCC
AACACATGCGCCCAAATCCAGGTTGCGGATATTGGCTTCATGGGTACAATGGCACGAGACGAAAAT
AGAAAGCCCGTGGAGGGGTTTGACATCTACCTTGGAGGTCGTATCGGCTCCGACTCACATTTGGGA
GAGCTCGTGGTGCCGGGCGTGCCTGCGACCAAGCTGCTCCCCGTTGTGCAAGACCTCATGATCCAG
CATTTCGGCGCCAAGCGTAAGACTTAA

SEQ ID NO: 49, Pinus taeda - TA3139
CGGCCGGGGGAGACAAGCCCTCATCATAGATTTAATTACTGATCTTTGCATCTTGGATTTGTAATC
GGAGTAGTCAGGATGAATCTCTCTAGTCCAGTCAGATTCGATGAGATTCGTCCCTTGGCCCATGTC
GTTTACAATCCTGTTTGCTGTGGGCATAAGCCGAATCGGCTCAGGTTGATGACAGCAATCCAGGTT
CGTGCTGTTAATCATGGTGGACGCAATTCTGAGATCAGTACAGATGGGAATAGCAAAGGGACAACA
GCCAAGGCTGTAGCCAGTCCTGCTGGCTCTCATGTGGCTGTAGATGCCTCAAGGCTGGAGGCTAGA
GTTGAGGAGAGGGATGGATACTGGGTTCTCAAAGAGGAATTCAGGGCTGGAATCAACCCTCAGGAG
AAGATTAAGTTGCAGAGGGAGCCCATGAAATTGTTCATGGAGAATGAGATCGAAGAACTTGCAAAG
AAGCCCTTCGCTGAAATTGAGAGTGAGAAGGTTAATAAAGATGATATAGATGTACGCCTGAAGTGG
TTGGGTCTCTTTCACCGAAGAAAACATCACTATGGGAGATTCATGATGAGACTTAAGCTTCCGAAT
GGAGTGACTACCAGTCTCCAAACTCGATATTTGGCAAGCGTGATTCAACAATATGGACCAGAGGGA
TGCGCAGATATAACAACTCGGCAGAATTGGCAGATTCGTGGAGTTGTGCTGGATGACGTGCCTGCC
ATATTGAAAGGGCTGAAGGAGGTTGGACTGTCTAGCTTGCAGAGTGGAATGGACAACGTTAGAAAC
CCTGTGGGAAATCCTTTAGCAGGGATTGATGCTGATGAAATCATTGACACAAGGCCATATACAAAG
GTTCTGACTGACTACATTGTCAACAATGGAAAGGGCAATCCATCCATAACCAACCTGCCACGTAAA
TGGAATGTCTGTGTTGTGGGTACACATGACTTGTTTGAGCATCCCCACATCAATGACCTCGCCTAC
ATTCCTGCAATGAATAGTGGGAGATTTGGTTTCAATCTGCTCGTTGGTGGATTCTTTAGTCCAAAA
CGCTGTGAAGAAGCAGTTCCACTTGATGCTTGGGTTGCTGGAGAGGATGTTGTACCAGTATGCAGA
GCCATTTTGGAGGTTTATAGAGATCTGGGCACCCGGGGAAATCGCCAGAAAACTCGAATGATGTGG
CTGATTGATGAGTTGGGCATAGAGGGCTTCCGTTCAGAAGTGGTGAAGAGAATGCCAGGAGAGAAG
TTGGAAAGAGCAGCAACAGAAGACATGTTAGATAAATCATGGGAGCGCAGGAGTTATCTTGGTGTG
CACCCACAGAAGCAGGAAGGCTTGAATTTCGTAGGTCTCCATGTTCCAGTGGGTCGACTTCAGGCA
GAAGATATGTTAGAACTGGCTCGTCTTGCAGAACAATATGGCACGCAGGAACTCCGCCTCACAGTA
GAACAAAATGCCATCATTCCAAACGTACCTACAGATAAGATAGAGGCACTTTTACAGGAACCCCTC

FIGURE 2 (continued)

CTCCAAAAAATTCTCCCCTTCCCCTCCTCTTCTTGTTAGCACATTAGTGGCTTGTACCGGCAACCAG
TTCTGTGGTCAGGCAATCATCGAAACAAAAGCAAGAGCCTTGAAAATCACAGAGGAATTGGATAGA
ACCATGGAAGTTCCCAAGCCTGTGAGAATGCACTGGACAGGATGCCCTAATACATGTGGACAAGTG
CAGGTTGCAGACATTGGCTTCATGGGTTGCATGACTAGGGATGAAAACAAGAAAGTTGTTGAGGGA
GTGGACATATTCATTGGAGGTAGGGTGGGAGCAGATTCACATCTAGGGGATTTAATCCACAAGGGA
GTACCTTGCAAGGACGTGGTACCTGTGGTTCAAGAACTACTTATTAAACACTTTGGAGCCATCAGG
AAAACAGACATGTGAAAATGAATTCCAATTTCTCATCCATCGCCATCTTCAGTGGAGGACAATCAC
CAGATTGCTAAGGTTCTGAGCGGGTATCCAACTCATTGAAATCTGAATAAATAAATGTAGAGATGC
AATGTATAGATGTATTGTTTACGAAGTCCAACGTGTTCAGAAATAAAATAGCTGATTACTGTGTTC
ACAGCAGGGTTTTTTTACATTAAACTCGTCTTGCACTTTTGAACAGTATGGAATACAAATAAAAAC
GGATTAGCCCAAAAAAAATAATGGAATAATAGAAATTCCAGTAAGATTATGATAAAATCTGTAGAAT
TTTTGAAAATCTGAGTTTCACTGGTG

SEQ ID NO: 50, Populus trichocarpa - 130.69

ACACTTCTCTAGAAACTATCTACCATCATTATGTCATCACTTTCAGTTCGTTTTCTCACGCCACAA
TTGTCACCCACAGTTCCAAGCTCCTCTGCAAGACCAAGAACAAGACTCTTTGCTGGACCTCCCACA
GTGGCTCAGCCAGCGGAGACGGGGGTGGATGCAGGGAGGTTGGAACCTAGAGTGGAGAAGAAAGAC
GGATACTATGTGTTGAAAGAGAAGTTTAGGCAAGGTATTAATCCTCAAGAGAAAGTGAAGATAGAG
AAAGAGCCAATGAAGCTTTTCATGGAAAATGGGATCGAGGAGCTTGCTAAATTGTCGATGGAAGAG
ATTGACAAAGAGAAGAGCACTAAAGATGATATTGATGTTAGACTCAAGTGGCTCGGTCTCTTTCAC
AGAAGGAAGCACCAATATGGTAGATTTATGATGAGACTAAAGCTACCAAATGGGGTAACAACAAGT
GCACAAACAAGATACTTGGCAAGCGTGATCAGGAAATATGGGAAAGATGGCTGTGCAGATGTAACA
ACAAGACAAAACTGGCAAATTCGTGGAGTGGTGTTGCCTGATGTGCCAGAAATACTAAGGGGTCTA
GCTGAAGTTGGTCTGACAAGCCTGCAGAGTGGCATGGACAACGTGAGAAACCCCGTCGGAAATCCG
CTTGCAGGAATTGATCCGGATGAGATTGTTGATACCAGACCTTATACCAACTTGTTGTCCCAATTT
ATCACTGCCAATTCTCGTGGAAATCCTGAGTTCACTAACTTGCCAAGGAAGTGGAATGTATGTGTC
GTGGGTTCTCATGATCTTTATGAGCATCCTCATATCAATGATCTTGCTTACATGCCTGCCATGAAG
GACGGGCGGTTTGGATTCAATTTGCTGGTTGGTGGGTTCTTTAGTCCCAAGCGATGTGCTGAGGCA
ATTCCTCTTGATGCTTGGGTTTCAGCTGATGATGTGCTCCCATCTTGCAAAGCAGTGTTAGAGGCC
TACAGAGATCTTGGCACCAGAGGGAACAGGCAAAAGACTAGAATGATGTGGCTGATCGACGAGCTT
GGCATTGAAGGATTCAGGTCAGAAGTAGTAAAAAGAATGCCACGTCAAGAGCTAGAGAGAGAATCT
TCTGAAGATTTGGTTCAAAAGCAATGGGAAAGGAGGGACTATTTCGGTGTCCATCCACAGAAGCAA
GAAGGCCTTAGCTATGCAGGTCTTCACATTCCTGTCGGTCGCGTCCAAGCAGATGACATGGATGAG
CTAGCTCGTTTAGCTGATATTTATGGCACTGGCGAACTCAGACTCACTGTGGAGCAGAACATCATA
ATTCCCAACATTGAGGACTCAAAGATTGAAGCCCTACTTAAAGAACCTCTATTAAAAGACAGGTTC
TCACCTGAGCCACCTCTTCTCATGCAAGGGTTGGTAGCATGCACTGGCAAAGAGTTTTGCGGGCAA
GCAATAATTGAAACAAAGGCTAGGGCCATGAAGGTAACTGAGGAGGTGCAGAGGTTAGTGTCGGTG
TCTAAACCAGTGAGAATGCACTGGACAGGCTGTCCTAATACCTGTGGGCAGGTACAAGTTGCCGAT
ATTGGGTTCATGGGTTGCATGGCAAGAGATGAAAATGGGAAAATCTGTGAAGGAGCAGATGTGTAC
GTAGGAGGAAGAGTTGGGAGTGACTCACATTTGGGAGAGCTTTATAAGAAAAGTGTTCCATGCAAG
GACTTGGTGCCTTTGGTTGTGGACATTTTAGTTAAACAATTCGGAGCTGTACCTAGGGAGAGGGAA
GAGGTGGATGATTAGTTCATTTAATCAAAATGTTCATTCTTGTTTCATTGCAAATTCGGAGGGGAT
CTAATGCATGCTTTTGGAATCGGAAATGA

SEQ ID NO: 51, Solanum lycopersicum - TA37687

CAACAATCAAGAGTCCACTAAACGTTTTGCCACACATCCATTTACTCCCACAGCTCTACAAAATGC
TCTGACATCTCTTTTGCAACTTCCAAAATGGCATCTTTTTCTATCAAATTTTTGGCACCTTCATTG
CCAAATCCAACTAGATTTTCCAAGAGTAGTATTGTCAAGCTCAATGCAACTCCGCCGCAGACAGTG

FIGURE 2 (continued)

GCTGCGGCGGGGCCTCCAGAGGTTGCTGCTGAGAGACTAGAACCAAGAGTTGAGGAAAAAGATGGA
TATTGGATACTAAAAGAGCAGTTTAGGCAAGGAATTAATCCTCAAGAGAAGGTGAAGATTGAGAAG
GAACCTATGAAGTTGTTCATGGAAAATGGTATTGAGGAGTTAGCTAAGATTCCAATTGAAGAGATA
GATCAATCAAAGCTTACTAAGGATGACATTGATGTTAGGCTCAAGTGGCTTGGCCTCTTCCATAGG
AGAAAGAATCAATATGGGAGATTCATGATGAGGTTGAAACTTCCAAATGGAGTAACAACAAGTGCT
CAGACTCGATATTTGGCGAGTGTGATAAGGAAATATGGAGAGGAAGGATGTGCTGATATTACGACA
AGGCAAAATTGGCAGATTCGTGGAGTAGTGCTGCCTGATGTGCCTGAGATTCTAAAGGGACTTGAA
GAAGTTGGCTTGACTAGTTTGCAGAGTGGCATGGATAATGTCAGGAATCCAGTTGGAAATCCTCTG
GCTGGAATTGATCCTGAAGAAATAGTTGACACAAGACCTTACACTAATTTGCTCTCCCAATTTATC
ACTGGTAATTCACGAGGCAATCCGGCTGTTTCTAACTTGCCAAGGAAGTGGAATCCGTGTGTAGTA
GGGTCTCATGATCTTTATGAGCACCCTCATATCAATGATCTTGCATACATGCCTGCCATAAAAGAT
GGACGATTTGGATTCAACCTGCTTGTGGGAGGGTTCTTCAGTGCCAAAAGATGTGATGAGGCAATT
CCTCTTGATGCATGGGTTCCAGCCGATGATGTTGTTCCGGTTTGCAAGGCAATACTGGAAGCTTTT
AGAGACCTTGGGTTCAGAGGGAACAGGCAGAAGTGTAGAATGATGTGGTTGATCGATGAACTGGGT
GTAGAAGGATTCAGGGCAGAGGTCGTAAAGAGAATGCCTCAGCAAGAGCTAGAGAGAGCATCTCCG
GAAGACTTGGTTCAGAAACAATGGGAAAGAAGAGATTATCTTGGTGTACATCCACAGAAACAGGAA
GGCTATAGCTTTATTGGTCTTCACATTCCAGTGGGTCGTGTACAAGCAGACGACATGGATGATCTA
GCTCGTTTGGCTGATGAGTACGGCTCAGGAGAGCTACGGCTGACTGTGGAACAGAACATTATTATT
CCCAACATTGAGAACTCAAAGATTGACGCACTGCTAAAAGAGCCTATTTTGAGCAAATTTTCACCT
GATCCACCTATTCTCATGAAAGGTTTAGTGGCTTGTACTGGTAACCAGTTTTGTGGACAAGCCATT
ATTGAAACGAAAGCTCGTTCCCTGAAGATCACCGAAGAGGTTCAAAGGCAAGTATCTCTAACGAGG
CCAGTAAGGATGCACTGGACAGGCTGCCCAAATACGTGTGCACAAGTTCAAGTTGCAGACATTGGA
TTCATGGGATGCCTGACTAGAGATAAAGACAAGAAGACTGTGGAAGGCGCCGATGTTTTCTTAGGA
GGCAGAATAGGGAGTGACTCACATTTGGGTGAAGTATACAAGAAGGCAGTTCCTTGTGATGAATTA
GTACCACTTATTGTGGACTTACTTATTAAGAACTTTGGTGCAGTTCCACGAGAAAGAGAAGAAACA
GAAGATTAATAAAATTTGGATTAGATCATAATGATGGAATGTGCAATTATGTTTAGTGATTATGGA
GGTATATAGCTAAGAGCTGGTTTGAATAATCAGAAATATGTTGTGTTCATATCATTTATTGTACGA
TAAATCAACACAAACATTCC

SEQ ID NO: 52, Solanum lycopersicum - TA37689

GACGATCACCGCTACCTCAATCGACTAAATTCTCAATTTTAAGTTGGTTTTGTAACTTAGTTGTTC
TTTTTAATTTGTCGAAATGACTTCTTTTTCTGTTAAATTTTCAGCTACTTCACTTCCAAATTCTAA
TAGATTTTCCAAACTTCATGCTACTCCACCGCAGACGGTGGCGGTACCGTCGTACGGGGCGGCGGA
GATAGCTGCTGAAAGACTAGAGCCTAGAGTTGAGCAAAGAGATGGGTATTGGGTAGTTAAGGATAA
GTTCAGACAAGGCATAAATCCAGCTGAAAAGGCGAAGATTGAAAAGGAACCAATGAAACTATTCAC
TGAAAATGGTATCGAAGATCTTGCTAAGATCTCGCTTGAAGAGATCGAGAAATCAAAGCTAACTAA
AGAAGATATTGATATTCGCCTCAAGTGGCTTGGACTCTTCCATCGGAGAAAACACCACTATGGTCG
ATTCATGATGCGATTGAAGCTTCCAAATGGAGTAACGACGAGTGATCAAACTCGATATTTAGGTAG
TGTGATTAGGAAATATGGGAAAGATGGATGTGGTGATGTGACTACAAGGCAAAATTGGCAGATTCG
TGGGGTTGTGTTACCTGATGTGCCTGAGATTCTAAAGGGGCTTGATGAAGTTGGCTTGACTAGTCT
GCAGAGTGGCATGGATAATGTTCGAAATCCGGTGGGGAATCCTCTCGCAGGGATTGATCTTCATGA
AATTGTAGACACAAGGCCTTACACTAATTTGCTGTCCCAATATGTCACCGCCAATTTTCGTGGCAA
TGTGGATGTGACTAACTTGCCAAGGAAGTGGAATGTATGTGTAATAGGGTCACATGATCTTTATGA
GCATCCGCATATCAATGATCTTGCGTATATGCCTGCAACCAAAGATGGACGATTTGGATTCAACCT
GCTTGTGGGTGGATTCTTCAGTCCGAAGCGATGTGCAGAGGCAATTCCTCTTGATGCATGGGTTCC
AGCTGATGATGTAGTCCCTGTTTGCAAAGCTATATTAGAAGCTTATAGAGATCTTGGTACCCGAGG
GAACAGGCAGAAAACAAGAATGATGTGGTTAATTGACGAACTGGGTGTTGAAGGATTCAGGGCAGA
AGTTGTGAAGAGAATGCCCCAAAAGAAGCTAGATAGAGAATCTTCAGAGGATTTGGTCCTGAAACA

FIGURE 2 (continued)

ATGGGAAAGGAGAGAGTACCTTGGCGTGCATCCGCAGAAACAGGAAGGATACAGCTTTGTTGGTCT
TCACATTCCGGTTGGTCGTGTCCAAGCAGATGACATGGACGAGCTAGCTCGTTTGGCTGATGAGTA
TGGTTCAGGAGAACTCCGGTTGACTGTTGAACAGAACATCATTATTCCCAACATCGAGAACTCAAA
GATCGATGCATTACTCAATGAGCCTCTCCTAAAGAACAGATTTTCACCTGATCCACCTATTCTCAT
GAGAAATTTGGTGGCTTGTACTGGTAACCAATTCTGTGGGCAAGCAATAATCGAGACTAAAGCACG
TTCAATGAAGATAACCGAGGAGGTTCAACGTCTAGTCTCTGTGACACAGCCAGTGAGGATGCACTG
GACAGGTTGCCCAAATACATGTGGACAAGTTCAAGTTGCCGATATCGGATTCATGGGATGCCTGAC
TAGAAAGGAAGGCAAAACTGTTGAAGGTGCTGATGTTTTCTTGGGTGGCAGAATAGGGAGCGACTC
GCATTTAGGAGAAGTTTATAAGAAGTCTGTACCATGTGAGGATTTGGTACCAATAATCGTCGACTT
ACTAATTAACAACTTTGGTGCTGTTCCAAGAGAAAGAGAAGAAACAGAGGAGTAATCTAAAATCTT
CAGAATGTACTTTTTATGATATTGAAATATTTCCAAGGTACAGCATTGTAAGTTAGTAAAATAATC
ACAACATGAGATGTTGTTAACATGTTCATGTGTGACATAGCATGATGCAAATACTTGAACTTGTTT
CAAAATATAATCACATTGTGTATTCTTTTGGAAATACTCATCCAAACTAAAAGGCTTTTGAATTGT
TGAATTCCTAATAATACATTTTTTAAAATGTAATTTGATATTCATTTGTTTTGATTATTATATTCT
TAAAATAATTTACTTATTCTCTC

SEQ ID NO: 53, Arabidopsis thaliana - AT2G15620

AAGAGCTCATCTCTTCCCTCTACAAAAATGGCCGCACGTCTCCAACCTTCTCCCAACTCCTTCTTC
CGCCATCATCATGACTTCTTTCTCTCTCACTTTCACATCTCCTCTCCTCCCTTCCTCCTCCACCAA
ACCCAAAAGATCCGTCCTTGTCGCCGCCGCTCAGACCACAGCTCCGGCCGAATCCACCGCCTCTGT
TGACGCAGATCGTCTCGAGCCAAGAGTTGAGTTGAAAGATGGTTTTTTTATTCTCAAGGAGAAGTT
TCGAAAAGGGATCAATCCTCAGGAGAAGGTTAAGATCGAGAGAGAGCCCATGAAGTTGTTTATGGA
GAATGGTATTGAAGAGCTTGCTAAGAAATCTATGGAAGAGCTTGATAGTGAAAAGTCTTCTAAAGA
TGATATTGATGTTAGACTCAAGTGGCTTGGTCTCTTTCACCGTAGAAAGCATCAGTATGGGAAGTT
TATGATGAGGTTGAAGTTACCAAATGGTGTGACTACAAGTGCACAGACTCGGTATTTAGCGAGTGT
GATTAGGAAGTATGGTGAAGATGGGTGTGCTGATGTGACTACTAGACAGAATTGGCAGATCCGTGG
TGTTGTGTTGCCTGATGTGCCTGAGATCTTGAAAGGTCTTGCTTCTGTTGGTTTAACGAGTCTTCA
AAGTGGTATGGATAACGTGAGGAACCCGGTTGGGAATCCTATAGCTGGGATTGATCCGGAGGAGAT
TGTTGACACGAGGCCTTACACGAATCTCCTTTCGCAGTTTATCACCGCTAATTCACAAGGAAACCC
CGATTTCACCAACTTGCCAAGAAAGTGGAATGTGTGTGTGGTGGGGACTCATGATCTCTATGAGCA
TCCACATATCAATGATTTGGCCTACATGCCTGCTAATAAAGATGGACGGTTTGGATTCAATTTGCT
TGTGGGAGGATTCTTTAGTCCCAAAAGATGTGAAGAAGCGATTCCTCTTGATGCTTGGGTCCCTGC
TGATGACGTTCTTCCACTCTGCAAAGCTGTTCTAGAGGCTTACAGAGATCTTGGAACTCGAGGAAA
CCGACAGAAGACAAGAATGATGTGGCTTATCGACGAACTTGGTGTTGAAGGATTTAGAACTGAGGT
AGAGAAGAGAATGCCAAATGGGAAACTCGAGAGAGGATCTTCAGAGGATCTTGTGAACAAACAGTG
GGAGAGGAGAGACTATTTCGGAGTCAACCCTCAGAAACAAGAAGGTCTTAGCTTCGTGGGGCTTCA
CGTTCCGGTTGGTAGGCTACAAGCTGATGACATGGATGAGCTTGCTCGGTTAGCTGATACCTACGG
GTCAGGTGAGCTAAGACTCACAGTAGAGCAAAACATCATCATCCCAAATGTAGAAACCTCGAAAAC
CGAAGCTTTGCTTCAAGAGCCGTTTCTCAAGAACCGTTTCTCCCCTGAACCATCTATCCTAATGAA
AGGCTTAGTTGCTTGTACCGGTAGCCAGTTCTGCGGACAAGCGATAATCGAGACTAAGCTAAGAGC
TTTAAAAGTGACAGAAGAAGTAGAGAGACTTGTATCTGTGCCAAGACCGATAAGGATGCATTGGAC
AGGATGTCCCAACACTTGCGGACAAGTCCAAGTAGCAGATATCGGATTCATGGGATGCTTAACACG
AGGCGAGGAAGGAAAGCCAGTCGAGGGTGCTGACGTGTACGTCGGGGGACGAATAGGAAGTGACTC
GCATATCGGAGAGATCTATAAGAAAGGTGTTCGTGTCACGGAGTTGGTTCCATTGGTGGCTGAGAT
TCTGATCAAAGAATTTGGTGCTGTGCCTAGAGAAAGAGAAGAGAATGAAGATTGATTCAAAAGCTA
TTGGATTCTTAATAAGTCAAGAGACCTATGAATGGTTCTCTCTCTGGTTTCAGACTTTGATACTTG
ATACTTGTATTTGTATTGTGCCCATAATTTTGGGTTTTGTAGCTCTCTCCTTTGTTGTAACCTGTA
ACTTTGTCCTTGGTTGTTTTGTAATATCTTGTTTTTAGTAATAGTAGTATAATCTGATTTTTTGT
CATATATTGTCTTGATTTCTCTGTGATATTTATAAGAAATAAACATTTGTTTCTTTTTACCTCC

FIGURE 2 (continued)

SEQ ID NO: 54, Vitis vinifera - GSVIVT00036600001
ATGGCTTCTATCTCTGTTCCTTTCCTCTCTCAGGCACCCACCCACCTTTCAAACTCCACTTCTCTC
CGTCTCAAAACCAGGATCTCTGCCACCCCGACTCCGACTCCAACTCCAACCACGGTTGCACCGTCG
TCCACGGCGGCGGTGGACGCCTCCAGGATGGAGCCCAGGGTGGAGGAGAGAGGGGGTTACTGGGTT
TTGAAGGAGAAGTTCAGGGAAGGTATAAATCCACAGGAGAAGGTGAAGATTGAGAAGGATCCTATG
AAGCTCTTCATAGAAGATGGGTTCAATGAGCTGGCCAGCATGTCTTTTGAAGAAATTGAAAAGTCT
AAGCATACTAAGGATGATATTGATGTGAGGCTCAAGTGGCTTGGACTGTTTCATAGGAGGAAGCAT
CAATATGGTAGATTTATGATGAGATTGAAGCTGCCAAATGGGGTGACATCAAGTGCACAAACTCGT
TACCTGGCCAGTGCAATAAGGCAATACGGGAAGGAGGGATGTGCCGATGTGACTACGCGGCAAAAC
TGGCAAATTCGAGGTGTGGTACTGCCTGATGTGCCTGAAATACTAAAGGGTCTTTCAGAGGTTGGT
TTGACGAGCCTGCAGAGTGGCATGGACAATGTGAGGAATCCTGTTGGAAATCCTCTTGCAGGCATT
GACCCTCATGAGATTGTTGATACACGACCTTACACCAACTTGTTATCCCAATTCATTACTGCCAAT
GCTCGTGGGAATACAGCCTTCACTAACTTGCCGAGGAAGTGGAATGTGTGTGTTGTAGGCTCCCAT
GATCTCTATGAGCATCCCCACATCAATGATCTGGCGTACATGCCTGCCACAAAGAAAGGAAGATTT
GGATTCAATCTGCTAGTAGGCGGGTTCTTTAGTCCCAAACGTTGTGCTGATGCTATTCCTCTCGAT
GCCTGGATCCCTGCCGACGATGTCCTCCCAGTTTGTCAAGCAGTACTAGAGGCTTACAGGGATCTT
GGTACCAGAGGAAACCGCCAAAAGACAAGAATGATGTGGTTAATTGATGAGCTGGGCATAGAGCAG
TTCCGGGCAGAGGTGGTGAAAAGAATGCCCCAACAAGAGCTGGAAAGATCATCTTCTGAAGACCTG
GTTCAGAAGCAATGGGAGAGGAGAGATTACCTTGGTGTCCATCCCCAGAAACAGGAAGGCTTTAGC
TTTGTGGGTATTCACATTCCAGTGGGTCGAGTCCAGGCAGATGACATGGACGAGCTAGCTCGATTG
GCAGACGAATATGGCTCAGGCGAGCTCCGGCTCACTGTAGAGCAGAACATCATAATTCCCAATGTG
GAGAACTCAAGACTTGAAGCCTTGCTCAAAGAGCCTCTCTTGAGAGACAGATTCTCTCCGGAGCCT
CCTATTCTCATGAAAGGCTTGGTGGCCTGCACCGGCAATCAGTTTTGTGGACAGGCCATTATCGAG
ACCAAGGCCAGAGCATTGAAGGTGACGGAGGATGTGGGGCGGCTGGTTTCAGTGACCCAGCCAGTG
AGGATGCACTGGACCGGCTGCCCAAACTCCTGCGGCCAGGTGCAAGTGGCGGATATCGGATTCATG
GGGTGCATGACAAGGGACGAGAATGGGAACGTTTGTGAAGGGGCAGATGTATTCTTAGGAGGTAGA
ATTGGGAGCGACTGTCATTTGGGAGAGGTTTATAAGAAGCGTGTTCCTTGCAAAGACTTAGTGCCC
TTGGTTGCTGAAATTTTGGTAAATCACTTTGGAGGAGTCCCCAGGGAGAGGGAAGAAGAAGCTGAA
GACTGA

SEQ ID NO: 55, Volvox sp. 83067
ATGCAGTCGCAGTCGCTGTCCCGCCGCACCTGCACCCGTACTCTTGGCCGCGGCCTCGTCACCCCT
GTCCTGGCAACCGCGGCACCGGCTTCAGCAGCGCAAGCGGCCGATGGCATCAACGCGCATAGCGGG
CTGAAGCACCTGCCAGAGGCTGCTCGCGTTCGCGCTCTCGACCGCAAGGCCAATAAGTTTGAGAAG
GTCAAGGTTGAGAAGTGCGGATCACGCGCATGGACAGATGTCTTCGAGCTGTCACGGCTGCTGAAA
GAGGGAAACACCAAGTGGGAGGATTTGGATTTGGACGACATAGACATCCGCATGAAGTGGGCGGGC
CTGTTCCATCGCGGAAAGCGCACGCCCGGCAAGTTCATGATGCGCCTCAAGGTTCCCAACGGCGAG
CTGGATGCCCGCCAGCTGCGCTTCCTCGCCTCGGCAATCGCGCCATACGGCGCCGACGGCTGCGCC
GACATCACCACGCGCGCCAACATCCAGCTCCGAGGCGTGACGCTGGCGGACGCCGACGCCATCATT
CGCGGTCTTTGGGACGTTGGCCTCACGTCCTTCCAGAGCGGTATGGACAGCGTACGGAACTTGACG
GGCAACCCCATCGCGGGTGTGGACCCCCATGAGCTCATAGATACCCGTCCGCTGCTGCGGGAAATG
GAGGCCATGCTGTTCAACAACGGCAAGGGCCGCGAGGAGTTTGCGAACCTGCCTCGCAAGCTCAAC
ATCTGCATTTCCTCAACCCGCGACGACTTCCCGCACACGCACATCAACGACGTGGGCTTCGAAGCG
GTGCGCCGCCCCGATGATGGCGAGGTGGTGTTCAATGTGGTCGTTGGCGGCTTCTTCTCCATCAAG
CGCAACGTTATGTCCATCCCTCTTGGCTGCTCTGTCACTCAAGACCAGCTGATGCCCTTCACGGAG
GCTCTGCTGCGGGTGTTCCGCGACCACGGGCCCCGCGGGGACCGCCAGCAGACTCGCCTGATGTGG
ATGGTAGATGCGATTGGCGTGGAGAAGTTCCGCCAGCTGCTTTCGGAGTACATGGGCGGCGCGGAG
CTGGCGCCGCCGGTGCACGTGCATCACGAGGGGCCCTGGGAGCGCCGTGACGTGCTGGGTGTGCAC

FIGURE 2 (continued)

CCCCAGAAGCAGCCGGGGCTGAATTGGGTGGGCGCCTGTGTTCCGGCTGGCAGGCTGCAGGCTGCC
GACTTTGACGAGTTCGCCCGCATCGCGGAGACGTACGGCGACGGCACCGTACGGATCACGTGCGAG
GAGAACGTGATCTTTACCAACGTCCCCGACGCCAAGCTGCCGGACATGCTTGCTGAGCCCCTGTTC
CAGCGCTTCAAAGTCAATCCGGGGCTGCTGCTCCGGGGGCTTGTGTCCTGCACGGGCAACCAGTTT
TGCGGCTTCGGTCTGGCGGAGACAAAGGCGCGGGCGGTCAAGGTAGTTGAGATGCTGGAGGAGCAG
TTGGAGCTCACCCGGCCTGTCAGGATCCACTTCACCGGATGCCCCAACAGTTGCGGCCAAGCGCAG
CAGGTTGGCGACATTGGGTTAATGGGAGCCCCCGCCAAGCTGGATGGCAAGGCGGTGGAGGGCTAC
AAGATCTTTTTGGGCGGGAAGATTGGGGAGAACCCGCAGCTGGCCACGGAATTCGCTCAAGGGATC
CCGGCTGTGGAGTCTCATCTGGTGCCCAAACTCAAGGAGATCCTTATTAAGGAGTTTGGTGCCAAG
GAAAAGGAGACTGCCGTTGTCGTCTAAATAGGCGTCGTTGCGTAATTAGGTGCTTATAACGGAGAA
GGGGGAATGATAGCTTGGTGTAAGTGTTACATAGGATTGGGGAGGGAGTGGTAGGCACGGGTTTGA
TGCGTGATATACTACATGTGACCTGATGTCGTATTTTGCATACAAGTATCTTGTCCGGCGCTTCTC
ATGCGTGTGCGTGTCTGTTTGTTCTGTTTCGGCTAGCAGGGCGGCCAAGTCGTTTATGTTCGGGGA
TTCCTACTACGGGCGCAATTGCAATGATAAAAGAAGGATGCGTGTCTTGTCTGGGGCCTGTGAATC
ACTCCTTCCGATATGCCGCGACGTTTGCTGTGCGCGCGGCGTGCAGGTCAGGGTTTGTCGATAGGT
AGCGTTTGCACGTCGCGTCCGTGAGTATCTATATCAGAGCAGCTTGCGCATGTATGTGTTAACCAA
GTTTTTTTTATTGGCGTGGGAACTGTGCTCCCGGGCGAATTATGCTCGCCAGCGCTGCCGGTGGTC
TGTGATTGATTAGGCATTGGTCATCTGTATCCATTCGACTTATCAGACTTATCATGTCTCGCGATC
GGATGTTGTGCTGCCTTGTTCCATTCTTTTGCACATCCGTTGTGTCGATGGCGTGGGAAGATGCCG
AGGCTACGATGAAGAGTGTAGATAGAGGGTCGCGTTCGTGGTGATGGTGCCGCACAG

SEQ ID NO: 56, Spinacia oleracea - X07568

CATCATCTTCATCTTCATCTTCATCATTCATAGTTGCAAGAAACAGAGCAACCAAAAAAAAATGGCA
TCACTTCCAGTCAACAAGATCATACCATCATCAACGACATTACTGTCATCGTCGAACAACAACAGA
AGAAGAAATAACTCATCAATTCGATGCCAGAAGGCGGTTTCACCCGCGGCAGAAACGGCTGCAGTG
TCGCCGTCTGTGGACGCGGCGAGGCTGGAGCCGAGAGTGGAGGAGAGAGATGGGTTTTGGGTATTG
AAGGAGGAATTTAGGAGTGGGATTAACCCAGCTGAGAAAGTTAAGATTGAGAAAGACCCAATGAAG
TTGTTTATTGAGGATGGGATTAGTGATCTTGCTACTTTGTCAATGGAGGAAGTTGATAAATCTAAG
CATAATAAGGATGATATTGATGTTAGACTCAAGTGGCTTGGACTTTTCCATCGCCGTAAACATCAC
TATGGGAGATTCATGATGAGGTTGAAGCTGCCGAATGGGGTAACAACGAGTGAGCAGACACGGTAC
CTAGCAAGCGTGATCAAGAAGTACGGAAAAGATGGATGTGCGGATGTAACAACAAGGCAAAACTGG
CAAATTAGAGGAGTTGTTCTGCCTGATGTGCCAGAGATCATCAAAGGGCTGGAATCCGTTGGTCTT
ACCAGCTTACAGAGTGGGATGGACAATGTAAGGAACCCTGTAGGTAACCCTCTTGCAGGGATTGAC
CCTCATGAAATTGTTGACACCCGACCTTTTACCAACCTAATTTCCCAATTTGTCACTGCCAATTCG
CGTGGAAACCTTTCTATTACCAATCTGCCAAGGAAGTGGAATCCATGTGTTATTGGGTCCCATGAT
CTTTATGAGCATCCACACATCAATGACCTTGCTTACATGCCTGCTACAAAGAATGGGAAATTCGGG
TTTAATTTGTTGGTTGGAGGATTCTTTAGCATCAAAAGATGTGAAGAGGCAATCCCACTAGACGCT
TGGGTCTCAGCAGAAGATGTGGTTCCTGTATGCAAAGCTATGCTTGAAGCTTTCAGGGACCTTGGC
TTTAGAGGAAACAGGCAGAAGTGCAGAATGATGTGGCTTATTGATGAGCTTGGTATGGAAGCATTC
AGGGGAGAGGTTGAGAAGAGAATGCCTGAGCAAGTTCTAGAAAGAGCATCCTCAGAAGAGCTGGTT
CAGAAGGACTGGGAGAGAAGAGAATACTTAGGAGTTCACCCTCAGAAACAACAAGGACTTAGCTTT
GTGGGTCTCCACATTCCTGTGGGCCGTCTGCAAGCTGATGAGATGGAAGAGTTAGCCCGTATAGCT
GATGTGTATGGATCAGGGGAGCTCCGTCTGACAGTAGAGCAGAACATAATCATCCCAAATGTTGAA
AACTCAAAGATAGATTCACTACTAAACGAGCCTCTGTTAAAAGAGCGTTACTCCCCTGAACCACCC
ATCTTGATGAAGGGGCTTGTGGCCTGTACGGGGAGCCAATTTTGTGGACAAGCCATTATCGAGACC
AAGGCTAGGGCACTCAAGGTGACAGAAGAGGTACAACGACTAGTGTCTGTAACACGGCCTGTTAGG
ATGCATTGGACCGGGTGTCCTAATAGTTGTGGTCAAGTACAAGTGGCTGATATTGGGTTCATGGGT
TGCATGACTAGGGATGAGAACGGTAAGCCTTGTGAAGGAGCTGATGTGTTTGTAGGAGGACGTATA

FIGURE 2 (continued)

GGAAGTGACTCGCATCTAGGAGACATTTACAAGAAGGCAGTCCCATGTAAAGATTTGGTGCCTGTT
GTTGCTGAGATATTGATCAACCAATTCGGTGCTGTTCCTAGGGAGAGGGAAGAGGCAGAGTAGTAG
CTAGACTGTTTTGGGTGCCTGTTCTTGTTAACTGTTATCGGTATTCGGTAATTACTTGTAATATTT
GCATTTTTTTTCAAGCATATAATTAAATTGCATAAAGATCCCTTGTATGTCTGCATAACAAGATAC
TCAGTTATGTAATGTCAATAGCAGGTTTACTTTGTTTATTCAATAGGCACTGTGAAAGGGAAAGTT
CATTATTCATTTCTCA

SEQ ID NO: 57, Nostoc sp. NC - 003272
ATGACAGATACAGTAACTACCCCCAAAGCCAGCCTCAATAAGTTTGAGAAATTCAAAGCCGAAAAA
GATGGACTTGCCATCAAGTCAGAGATCGAAAAAATTGCCTCTTTGGGATGGGAAGCAATGGACGCA
ACAGACCGAGATCATCGCCTCAAATGGGTGGGTGTATTCTTTCGCCCAGTCACCCCTGGTAAATTT
ATGATGCGGATGCGGATGCCGAATGGTATCCTCACCAGCGATCAGATGCGTGTTTTAGCCGAAGTG
GTGCAGCGTTACGGAGATGACGGCAACGCTGATATTACAACTAGGCAGAATATTCAACTACGAGGT
ATCAGAATAGAAGACTTACCGCACATATTCAATAAATTTCATGCAGTAGGTTTAACCAGTGTGCAG
TCAGGGATGGACAACATCCGTAACATCACAGGCGACCCGATAGCGGGGTTAGATGCGGATGAGTTG
TATGACACCCGTGAGTTAGTGCAGCAAATTCAGGATATGCTCACCAACAAAGGAGAAGGCAATCGA
GAGTTTAGTAATTTGCCTCGTAAATTTAATATTGCGATCGCCGGTGGACGGGATAATTCAGTTCAT
GCGGAAATCAACGATTTAGCCTTTGTTCCAGCATTTAAAGAAGGGATTGGAGATTGGGTATTGGGG
AATGGGGAAGAATCATCTACTTACCAAAAAGTCTTTGGATTTAACGTGTTAGTTGGTGGTTTCTTT
TCTGCTAAACGCTGTGAGGCGGCGATTCCTTTGAATGCTTGGGTAACTCCGGAAGAAGTCTTACCC
TTATGTAGAGCAATTTTAGAGGTCTATCGTGACAATGGACTCAGGGCTAATCGGCTCAAGTCTCGC
TTGATGTGGCTAATTGATGAATGGGGTATAGATAAGTTTCGGGCAGAAGTCGAACAGCGTTTGGGT
AAATCCTTACTCCCCGCAGCCCCCAAAGACGAAATTGATTGGGAAAAACGCGACCATATCGGAGTC
TATAAGCAAAAGCAAGAGGGATTGAACTATGTAGGGTTACACATCCCTGTAGGTAGATTGTATGCC
GAGGATATGTTTGAATTGGCTCGGATAGCCGATGTATACGGTAGCGGTGAAATCCGCATGACTGTT
GAACAAAACATCATCATTCCCAACATTACCGACTCGCGGTTAAGGACTTTGTTGACAGATCCCTTA
CTAGAGAGATTTTCTCTTGATCCTGGAGCATTGACGCGATCGCTAGTTTCCTGCACGGGCGCACAA
TTTTGCAACTTCGCCCTCATCGAAACCAAAAACCGCGCCCTAGAAATGATTAAAGGCTTAGAAGCA
GAATTGACCTTTACTCGTCCAGTGCGAATCCATTGGACAGGTTGCCCCAACTCCTGCGGACAGCCC
CAAGTTGCAGACATTGGCTTAATGGGAACAAAAGCTCGTAAAAACGGTAAAGCCGTGGAAGGTGTT
GACATCTATATGGGTGGCAAAGTCGGCAAAGATGCACATTTAGGTAGCTGTGTACAAAAAGGCATC
CCCTGCGAAGACTTGCACCTAGTATTACGAGACTTACTCATTACTAATTTTGGAGCCAAACCCAGA
CAGGAAGCCTTAGTTACCAGCCAATAA

SEQ ID NO: 58, Plectonema boryanum - D31732
AACACTGCCGGAACTCGACTCATGACCCATCCAACGCTTGCCCACGATAGAAATGTTCTCCGACGC
ATGAGGTTCTCCTAAAGAACGATAGAGGAATAGTGAGTAGGGAGTGGGGAGTAGGGTAAATCCTTT
CTATCTCCCACTCCTCCCCCGCTCCCCACCAAATTACAACTATTTCTAAAGTACGCCCTTCCCCCT
CTTCCCGCCGACAGATGACGAAAACGAATCGGCTTTATGCAGAAACGTCATATTATGAAAAGTTTT
GTAACAACAGATACGAATGTCCTCTGTGATCCCGATTACCTTTACTCAGTAATCACCGCGAATCAT
CAAACGGTTCCGCAGTTGATATCGATTTGTGTTCGCTCTGGAACACCTTATATTCATAGGCTCAAT
CCATGACAGACACCCTTGCAGCACCGACCCTCAATAAGTTTGAAAAACTCAAAGCAGAGAAAGATG
GTCTTGCGGTGAAAGCAGAACTCGAGCACTTTGCTCGGCTCGGCTGGGAAGCAATGGATGAAACCG
ATCGTGATCATCGCTTGAAGTGGCTCGGTGTGTTCTTTCGCCCCGTAACTCCTGGCAAATTTATGC
TGAGAATGCGGGTTCCGAATGGCATTATCACGAGCGGACAAACCCGGGTGCTAGGAGAAATCCTTC
AGCGCTATGGAGATGATGGCAATGCAGACATCACGACTCGCCAGAACTTTCAACTGCGAGGAATTC
GGATTGAAGACCTTCCCGAAATTTTTCGTAAGTTTGACCAAGCTGGATTGACGAGCATTCAATCCG
GGATGGATAACGTTCGTAACATTACCGGATCGCCTGTTGCTGGCATTGATGCAGATGAGCTAATTG

FIGURE 2 (continued)

ATACTCGTGGGCTAGTTCGCAAAGTTCAAGACATGATCACGAACAATGGTCGTGGTAATTCGAGCT
TTAGTAACTTGCCTCGGAAATTCAATATTGCGATCGCAGGGTGCCGCGATAACTCAGTTCATGCTG
AAATCAATGACATTGCTTTCGTTCCCGCTTTCAAAGATGGCACATTAGGATTCAATATCCTAGTTG
GCGGATTCTTCTCTGGGAAACGCTGCGAAGCTGCAATTCCACTCAATGCTTGGGTTGACCCGCGCG
ATGTCGTTGCGGTCTGCGAAGCAATTTTAACGGTCTATCGGAACTTGGGACTGAGAGCAAATCGTC
AAAAAGCTCGCTTAATGTGGCTGATTGATGAGATGGGATTGGAACCGTTCCGCGAAGCGGTTGAAA
AACAATTGGGATATGCTTTTACGCCTGCTGCTGCCAAAGACGAGATCCTTTGGGACAAGCGAGATC
ACATTGGGATTCATGCCCAAAAACAGCCTGGATTAAACTATGTGGGCTTGCATGTTCCAGTGGGAC
GGTTATACGCGCAAGATTTGTTTGATTTAGCTCGGATCGCTGAAGTTTACGGCAGTGGTGAAATTC
GCTTAACTGTCGAGCAGAATGTGATCATTCCGAATGTTCCGGATTCACGAGTTTCTGCATTGCTCA
GAGAACCCATTGTCAAACGGTTCTCGATCGAGCCTCAGAATCTTTCACGGGCATTAGTGTCTTGTA
CTGGCGCACAGTTTTGTAACTTCGCACTGATTGAAACTAAAAATCGTGCGGTTGCTTTAATGCAAG
AGCTAGAACAAGACCTGTACTGTCCTCGTCCAGTGCGCATTCATTGGACAGGTTGCCCGAACTCTT
GTGGACAACCTCAAGTTGCAGATATCGGACTGATGGGCACAAAAGTCCGCAAAGATGGCAAAACAG
TCGAAGGCGTGGATCTCTATATGGGGGGCAAAGTTGGCAAACATGCTGAACTTGGAACCTGTGTGA
GAAAAAGCATTCCCTGTGAAGATCTCAAACCGATTCTGCAAGAGATTTTGATCGAGCAATTTGGGG
CGCGTCTCTGGTCAGACCTGCCCGAATCCGCTCGTCCAAATCCGACCGCCTTGATCACGCTCGATC
GTCCCACGGTGGAAACACCGAACGGGAAATCAACAACCGTGCAAGAGCTTAATGCACAAGAGTTTG
ACTATGTGCTGAGTGCGCCACCTGTTGTAAAAGCGCCAACAGAAATCGCAGCTCCAGCAACGATTC
GTTTTGCTCAGTCAGGAAAAGAAATCACCTGCACCCAGGATGATTTGATTCTAGACATTGCAGACC
AAGCCGAAGTCGCGATCGAAAGTTCTTGCCGATCAGGAACGTGTGGAAGTTGTAAATGCACCTTAC
TCGAAGGTGAAGTCAGCTATGACAGCGAACCCGATGTGCTCGATGAGCACGATCGCGCTTCGGGTC
AGATTCTCACCTGTATTGCTCGTCCTGTCGGTCGTATCTTGCTCGATGCTTGATCCCTAAGTTTTG
TTGCTCCGCTCATTGTTCTCACATGCGCCAGCTTTTTGCTGTGCTTCCTTTTCCTTCAGTACATTC
TCTAAAAAGGACGATCCATGTCTTCTAATCTTTCAAGACGTAAGTTCATTTTGACCGCAGGCGCAA
CCGCAGCAGGCGCAGTGATTGTGAATGGTTGTAGCACAGGTCTAAATAAAAGTGCTTCTAGCGGTG
CGTCCTCTCCTGCTGCCTCTCCTGCTGCAAATATCAGTGCGGCAGATGCACCAGAAGTCACAACGG
CTAAATTAGGCTTTATCGCCCTGACCGATTCGGCTCCATTGATCATTGCGTTAGAGAAAG

SEQ ID NO: 59, Anabaena variabilis - CP000117

ATGACAGATACAGCAACTACCCCCAAAGCCAGTCTCAATAAGTTTGAGAAATTCAAAGCCGAAAAA
GATGGCCTTGCCATCAAGTCAGAGATTGAAAAAATTGCCTCTTTGGGATGGGAAGCAATGGACGAA
ACAGACCGAGACCATCGCCTCAAATGGGTGGGTGTATTCTTTCGTCCAGTCACCCCTGGCAAATTC
ATGATGCGGATGCGGATGCCTAATGGTATTCTCACCAGCGATCAAATGCGTGTTTTAGCTGAAGTG
GTGCAGCGTTACGGAGATGATGGCAACGCTGATATTACAACTAGGCAGAATATCCAACTACGGGGA
ATCAGAATAGAAGACTTACCGCACATATTCAATAAATTTCATGCAGTAGGTTTAACTAGTGTGCAG
TCGGGGATGGACAATATCCGCAATATTACAGGCGACCCCATAGCAGGGTTGGATGCAGATGAATTG
TATGATACCCGTGAGTTAGTGCAGCAAATCCAAGATATGCTCACCAACAAGGGAGAAGGTAATCGA
GAGTTTAGTAATTTACCACGGAAATTTAATATTGCGATCGCTGGTGGACGGGATAATTCAGTTCAT
GCAGAAATCAACGATTTAGCTTTTGTTCCCGCATTCAAAGAAGGGATTGGGGATTGGGTATTGGGA
GGTGGTGAAGAATCTTCTACTCACCAAAAAGTCTTTGGATTTAACGTGTTAGTTGGTGGCTTCTTT
TCTGCCAAACGTTGTGAAGCGGCAATTCCTTTAAATGCTTGGGTAACAGCTGAAGAAGTCGTAGCC
TTATGTAGAGCAGTTCTGGAAGTCTATCGTGACAACGGACTTAGAGCTAATCGGCTTAAGTCTCGC
TTGATGTGGCTAATTGATGAATGGGGTATAGATAAGTTCCGTGCAGAAGTCGAACAGCGTTTGGGT
AAATCCTTACTATACGCTGCACCCAAAGACGAAATTGATTGGGAAAAACGCGACCATATCGGAGTC
TATAAACAAAAGCAAGAGGGATTGAACTATGTAGGCTTACACATACCCGTAGGTAGATTGTATGCC
GAAGATATGTTTGAACTAGCTCGGATAGCCGATGTTTACGGTAGCGGTGAAATCCGTATGACTGTT
GAACAAAACATCATCATTCCCAACATTACCGACTCGCGGTTAAAGACTTTGTTGACAGATCCTTTA

FIGURE 2 (continued)

CTAGAGAGATTTTCTCTTGATCCGGGAGCATTGACGCGATCGCTAGTTTCCTGCACAGGCGCACAA
TTTTGCAACTTCGCCCTCATCGAAACCAAAAACCGCGCCCTAGAAATGATTAAAGGCTTAGAAGCA
GAGTTAACATTCACCCGTCCAGTGCGAATCCATTGGACAGGTTGCCCCAACTCCTGCGGACAACCC
CAAGTTGCAGACATCGGTTTAATGGAACAAAAGCCCGTAAGAACGGTAAAGCCGTCGAAGGTGTT
GACATCTATATGGGGGGCAAAGTCGGCAAAGACGCACATTTAGGTAGTTGTGTACAAAAAGGCATC
CCCTGCGAAGACTTGCACCTAGTATTACGAGACTTGCTGATTACTAATTTTGGAGCCAAACCCAGG
CAGGAAGCCTTAGTTAGTAGCCAGTAG

SEQ ID NO: 60, Synechococcus sp. CP000239
ATGGCGAACCAATTTGAACGCCTCAAAAGCGAAAAGGATGGGCTGGCGGTCAAGGCCGAGCTGGAG
GCGTTTGCCCGGATGGGTTGGGAGAACATTCCTGAAGACGACCGGGATCACCGCCTCAAGTGGCTG
GGGATCTTCTTTCGCAAGCGCACCCCAGGTCAGTTCATGCTGCGGCTGCGCCTGCCCAATGGGATC
CTAACCAGCGGCCAAATGCGGATGTTGGGCGCAATCATCCACCCCTATGGAGAACAGGGCGTAGCC
GACATCACCACCCGGCAGAACCTGCAACTGCGCGGCATCCCCATTGAAGAAATGCCCCAGATCCTG
GGCTACCTGAAAGAGGTAGGCCTGACCAGCATCCAGTCGGGCATGGACAACGTGCGCAACATCACG
GGATCCCCTCTGGCCGGTATTGACCCGGATGAGCTGATCGATGTGCGCGGTCTCACCCGCAAGGTG
CAGGACATGGTTACCAACAACGGCGAGGGCAACCCTTCCTTCAGCAACCTGCCGCGCAAGTTCAAC
ATCGCCATCTGCGGTTGTCGCGACAACTCCGTGCATGCGGAGATCAACGACCTGGCCTTTGTGCCC
GCCTTCAAAAATGGCCGCCTGGGCTTCAACGTCCTGGTGGGCGGCTTTTTCTCGGCTCGCCGCTGC
GCCGAGGCAATTGGCCTAGATGTCTGGGTGGATCCCCGCGATGTAGTTCCCCTGTGCGAGGCGGTG
CTGCTGGTCTACCGGGATCACGGCCTGCGGGCCAACCGGCAAAAGGCGCGGTTGATGTGGCTCATT
GACGAGTGGGGCCTAGAGAAGTTCCGGGCGGCTGTGGAGCGCCAGATAGGCCACCCTCTGCCCAGG
GCAGCGGAAAAAGACGAGGTGGTCTGGCACAAGCGGGATCTGCTGGGGGTGCATGCCCAGAAGCAG
CCGGGCCTCAACTTTGTCGGCCTGCATGTGCCGGTGGGGCGGCTCAACGCCCTGGAGATGATGGAG
CTGGCCCGCTTGGCGGAGGTGTACGGCTCCGGGGAGCTGCGGCTGACAGTGGAGCAGAACGTGCTC
ATCCCCAATGTGCCCGACTCCCGAGTGGCCCCGCTCCTCAAAGAGCCGCTCTTGAAGAAGTTCTCC
CCCAACCCAGGGCCCTTGCAGCGGGGGTTGGTGTCCTGCACGGGCAACCAGTTCTGCAACTTTGCC
CTTATCGAGACCAAAAACCGGGCTGTGGCCTTGATGGAGGAGCTGGAGGCGGAGCTGGAGATCCCC
CAAACGGTGCGCATCCACTGGACGGGCTGCCCCAACTCCTGCGGCCAACCCCAAGTAGCCGATATC
GGCCTTATGGGCACCACTGCTCGCAAGGACGGCAGGGTGGTGGAGGCCGTGGACATCTACATGGGG
GGAGAGGTGGGCAAAGACGCCAAGCTGGGCGAATGCGTGCGCAAAGGGATCCCTTGCGAAGACCTC
AAGCCGGTCTTGGTGGAGCTGCTCATTGAACACTTTGGGGCCAAGCCGCGTCAGCATCCGTCCGCC
GCCCAGGCTTCTGTTTTGGTAACCCGCTAG

SEQ ID NO: 61, Arabidopsis thaliana - AT5G04590
TCTCACCCACCCAAAGCCACTCACTCTCTCTTCTCTCTCTCTGAAGCGATGTCATCGACGTTTCGA
GCTCCGGCGGGAGCCGCTACTGTGTTTACGGCGGATCAGAAGATCAGACTTGGGAGGCTCGACGCT
CTGAGATCCTCTCATTCTGTTTTCTTAGGAAGATATGGACGCGGCGGCGTCCCGGTTCCTCCTTCC
GCTTCTTCGTCGAGTTCTTCGCCTATTCAAGCCGTCTCCACTCCTGCGAAGCCTGAGACTGCGACC
AAGCGGAGCAAAGTCGAAATTATCAAGGAGAAGAGTAATTTCATAAGGTATCCTTTGAACGAGGAG
CTTTTAACAGAGGCTCCAAATGTCAACGAGTCAGCCGTGCAGCTTATCAAGTTCCACGGTAGCTAC
CAACAGTACAACAGAGAAGAACGTGGTGGAAGATCTTACTCCTTCATGCTTCGAACTAAGAATCCA
TCTGGGAAGGTCCCTAACCAGCTCTATTTGACTATGGATGACTTAGCTGATGAGTTTGGAATTGGT
ACTCTTCGTTTGACCACAAGGCAGACGTTTCAGCTTCATGGTGTTCTGAAGCAGAATCTTAAGACT
GTGATGAGCTCGATTATTAAAAATATGGGTAGCACGCTTGGTGCATGTGGTGATCTGAACAGAAAT
GTTCTTGCTCCTGCTGCACCTTATGTGAAGAAAGACTATCTCTTTGCACAAGAAACTGCTGACAAC
ATTGCGGCTCTTCTTTCTCCTCAATCAGGGTTCTATTATGATATGTGGGTTGATGGAGAGCAGTTC
ATGACTGCTGAACCTCCAGAGGTAGTGAAGGCTCGAAATGATAACTCCCATGGAACTAACTTTGTC

FIGURE 2 (continued)

GACTCTCCTGAGCCCATCTATGGCACCCAGTTCTTGCCTAGAAAGTTCAAGGTCGCTGTAACTGTT
CCTACAGATAATTCCGTCGACCTCCTCACCAATGACATTGGCGTTGTTGTTGTTTCAGATGAAAAT
GGGGAACCACAGGGTTTCAATATTTATGTTGGTGGGGGTATGGGAAGAACACACAGAATGGAGTCT
ACTTTTGCCCGCCTGGCAGAACCAATAGGTTATGTTCCAAAGGAAGATATTTTGTATGCTGTGAAG
GCCATTGTAGTCACACAGCGAGAACACGGGAGACGAGATGATCGTAAATATAGCAGAATGAAATAT
TTGATCAGCTCCTGGGGAATTGAGAAGTTCAGAGATGTTGTTGAGCAATATTATGGTAAAAAGTTT
GAGCCTTCCCGTGAACTTCCAGAGTGGGAGTTCAAGAGTTACTTGGGATGGCATGAACAGGGAGAT
GGTGCATGGTTTTGTGGGCTTCACGTAGACAGTGGTCGTGTTGGAGGTATAATGAAGAAGACGCTG
AGAGAAGTAATAGAGAAATACAAAATTGATGTCCGCATCACACCAAACCAAAACATTGTCTTGTGT
GATATAAAGACTGAATGGAAGCGTCCCATCACCACAGTACTTGCTCAGGCCGGCTTACTGCAACCT
GAGTTTGTCGACCCATTAAACCAAACTGCAATGGCTTGCCCAGCTTTTCCTTTGTGCCCTCTGGCA
ATAACTGAGGCAGAGCGCGGGATCCCCAGCATTCTAAAGAGAGTTAGGGCAATGTTTGAAAAGGTT
GGTCTGGACTACGACGAGTCTGTTGTGATAAGAGTAACCGGTTGTCCAAACGGCTGTGCAAGACCG
TACATGGCTGAGCTCGGTCTAGTCGGGGATGGTCCCAACAGCTATCAGGTTTGGCTAGGAGGAACA
CCGAACCTGACCCAGATAGCGAGAAGTTTCATGGATAAGGTTAAGGTTCACGACTTAGAGAAAGTC
TGCGAGCCATTGTTCTATCACTGGAAACTAGAGAGGCAAACTAAAGAATCATTTGGAGAATACACA
ACCCGCATGGGATTCGAGAAACTGAAGGAGCTGATAGATACATACAAAGGAGTTTCTCAATGAGCA
CAACAGAGATCATCTTTCGTTTTATAATTCATGTAATGTAATGTCTCTGTCTGAACTGTTACTCTT
CGGTAACTCTGATGGAGAACTTGTTCTCGTTTTGGTTTGATTTTGTACCCTCTTTTTTTTTTTTGT
TTTTTTGGATTGCTTTGTCTTTGATTGGATAATGAAGCATTACTGTATCAAGGCTAATTAGCCCAT
CAATAAGCCTTTTTAAAGCTCTGGA

FIGURE 2 (continued)

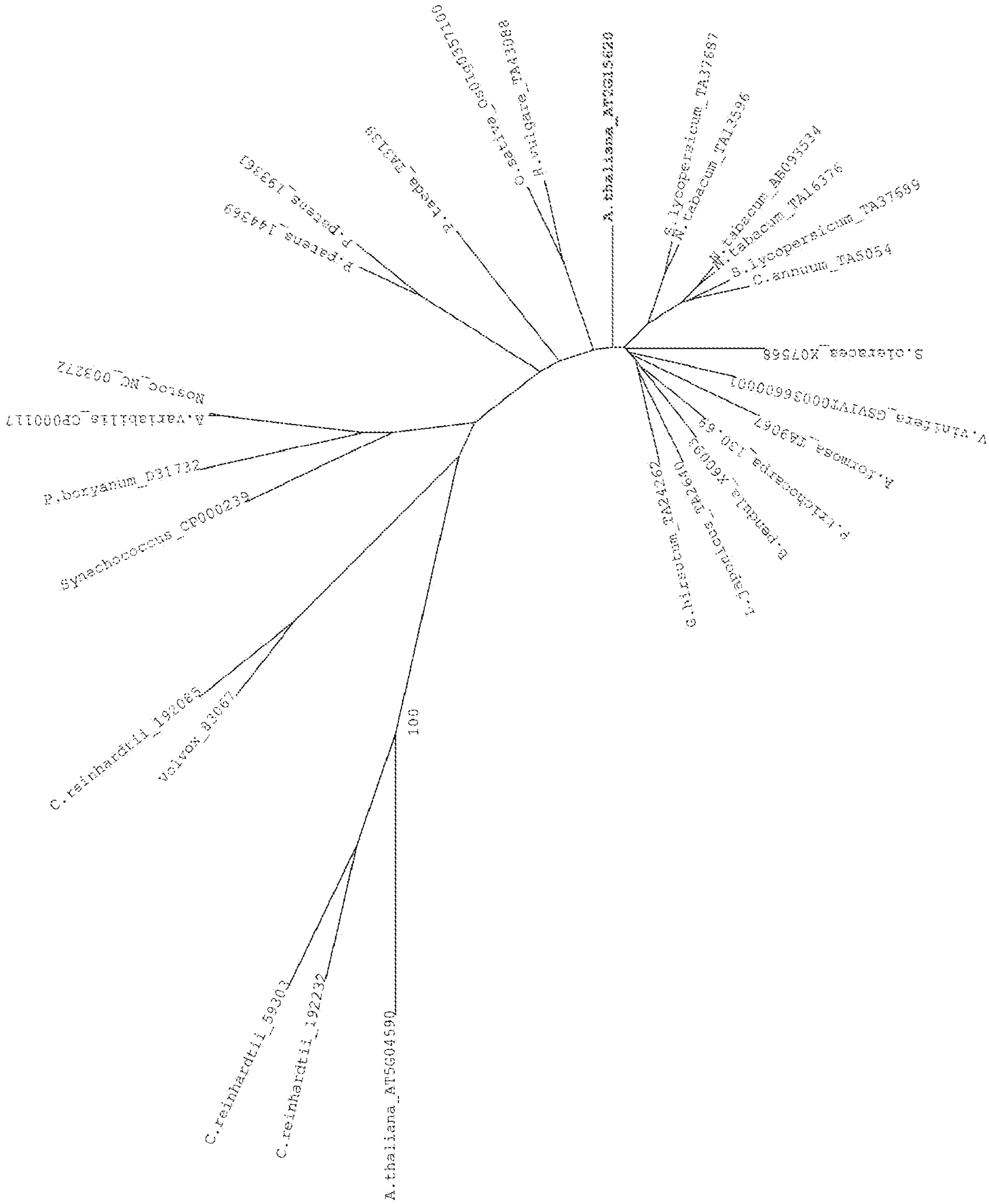
O.sativa_Os01g0357100
H.vulgare_TA43088
S.lycopersicum_TA37687
N.tabacum_TA13596
N.tabacum_AB093534
N.tabacum_TA16376
S.lycopersicum_TA37689
C.annuum_TA5054
S.oleracea_X07568
V.vinifera_GSVIVT00036600001
A.formosa_TA9067
P.trichocarpa_130_69
B.pendula_X60093
L.japonicus_TA2640
G.hirsutum_TA24262
P.taeda_TA3139
P.patens_193363
P.patens_144369
Nostoc_NC_003272
A.variabilis_CP000117
P.boryanum_D31732
Synechococcus_CP000239
C.reinhardtii_192085
Volvox_83067
100
C.reinhardtii_59303
C.reinhardtii_192232
A.thaliana_AT5G04590

FIGURE 3

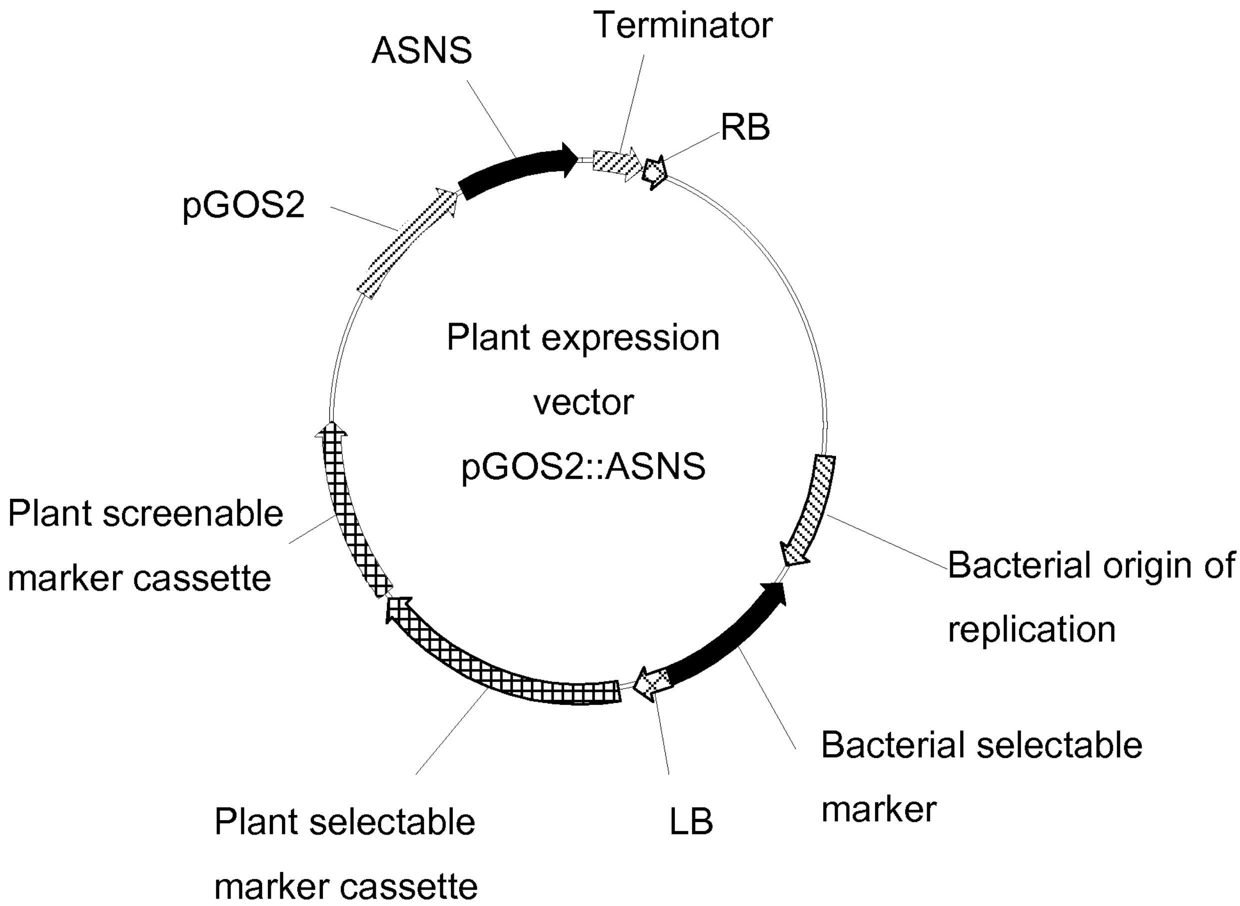
ASNS
Terminator
RB
pGOS2
Plant expression vector pGOS2::ASNS
Plant screenable marker cassette
Bacterial origin of replication
Bacterial selectable marker
LB
Plant selectable marker cassette

FIGURE 4

SEQ ID NO: 62, DNA – *Oryza sativa*, ORF 44 - 1816

TTTTTTATAATGCCAACTTTGTACAAAAAAGCAGGCTTAAACAATGTGTGGCATCCTCGCCGTGCT
CGGCGTCGCAGACGTCTCCCTCGCCAAGCGCTCCCGCATCATCGAGCTATCCCGCCGGTTACGTCA
TAGAGGCCCTGATTGGAGTGGTATACACTGCTATCAGGATTGCTATCTTGCACACCAGCGGTTGGC
TATTGTTGATCCCACATCCGGAGACCAGCCGTTGTACAATGAGGACAAATCTGTTGTTGTGACGGT
GAATGGAGAGATCTATAACCATGAAGAATTGAAAGCTAACCTGAAATCTCATAAATTCCAAACTGC
TAGCGATTGTGAAGTTATTGCTCATCTGTATGAGGAATATGGGGAGGAATTTGTGGATATGTTGGA
TGGGATGTTCGCTTTTGTTCTTCTTGACACACGTGATAAAAGCTTCATTGCAGCCCGTGATGCTAT
TGGCATTTGTCCTTTATACATGGGCTGGGGTCTTGATGGTTCGGTTTGGTTTTCGTCAGAGATGAA
GGCATTAGGTGATGATTGCGAGCGATTCATATCCTTCCCCCCTGGGCACTTGTACTCCAGCAAAAC
AGGTGGCCTAAGGAGATGGTACAACCCACCATGGTTTTCTGAAAGCATTCCCTCCACCCCGTACAA
TCCTCTTCTTCTCCGACAGAGCTTTGAGAAGGCTATTATTAAGAGGCTAATGACAGATGTGCCATT
TGGTGTTCTCTTGTCTGGTGGACTGGACTCTTCTTTGGTTGCATCTGTTGTTTCGCGGCACTTGGC
AGAGGCAAAAGTTGCCGCACAGTGGGGAAACAAACTGCATACATTTTGCATTGGTTTGAAAGGTTC
TCCTGATCTTAGAGCTGCTAAGGAAGTTGCAGACTACCTTGGTACTGTTCATCACGAACTCCACTT
CACAGTGCAGGAAGGCATTGATGCACTGGAGGAAGTCATTTACCATGTTGAGACATATGATGTAAC
GACAATTAGAGCAAGCACCCCAATGTTCTTGATGTCACGTAAAATTAAATCTTTGGGGGTGAAGAT
GGTTCTTTCGGGAGAAGGTTCTGATGAGATATTTGGCGGTTACCTTTATTTTCACAAGGCACCAAA
CAAGAAGGAATTCCATGAGGAAACATGTCGGAAGATAAAAGCCCTTCATTTATATGATTGCTTGGG
AGCGAACAAATCAACTTCTGCATGGGGTGTTGAGGCCCGTGTTCCGTTCCTTGACAAAAACTTCAT
CAATGTAGCTATGGACATTGATCCTGAATGGAAAATGATAAAACGTGATCTTGGCCGTATTGAGAA
ATGGGTTCTCCGGAATGCATTTGATGATGAGGAGAAGCCCTATTTACCTAAGCACATTCTATACAG
GCAAAAGGAGCAATTCAGTGATGGTGTTGGGTACAGTTGGATTGATGGATTGAAGGATCATGCAAA
TGAACATGTATCAGATTCCATGATGATGAACGCTAGCTTTGTTTACCCAGAAAACACTCCAGTTAC
AAAAGAAGCGTACTATTATAGGACAATATTCGAGAAATTCTTTCCCAAGAATGCTGCTAGGTTGAC
AGTACCTGGAGGTCCTAGCGTCGCGTGCAGCACTGCTAAAGCTGTTGAATGGGACGCAGCCTGGTC
CAAAAACCTTGATCCATCTGGTCGTGCTGCTCTTGGTGTTCATGATGCTGCATATGAAGATACTCT
ACAAAAATCTCCTGCCTCTGCCAATCCTGTCTTGGATAACGGCTTTGGTCCAGCCCTTGGGGAAAG
CATGGTCAAAACCGTTGCTTCAGCCACTGCCGTTTAACTTTCTATCGTCGCACCCAGCTTTCTTGT
ACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTC
AAAATAATATCATTATTTGCCATCCAGCTGCAGCTCTGGCCCGTGTCTCAAAATCTCTGATGTTAC
ATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACA
AGGGGTGTTATGAGCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACATGGATGCT
GATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGCTTG
TATGGGAAGCCCGATGA

SEQ ID NO: 63, protein – *Oryza sativa*

MCGILAVLGVADVSLAKRSRIIELSRRLRHRGPDWSGIHCYQDCYLAHQRLAIVDPTSGDQPLYNE
DKSVVVTVNGEIYNHEELKANLKSHKFQTASDCEVIAHLYEEYGEEFVDMLDGMFAFVLLDTRDKS
FIAARDAIGICPLYMGWGLDGSVWFSSEMKALGDDCERFISFPPGHLYSSKTGGLRRWYNPPWFSE
SIPSTPYNPLLLRQSFEKAIIKRLMTDVPFGVLLSGGLDSSLVASVVSRHLAEAKVAAQWGNKLHT
FCIGLKGSPDLRAAKEVADYLGTVHHELHFTVQEGIDALEEVIYHVETYDVTTIRASTPMFLMSRK
IKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLGANKSTSAWGVEARV
PFLDKNFINVAMDIDPEWKMIKRDLGRIEKWVLRNAFDDEEKPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHANEHVSDSMMMNASFVYPENTPVTKEAYYYRTIFEKFFPKNAARLTVPGGPSVACSTAKA
VEWDAAWSKNLDPSGRAALGVHDAAYEDTLQKSPASANPVLDNGFGPALGESMVKTVASATAV

FIGURE 5

SEQ ID NO: 64, DNA - *Oryza sativa* - GOS2 promoter sequence
AATCCGAAAAGTTTCTGCACCGTTTTCACCCCCTAACTAACAATATAGGGAACGTGTGCTAAATAT
AAAATGAGACCTTATATATGTAGCGCTGATAACTAGAACTATGCAAGAAAAACTCATCCACCTACT
TTAGTGGCAATCGGGCTAAATAAAAAAGAGTCGCTACACTAGTTTCGTTTTCCTTAGTAATTAAGT
GGGAAAATGAAATCATTATTGCTTAGAATATACGTTCACATCTCTGTCATGAAGTTAAATTATTCG
AGGTAGCCATAATTGTCATCAAACTCTTCTTGAATAAAAAAATCTTTCTAGCTGAACTCAATGGGT
AAAGAGAGAGATTTTTTTTAAAAAAATAGAATGAAGATATTCTGAACGTATTGGCAAAGATTTAAA
CATATAATTATATAATTTTATAGTTTGTGCATTCGTCATATCGCACATCATTAAGGACATGTCTTA
CTCCATCCCAATTTTTATTTAGTAATTAAAGACAATTGACTTATTTTTATTATTTATCTTTTTTCG
ATTAGATGCAAGGTACTTACGCACACACTTTGTGCTCATGTGCATGTGTGAGTGCACCTCCTCAAT
ACACGTTCAACTAGCAACACATCTCTAATATCACTCGCCTATTTAATACATTTAGGTAGCAATATC
TGAATTCAAGCACTCCACCATCACCAGACCACTTTTAATAATATCTAAAATACAAAAAATAATTTT
ACAGAATAGCATGAAAAGTATGAAACGAACTATTTAGGTTTTTCACATACAAAAAAAAAAAGAATT
TTGCTCGTGCGCGAGCGCCAATCTCCCATATTGGGCACACAGGCAACAACAGAGTGGCTGCCCACA
GAACAACCCACAAAAAACGATGATCTAACGGAGGACAGCAAGTCCGCAACAACCTTTTAACAGCAG
GCTTTGCGGCCAGGAGAGAGGAGGAGAGGCAAAGAAAACCAAGCATCCTCCTTCTCCCATCTATAA
ATTCCTCCCCCCTTTTCCCCTCTCTATATAGGAGGCATCCAAGCCAAGAAGAGGGAGAGCACCAAG
GACACGCGACTAGCAGAAGCCGAGCGACCGCCTTCTCGATCCATATCTTCCGGTCGAGTTCTTGGT
CGATCTCTTCCCTCCTCCACCTCCTCCTCACAGGGTATGTGCCTCCCTTCGGTTGTTCTTGGATTT
ATTGTTCTAGGTTGTGTAGTACGGGCGTTGATGTTAGGAAAGGGGATCTGTATCTGTGATGATTCC
TGTTCTTGGATTTGGGATAGAGGGGTTCTTGATGTTGCATGTTATCGGTTCGGTTTGATTAGTAGT
ATGGTTTTCAATCGTCTGGAGAGCTCTATGGAAATGAAATGGTTTAGGGATCGGAATCTTGCGATT
TTGTGAGTACCTTTTGTTTGAGGTAAAATCAGAGCACCGGTGATTTTGCTTGGTGTAATAAAGTAC
GGTTGTTTGGTCCTCGATTCTGGTAGTGATGCTTCTCGATTTGACGAAGCTATCCTTTGTTTATTC
CCTATTGAACAAAAATAATCCAACTTTGAAGACGGTCCCGTTGATGAGATTGAATGATTGATTCTT
AAGCCTGTCCAAAATTTCGCAGCTGGCTTGTTTAGATACAGTAGTCCCCATCACGAAATTCATGGA
AACAGTTATAATCCTCAGGAACAGGGGATTCCCTGTTCTTCCGATTTGCTTTAGTCCCAGAATTTT
TTTTCCCAAATATCTTAAAAAGTCACTTTCTGGTTCAGTTCAATGAATTGATTGCTACAAATAATG
GTGCAAATCAGGTCTATATGATTGATTTTGGGCTGGCCAAGAAGTATAGAGACTCATCAACTCATC
AGCATATTCCGTATAGAGAAAACAAAAATTTGACAGGAACTGCTAGATACGCAAGCATGAATACTC
ATCTTGGCATTGAACAAAGTCGAAGGGATGATTTGGAATCGCTGGGTTATGTTTTAATGTACTTCT
TAAGAGGAAGTCTCCCTTGGCAGGGGCTGAAAGCAGGCACTAAGAAACAGAAGTATGAGAAGATCA
GTGAGAAGAAAGTATCAACATCAATAGAGACCTTGTGTAGGGGATATCCTGCAGAGTTTGCATCAT
ATTTTCATTACTGTCGATCACTAAGATTTGATGATAAACCAGATTATGCTTATCTGAAGAGAATTT
TCCGTGATCTTTTCATTCGTGAAGGGTTTCAATTTGATTATATATTTGACTGGACCATTTTGAAAT
ATCAGCAATCACAGCTTGCCAATCCTCCATCTCGTGCTCTTGGTGGTACTGCTGGGCCAAGCTCAG
GGATGCCTCATGCTCTTGTTAATGTTGAGAGGCAATCAGGTGGAGATGAAGGTCGACCAACTGGTT
GGTCTTCATCAAATCTTACACGTAATAAGAGCACGGGGCTGCATTTCAATTCTGGAAGCTTATTGA
AGCAAAAAGGCACAGTTGCTAATGATTTATCCATGGGTAAAGAGTTATCCAGTTCTAATTTTTTCC
GGTCAAGTGGACCATTGAGGCGTCCAGTTGTCTCTAGCATCCGAGACCCAGTGATTGCAGGGGGTG
AACCTGACCCCTCCGGCACTCTGACAAAAGATGCAAGCCCGGGACCATTGCGTAAAGTATCCAGTG
CTGCACGGAGGAGTTCACCAGTTGTGTCCTCAGATCACAAGCGCAGCTCCTCTATCAAAAATGCCA
ACATAAAGAATTTAGAGTCCACCGTCAAGGGAATAGAGGGTTTAAGTTTTCGATGATGAGGGACTG
CATTAGTAGCTGTGCTTTGTCTCAGTTCTCCGTTCACTGTAAATTTTGGCACACCAACTTGGGGAG
TAAGAGTTCTGATATTAGTTGCTGTCAGGAAGTACCATAAAGCTGAATTATACAATTAAAATTTGG
GATCCAATCGCAAAAGCACATTAAGGATATGATGGGGTTGCAGATCCAAACTCACAGATTCCAGTT

FIGURE 5 (continued)

TATGCTCGTCCATACAGTTATAGGCACTTTCCATATTCTTTTCTTTAATCTCTGTCTCTTGCTTGT
TATTGTTATGTCGTGGTATTCTTGTTGAGGTCATGTTTGTGAATTGCGAAGATGGTCATGTATAAT
TGCCGAGAAATCATGTACTAGTTTGTTTTAAACATGAGCAAACTGTTATTTTGTTCAAGCTACTTT
AATATCAAAAAAAAAAAAAAAAAAGGGCGGCCGCTCTAGAGTATCCCTCGAGGGGCCCAAGCTTACGC
GTACCCAGCTTT

SEQ ID NO: 65, Artificial sequence - prm06049
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAACAATGTGTGGCATCCTCGCCGTGCTCG

SEQ ID NO: 66, Artificial sequence - prm06050
GGGGACCACTTTGTACAAGAAAGCTGGGTGCGACGATAGAAAGTTAAACGGCAG

SEQ ID NO: 67, Oryza sativa - Os06g0265000#1
MCGILAVLGVADVSLAKRSRIIELSRRLRHRGPDWSGIHCYQDCYLAHQRLAIVDPTSGDQPLYNE
DKSVVVTVNGEIYNHEELKANLKSHKFQTASDCEVIAHLYEEYGEEFVDMLDGMFAFVLLDTRDKS
FIAARDAIGICPLYMGWGLDGSVWFSSEMKALSDDCERFISFPPGHLYSSKTGGLRRWYNPPWFSE
SIPSTPYNPLLLRQSFEKAIIKRLMTDVPFGVLLSGGLDSSLVASVVSRHLAEAKVAAQWGNKLHT
FCIGLKGSPDLRAAKEVADYLGTVHHELHFTVQEGIDALEEVIYHVETYDVTTIRASTPMFLMSRK
IKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLRANKSTSAWGVEARV
PFLDKNFINVAMDIDPEWKMIKRDLGRIEKWVLRNAFDDEEKPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHANEHVSDSMMMNASFVYPENTPVTKEAYYYRTIFEKFFPKNAARLTVPGGPSVACSTAKA
VEWDAAWSKNLDPSGRAALGVHDAAYEDTLQKSPASANPVLDNGFGPALGESMVKTVASATAV

SEQ ID NO: 68, Aquilegia formosa - TA8085_338618#1
MCGILAVLGCSDDSQAKRVRVLELSRRLKHRGPDWSGLYQHGDNFLSHQRLAVIDPASGDQPLYNE
DKSIVVTVNGEIYNHEALRKRLPNHKFRTGSDCDVIAHLYEEFGEDFVDMLDGMFSFVLLDTRDNS
FLVARDAIGITSLYIGWGLDGSIWISSEMKGLNDDCEHFECFPPGHLYSSKNSGFRRWYNPSWFSE
AVPSTPYDPLVLRRAFENAVVKRLMTDVPFGVLLSGGLDSSLVASITARHLAETKAAKQWGAQLHS
FCVGLEGSPDLKAGKEVADYLGTVHHEFHFTVQDGIDAIEDVIYHVETYDVTTIRASTPMFLMSRK
IKSLGVKMVISGEGSDEIFGGYLYFHKAPNKEEFHRETCHKIKALHQYDCLRANKSTSAWGLEARV
PFLDKEFINVAMAIDPEWKMIKRDQGRIEKWVLRRAFDDEDHPYLPKHILYRQKEQFSDGVGYSWI
DGLKAHAASHVTDKMMRNAKNIFLHNTPTTKEAYYYRMIFERFFPQNSAKLTVPGGPSVACSTAKA
VEWDASWSNNLDPSGRAALGVHASAYEAQLSAPLANGNVPVKIFNNVPRMVEVGAPASLTIRS

SEQ ID NO: 69, Asparagus officinalis - AOASPSYNM#1
MCGILAVLGCSDDSQAKRVRVLELSRRLKHRGPDWSGLCQHGDCFLSHQRLAIIDPASGDQPLYNE
DKSIVVTVNGEIYNHEELRRRLPDHKYRTGSDCEVIAHLYEEHGEDFVDMLDGMFSFVLLDTRNNC
FVAARDAVGITPLYIGWGLDGSVWLSSEMKGLNDDCEHFEVFPPGNLYSSRSGSFRRWYNPQWYNE
TIPSAPYDPLVLRKAFEDAVIKRLMTDVPFGVLLSGGLDSSLVAAVTARHLAGSKAAEQWGTQLHS
FCVGLEGSPDLKAAKEVAEYLGTVHHEFHFTVQDGIDAIEDVIFHIETYDVTTIRASTPMFLMARK
IKSLGVKMVISGEGSDEIFGGYLYFHKAPNKEEFHHETCRKIKALHQYDCLRANKATSAWGLEARV
PFLDKEFMDVAMSIDPESKMIKPDLGRIEKWVLRKAFDDEENPYLPKHILYRQKEQFSDGVGYSWI
DGLKAHAAKHVTDRMMLNAARIYPHNTPTTKEAYYYRMIFERFFPQNSARFTVPGGPSIACSTAKA
IEWDARWSNNLDPSGRAALGVHDSAYDPPLPSSISAGKGAAMITNKKPRIVDVATPGVVIST

FIGURE 5 (continued)

SEQ ID NO: 70, Brassica oleracea - TA5921_3712#1
MCGILALLGCSDDSQAKRVRVLELSRRLRHRGPDWSGIYQNGFNYLAHQRLAIIDPDSGDQPLFNE
DKSIVVTVNGEIYNHEELRKGLKNHKFHTGSDCDVIAHLYEEHGENFVDMLDGIFSFVLLDTRDNS
FMVARDAVGVTSLYIGWGLDGSLWVSSEMKGLHEDCEHFEAFPPGHLYSSKSGGGFKQWYNPPWFN
ESVPSTPYEPLAIRSAFEDAVIKRLMTDVPFGVLLSGGLDSSLVASITARHLAGTKAAKRWGPQLH
SFCVGLEGSPDLKAGKEVAEYLGTVHHEFHFTVQDGIDAIEDVIYHVETYDVTTIRASTPMFLMSR
KIKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKQEFHQETCRKIKALHKYDCLRANKATSAFGLEAR
VPFLDKEFINTAMSLDPESKMIKPEEGRIEKWVLRRAFDDEERPYLPKHILYRQKEQFSDGVGYSW
IDGLKAHAAENVNDKMMSKAAFIFPHNTPLTKEAYYYRMIFERFFPQNSARLTVPGGATVACSTAK
AVEWDASWSNNMDPSGRAAIGVHLSAYDGSKVALPLPAPHKAIDDIPMMMGQEVVIQT

SEQ ID NO: 71, Chlamydomonas reinhardtii - 140252#1
MCGILAVLNTTDDSQAMRSRVLALSRRQRHRGPDWSGMHQFGNNFLAHERLAIMDPASGDQPLFNE
DRTIVVTVNGEIYNYKELRQQITDACPGKKFATNSDCEVISHLYELHGEKVASMLDGFFAFVVLDT
RNNTFYAARDPIGITCMYIGWGRDGSVWLSSEMKCLKDDCTRFQQFPPGHFYNSKTGEFTRYYNPK
YFLDFEAKPQRFPSAPYDPVALRQAFEQSVEKRMMSDVPFGVLLSGGLDSSLVASIAARKIKREGS
VWGKLHSFCVGLPGSPDLKAGAQVAEFLGTDHHEFHFTVQEGIDAISEVIYHIETFDVTTIRASTP
MFLMSRKIKALGVKMVLSGEGSDEVFGGYLYFHKAPNKEEFQSETVRKIQDLYKYDCLRANKSTMA
WGVEARVPFLDRHFLDVAMEIDPAEKMIDKSKGRIEKYILRKAFDTPEDPYLPNEVLWRQKEQFSD
GVGYNWIDGLKAHADSQVSDDMMKTAAHRYPDNTPRTKEAYWYRSIFETHFPQRAAVETVPGGPSV
ACSTATAALWDATWAGKEDPSGRAVAGVHDSAYDAAAAANGEPAAKKAKK

SEQ ID NO: 72, Glycine max - TA41694_3847#1
MCGILAVLGCSDSSQAKRVRVLELSRRLKHRGPDWSGLHQYGDNYLAHQRLAIVDPASGDQPLFNE
DKTVVVTVNGEIYNHEELRKQLPNHTFRTGSDCDVIAHLYEEHGENFVDMLDGIFSFVLLDTRDNS
FIVARDAIGVTSLYIGWGLDGSVWISSELKGLNDDCEHFESFPPGHLYSSKERAFRRWYNPPWFSE
AIPSAPYDPLALRHAFEKAVVKRLMTDVPFGVLLSGGLDSSLVAAVTARYLAGTNAAKQWGTKLHS
FCVGLEGAPDLKAAKEVADYIGTVHHEFHYTVQDGIDAIEDVIYHIETYDVTTIRASIPMFLMSRK
IKSLGVKWVISGEGSDEIFGGYLYFHKAPNKEEFHQETCRKIKALHKYDCLRANKSTFAWGLEARV
PFLDKDFIRVAMNIDPDYKMIKKEEGRIEKWVLRRAFDDEEHPYLPKHILYRQKEQFSDGVGYGWI
DGLKAHAEKHVTDRMMLNAANIFPFNTPTTKEAYYYRMIFERFFPQNSARLSVPGGPSVACSTAKA
VEWDAAWSNNLDPSGRAALGVHASAYGNQVKAVEPEKIIPKMEVSPLGVAI

SEQ ID NO: 73, Glycine max - U77679#1
MCGILAVLGCSDSSQAKRVRVLELSRRLKHRGPDWSGLHQYGDNYLAHQRLAIVDPASGDQPLFNE
DKTVVVTVNGEIYNHEELRKQLPNHTFRTGSDCDVIAHLYEEHGENFMDMLDGISSFVLLDTRDNS
FIVARDAIGVTSLYIGWGLDGSVWISSELKGLNDDCEHFESFPPGHLYSSKERAFRRWYNPPWLSL
AIPSAPYDPLALRHAFEKLWIKRLMTDVPFGVLLSGGLDSSLVAAVTARYLAGTKAAKQWGTKLHS
FCVGLEGAPDLKATKEVAEYIGTVHHEFHYTVQDGIDAIEDVIYHIETYDVTTIRASIPMFLMSRK
IKSLGVKWVISGEGSDVFFGGYLYFHKAPNKEEFHQETCRTIIVLHRYDCSRANKSTFVWGLEARV
PFLDKEFIRVAMNIDPECKMIKKEEGRIEKWALRRAFDDEEHPYLPKHILYRQKEQFSDGVGYGWI
DGLKAHAEKHVTDRMMLNAANIFPFNTPTTKEAYHYRMIFERFFPQNSCRLTVPGGTSVACSTAKA
VEWDAAWSNNLDPSGRAALGVHASAYGNQVKAVEPEKIIPKMEVSPLGVAI*

SEQ ID NO: 74, Glycine max - TA41698 - 3847#1
MCGILAVLGCSDDSRAKRVRVLELSRRLKHRGPDWSGLHQHGDCFLAHQRLAIVDPASGDQPLFNE
DKSVIVTVNGEIYNHEELRKQLPNHNFRTGSDCDVIAHLYEEHGEDFVDMLDGIFSFVLLDTRDNS
FIVARDAIGVTSLYIGWGLDGSVWISSEMKGLNDDCEHFECFPPGHLYSSKERGFRRWYNPPWFSE

FIGURE 5 (continued)

AIPSAPYDPLVLRHAFEQAVIKRLMTDVPFGVLLSGGLDSSLVASITSRYLANTKAAEQWGSKLHS
FCVGLEGSPDLKAAKEVADYLGTVHHEFTFTVQDGIDAIEDVIYHIETYDVTTIRASTPMFLMSRK
IKSLGVKWVISGEGSDEIFGGYLYFHKAPNKEEFHRETCRKIKALHQYDCLRANKSTFAWGLEARV
PFLDKAFINAAMSIDPEWKMIKRDEGRIEKWILRRAFDDEEHPYLPKHILYRQKEQFSDGVGYSWI
DGLKAHAAKHVTEKMMLNAGNIYPHNTPKTKEAYYYRMIFERFFPQNSARLTVPGGASVACSTAKA
VEWDAAWSNNLDPSGRAALGVHISAYENQNNKGVEIEKIIPMDAAPLGVAIQG

SEQ ID NO: 75, Glycine max - TA51197_3847#1
MCGILAVLGCVDNSQTKRARIIELSRRLRHRGPDWSGIHCYEDCYLAHQRLAIVDPTSGDQPLYNE
DKTIIVTVNGEIYNHKQLRQKLSSHQFRTGSDCEVIAHLYEEHGEEFVNMLDGMFAFILLDTRDKS
FIAARDAIGITPLYLGWGHDGSTWFASEMKALSDDCERFISFPPGHIYSSKQGGLRRWYNPPWFSE
DIPSTPYDPTLLRETFERAVVKRMMTDVPFGVLLSGGLDSSLVAAVVNRYLAESESARQWGSQLHT
FCIGLKGSPDLKAAKEVADYLGTRHHELYFTVQEGIDALEEVIYHIETYDVTTIRASTAMFLMSRK
IKALGVKMVLSGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLRANKSTAAWGVEARV
PFLDKEFINVAMSIDPEWKMIRPDLGRIEKWVLRNAFDDDKNPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHANKQVTDATMMAANFIYPENTPTTKEGYLYRTIFEKFFPKNAAKATVPGGPSVACSTAKA
VEWDAAWSKNLDPSGRAALGIHDAAYDAVDTKIDEPKNGTL

SEQ ID NO: 76, Physcomitrella patens – 173106_estExt_fgenesh1 pg.C_4100023#1
MCGILAILGSHDASPARRDRILELSRRLRHRGPDWSGLFAGQKCWCYLAHERLAIIDPASGDQPLY
NENKDIVVAANGEIYNHEALKKSMKPHKYHTQSDCEVIAHLFEDVGEDVVNMLDGMFSFVLVDNRD
NSFIAARDPIGITPLYYGWGADGSVWFASEMKALKDDCERFEIFPPGHIYSSKAGGLRRYYNPAWF
SETFVPSTPYQSLVLRAAFEKAVIKRLMTDVPFGVLLSGGLDSSLVAAVASRHIAGTKAANIWGKQ
LHSFCVGLQGSPDLKAAREVANYIGTQHHEFHFTVQEGLDALSDVIYHVETYDVTTIRASTPMFLM
TRKIKALGVKMVLSGEGSDEIFGGYLYFHKAPNREEFHHELVRKIKALHMYDCQRANKSTSAWGLE
ARVPFLDKEFMEVAMAIDPAEKLIRKDQGRIEKWVLRKAFYDEKNPYLPKHILYRQKEQFSDGVGY
SWIDGLKAHAQSHVSDQMLKHAKHVYPYNTPQTKEAYYYRMLFEKHFPQQSARLTVPGGASVACST
ATAVAWDKSWAGNLDPSGRAALGCHDAAYTENSAAMSYITKNMSNVGQKMTIH

SEQ ID NO: 77, Physcomitrella patens - 180723_estExt_gwp gw1.C_440158#1
MCGILAILGADGAVPSAGRDRALALSRRLRHRGPDWSGLFEGKDSWCYLAHERLAIIDPASGDQPL
YNGTKDIVVAANGEIYNHELLKKNMKPHEYHTQSDCEVIAHLYEDVGEEVVNMLDGMWSFVLVDSR
DNSFIAARDPIGITPLYLGWGADGRTVWFASEMKALKDDCERLEVFPPGHIYSSKAGGLRRYYNPQ
WFSETFVPETPYQPLELRSAFEKAVVKRLMTDVPFGVLLSGGLDSSLVASVAARHLAETKAVRIWG
NELHSFCVGLEGSPDLKAAREVAKYIGTRHHEFNFTVQEGLDALSDVIYHVETYDVTTIRASTPMF
LMTRKIKALGVKMVLSGEGSDEIFGGYLYFHKAPNREEFHHELVRKIKALHLYDCQRANKSTSAWG
LEARVPFLDKEFMDVAMMIDPSEKMIRKDLGRIEKWVLRKAFDDEERPYLPKHILYRQKEQFSDGV
GYSWIDGLKEYAESHVTDQMMKHAKHVYPFNTPNTKEGYYYRMIFEKHFPQQSARMTVPGGPSVAC
STATAVAWDEAWANNLDPSGRAALGCHDSAYTDKHSEKAAPAAEANGTASHENGHTFSKPKSTLDA
TILKTQAVH

SEQ ID NO: 78, Physcomitrella patens - 226188_estExt_Genewise1.C 3500008#1
MCGILAILGCHDKSVTRRHRCLELSRRLRHRGPDWSGLFVDEASGCYLAHERLAIIDPTSGDQPLF
NENKDIVVAVNGEIYNHEALKASMKAHKYHTQSDCEVIAHLYEEIGEEVVEKLDGMFSFVLVDLRD
KSFIAARDPLGITPLYLGWGNDGSVWFASEMKALKDDCERFESFPPGHMYSSKQGGLRRYYNPPWF

FIGURE 5 (continued)

NESIPAEPYDPLILRHAFEKSVIKRLMTDVPFGVLLSGGLDSSLVAAVAQRHLAGSTAAKQWGNKL
HSFCVGLEGSPDLKAGREVADYIGTVHKEFHFTVQEGLDAISDVIYHIETYDVTTIRASTPMFLMS
RKIKALGVKMVLSGEGSDEIFGGYLYFHKAPNKEEFHKETCRKLKALHLYDCLRANKSTSAWGLEA
RVPFLDRDFVNLAMSIDPAEKMINKKEGKIEKWIIRKAFDDEENPYLPKHILYRQKEQFSDGVGYS
WIDGLKDHAASQVSDQMLANAKHIYPHNTPGTKEGYYYRMIFERCFPQESARLTVPGGPSVACSTA
AAIAWDKAWANNLDPSGRAATGVHDSAYEGGEVESSAVSHKEGGEDGLANSKVGDKVQEAIAVA

SEQ ID NO: 79, Populus trichocarpa - 722643#1
MCGILAVLGCSDDSQAKRVRVLELSRRLKHRGPDWSGLYQCGDFYLAHQRLAIIDPASGDQPLFNE
DQAIVVTVNGEIYNHEELRKRLPNHKFRTGSDCDVIAHLYEEYGENFVDMLDGMFSFVLLDTRDNS
FIVARDAIGITPLYIGWGLDGSVWISSELKGLNDDCEHFECFPPGHLYSSKSGGLRRWYNPPWFCE
AIPSTPYDPLVLRRAFEKAVIKRLMTDVPFGVLLSGGLDSSLVAAVTARHLAGTKAARQWGAQLHS
FCVGLENSPDLKAAREVADYLGTVHHEFYFTVQDGIDAIEDVIYHIETYDVTTIRASTPMFLMARK
IKALGVKMVISGEGSDEIFGGYLYFHKAPNKEELHRETCRKIKALHQYDCLRANKATSAWGLEARV
PFLDKDFINVAMAIDPEWKMIKPGQGHIEKWVLRKAFDDEEHPYLPKHILYRQKEQFSDGVGYSWI
DGLKAHAAQHVTDKMMQNAEHIFPHNTPTTKEAYYYRMIFERFFPQNSARLSVPGGASVACSTAKA
VEWDAAWSNNLDPSGRAALGVHLSDYDQQAALANAGVVPPKIIDTLPRMLEVSASGVAIHS

SEQ ID NO: 80, Populus trichocarpa - 829702#1
MCGILAVLGCSDDSQAKRFRVLELSRRLKHRGPDWSGLFQHGDFYLAHQRLAIIDPASGDQPLFNE
DQAIVVTVNGEIYNHEELRKRLPNHKFRTGSDCDVISHLYEEYGENFVDMLDGMFSFVLLDTRDNS
FIVARDAIGITSLYIGWGLDGSVWISSELKGLNDDCEHFKCFPPGHIYSSKSGGLRRWYNPLWFSE
AIPSTPYDPLALRRAFEKAVIKRLMTDVPFGVLLSGGLDSSLVAAVTARHLAGTQAARQWGAHLHS
FCVGLENSPDLKAAREVADYLGTIHHEFHFTVQDGIDAIEDVIYHVETYDVTTIRASTPMFLLARK
IKALGVKMVISGEGSDEIFGGYLYFHKAPNKEELHGETCRKIKALHQYDCLRANKATSAWGLEARV
PFLDKDFINVAMAIDPEWKMIKPGRIEKWVLRKAFDDEEHPYLPKHILYRQKEQFSDGVGYSWIDG
LKAHAELHVHDKMMQNAEHIFPHNTPTTKEAYYYRMIFERFFPQNSARLTVPGGASVACSTAKAVE
WDASWSNNLDPSGRAALGVHLSAYEQQAALASAGVVPPEIIDNLPRMMKVGAPGVAIQS

SEQ ID NO: 81, Arabidopsis thaliana - AT5G65010.1#1
MCGILAVLGCIDNSQAKRSRIIELSRRLRHRGPDWSGLHCYEDCYLAHERLAIIDPTSGDQPLYNE
DKTVAVTVNGEIYNHKILREKLKSHQFRTGSDCEVIAHLYEEHGEEFIDMLDGMFAFVLLDTRDKS
FIAARDAIGITPLYIGWGLDGSVWFASEMKALSDDCEQFMSFPPGHIYSSKQGGLRRWYNPPWYNE
QVPSTPYDPLVLRNAFEKAVIKRLMTDVPFGVLLSGGLDSSLVAAVALRHLEKSEAARQWGSQLHT
FCIGLQGSPDLKAGREVADYLGTRHHEFQFTVQDGIDAIEEVIYHIETYDVTTIRASTPMFLMSRK
IKSLGVKMVLSGEGSDEILGGYLYFHKAPNKKEFHEETCRKIKALHQFDCLRANKSTSAWGVEARV
PFLDKEFLNVAMSIDPEWKLIKPDLGRIEKWVLRNAFDDEERPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHANKHVSDTMLSNASFVFPDNTPLTKEAYYYRTIFEKFFPKSAARATVPGGPSIACSTAKA
VEWDATWSKNLDPSGRAALGVHVAAYEEDKAAAAAKAGSDLVDPLPKNGT

SEQ ID NO: 82, Arabidopsis thaliana - AT3G47340.1#1
MCGILAVLGCSDDSQAKRVRVLELSRRLRHRGPDWSGLYQNGDNYLAHQRLAVIDPASGDQPLFNE
DKTIVVTVNGEIYNHEELRKRLKNHKFRTGSDCEVIAHLYEEYGVDFVDMLDGIFSFVLLDTRDNS
FMVARDAIGVTSLYIGWGLDGSVWISSEMKGLNDDCEHFETFPPGHFYSSKLGGFKQWYNPPWFNE
SVPSTPYEPLAIRRAFENAVIKRLMTDVPFGVLLSGGLDSSLVASITARHLAGTKAAKQWGPQLHS
FCVGLEGSPDLKAGKEVAEYLGTVHHEFHFSVQDGIDAIEDVIYHVETYDVTTIRASTPMFLMSRK
IKSLGVKMVLSGEGADEIFGGYLYFHKAPNKKEFHQETCRKIKALHKYDCLRANKSTSAFGLEARV

FIGURE 5 (continued)

PFLDKDFINTAMSLDPESKMIKPEEGRIEKWVLRRAFDDEERPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHAAQNVNDKMMSNAGHIFPHNTPNTKEAYYYRMIFERFFPQNSARLTVPGGATVACSTAKA
VEWDASWSNNMDPSGRAAIGVHLSAYDGKNVALTIPPLKAIDNMPMMMGQGVVIQS

SEQ ID NO: 83, Arabidopsis thaliana - AT5G10240.1#1
MCGILAVLGCVDNSQAKRSRIIELSRRLRHRGPDWSGLHCYEDCYLAHERLAIVDPTSGDQPLYNE
DKTIAVTVNGEIYNHKALRENLKSHQFRTGSDCEVIAHLYEEHGEEFVDMLDGMFAFVLLDTRDKS
FIAARDAIGITPLYIGWGLDGSVWFASEMKALSDDCEQFMCFPPGHIYSSKQGGLRRWYNPPWFSE
VVPSTPYDPLVVRNTFEKAVIKRLMTDVPFGVLLSGGLDSSLVASVALRHLEKSEAACQWGSKLHT
FCIGLKGSPDLKAGREVADYLGTRHHELHFTVQDGIDAIEEVIYHVETYDVTTIRASTPMFLMSRK
IKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHQYDCLRANKSTSAWGVEARV
PFLDKEFINVAMSIDPEWKMIRPDLGRIEKWVLRNAFDDEKNPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHANKHVSETMLMNASFVFPDNTPLTKEAYYYRTIFEKFFPKSAARATVPGGPSVACSTAKA
VEWDAAWSQNLDPSGRAALGVHVSAYGEDKTEDSRPEKLQKLAEKTPAIV

SEQ ID NO: 84, Triticum aestivum - TA71252 - 4565#1
MCGILAVLGCGDESQGKRVHVLELSRRLKHRGPDWSGLHQVADNYLCHQRLAIIDPASGDQPLYNE
DKSIAVAVNGEVYNHEELRARLSGHRFRTGSDCEVIAHLYEEYGESFIDMLDGVFSFVLLDARDNS
FIAARDAIGVTPLYIGWGIDGSVWISSEMKGLNDDCEHFEIFPPGNLYSSKEKSFKRWYNPPWFSE
VIPSVPYDPLRLRSAFEKAVIKRLMTDVPFGVLLSGGLDSSLVAAVAARHFAGTKAAKRWGTRLHS
FCVGLEGSPDLKAAKEVADHLGTVHHEFNFTVQDGIDAIEDVIYHIETYDVTTIRASTLMFQMSRK
IKALGVKMVISGEGADEIFGGYLYFHKAPNKEEFHQETCRKIKALHQYDCLRANKATSAWGLEVRV
PFLDKEFINEAMSIDPEWKMIRPDLGRIEKWILRKAFDDEERPFLPKHILYRQKEQFSDGVGYSWI
DGLKDHAASNVSDKMMSNAKFIYPHNTPTTKEAYYYRMIFERYFPQSSAILTVPGGPSVACSTAKA
IEWDAQWSGNLDPSGRAALGVHLSAYEQDTVAVGGSNKPGVMNTVVPGVAIET

SEQ ID NO: 85, Triticum aestivum - TA54599 - 4565#1
MCGILAVLGCADDTQGKRVRVLELSRRLKHRGPDWSGMHQVGDCYLSHQRLAIIDPASGDQPLYNE
DKSIVVTVNGEIYNHEQLRAQLSSHTFRTGSDCEVIAHLYEEHGENFIDMLDGVFSFVLLDTRDNS
FIAARDAIGVTPLYIGWGIDGSVWISSEMKGLNDDCEHFEIFPPGHLYSSKQGGFKRWYNPPWFSE
VIPSVPYDPLALRKAFEKAVIKRLMTDVPFGVLLSGGLDSSLVAAVTVRHLAGTKAAKRWGTKLHS
FCVGLEGSPDLKAAKEVANYLGTMHHEFTFTVQDGIDAIEDVIYHTETYDVTTIRASTPMFLMSRK
IKSLGVKMVISGEGSDEIFGGYLYFHKAPNKEELHRETCQKIKALHQYDCLRANKATSAWGLEARV
PFLDKEFINEAMSIDPEWKMIRPDLGRIEKWMLRKAFDDEEQPFLPKHILYRQKEQFSDGVGYSWI
DGLKAHAESNVTDKMMSNAKFIYPHNTPTTKEAYCYRMIFERFFPQNSAILTVPGGPSVACSTAKA
VEWDAQWSGNLDPSGRAALGVHLSAYEQEHLPATIMAGTSKKPRMIEVAAPGVAIES

SEQ ID NO: 86, Vitis vinifera - GSVIVT00024074001#1
MCGILAVLGCSDDSQAKRVRLFYHCYLCFCDRLKHRGPDWSGLYQHGDCYLAHQRLAIIDPASGDQ
PLYNENQAIVVTVNGEIYNHEELRKSMPNHKFRTGSDCDVIAHLYEEHGENFVDMLDGMFSFVLLD
TRDDSFIVARDAIGITSLYIGWGLDGSSVWISSELKGLNDDCEHFESFPPGHMYSSKEGGFKRWYN
PPWFSEAIPSAPYDPLVLRRAFENAVIKRLMTDVPFGVLLSGGLDSSLVASITARHLAGTKAAKQW
GAQLHSFCVGLEGSPDLKAAKEVADYLGTVHHEFHFTVQDGIDAIEDVIYHIETYDVTTIRASTPM
FLMSRKIKSLGVKMVISGEGSDEIFGGYLYFHKAPNKEEFHRETCRKIKALYQYDCLRANKSTSAW
GLEARVPFLDKEFIKVAMDIDPEWKMIKPEQGRIEKWVLRRAFDDEEQPYLPKHILYRQKEQFSDG
VGYSWIDGLKAHASQHVTDKMMLNASHIFPHNTPTTKEAYYYRMIFERFFPQNSARLTVPGGASVA
CSTAKAVEWDSAWSNNLDPSGRAALGVHLSAYDQKLTTVSAANVPTKIIDNMPRIMEVTAP

FIGURE 5 (continued)

SEQ ID NO: 87, Volvox carteri - 65699 - e - gw1.50.7.1#1
MCGILAVLNSTDDSPAMRAKVLALSRRQKHRGPDWSGMHQFGNNFLAHERLAIMDPSSGDQPLYNE
DKSIVVTVNGEIYNYKELRKEISDKCPGKKFRTNSDCEVISHLYELYGEAVANKLDGFFAFVLLDT
RNNTFFAARDPLGVTCMYIGWGRDGSVWLSSEMKCLKDDCARFQQFPPGHYYSSKTGEFVRYFNPQ
FYLDFEAEPQVFPSVPYDPVTLRTAFEAAVEKRMMSDVPFGVLLSGGLDSSLVASIAARKIKREGS
VWGKLHSFCVGLEGSPDLKAGAAVAEFLGTDHHEFHFTVQEGIDAISEVIYHIETFDVTTIRASTP
MFLMSRKIKALGVKMVLSGEGSDEVFGGYLYFHKAPSKDEFHSETVRKLKDLFKYDCLRANKATMA
WGVEARVPFLDRAFLDVAMSIDPAEKMIDKSKGRIEKYILRKAFDTPEDPYLPKEVLWRQKEQFSD
GVGYNWIDGLKAHAESQVSDEMLKNAVHRFPDNTPRTKEAYWYRSIFESHFPQRAAMETVPGGPSV
ACSTATAALWDAAWAGKEDPSGRAVAGVHDAAYEEGAEANGEPASKKQKV

SEQ ID NO: 88, Zea mays - TA174465 - 4577#1
MCGILAVLGCSDWSQAKRARILACSRRLKHRGPDWSGLYQHEGNFLAQQRLAVVSPLSGDQPLFNE
DRTVVVVANGEIYNHKNVRKQFTGTHNFSTGSDCEVIIPLYEKYGENFVDMLDGVFAFVLYDTRDR
TYVAARDAIGVNPLYIGWGSDGSVWIASEMKALNEDCVRFEIFPPGHLYSSAGGGFRRWYTPHWFQ
EQVPRMPYQPLVLREAFEKAVIKRLMTDVPFGVLLSGGLDSSLVASVTKRHLVETEAAEKFGTELH
SFVVGLEGSPDLKAAREVADYLGTIHHEFHFTVQDGIDAIEEVIYHDETYDVTTIRASTPMFLMAR
KIKSLGVKMVLSGEGSDELLGGYLYFHFAPNKEEFHRETCRKVKALHQYDCLRANKATSAWGLEVR
VPFLDKEFINVAMGMDPEWKMYDKNLGRIEKWVMRKAFDDDEHPYLPKHILYRQKEQFSDGVGYNW
IDGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMIFERLFPQDSARETVPWGPSIACSTPA
AIEWVEQWKASNDPSGRFISSHDSAATDHTGGKPAVANGGGHGAANGTVNGKDVAVAIAV

SEQ ID NO: 89, Zea mays - X82849#1
MCGILAVLGVVEVSLAKRSRIIELSRRLRHRGPDWSGLHCHEDCYLAHQRLAIIDPTSGDQPLYNE
DKTVVVTVNGEIYNHEELKAKLKTHEFQTGSDCEVIAHLYEEYGEEFVDMLDGMFSFVLLDTRDKS
FIAARDAIGICPLYMGWGLDGSVWFSSEMKALSDDCERFITFPPGHLYSSKTGGLRRWYNPPWFSE
TVPSTPYNALFLREMFEKAVIKRLMTDVPFGVLLSGGLDSSLVASVASRHLNETKVDRQWGNKLHT
FCIGLKGSPDLKAAREVADYLSTVHHEFHFTVQEGIDALEEVIYHIETYDVTTIRASTPMFLMSRK
IKSLGVKMVISGEGSDEIFGGYLYFHKAPNKKEFLEETCRKIKALHLYDCLRANKATSAWGVEARV
PFLDKSFISVAMDIDPEWNMIKRDLGRIEKWVMRKAFDDDEHPYLPKHILYRQKEQFSDGVGYNWI
DGLKSFTEQQVTDEMMNNAAQMFPYNTPVNKEAYYYRMIFERLFPQDSARETVPWGPSIACSTPAA
IEWVEQWKASNDPSGRFISSHDSAATDHTAVSRRWPTAAARPANGTVNGKDVPVPIAV

SEQ ID NO: 90, Zea mays - TA182904 - 4577#1
MCGILAVLGCADEAKGSSKRSRVLELSRRLKHRGPDWSGLRQVGDCYLSHQRLAIIDPASGDQPLY
NEDQSVVVAVNGEIYNHLDLRSRLAGAGHSFRTGSDCEVIAHLYEEHGEEFVDMLDGVFSFVLLDT
RHGDRAGSSFFMAARDAIGVTPLYIGWGVDGSVWISSEMKALHDECEHFEIFPPGHLYSSNTGGFS
RWYNPPWYDDDDDEEAVVTPSVPYDPLALRKAFEKAVVKRLMTDVPFGVLLSGGLDSSLVATVAVR
HLARTEAARRWGTKLHSFCVGLEGSPDLKAAREVAEYLGTLHHEFHFTVQDGIDAIEDVIYHTETY
DVTTIRASTPMFLMSRKIKSLGVKMVISGEGSDELFGGYLYFHKAPNKEELHRETCRKVKALHQYD
CLRANKATSAWGLEARVPFLDKEFINAAMSIDPEWKMVQPDLGRIEKWVLRKAFDDEEQPFLPKHI
LYRQKEQFSDGVGYSWIDGLKAHATSNVTDKMLSNAKFIFPHNTPTTKEAYYYRMVFERFFPQKSA
ILTVPGGPSVACSTAKAIEWDAQWSGNLDPSGRAALGVHLAAYEHQHDPEHVPAAIAAGSGKKPRT
IRVAPPGVAIEG

FIGURE 5 (continued)

SEQ ID NO: 91, Zea mays - TA11549 - 4577999#1
MCGILAVLGCSDCSQARRARILACSRRLKHRGPDWSGLYQHEGNFLAQQRLAIVSPLSGDQPLFNE
DRTVVVVANGEIYNHKNVRKQFTGAHSFSTGSDCEVIIPLYEKYGENFVDMLDGVFAFVLYDTRDR
TYVAARDAIGVNPLYIGWGSDGSVWMSSEMKALNEDCVRFEIFPPGHLYSSAAGGFRRWYTPHWFQ
EQVPRTPYQPLVLREAFEKAVIKRLMTDVPFGVLLSGGLDSSLVASVTKRHLVKTDAAGKFGTELH
SFVVGLEGSPDLKAAREVADYLGTTHHEFHFTVQDGIDAIEEVIYHDETYDVTTIRASTPMFLMAR
KIKSLGVKMVLSGEGSDELLGGYLYFHFAPNREELHRETCRKVKALHQYDCLRANKATSAWGLEVR
VPFLDKEFVDVAMGMDPEWKMYDKNLGRIEKWVLRKAFDDEEHPYLPEHILYRQKEQFSDGVGYNW
IDGLKAFTEQQVDGRRRS*LTSADVPPHVQVTDEMMNSAAQMFPYNTPVNKEAYYYRMIFERLFPQ
DSARETVPWGPSIACSTPAAIEWVEQWKASNDPSGRFISSHDSAATDRTGDKLAVVNGDGHGAANG
TVNGNDVAVAIAV

SEQ ID NO: 92, Zea mays - TA15078 - 4577999#1
MCGILAVLGVAEVSLAKRSRIIELSRRLRHRGPDWSGLHCHEDCYLAHQRLAIIDPTSGDQPLYNE
DKTVVVTVNGEIYNHEELKAKLKTHEFQTGSDCEVIAHLYEEYGEEFVDMLDGMFSFVLLDTRDKS
FIAARDAIGICPLYMGWGLDGSVWFSSEMKALSDDCERFITFPPGHLYSSKTGGLRRWYNPPWFSE
TVPSTPYNALFLREMFEKAVIKRLMTDVPFGVLLSGGLDSSLVASVASRHFNETKGDRQWGNKLHT
FCIGLKGSPDLKAAREVADYLSTVHHEFHFTVQEGIDALEEVIYHIETYDVTTIRASTPMFLMSRK
IKSLGVKMVISGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLRANKATSAWGVEARV
PFLDKSFISVAMDIDPDWKMIKRDLGRIEKWVIRNAFDDDERPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHASQHVSDSMMMNAGFVYPENTPTTKEGYYYRMIFEKFFPKPAARSTVPGGPSVACSTAKA
VEWDASWSKNLDPSGRAALGVHDAAYEDTAGKTPASADPVSDKGLRPAIGESLGTPVASATAV

SEQ ID NO: 93, Brassica napus - P3_BPS4LI_BN06M@BN06MC14360 43814276@14314#1
MCGILAVLGCVDNSQATRSRIIKLSRRLRHRGPDWSGLHCYEDCYLAHERLAIIDPISGDQPLYSE
DKTVVVTVNGEIYNHKALRESESLKSHKYHTGSDCEVLAHLYEEHGEEFINMLDGMFAFVLLDTKD
KSYIAVRDAIGVIPLYIGWGLDGSVWFASEMKALSDDCEQFMAFPPGHIYSSKQGGLRRWYNPPWF
SELVPSTPYDPLVLRDTFEKAVIKRLMTDVPFGVLLSGGLDSSLVASVAIRHLEKSDARQWGSKLH
TFCIGLKGSPDLKAGKEVADYLGTRHHELHFTVQEGIDAIEEVIYHVETYDVTTIRASTPMFLMSR
KIKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKKELHEETCRKIKALYQYDCLRANKSTSAWGVEAR
VPFLDKAFLDVAMGIDPEWKMIRPDLGRIEKWVLRNAFDDEKNPYLPKHILYRQKEQFSDGVGYSW
IDGLKDHANKHVSDAMLTNANFVFPENTPLTKEAYYYRAIFEKFFPKSAARATVPGGPSVACSTAK
AVEWDAAWKGNLDPSGRAALGVHVAAYEGDKAEDPRPEKVQKLAEKTAEAIV

SEQ ID NO: 94, Triticum aestivum - BPS_Hyseq_TA Wheat@c54713691@13255#1
MCGILAVLGVGDVSLAKRSRIIELSRRLRHRGPDWSGIHSFEDCYLAHQRLAIVDPTSGDQPLYNE
DKTVVVTVNGEIYNHEELKAKLKSHQFQTGSDCEVIAHLYEEYGEEFVDMLDGMFSFVLLDTRDKS
FIAARDAIGICPLYMGWGLDGSVWFSSEMKALSDDCERFISFPPGHLYSSKTGGLRRWYNPPWFSE
SIPSAPYDPLLIRESIEKAAIKRLMTDVTFGVLLSGGLDSSLVASVVSRYLAETKVARQWRNKLHT
FCIGMKGSPDLKAAKEVADYLGTVHHELHFTVQEGIDALEEVIYHIETYDVTTIRASTPMFLMSRK
IKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKKELHEETCRKIKALHLYDCLRANKATSAWGLEARV
PFLDKNFINVAMDLDPECKMIRRDLGRIEKWVLRNAFDDEEKPYLPKHILYRQKEQFSDGVGYSWI
DGLKDHAKAHVSDSMMTNASFVYPENTPTTKEAYYYRTVFEKFYPKNAARLTVPGGPSIACSTAKA
VEWDAAWSKLLDPSGRAALGVHDAAYKEKAPASVDPAVDNVSRSPAHDVKRLKTAISAAAV

FIGURE 5 (continued)

SEQ ID NO: 95, Oryza sativa - Os06g0265000#1
GCGGATTCCATTCTCCTCTTGGCATCACGAGGCGGCGCCGCTTGGTCTAGCTAGTAGCCACAGGGA
GAGGTGGTAGCCGCAGCCGCCGCCGACGAGACCTCGCCGCCGGGGGGAGGGCACCATGTGTGGCAT
CCTCGCCGTGCTCGGCGTCGCAGACGTCTCCCTCGCCAAGCGCTCCCGCATCATCGAGCTATCCCG
CCGGTTACGTCATAGAGGCCCTGATTGGAGTGGTATACACTGCTATCAGGATTGCTATCTTGCACA
CCAGCGGTTGGCTATTGTTGATCCCACATCCGGAGACCAGCCGTTGTACAATGAGGACAAATCTGT
TGTTGTGACGGTGAATGGAGAGATCTATAACCATGAAGAATTGAAAGCTAACCTGAAATCTCATAA
ATTCCAAACTGCTAGCGATTGTGAAGTTATTGCTCATCTGTATGAGGAATATGGGGAGGAATTTGT
GGATATGTTGGATGGGATGTTCGCTTTTGTTCTTCTTGACACACGTGATAAAAGCTTCATTGCAGC
CCGTGATGCTATTGGCATTTGTCCTTTATACATGGGCTGGGGTCTTGATGGTTCGGTTTGGTTTTC
GTCAGAGATGAAGGCATTAAGTGATGATTGCGAGCGATTCATATCCTTCCCCCCTGGGCACTTGTA
CTCCAGCAAAACAGGTGGCCTAAGGAGATGGTACAACCCACCATGGTTTTCTGAAAGCATTCCCTC
CACCCCGTACAATCCTCTTCTTCTCCGACAGAGCTTTGAGAAGGCTATTATTAAGAGGCTAATGAC
AGATGTGCCATTTGGTGTTCTCTTGTCTGGTGGACTGGACTCTTCTTTGGTTGCATCTGTTGTTTC
GCGGCACTTGGCAGAGGCAAAAGTTGCCGCACAGTGGGGAAACAAACTGCATACATTTTGCATTGG
TTTGAAAGGTTCTCCTGATCTTAGAGCTGCTAAGGAAGTTGCAGACTACCTTGGTACTGTTCATCA
CGAACTCCACTTCACAGTGCAGGAAGGCATTGATGCACTGGAGGAAGTCATTTACCATGTTGAGAC
ATATGATGTAACGACAATTAGAGCAAGCACCCCAATGTTCTTGATGTCACGTAAAATTAAATCTTT
GGGGGTGAAGATGGTTCTTTCGGGAGAAGGTTCTGATGAAATATTTGGCGGTTACCTTTATTTTCA
CAAGGCACCAAACAAGAAGGAATTCCATGAGGAAACATGTCGGAAGATAAAAGCCCTTCATTTATA
TGATTGCTTGAGAGCGAACAAATCAACTTCTGCATGGGGTGTTGAGGCCCGTGTTCCGTTCCTTGA
CAAAAACTTCATCAATGTAGCTATGGACATTGATCCTGAATGGAAAATGATAAAACGTGATCTTGG
CCGTATTGAGAAATGGGTTCTCCGGAATGCATTTGATGATGAGGAGAAGCCCTATTTACCTAAGCA
CATTCTATACAGGCAAAAGGAGCAATTCAGTGATGGTGTTGGGTACAGTTGGATTGATGGATTGAA
GGATCATGCAAATGAACATGTATCAGATTCCATGATGATGAACGCTAGCTTTGTTTACCCAGAAAA
CACTCCAGTTACAAAAGAAGCGTACTATTATAGGACAATATTCGAGAAATTCTTTCCCAAGAATGC
TGCTAGGTTGACAGTACCTGGAGGTCCTAGCGTCGCGTGCAGCACTGCTAAAGCTGTTGAATGGGA
CGCAGCCTGGTCCAAAAACCTTGATCCATCTGGTCGTGCTGCTCTTGGTGTTCATGATGCTGCATA
TGAAGATACTCTACAAAAATCTCCTGCCTCTGCCAATCCTGTCTTGGATAACGGCTTTGGTCCAGC
CCTTGGGGAAAGCATGGTCAAAACCGTTGCTTCAGCCACTGCCGTTTAACTTTCTATCGTCGCATA
AAACTCCGTAGTTTGTTGTTCTTGGTTCAATCCCAGCTTCTTTCAGATGTCGTTAGTTTCTTCAAA
CATGTAATGGAGATGCGTGCTTTTCCTGGCTTGTTAGTTACTGTATGCTTGTCATCGTGTATGTTT
TCTTTTCTTTTCCAATATGCAAACTGTTTGGTCGTGGACTGATCAGAACATTGTAAATATGAATAA
CCGCGACTGATATCCTCAAGTTGCTTTTGGTTTGCAATAGTTCTAATCTTGATGTTCTGCTGGGAA
TCGGAAGATGTTATGCAGTATGCGTATTGTTGGGGTGTAACCGTGTAAGTGCATCTGAAATGAAGT
TACGGGCGATGGTAACTGGG

SEQ ID NO: 96, Aquilegia formosa - TA8085_338618#1
CAAGTGATTAAATCATCCACATTTCTTCTTTCTTTCTTTCTTTCTTTTTTTTTCTTTTTTTTCTTTG
TTCTCCTTGTTTATAGAATCTTATTATTTATTAACAGAGCAAAAGCATTTCTCTAGCTAGCTAGCT
TTATTTCTTTGTGATCATCAATCAATAAATATATATAATTCATCATCATGTGTGGAATTCTAGCTG
TTTTGGGTTGTTCTGATGATTCTCAAGCCAAAAGAGTTCGTGTTCTTGAGCTTTCTCGCAGATTGA
AGCACCGTGGGCCTGATTGGAGTGGTCTGTATCAGCATGGTGACAATTTTCTATCTCATCAAAGGC
TTGCAGTCATTGATCCTGCTTCTGGGGATCAGCCTCTTTATAATGAAGACAAATCAATTGTCGTAA
CTGTGAATGGAGAAATTTATAACCATGAAGCCTTGAGGAAGCGCTTGCCAAATCACAAATTTCGAA
CTGGAAGTGACTGTGATGTTATTGCTCATCTGTATGAAGAATTCGGGGAGGATTTTGTTGACATGT
TGGACGGGATGTTCTCATTTGTTTTATTGGACACCCGCGATAACAGCTTCCTTGTCGCCCGGGATG
CCATTGGGATTACCTCCCTTTATATTGGTTGGGGACTTGATGGTTCAATTTGGATTTCATCTGAGA

FIGURE 5 (continued)

TGAAAGGACTAAATGATGACTGTGAACACTTTGAATGCTTTCCTCCTGGTCACCTTTACTCGAGCA
AAAATAGTGGTTTTCGTAGGTGGTACAATCCCTCATGGTTCTCAGAAGCTGTTCCATCTACACCAT
ATGATCCACTCGTCCTCAGACGTGCATTTGAAAATGCTGTAGTTAAGAGGCTAATGACTGATGTAC
CATTTGGAGTTCTCCTATCTGGTGGCCTTGATTCATCATTAGTTGCCTCCATCACGGCACGCCACT
TGGCAGAGACAAAGGCTGCCAAGCAATGGGGGGCACAACTTCATTCCTTCTGTGTTGGTCTGGAGG
GCTCACCTGATTTAAAGGCTGGAAAAGAGGTTGCCGATTATTTGGGTACCGTTCACCATGAGTTTC
ACTTCACTGTTCAGGATGGTATCGATGCCATTGAAGATGTGATTTACCATGTAGAAACATATGATG
TAACGACTATCCGGGCGAGCACACCTATGTTTCTTATGTCTCGCAAGATCAAGTCACTAGGAGTGA
AGATGGTTATCTCTGGAGAAGGCTCCGATGAAATATTTGGTGGGTACTTATATTTCCACAAGGCTC
CTAACAAGGAGGAGTTTCATCGCGAGACATGTCATAAGATAAAGGCTCTTCATCAGTATGATTGCT
TGAGAGCTAATAAATCGACCTCTGCTTGGGGTCTGGAAGCTCGGGTGCCATTCTTAGACAAAGAAT
TCATCAATGTTGCAATGGCTATTGACCCTGAATGGAAGATGATTAAACGTGATCAAGGCCGTATTG
AAAAGTGGGTACTCAGGAGGGCTTTTGATGATGAGGACCACCCCTACCTGCCAAAGCACATTCTCT
ACAGGCAGAAAGAACAATTTAGTGATGGTGTTGGATATAGTTGGATCGATGGACTCAAGGCCCACG
CTGCATCACATGTTACGGATAAGATGATGCGCAATGCCAAGAACATTTTCCTACACAACACACCAA
CTACCAAAGAAGCCTACTACTACAGAATGATTTTTGAGAGGTTTTTCCCTCAGAACTCGGCAAAAT
TAACAGTTCCAGGTGGTCCAAGTGTTGCTTGCAGCACTGCCAAGGCTGTCGAATGGGATGCTTCTT
GGTCAAATAATTTGGACCCTTCTGGCAGGGCTGCATTAGGTGTCCATGCTTCAGCATATGAAGCAC
AACTGTCTGCTCCTCTTGCTAATGGTAATGTTCCAGTTAAGATTTTTAACAATGTACCAAGAATGG
TTGAAGTAGGTGCTCCAGCTAGCCTCACGATCCGCAGCTAATATTTCTGGTGAATGTGCCTTATTT
TGTATGGATTTGAAGTTAAGAGGCCATAGTATGCAAGGTTCTTTTTTTTTCTTTTTTTTTTTCAGT
GTGCAGTGTGTATATGTACTAGTAGTCCATATGTGAAGGAAGATGAAACAAAACTATGTAAAAGTC
CATGTCTTTTATATTTCTGAAAAAAGAAGGTTCTTGTGATTTCTTTTTTGCTACAAATAGGCATAA
AATAGCTGATTCCATGTATCGGGCACCCCTGGCAAACACCAATGTATGCAGTCTGCATAGCGTTGT
GGATCAGCCTTCTGCTCATCGGTCAACACTTTCCCTTGTTGTTCTGTGTAAACTGATGTATGTGCA
TCAATCCGATATTCAGATATTT

SEQ ID NO: 97, Asparagus officinalis - AOASPSYNM#1

TCTGCTTGCACCTTTTGAGAGAGAGGGAGAGAGAGAGAGAGAGAGAGAGGATCATGTGTGGGAT
ACTTGCAGTGCTCGGTTGCTCCGATGACTCTCAGGCGAAGAGGGTTCGAGTTCTCGAGCTCTCTCG
CAGGTTGAAGCACAGGGGCCCAGATTGGAGCGGGCTTTGCCAACATGGAGATTGTTTCTTGTCTCA
TCAGAGATTGGCGATCATTGATCCCGCCTCTGGTGATCAACCCCTGTACAACGAGGACAAGTCCAT
CGTTGTCACGGTAAACGGAGAGATTTACAACCACGAAGAGCTAAGGCGACGCCTGCCTGATCATAA
ATACAGAACTGGAAGCGACTGTGAAGTCATCGCTCATCTGTATGAGGAACACGGAGAAGATTTCGT
CGATATGTTGGATGGAATGTTCTCCTTCGTTCTATTGGACACCCGAAACAATTGCTTCGTTGCGGC
AAGGGATGCAGTGGGAATAACCCCCCTCTACATTGGCTGGGGATTAGACGGCTCTGTTTGGCTCTC
GTCGGAAATGAAAGGATTAAACGATGACTGCGAACATTTTGAAGTATTTCCACCTGGAAACCTGTA
CTCAAGCAGATCAGGCAGCTTCAGAAGATGGTATAATCCTCAGTGGTACAATGAGACTATCCCTTC
GGCCCCCTATGATCCTCTTGTTCTGAGGAAAGCTTTTGAGGATGCTGTTATAAAGAGGCTGATGAC
TGATGTGCCATTTGGGGTTCTGTTATCTGGTGGCCTCGATTCCTCGTTGGTCGCCGCTGTTACTGC
TCGGCATCTTGCAGGAAGTAAAGCTGCAGAGCAATGGGGAACTCAGCTCCATTCTTTCTGTGTTGG
CTTAGAGGGATCACCAGATCTCAAGGCTGCAAAAGAGGTTGCAGAGTATCTGGGTACTGTCCACCA
TGAGTTTCACTTCACAGTTCAGGATGGAATTGATGCCATTGAGGATGTAATCTTCCACATTGAAAC
GTACGATGTGACAACAATCAGGGCAAGCACTCCAATGTTCCTCATGGCCAGAAAAATCAAGTCCTT
AGGAGTAAAAATGGTGATCTCAGGCGAAGGCTCGGATGAAATCTTTGGCGGGTACTTGTATTTTCA
CAAAGCACCTAACAAAGAAGAATTCCATCACGAAACATGTCGAAAGATCAAAGCTCTGCATCAGTA
TGACTGCCTCAGAGCCAACAAAGCAACATCAGCATGGGGGCTGGAAGCTCGAGTGCCATTTTTAGA
CAAGGAGTTCATGGATGTTGCTATGAGTATAGATCCTGAATCGAAAATGATTAAGCCTGATCTCGG

FIGURE 5 (continued)

GAGGATCGAGAAGTGGGTACTGAGGAAAGCTTTTGATGATGAAGAGAATCCCTATCTTCCAAAGCA
TATTCTCTATAGGCAAAAGGAGCAGTTCAGTGATGGTGTTGGATATAGTTGGATTGATGGGCTGAA
GGCTCATGCTGCAAAACATGTAACTGATAGAATGATGCTGAATGCAGCACGTATTTACCCCCACAA
CACACCAACCACAAAAGAGGCTTATTACTACAGAATGATCTTTGAAAGGTTCTTCCCTCAGAACTC
GGCGAGATTTACTGTCCCTGGAGGTCCAAGCATTGCTTGCAGCACGGCGAAGGCTATCGAATGGGA
CGCTCGCTGGTCGAACAATTTGGATCCGTCGGGGAGAGCAGCTCTCGGCGTCCATGACTCTGCCTA
CGATCCTCCTCTTCCTTCTTCGATTTCTGCAGGAAAAGGAGCTGCAATGATCACTAACAAGAAGCC
GAGGATTGTGGATGTAGCAACTCCGGGAGTTGTTATTAGTACCTGATGTTGGTTTGGTTTGGTTTG
GTTTTGATGTACAAGTTAAAATAAATGTGTGGGGCGTTGTATTTTGGATGGAGGGTACTAAAGCGT
GTAATTTGCTG

SEQ ID NO: 98, Brassica oleracea - TA5921 - 3712#1

TTGCGATTAAATAAGAAAAATGTGTGGAATACTTGCCCTTTTAGGATGCTCCGACGATTCTCAGGC
CAAGAGAGTACGCGTTCTTGAGCTTTCTCGCAGATTGAGGCACAGAGGACCTGATTGGAGCGGAAT
ATATCAGAACGGGTTCAATTACTTGGCCCATCAACGTCTTGCTATCATCGATCCTGATTCCGGTGA
TCAACCTCTCTTTAACGAGGACAAGTCCATTGTTGTCACGGTGAACGGAGAGATTTATAACCATGA
GGAGCTGAGAAAGGGTTTGAAGAATCACAAGTTCCACACCGGTAGTGATTGTGACGTCATAGCTCA
CCTGTACGAGGAGCATGGTGAGAATTTTGTGGACATGTTGGATGGAATCTTCTCCTTTGTGTTGCT
GGACACAAGAGATAACTCATTCATGGTTGCTCGTGACGCGGTTGGTGTCACTTCGCTCTACATTGG
TTGGGGATTAGATGGATCTCTGTGGGTCTCTTCCGAGATGAAAGGCTTACACGAAGATTGTGAGCA
TTTCGAAGCCTTTCCTCCAGGTCATTTGTATTCAAGCAAATCAGGAGGAGGGTTTAAGCAATGGTA
CAATCCTCCTTGGTTCAATGAATCTGTTCCTTCTACGCCTTATGAGCCTCTCGCAATTAGAAGCGC
CTTTGAAGACGCTGTGATAAAGCGGTTGATGACTGATGTCCCATTTGGAGTTTTGCTATCTGGTGG
TCTTGATTCTTCTCTTGTTGCATCCATCACTGCCCGTCACTTGGCCGGTACTAAGGCCGCTAAGCG
ATGGGGTCCTCAGCTCCATTCCTTTTGTGTCGGTCTTGAGGGCTCGCCGGACTTGAAGGCGGGGAA
AGAAGTGGCGGAGTATTTGGGGACGGTGCACCATGAGTTCCATTTCACGGTGCAAGACGGGATTGA
TGCGATTGAGGATGTGATCTACCATGTCGAGACATATGATGTGACGACAATTAGAGCTAGCACACC
CATGTTCTTGATGTCCAGGAAAATCAAGTCTCTAGGTGTTAAGATGGTTCTTTCCGGTGAAGGTTC
TGATGAGATCTTTGGAGGGTATCTTTACTTCCACAAGGCACCTAACAAGCAAGAATTTCACCAAGA
AACTTGTCGCAAGATCAAGGCTCTTCACAAATACGATTGTTTAAGAGCCAACAAAGCTACCTCTGC
TTTTGGTCTAGAGGCGCGTGTTCCTTTTCTGGACAAGGAGTTTATCAACACCGCTATGTCTCTCGA
CCCTGAATCCAAGATGATCAAACCAGAGGAAGGGAGGATCGAGAAGTGGGTTCTAAGGAGAGCCTT
TGATGATGAAGAACGTCCTTATTTGCCAAAACACATTCTCTACAGACAGAAAGAGCAGTTTAGTGA
TGGTGTTGGCTACAGCTGGATCGATGGCCTCAAAGCCCACGCTGCTGAAAATGTTAATGACAAGAT
GATGTCGAAAGCTGCTTTTATCTTCCCTCACAACACCCCACTCACCAAAGAAGCATACTATTACAG
AATGATCTTTGAGAGGTTCTTCCCACAGAACTCGGCAAGGCTAACTGTTCCCGGAGGTGCGACCGT
GGCTTGCTCGACCGCAAAAGCGGTGGAGTGGGATGCAAGCTGGTCCAACAATATGGATCCATCTGG
AAGAGCTGCGATTGGAGTTCACCTCTCGGCCTACGACGGCAGCAAAGTGGCATTGCCCTTGCCGGC
GCCACATAAGGCAATCGACGACATCCCAATGATGATGGGACAAGAAGTTGTGATTCAGACATGAGT
TTGAAGGATATATAGGGGAATTGGAGTTCTTTAAAGTTGTCCTAATGGGTTTAAGTGTTTTTGTAT
GATTTCAAAATAAAATTGGTTTCGTGTTCTTAGGGAAATATGAATGCATAAATTATTTTTCTTGTA
CTATTAGTAAATATTCGAATGTACTGTTTCTGCAAAATCGATGTACATCAATCTTATTATAATTAT
ATGTATTGTAATATGATATGAAAAATGTGATTTTGCTTGTTTTCAC

FIGURE 5 (continued)

SEQ ID NO: 99, Chlamydomonas reinhardtii - 140252#1

CCCCTCCCGCTCCCTCCCCGACATATGATCCAGCATTGATGGGTGATACAGACGAAGCGCAGAAGC
AGCAATCCGGTGTGTACGCATATGGGCACGGCAGCAGCTGCTGGCAGCCCCGGACGAAATCCCTAG
CTGCACTTTCGGGCCCGCGCCAGTCCCTTCCAAGCGCTTTGTGACGTCTTCTGGCTACTTACTTGC
TCAGCGTATCGCGCACGCCCGGCTGCGCCCCGCTTTGCCCTTGCGCCACTTCCGCACGAAGGGTCT
GCACCTTCTCCAGGTCATCCGCTGCATCGTCTGCTTCCCTGTCCGAGTACGTTGCCCTTATATAAG
TCAGCAGCGGTGTTTTGATGTCCACAGTCTCCGTCTTCTTGCAATGTATCGCTAACATAACCGATT
GAGCGGTCGGCATTTTTCAAGAGGCCCTTCGTGAGCGTGCCTTGCTAGATCTGGCTAGAGGTTGCA
GCGCGGGTGTGAAAACGCAGTGAGGGTTTGGTTGAATCGACATGCAGCCCCGTGCGCCCATGCAAC
TGTCTTTCCGCGCGCAGCAGGGCCGATGGATTCCTTTCCTTTACGCCCAAACTACGCTGGGCACAC
ACATCTTTTTGGGTAGGGCTCTTACGGTAGCCAAATTCTTATAGAGTTTGGGGAGTGCGGGTAGCA
CTCAAAAATGTGCGGCATTCTTGCCGTCCTCAACACGACGGATGACAGCCAGGCTATGCGCTCGAG
GGTGCTGGCCCTGAGCCGTCGCCAGCGTCACCGTGGCCCCGACTGGTCTGGCATGCACCAGTTCGG
CAACAACTTCCTTGCCCATGAGCGCCTTGCGATTATGGACCCCGCCTCGGGTGACCAGCCCCTGTT
CAACGAGGACCGCACAATCGTGGTCACCGTGAACGGTGAGATCTACAACTACAAGGAGCTGCGCCA
GCAGATCACGGATGCCTGCCCCGGCAAGAAGTTCGCCACCAACAGCGATTGCGAGGTGATTAGCCA
CCTGTACGAGCTGCACGGCGAGAAGGTGGCCTCCATGCTGGACGGCTTCTTCGCCTTCGTGGTGCT
GGACACCCGCAACAACACCTTCTACGCCGCGCGCGACCCCATTGGCATCACCTGCATGTACATCGG
CTGGGGCCGTGACGGCAGCGTGTGGCTGTCGAGCGAGATGAAGTGCCTGAAGGATGACTGCACCCG
CTTCCAGCAGTTCCCTCCCGGGCACTTCTACAACTCCAAGACGGGTGAGTTCACCCGCTACTACAA
CCCCAAGTACTTCCTGGACTTCGAGGCCAAGCCGCAGCGTTTCCCCAGCGCTCCCTACGACCCCGT
CGCGCTGCGTCAGGCGTTCGAGCAGTCCGTGGAGAAGCGCATGATGTCGGATGTGCCGTTCGGCGT
GCTGCTGTCGGGCGGCCTGGACAGCTCGCTGGTGGCGTCCATCGCGGCGCGCAAGATTAAGCGTGA
GGGCAGCGTGTGGGGCAAGCTGCACAGCTTCTGCGTGGGCCTGCCCGGCAGCCCTGACCTGAAGGC
TGGCGCCCAGGTGGCTGAGTTCCTGGGCACCGACCACCACGAGTTCCACTTCACGGTGCAGGAGGG
CATTGACGCCATCAGCGAGGTCATCTACCACATTGAGACCTTTGACGTCACCACCATCCGCGCCTC
CACGCCCATGTTCCTGATGAGCCGCAAGATCAAGGCGCTGGGCGTGAAGATGGTGCTGTCAGGCGA
GGGTTCCGACGAGGTGTTCGGCGGCTACCTGTACTTCCACAAGGCGCCCAACAAGGAGGAGTTCCA
GTCGGAGACTGTGCGCAAGATCCAGGACCTGTACAAGTACGACTGCCTGCGCGCCAACAAGTCCAC
CATGGCTTGGGGCGTGGAGGCGCGCGTGCCGTTCCTGGACCGCCACTTCCTGGACGTGGCCATGGA
GATCGACCCCGCCGAGAAGATGATTGACAAGAGCAAGGGCCGCATCGAGAAGTACATCCTCCGGAA
AGCCTTCGATACCCCCGAGGACCCCTACCTGCCCAACGAGGTGCTCTGGCGCCAGAAGGAGCAGTT
CAGCGACGGCGTGGGCTACAACTGGATCGACGGCCTCAAGGCGCACGCGGACAGCCAGGTCAGCGA
CGACATGATGAAGACGGCCGCGCATCGGTACCCCGACAACACGCCCCGCACCAAGGAGGCGTACTG
GTACCGCAGCATCTTCGAGACCCACTTCCCCCAGCGTGCCGCCGTGGAGACGGTGCCGGGCGGCCC
CTCGGTGGCCTGCTCCACCGCCACCGCCGCGCTGTGGGACGCCACCTGGGCTGGCAAGGAGGACCC
CTCGGGCCGCGCCGTGGCCGGCGTGCACGACTCGGCCTACGACGCCGCCGCCGCCGCCAACGGCGA
GCCGGCTGCCAAGAAGGCCAAGAAGTAAACGGGCCTTGTCCACCACTTGCGGTCCCGACTGCGGCA
GCTGAGACTAGCTGTCAGAGGTTGCTGCGCATGGGGCCGCGGCGTGCGTCGCTACCGGGAAGCAGC
GTGCTGTGGGGGAGTTTGATGTGCTTCCTGATCAGCATCGTGCTCGCGGAGTAGCGAGAGCGAGTC
CGGATCATGCACGCGATGCGGCTGCATGCATAAAGAGCAGCACCTCAGCTGCACCGCCGTCTGTGC
ATGCATGGCCAGTGATTCCACCAGGTGCACGGCCTTGCGTTTTTGAGCGAAGAGCACACGTCACGG
ATGTCAACGCGTTATTCGGGGGCTACGAGCCTGCGCGCTATTGTGTCGTGTTTTACTGGCGTGGAG
TGTCGTGGATGCTGTTTCTGACAGATGTCTTTCACTGCGAGTGTGAATCATAGGGGTGACTTGACG
GTCAATGTAGACGAGGAACGGGGAGACGACATGCCCATTGACAGGATGACTAGGTCTTGACGGTGG
AGGATGGGTCACGGGCGGCACAAGACGCGGGGGAACAGGCGGTGCGAAGTCCAGCACATGGATTAA
TTAGATAAAGGGGCGCCAGCAACTTGGCGCCCGCGTAGAAAGTCATGAAGCCATGCTAGGCGGTAG
TCGCAAGGAAGCGAGAACGGGATGGGACGCAGCTGCACACGTGCGGCGGTGGGGAGCCGCTGAAGC
TCTTTAAGAAGACGTTCCGCAGACTCTCTGATCCCAACTGCCATTCTGCCAACCCGTTTTGCACGC
CGAAAACCTGGCACACTGGAAGCGCTCATCACGCT

FIGURE 5 (continued)

SEQ ID NO: 100, Glycine max - TA41694_3847#1
TGGAACCCTTCTACGTGTTCTCCATTCCCTCTCTCACTCCTCCATCTACGTTTCTTAAATCATTTC
CTTCTTTCTCTCTTTCTTTATCTTCTCATTTTCCTCATTACACTCTTTTTTTTTTCTCTCAACTTT
TCTCTTATTAACCATAGTTCACATATTATATCATCACATATCATAGTGATATATTATATCATATCA
CAATGTGTGGCATACTTGCTGTGCTTGGTTGCTCTGATTCATCTCAAGCCAAAAGGGTTCGCGTCC
TTGAGCTTTCTCGCAGATTGAAGCACCGTGGTCCTGACTGGAGTGGGCTCCACCAATATGGTGATA
ACTATTTGGCTCATCAAAGGTTAGCCATAGTTGATCCAGCTTCTGGTGATCAACCCCTCTTCAATG
AAGACAAAACTGTCGTGGTTACGGTGAATGGAGAGATCTACAATCATGAAGAACTCAGGAAACAGT
TGCCTAATCACACCTTCCGTACAGGAAGTGACTGTGATGTTATTGCTCACCTGTATGAGGAGCACG
GAGAAAACTTTGTGGACATGCTTGATGGTATATTTTCGTTTGTTCTGCTAGATACTCGTGACAACA
GTTTTATAGTGGCACGAGATGCAATTGGGGTCACTTCCTTGTACATTGGTTGGGGTCTAGATGGCT
CTGTCTGGATTTCATCAGAATTGAAGGGGTTGAATGATGATTGCGAACATTTTGAGTCTTTTCCAC
CTGGTCACTTGTACTCTAGCAAAGAGAGAGCGTTCCGCAGATGGTACAATCCTCCATGGTTCTCTG
AGGCTATTCCCTCAGCACCTTATGATCCTCTTGCTTTGAGGCATGCCTTTGAGAAGGCTGTGGTAA
AAAGGTTGATGACTGATGTTCCCTTTGGTGTTTTGCTCTCTGGAGGTTTGGACTCTTCATTGGTTG
CAGCCGTCACGGCTCGCTACCTGGCAGGCACAAATGCTGCCAAGCAATGGGGAACCAAATTACACT
CTTTCTGTGTAGGCCTTGAGGGTGCACCTGACCTAAAGGCAGCAAAGGAAGTAGCAGACTACATAG
GAACTGTACATCATGAATTTCACTACACTGTTCAGGATGGCATAGATGCCATTGAGGATGTGATCT
ATCACATTGAAACATATGATGTGACAACAATTAGAGCAAGCATTCCCATGTTTCTTATGTCTCGTA
AGATCAAGTCATTGGGAGTCAAATGGGTTATATCTGGAGAAGGATCTGATGAGATCTTTGGAGGGT
ATCTATATTTCCACAAGGCACCAAACAAAGAAGAGTTTCATCAAGAAACATGCCGCAAGATTAAAG
CACTCCACAAATATGATTGCTTGCGAGCCAATAAATCGACCTTTGCCTGGGGTCTAGAAGCCAGAG
TGCCATTTTTGGACAAAGATTTTATCAGAGTTGCAATGAACATTGATCCTGATTATAAAATGATTA
AAAAGGAAGAAGGGCGAATTGAGAAATGGGTACTGAGGAGGGCCTTTGATGATGAAGAACATCCTT
ATCTGCCAAAGCACATTTTATACAGGCAGAAAGAACAATTCAGTGATGGAGTTGGCTATGGTTGGA
TTGATGGCCTTAAAGCTCATGCTGAGAAACATGTGACTGACAGAATGATGCTCAATGCTGCTAACA
TTTTCCCCTTCAACACACCAACCACCAAAGAAGCATACTACTATAGAATGATATTTGAGAGGTTCT
TCCCTCAGAACTCAGCCAGGCTGAGTGTTCCTGGAGGACCAAGTGTTGCATGTAGCACAGCCAAAG
CTGTAGAGTGGGATGCTGCTTGGTCTAACAACCTTGATCCATCTGGTAGGGCAGCACTTGGAGTGC
ATGCATCAGCTTATGGAAATCAGGTCAAAGCTGTAGAACCAGAGAAGATCATACCAAAGATGGAAG
TTTCCCCACTAGGAGTTGCCATATAGAGCTAGTATGAGCCATAGCAAAAACTAGTAGTTGCCCTAG
AACCAAAATATATTATTATACTAGTCATCAATGACTCATTAATCATCATAAATGAAAATTTGGCCT
GCTGTGTAGTTTATTCAGGCAAGGCTATATATAAATAGATAAGGCTCTCTATCTAGCTGTCTTAAG
TGTTGTTCCATCCACATCTTGTCTTCGTTTTCTATTTATGTCATCTGAGCACTATCATGATGTACT
GGATTTCCAAGAAAATGTTCAGTTAAATTTGAATGCAAAGTTCACTATTTCAGACTTTCA

SEQ ID NO: 101, Glycine max - U77679#1
GGGGCATTGGATTCTCACCAACGTTTGCGTTACTCAAGCCGACATTCTCGCTTCCGTTGGAACCGT
TCTTCGTGTTCTCCATTCCCTCTCTCACTCCTTCATCTACTTCACATATTATATCATCACATATCA
TAGTGATATCATATCACAATGTGTGGCATACTTGCTGTGCTTGGTTGCTCTGATTCATCTCAAGCC
AAAAGGGTCCGCGTCCTTGAGCTTTCTCGCAGATTGAAGCACCGTGGTCCTGACTGGAGTGGGCTC
CACCAATATGGTGATAACTATTTGGCTCATCAACGGTTAGCCATAGTTGATCCAGCTTCTGGTGAT
CAACCCCTCTTCAATGAAGACAAAACTGTTGTTGTTACGGTGAATGGAGAGATCTACAATCATGAA
GAACTCAGGAAACAATTGCCTAATCACACCTTCCGTACAGGAAGTGATTGTGATGTTATTGCTCAC
CTGTATGAGGAGCACGGAGAAAACTTTATGGACATGCTTGATGGTATATCTTCATTTGTTCTGCTG
GATACTCGTGACAACAGTTTTATAGTGGCGCGGGATGCAATTGGGGTCACTTCCTTGTACATTGGT
TGGGGTTTAGATGGCTCTGTCTGGATTTCCTCTGAATTGAAGGGGTTGAATGATGATTGCGAACAT
TTTGAGTCTTTTCCACCTGGTCACTTGTATTCTAGCAAAGAGAGAGCGTTCCGCAGATGGTACAAT

FIGURE 5 (continued)

CCTCCATGGTTGTCTCTGGCTATTCCATCTGCCCCTTATGATCCTCTTGCTTTGAGACATGCCTTT
GAGAAGCTGTGGATAAAAAGGTTGATGACTGATGTGCCCTTTGGTGTTTTGCTCTCTGGAGGTTTG
GACTCTTCATTGGTTGCAGCCGTCACGGCTCGCTACCTGGCAGGCACAAAAGCTGCGAAGCAATGG
GGAACTAAATTACACTCTTTCTGTGTAGGCCTTGAGGGTGCACCCGACCTAAAGGCTACAAAGGAA
GTAGCAGAGTACATAGGAACTGTCCATCATGAATTTCACTACACTGTTCAGGATGGCATAGATGCC
ATCGAAGATGTGATCTATCACATTGAGACATATGATGTGACAACAATTAGAGCAAGCATTCCCATG
TTTCTTATGTCTCGGAAGATCAAGTCATTGGGAGTCAAATGGGTTATCTCTGGAGAAGGATCTGAT
GTTTTTTTGGAGGGTATCTATATTTCCACAAGGCACCCAACAAAGAAGAGTTCCACCAAGAAACA
TGCCGCACAATTATTGTACTCCACAGGTATGATTGCTCGCGAGCCAATAAATCGACCTTTGTCTGG
GGTCTAGAAGCCAGAGTACCATTTTTGGACAAAGAGTTTATCAGAGTTGCAATGAACATTGATCCT
GAGTGTAAAATGATAAAAAAGGAAGAAGGGCGAATTGAGAAATGGGCACTGAGGAGGGCCTTTGAT
GATGAAGAACATCCTTATCTGCCAAAGCACATTTTATATAGGCAGAAAGAACAATTCAGTGATGGA
GTTGGCTATGGTTGGATTGATGGCCTTAAAGCTCATGCTGAGAAACATGTGACTGACAGAATGATG
CTCAATGCTGCCAACATTTTCCCCTTCAACACTCCAACCACCAAAGAAGCATACCACTATAGAATG
ATATTTGAGAGGTTCTTCCCTCAGAACTCATGCAGGCTCACTGTTCCTGGAGGAACAAGTGTTGCA
TGTAGCACAGCAAAAGCTGTTGAGTGGGATGCTGCTTGGTCTAACAACCTTGATCCATCAGGTAGA
GCAGCACTTGGAGTGCATGCATCAGCTTATGGAAACCAGGTCAAAGCTGTAGAACCAGAGAAGATC
ATACCCAAGATGGAAGTTTCTCCACTAGGAGTTGCCATATAGAGCTAGTATGAGCCATAGCAAGGA
CTAGTAGTTGCCCTAGAACCAGCATATATTATTATTATACTAATCATCAAATCATGAAACATCAGG
TTGCTTTGTAGTTATCCAGGGAATGGTATATAAATAGATAAGGATCTCTATCTATCTGGCTCTCTT
TCTGGGCCACCCAGATCTAGCCTCAACTTGCTTTCGATGTCACCTGATGCACAATCATAAAG

SEQ ID NO: 102, Glycine max - TA41698_3847#1

GGCACGAGCTTCAACTTCACCCATTCATACGTGGTGTTGTTACTGCTGCTCTTTTCTCTTTTCTTT
TCTCTTTAGTTCTCTCTTCCCCTTTCTTTTTCTTTTTCTTCTTCTTCTGAGCTTGTTTAAGCTTTT
CTTCCATTAACATATTATCACAATGTGTGGTATTCTTGCTGTTCTTGGTTGTTCTGATGACTCTCG
AGCCAAAAGGGTCCGCGTGCTTGAGCTCTCTCGCAGATTGAAGCACCGTGGCCCTGACTGGAGTGG
GCTCCATCAACATGGTGACTGCTTTTTGGCACATCAACGGTTAGCCATAGTTGATCCTGCTTCTGG
GGATCAACCTCTCTTTAACGAGGACAAATCCGTCATTGTTACGGTAAATGGAGAGATTTACAACCA
TGAAGAGCTCAGGAAACAGCTGCCTAATCACAACTTCCGAACTGGAAGTGATTGTGATGTTATTGC
ACACCTGTACGAGGAACATGGAGAAGACTTTGTGGACATGCTGGATGGTATCTTCTCATTTGTTCT
ACTGGACACCCGTGACAACAGTTTTATAGTGGCTCGGGATGCTATTGGGGTCACTTCCTTGTACAT
TGGATGGGGGTTAGATGGCTCTGTTTGGATTTCATCAGAAATGAAAGGCCTGAATGATGATTGTGA
ACACTTTGAGTGTTTTCCACCTGGTCACTTGTACTCTAGCAAAGAAAGAGGGTTCCGCAGATGGTA
CAATCCTCCTTGGTTCTCTGAGGCTATTCCATCTGCCCCTTATGATCCTCTTGTTTTAAGACACGC
CTTTGAGCAGGCAGTCATAAAAAGGTTGATGACTGATGTGCCTTTTGGTGTTCTACTCTCTGGAGG
TTTGGACTCTTCTTTGGTTGCATCCATCACTTCTCGTTACTTGGCCAACACAAAGGCTGCTGAGCA
GTGGGGATCAAAGTTACATTCATTCTGTGTAGGCCTTGAGGGCTCACCAGATTTGAAGGCTGCAAA
AGAGGTTGCTGACTATCTAGGCACTGTCCACCATGAGTTTACCTTCACTGTTCAGGATGGAATAGA
TGCCATTGAAGATGTTATCTACCATATTGAAACATATGATGTGACTACAATTAGAGCAAGCACACC
TATGTTTCTCATGTCTCGGAAGATTAAATCACTTGGTGTCAAATGGGTTATCTCAGGAGAAGGATC
TGATGAGATCTTTGGAGGGTATTTGTACTTCCACAAGGCACCCAACAAGGAGGAGTTCCACAGAGA
AACATGCCGCAAGATCAAAGCACTTCACCAATATGATTGCTTGCGAGCCAATAAATCAACATTTGC
TTGGGGTCTAGAAGCCCGTGTACCATTTTTGGACAAGGCGTTTATCAATGCTGCAATGAGTATTGA
CCCTGAGTGGAAGATGATAAAAAGAGATGAAGGACGAATTGAGAAGTGGATTCTGAGGAGAGCCTT
TGATGATGAAGAGCATCCTTATCTGCCAAAGCACATTTTATACAGGCAGAAAGAACAATTCAGTGA
TGGAGTTGGCTATAGTTGGATTGATGGCCTTAAGGCCCATGCTGCAAAACATGTGACTGAAAAAAT
GATGCTTAATGCTGGTAACATTTACCCCCACAACACCCCAAAAACCAAGGAAGCATATTACTACAG

FIGURE 5 (continued)

AATGATCTTTGAGAGGTTCTTCCCTCAGAACTCAGCTAGGCTCACTGTTCCTGGAGGAGCAAGTGT
TGCATGTAGCACAGCCAAAGCTGTTGAGTGGGATGCTGCTTGGTCTAACAACCTTGATCCCTCTGG
TAGAGCAGCACTTGGAGTGCACATTTCAGCCTATGAAAACCAGAACAACAAGGGTGTAGAAATTGA
GAAGATAATACCTATGGATGCTGCTCCCCTTGGTGTTGCCATCCAGGGCTAATACAAAGATGTGAC
AAAGAATAATTTGGGCGACAATGAAGATAACTAAGCTAAAGGTGAATGAAAAATTTGCCTGCAGTG
TAATTTCATCTGGGCAAAGCTTTTATAGTTTATAGTTATAAGGCTTTCTAAAAAGTGTTGCGTATT
GTATTATCTTGAATGCTGTGATTTGAAGTCTTAATAAAAGTGTTTCCTTTATCAGTTCATAATGAA
TGCAAAGTCCATTATTTTAAAA

SEQ ID NO: 103, Glycine max - TA51197_3847#1
AGCAGTGGTATCAACGCAGAGTACGCGGGAGTTCTGTTGTTGTGTTGTGTTGTGTGTCTTCCCTTG
TGTGTTCCAGTTTTTATTTGCAGCCGCCATGTGCGGAATCCTCGCAGTGTTGGGTTGCGTCGACAA
CTCTCAGACCAAGCGCGCTCGCATCATCGAATTGTCTCGCAGGTTGCGGCATAGAGGTCCTGATTG
GAGTGGCATACATTGCTATGAGGATTGTTACCTAGCTCATCAACGCCTTGCTATTGTTGACCCTAC
TTCAGGGGACCAACCTTTGTACAACGAAGACAAAACTATTATTGTCACTGTAAATGGGGAGATATA
CAATCACAAGCAATTGAGGCAGAAACTGAGTTCCCATCAATTTCGAACTGGTAGTGATTGTGAAGT
GATTGCCCATCTTTATGAAGAACATGGAGAAGAATTTGTTAATATGCTGGATGGGATGTTTGCCTT
TATTCTTCTTGATACTAGGGATAAAAGTTTTATTGCTGCTCGTGATGCTATTGGCATTACCCCTCT
ATACTTGGGCTGGGGTCATGATGGATCAACATGGTTTGCATCTGAAATGAAAGCTCTGAGTGATGA
TTGTGAGAGATTCATATCTTTTCCTCCAGGGCACATCTATTCCAGCAAACAGGGAGGATTAAGAAG
GTGGTACAATCCACCATGGTTTTCAGAGGATATTCCATCAACTCCCTATGATCCAACCCTTTTGCG
TGAGACCTTCGAGAGGGCTGTAGTTAAGAGAATGATGACTGATGTACCTTTTGGAGTTCTTTTGTC
TGGAGGATTGGACTCATCACTTGTTGCTGCAGTGGTCAATCGTTATTTGGCTGAATCTGAATCTGC
TCGTCAATGGGGATCACAGTTACATACTTTCTGCATTGGTTTAAAGGGCTCTCCTGACTTGAAAGC
TGCAAAAGAGGTAGCAGATTACCTTGGTACTCGTCACCATGAACTTTATTTCACGGTTCAGGAAGG
TATAGATGCACTTGAAGAAGTCATTTACCATATTGAAACATATGATGTAACGACTATCAGAGCAAG
TACTGCAATGTTTCTTATGTCCAGAAAAATTAAAGCCTTGGGAGTGAAAATGGTACTTTCTGGAGA
AGGTTCAGATGAAATATTTGGAGGTTACCTGTATTTTCACAAGGCACCTAATAAGAAAGAGTTTCA
TGAAGAAACATGTCGAAAAATTAAAGCTCTTCATCTTTATGACTGCCTGAGAGCCAATAAATCAAC
TGCAGCATGGGGTGTAGAGGCACGTGTACCATTCTTGGATAAAGAATTTATCAACGTAGCCATGAG
TATAGATCCGGAATGGAAAATGATAAGGCCTGATCTTGGAAGGATAGAGAAGTGGGTATTACGCAA
TGCATTTGATGACGATAAGAATCCATATTTACCAAAGCACATATTGTACAGGCAGAAGGAACAATT
CAGTGATGGGGTTGGTTACAGCTGGATTGATGGCTTGAAGGATCACGCAAACAAACAAGTCACAGA
TGCGACGATGATGGCTGCCAATTTTATTTACCCTGAAAACACTCCTACCACAAAAGAAGGATACCT
CTACAGGACAATTTTTGAGAAGTTCTTTCCAAAGAATGCAGCAAAGGCAACAGTGCCAGGAGGTCC
TAGTGTGGCATGCAGTACTGCAAAAGCTGTGGAATGGGATGCAGCATGGTCAAAAAATCTTGATCC
TTCTGGTCGTGCCGCACTTGGTATTCATGATGCTGCATATGATGCAGTGGATACCAAAATTGACGA
GCCCAAAAATGGAACCCTTTAAGGCCCATAATCGATTGTCAAGAGAAAAAAATGTATGCAACAACT
GTCTAGTGGGGATTTAAACTTCTAGTAGGCAAAACTAATGAGAAGTGGGATTGTTTTTATTTTCAG
CTCAAATTAATATGTAGGTTTTGAACTGTTTGTGGGTTATTTTAAATAAATATCTATATTTAAATT
TTGTAG

SEQ ID NO: 104, Physcomitrella patens - 173106_estExt_fgenesh1 pg.C_4100023#1
ATGTGCGGAATTTTGGCTATTCTCGGTTCCCACGACGCGTCGCCTGCGCGACGTGATCGCATTCTG
GAGCTTTCCCGCAGGCTGCGCCACCGCGGTCCCGACTGGAGTGGGCTGTTCGCAGGGCAGAAGTGC
TGGTGTTATCTGGCTCATGAGCGCTTGGCCATCATTGATCCCGCCTCGGGCGACCAACCTCTGTAC
AATGAGAACAAAGATATCGTCGTCGCTGCCAATGGAGAAATCTACAACCACGAGGCCTTGAAGAAG

FIGURE 5 (continued)

AGCATGAAGCCTCACAAGTATCACACGCAGTCCGACTGTGAAGTTATTGCTCATCTCTTTGAAGAT
GTCGGCGAGGACGTGGTCAACATGCTGGACGGCATGTTCTCATTCGTGTTGGTCGACAACCGCGAT
AATTCCTTCATCGCCGCCCGGGATCCCATTGGCATCACCCCTCTCTACTACGGCTGGGGTGCGGAT
GGAAGTGTTTGGTTTGCATCGGAGATGAAGGCCTTGAAGGACGATTGCGAGCGGTTCGAGATTTTC
CCACCCGGTCACATCTACTCTAGCAAAGCTGGAGGGCTTCGGCGATATTACAACCCAGCTTGGTTC
TCTGAGACTTTTGTCCCCAGCACCCCTTACCAGTCTCTTGTTCTCCGCGCAGCCTTCGAGAAGGCT
GTAATCAAGAGACTGATGACCGACGTGCCCTTCGGTGTACTCCTATCCGGAGGGCTGGATTCTTCA
TTAGTGGCAGCAGTGGCATCCCGTCATATCGCAGGAACTAAAGCTGCCAACATCTGGGGCAAGCAG
CTTCACTCTTTCTGCGTCGGACTTCAGGGTTCTCCTGACCTGAAGGCTGCTCGGGAAGTCGCCAAC
TACATCGGCACCCAGCACCACGAGTTCCACTTTACTGTCCAAGAAGGTTTGGACGCTCTGTCGGAT
GTGATCTATCATGTGGAGACTTACGACGTGACCACCATCCGAGCTAGCACGCCCATGTTCCTCATG
ACACGCAAGATTAAGGCTCTGGGTGTAAAGATGGTGTTGTCTGGGGAGGGATCCGATGAAATTTTT
GGTGGTTACCTCTATTTCCATAAAGCGCCCAACAGGGAGGAGTTCCACCATGAGCTTGTTCGCAAG
ATCAAGGCGCTGCATATGTATGATTGCCAGAGAGCCAATAAGTCGACGTCTGCCTGGGGTTTGGAG
GCGCGTGTTCCCTTCCTAGACAAAGAATTTATGGAAGTTGCCATGGCTATCGATCCTGCGGAAAAG
CTGATCAGGAAGGACCAAGGAAGAATAGAGAAGTGGGTGCTCCGAAAAGCTTTCTACGACGAAAAG
AATCCTTACCTGCCCAAGCACATTTTGTATCGCCAGAAGGAGCAATTCAGCGATGGCGTTGGCTAC
AGCTGGATTGACGGCCTCAAGGCTCATGCACAGAGCCATGTATCCGACCAAATGCTGAAGCATGCA
AAGCACGTGTACCCCTACAACACGCCGCAGACTAAAGAAGCATACTATTACCGAATGCTCTTCGAG
AAACACTTCCCGCAGCAATCCGCTCGCTTGACGGTCCCCGGAGGTGCTAGCGTCGCATGTAGCACG
GCCACAGCAGTTGCATGGGACAAGTCCTGGGCGGGCAACCTGGACCCATCTGGCCGAGCAGCATTG
GGATGCCACGACGCGGCCTACACGGAAAACAGCGCTGCAATGAGTTACATAACAAAAAACATGTCA
AATGTTGGACAAAAAATGACCATACATTGA

SEQ ID NO: 105, Physcomitrella patens - 180723_estExt_gwp_gw1.C440158#1

ATGTGTGGAATTCTAGCGATTCTCGGTGCCGACGGCGCCGTTCCGTCTGCCGGACGTGATCGCGCT
CTAGCGCTGTCCCGAAGGCTGCGCCATCGAGGACCTGACTGGAGTGGACTCTTTGAGGGCAAGGAT
TCCTGGTGTTACCTCGCTCATGAGCGCCTGGCTATCATCGATCCGGCTTCGGGTGATCAACCCCTC
TACAATGGCACTAAGGACATCGTTGTCGCTGCTAACGGAGAGATTTACAACCACGAGTTGTTGAAG
AAGAACATGAAACCACACGAGTACCACACGCAGTCCGATTGCGAAGTCATTGCTCATCTTTATGAG
GATGTAGGTGAGGAGGTTGTGAACATGCTTGACGGCATGTGGTCGTTCGTGCTGGTGGACAGCCGA
GACAACTCCTTCATCGCAGCCCGCGACCCCATCGGCATCACTCCTCTCTATCTTGGTTGGGGAGCC
GATGGTAGAACTGTGTGGTTTGCCTCGGAGATGAAAGCCTTGAAGGACGATTGCGAACGGCTTGAG
GTCTTTCCACCAGGCCACATCTACTCAAGCAAAGCTGGAGGGCTCCGTCGCTACTACAACCCACAG
TGGTTCTCAGAGACTTTTGTTCCCGAAACTCCTTACCAGCCTCTGGAACTACGTTCAGCCTTCGAG
AAGGCTGTGGTAAAGAGGCTCATGACCGACGTCCCCTTCGGTGTGCTCCTTTCCGGAGGCTTGGAT
TCTTCCTTGGTGGCATCAGTGGCAGCCCGACATCTTGCCGAAACCAAAGCTGTCAGAATCTGGGGC
AACGAGCTCCACTCCTTCTGTGTTGGCCTTGAGGGTTCTCCCGACCTGAAGGCTGCGAGGGAAGTT
GCCAAGTACATCGGCACCCGCCACCACGAATTTAACTTCACCGTCCAGGAAGGATTGGACGCTCTG
TCTGACGTGATCTACCATGTGGAGACCTACGACGTGACCACCATTAGGGCGAGCACACCAATGTTC
CTCATGACACGGAAGATCAAGGCTCTGGGTGTGAAGATGGTGTTGTCTGGGGAGGGATCCGACGAG
ATCTTTGGTGGTTACCTCTACTTCCACAAAGCTCCCAACAGGGAGGAGTTTCACCACGAACTAGTC
CGCAAGATCAAGGCGCTACACTTGTACGATTGCCAGAGAGCCAACAAATCAACCTCTGCTTGGGGT
CTGGAAGCTCGTGTTCCCTTCCTTGACAAGGAGTTCATGGACGTTGCGATGATGATCGACCCTAGC
GAGAAGATGATCAGGAAGGACCTGGGCAGAATTGAGAAGTGGGTGCTGCGTAAAGCTTTCGATGAC
GAAGAGAGACCATACTTGCCCAAGCACATTTTGTACAGGCAAAAGGAGCAATTCAGCGATGGAGTG
GGCTACAGCTGGATTGATGGACTCAAGGAATATGCGGAGAGCCATGTGACGGATCAGATGATGAAG

FIGURE 5 (continued)

CACGCGAAGCATGTGTACCCCTTCAACACGCCCAACACCAAAGAAGGATATTACTACCGAATGATC
TTCGAGAAGCATTTCCCCCAACAATCCGCCCGGATGACGGTCCCCGGAGGTCCTTCGGTAGCATGC
AGCACCGCCACAGCTGTGGCATGGGACGAAGCATGGGCCAACAACTTGGACCCCTCCGGCAGAGCA
GCATTGGGATGCCATGACTCAGCTTACACAGACAAACACAGTGAGAAAGCTGCACCAGCGGCAGAA
GCTAACGGCACGGCTTCTCACGAGAACGGCCACACATTCTCCAAGCCCAAATCCACACTGGATGCC
ACCATTCTGAAAACTCAGGCCGTGCACTAATCTCTAGCAAGACACACGTTTCAGTAGTTATCTAAG
TGGCAGCAACTGCAACCAAGCCTCAGAATGGGCTCCCAACAAGCTGGGTTTCCATGTGAAGAGCTG
GAGCTTGAATTGCAACATGCGCCCTGTAACAATAATAGAAAACTCGCTCAAAACAAACGTAGAAAA
ATAGAATAAAGAGTACTGGACTGAAAGACCGAAGACCTTTGCTTGAGTCCTCTGAGGCGCTGGTAT
GGATATAAACCGGACAGTGTATGGCAAATAGTGCGAGGAAAGTAATTTTAATAAGTTAGCAGCTAT
AGTTTGAGCTATGGCAGTCACAGACCCATATCTGTACAAGCTTCACTTCCCCTAAGTTATGAATTC
CCTCGTTTCCAGTTTCATATA

SEQ ID NO: 106, Physcomitrella patens – 226188_estExt_Genewise1.C3500008#1

ACTGTGTGGGCTTGGGTGGTTGTGGTGAAGGAGGACGAGGAAGAGTAAGAGGAAGAGGCGGATTCT
GCATCAAGGGTTTATGATGCTCTTTGCACGACAAACCTACGAATCCTGACCCAGCTGGTCGCTTGT
CGTCCCCCCTCCTTCCTTTTTGGCTTCTCTCTTGTCTTTCCGTTAGCGCTTTTGAGGAGACTTGAG
CCGCCGTCACAATGTGTGGAATTTTAGCCATCCTTGGGTGCCATGACAAGAGCGTCACGCGGCGGC
ATCGCTGCCTGGAGCTCTCTCGCAGGTTGCGGCACCGGGGACCTGACTGGAGTGGTTTGTTCGTGG
ACGAGGCGTCGGGATGTTATCTGGCGCACGAAAGGTTGGCAATTATCGATCCCACGTCGGGCGACC
AGCCGTTGTTCAACGAGAACAAGGACATTGTCGTCGCGGTGAATGGCGAGATTTACAACCATGAGG
CCCTCAAGGCGAGCATGAAGGCACATAAATACCACACTCAGAGTGATTGTGAAGTTATTGCACATC
TGTACGAGGAAATTGGGGAGGAGGTGGTTGAGAAGCTGGATGGCATGTTTTCATTTGTATTGGTAG
ACTTGCGCGATAAGTCATTCATTGCTGCTCGCGATCCCCTTGGAATCACACCACTCTACCTCGGGT
GGGGCAATGATGGGTCTGTATGGTTTGCGTCTGAGATGAAGGCTTTGAAGGACGATTGTGAGCGCT
TTGAGTCGTTCCCTCCAGGTCACATGTATTCCAGCAAGCAAGGTGGTCTGCGTAGGTATTACAACC
CACCTTGGTTCAACGAAAGCATCCCAGCAGAACCTTATGACCCGCTCATACTACGACATGCCTTTG
AGAAATCAGTCATCAAACGGTTAATGACGGATGTGCCGTTTGGAGTGCTGCTGTCGGGTGGCCTTG
ATTCCTCGTTGGTAGCTGCGGTTGCTCAACGACATCTAGCCGGCAGTACAGCAGCCAAGCAATGGG
GGAATAAGCTTCATTCTTTCTGTGTTGGACTGGAGGGCTCTCCCGATTTGAAGGCTGGACGGGAAG
TTGCTGATTACATCGGTACGGTGCACAAAGAGTTTCATTTCACTGTCCAGGAAGGTCTGGATGCCA
TTTCTGATGTAATATATCACATTGAAACGTATGATGTCACTACAATTCGAGCTAGTACACCCATGT
TCCTCATGTCTCGAAAAATCAAAGCCCTTGGCGTGAAGATGGTTCTTTCTGGAGAGGGTTCAGACG
AGATATTTGGGGGTTACCTTTACTTCCACAAAGCTCCTAACAAGGAGGAGTTTCACAAGGAAACTT
GTAGGAAGTTGAAGGCACTGCACTTGTACGATTGTTTGAGGGCAAACAAATCAACATCAGCCTGGG
GTTTGGAAGCTCGTGTACCATTCTTGGATAGGGACTTCGTAAACCTCGCCATGTCGATCGACCCTG
CTGAGAAAATGATAAACAAGAAGGAAGGGAAAATCGAGAAGTGGATCATCCGTAAAGCTTTTGATG
ATGAAGAGAACCCATACCTGCCCAAGCATATTTTGTACAGACAGAAGGAGCAGTTCAGTGACGGTG
TTGGCTACAGTTGGATTGATGGCTTGAAGGACCATGCAGCCAGTCAGGTTTCTGACCAGATGCTGG
CAAATGCTAAACACATTTATCCCCACAACACTCCAGGAACAAAGGAAGGTTACTACTACCGCATGA
TCTTCGAGAGATGCTTCCCACAGGAGTCAGCAAGGCTTACAGTTCCAGGAGGACCTAGTGTAGCTT
GCAGTACTGCTGCTGCCATTGCCTGGGACAAGGCATGGGCCAATAACTTGGATCCCTCAGGCAGGG
CAGCTACAGGTGTTCACGATTCCGCATATGAAGGTGGTGAGGTGGAGAGCTCAGCAGTGAGCCACA
AAGAAGGTGGTGAGGATGGTTTGGCCAACTCGAAAGTGGGCGACAAGGTTCAGGAAGCCATAGCTG

FIGURE 5 (continued)

TTGCCTGAGGTGACGCATGGTGTTCTTTGATTAGGATGCTCATTGTAAGCTGACCCACCTACTGTA
CTGCAAGCAATTGTAGCTTTATATGTATTGGTGAACAATTGCCATTTTAGAGTGATCAGTTTTCAT
TTCCGTTTACTTTGAGATAAATGCCTTATGTGTATTTGAGTAGGAACTGGTTAAAGGACTTTTAAA
TTTGTTGTTGACCGTGAAAGAGATCAACCTTCAGGTATATATTGTTTTCGAATGAGCTTGTTTTTC
AAACCCTC

SEQ ID NO: 107, Populus trichocarpa - 722643#1

CTCATTCAACAATAACAAAACAAGCTCTTGCTCTACGTGTTGGTGTTTCCTATTAACAGCCCATCT
CCTTCTCCTGCCACCTCGCTTTCCTTTTTATTACCAGATTTTCTTCTTTCATTACTACCCAATTTC
ATCTCTATAGTTTATCCATCCATTTTTCTCTGTCTTTGTTTTTAAGATATACATATCTAGCAAAAT
CTTCTTTTATCTGCTATATCGTTTTTTTTTAAGAAACGACGATGTGTGGGATACTTGCTGTTTTGG
GTTGTTCTGACGACTCTCAGGCCAAGAGGGTTCGGGTGCTAGAGCTCTCTCGCAGGTTGAAGCACC
GTGGTCCAGATTGGAGTGGGCTCTATCAGTGCGGTGACTTTTACTTGGCTCATCAAAGGCTGGCTA
TTATCGATCCTGCTTCTGGTGACCAGCCACTCTTTAATGAGGACCAAGCCATCGTTGTCACGGTGA
ACGGAGAAATTTACAACCATGAAGAACTAAGGAAGCGTTTGCCAAATCACAAGTTCCGAACAGGCA
GTGACTGTGATGTTATCGCCCATCTGTACGAGGAATATGGCGAAAATTTTGTGGACATGTTGGATG
GAATGTTTTCATTTGTTCTGCTGGATACTCGTGACAACAGTTTCATTGTTGCTCGTGACGCCATTG
GGATCACCCCCCTCTATATTGGCTGGGGACTTGATGGGTCCGTGTGGATTTCATCTGAACTGAAAG
GTCTGAATGACGACTGTGAACATTTTGAGTGCTTTCCTCCTGGTCATTTGTACTCGAGTAAATCGG
GTGGATTACGTCGTTGGTACAATCCTCCTTGGTTCTGCGAGGCCATTCCCTCAACCCCATATGATC
CACTTGTTCTGAGACGTGCATTTGAAAAGGCTGTGATTAAAAGGCTAATGACTGATGTGCCTTTTG
GAGTTCTTTTATCTGGAGGCCTAGATTCATCACTGGTTGCTGCTGTTACTGCTCGCCATTTGGCAG
GTACAAAGGCTGCCAGACAATGGGGGGCACAACTCCATTCCTTCTGTGTTGGCCTAGAGAATTCAC
CAGATTTGAAGGCTGCAAGAGAAGTTGCAGATTATCTGGGAACCGTCCACCATGAATTTTACTTCA
CGGTTCAGGATGGTATAGATGCCATTGAGGATGTCATATACCATATAGAAACATATGATGTTACAA
CCATCAGAGCAAGTACCCCTATGTTCCTAATGGCTCGTAAGATCAAGGCACTAGGAGTGAAGATGG
TTATTTCTGGTGAAGGTTCTGATGAGATTTTTGGTGGGTATTTGTACTTTCATAAGGCACCTAACA
AAGAAGAGTTACACCGCGAAACATGTCGCAAGATAAAGGCCCTTCATCAATATGATTGCTTGAGAG
CTAACAAGGCAACATCTGCTTGGGGTTTAGAAGCCCGTGTCCCCTTCTTGGACAAGGATTTTATTA
ATGTTGCAATGGCTATTGATCCTGAATGGAAGATGATCAAACCTGGACAAGGCCATATTGAGAAAT
GGGTCCTTAGGAAAGCCTTTGACGACGAGGAGCATCCTTATCTGCCTAAGCATATTCTTTACAGGC
AGAAAGAGCAATTTAGCGATGGTGTTGGCTATAGCTGGATCGATGGTCTCAAAGCTCATGCTGCCC
AACATGTGACTGACAAGATGATGCAAAATGCTGAGCACATCTTTCCACATAATACCCCTACCACCA
AAGAAGCCTATTACTACAGAATGATTTTTGAGAGGTTCTTCCCACAGAACTCAGCCAGGCTGTCTG
TTCCTGGAGGAGCCAGTGTAGCATGCAGCACAGCTAAAGCTGTTGAATGGGATGCTGCCTGGTCCA
ATAATCTGGATCCTTCTGGACGGGCTGCATTGGGTGTACATCTCTCTGATTATGATCAGCAGGCAG
CTCTTGCCAATGCAGGAGTGGTGCCACCAAAAATTATTGACACTCTTCCTCGAATGTTGGAAGTTA
GTGCTTCGGGAGTTGCGATCCACAGTTAGCGCCTGCTGGAGGACTAAGTATTGGTGAATTTGATAT
CTATAGCCTTGGTATTATTTAAACTTGTGTTGCCTTGTATATGTAAAAATCTTAGAGGTCATATGT
AGATGTTACAAATAATGATCCGTGGTCCTTTGAAGTCGTGTGTTGTCATTACTTTGTGGTTTTTGT
ACAAGGTAATTCATGTATGTTATCAATGCCCTGTAGCTGTTTAAAGCTGCAAGGCAACCTTTCCTA
CTGTTTAAAGCTGTAATGCAACCTTTCCTATGGTTTCTTTGCTTC

SEQ ID NO: 108, Populus trichocarpa - 829702#1

GCTCATTCAACAATAACATAACAGGCTATTACTCTACGGATTATGGTTTCCTGTTAACACTCCATC
TCCCTCTCCTCCTGCTTCTTTGTTTTCCCTTTTTTTTTCCCAGTATTATTCTCTCGTTATTACCTG
GTTCCATCTTTATCTTCGATCTTAAGATATACTTAAGCTACTTCTATCTTCAATATCGAACGTTTT
ATTTTTGAAAAACAAAGAAGGATGTGTGGGATACTTGCTGTTTTGGGTTGTTCTGATGACTCTCAG

FIGURE 5 (continued)

GCCAAAAGGTTTCGAGTGCTTGAGCTCTCTCGCAGATTGAAGCACCGTGGTCCTGATTGGAGTGGG
CTCTTTCAGCACGGTGACTTCTACTTGGCTCATCAAAGGCTAGCCATTATTGATCCGGCTTCTGGT
GATCAGCCTCTCTTTAATGAAGACCAAGCCATCGTTGTCACGGTGAACGGAGAAATTTACAATCAT
GAAGAACTGAGGAAGCGCTTGCCAAATCACAAGTTTCGAACAGGCAGTGACTGTGATGTTATCTCC
CATTTGTACGAGGAATATGGCGAGAATTTTGTGGACATGTTGGATGGAATGTTTTCATTTGTTCTG
CTGGATACTCGTGACAACAGTTTCATTGTCGCCCGAGACGCCATTGGGATCACCTCCCTCTACATT
GGCTGGGGACTTGATGGGTCTGTGTGGATTTCGTCGGAATTGAAAGGTCTGAATGATGACTGCGAA
CATTTCAAGTGCTTTCCACCTGGTCATATATACTCGAGCAAATCCGGTGGATTAAGGCGTTGGTAT
AATCCTCTTTGGTTCTCTGAGGCTATTCCCTCGACCCCATATGACCCACTTGCTCTGAGAAGGGCA
TTTGAAAAGGCTGTGATTAAGAGGCTGATGACTGATGTTCCTTTTGGAGTGCTTTTATCCGGGGGA
CTAGATTCGTCATTGGTTGCTGCTGTGACTGCCCGGCATTTGGCAGGTACACAGGCTGCCAGACAA
TGGGGGGCACATCTCCATTCCTTCTGTGTAGGCCTAGAGAATTCTCCAGATCTGAAGGCTGCTAGA
GAAGTTGCAGATTATTTGGGCACCATCCACCATGAATTTCACTTCACAGTTCAGGATGGTATTGAT
GCCATTGAAGATGTCATATACCATGTTGAAACATATGATGTTACAACCATCAGAGCAAGTACCCCT
ATGTTCCTTTTGGCTCGTAAGATCAAGGCGCTAGGAGTGAAGATGGTTATTTCCGGTGAAGGTTCT
GATGAGATTTTTGGTGGGTATTTGTACTTTCACAAGGCACCTAATAAGGAAGAGCTCCACGGCGAA
ACATGTCGCAAGATAAAAGCCCTTCATCAATATGACTGCTTGAGAGCTAACAAAGCAACATCTGCT
TGGGGTCTAGAAGCCCGCGTCCCCTTCTTGGACAAGGATTTTATTAATGTTGCAATGGCTATTGAT
CCTGAATGGAAGATGATCAAACCTGGACGTATCGAGAAATGGGTTCTTAGGAAAGCCTTTGACGAC
GAGGAGCATCCTTATCTGCCAAAGCATATTCTGTACAGGCAGAAAGAGCAATTTAGTGATGGCGTT
GGCTACAGTTGGATTGATGGTCTCAAAGCTCATGCTGAATTACATGTGCACGACAAGATGATGCAA
AATGCTGAGCACATCTTTCCACATAATACCCCTACCACCAAAGAGGCCTATTACTACAGAATGATT
TTTGAGAGGTTCTTCCCACAGAACTCAGCGAGGCTGACTGTTCCTGGAGGAGCCAGTGTAGCATGC
AGCACAGCTAAAGCTGTTGAATGGGATGCTTCCTGGTCCAACAATCTCGATCCTTCCGGCCGTGCT
GCATTGGGTGTGCATCTTTCTGCTTATGAACAGCAGGCAGCTCTTGCCAGTGCTGGAGTGGTGCCA
CCGGAGATTATTGACAATCTTCCTCGAATGATGAAAGTTGGTGCTCCAGGAGTTGCAATCCAAAGT
TAGCTTCTGCTGGAGGACCGAAGTACATGCCTTGTACATGTATAAATCATATAGATCATGTGTAGA
AGTTACGAATAATAATCTCTGCTCGTTTGTAGTAGTGTTGGCACCTTGTTGTTTCTGTACAAGGCA
ATTCAAGTGTGCAATCGATGTTCTGTAGCTGTTTAAAGTTGTAATGCAACCTTTCCTCTGGTTTCC
TTACTTCATAGACGAATCCTTTGTTTT

SEQ ID NO: 109, Arabidopsis thaliana - AT5G65010.1#1

CCATTGTTATTTGTTTTCGTTGCCACTCTAACACAATGTGTGGGATTCTCGCTGTTCTTGGTTGCA
TCGACAACTCTCAAGCTAAACGTTCTCGTATCATCGAACTCTCTCGCAGATTGAGGCACAGAGGTC
CTGATTGGAGTGGACTCCATTGTTATGAAGATTGTTATCTTGCCCATGAGCGTTTAGCCATCATTG
ACCCTACTTCAGGAGACCAACCTCTCTATAACGAAGACAAGACCGTCGCTGTCACTGTAAATGGAG
AGATATACAACCACAAGATTTTGCGTGAAAAGCTTAAGTCTCATCAGTTCCGTACTGGTAGTGACT
GTGAAGTGATTGCACATCTTTACGAAGAACATGGAGAGGAATTTATCGACATGTTGGATGGAATGT
TCGCGTTTGTCCTTCTTGATACTCGCGACAAAAGTTTTATTGCTGCAAGGGACGCTATTGGTATCA
CTCCACTTTACATTGGATGGGGTCTTGATGGTTCTGTCTGGTTTGCTTCGGAGATGAAAGCGCTTA
GTGATGATTGTGAACAGTTTATGTCTTTTCCTCCTGGCCACATCTACTCAAGTAAACAAGGAGGGC
TTAGGAGGTGGTACAATCCTCCGTGGTACAATGAGCAGGTTCCTTCAACCCCATATGATCCTTTAG
TTCTGCGCAATGCTTTCGAGAAGGCTGTAATAAAGAGACTTATGACTGATGTGCCTTTTGGAGTTC
TCCTATCTGGAGGATTGGACTCGTCTCTCGTTGCTGCAGTAGCATTACGCCATTTGGAAAAAATCAG
AAGCTGCTCGTCAATGGGGTTCACAATTGCACACGTTTTGCATCGGTTTGCAGGGATCGCCAGATC
TTAAAGCTGGCAGAGAAGTTGCTGACTATCTTGGAACACGCCACCACGAGTTTCAGTTTACAGTTC
AGGACGGGATAGACGCGATAGAGGAAGTCATTTACCATATTGAGACTTATGACGTTACTACAATAA
GAGCTAGCACCCCAATGTTTCTTATGTCCAGAAAAATTAAATCTTTAGGTGTAAAGATGGTTCTTT

FIGURE 5 (continued)

CTGGGGAAGGTTCTGATGAAATACTGGGGGGATACTTGTACTTCCATAAGGCTCCCAACAAGAAAG
AATTTCATGAAGAAACATGCCGAAAGATCAAAGCTCTCCACCAATTTGATTGTTTGAGAGCTAACA
AATCAACTTCTGCGTGGGGTGTCGAAGCTCGTGTGCCTTTCCTAGATAAAGAATTTTTAAATGTTG
CAATGAGCATCGATCCAGAGTGGAAGTTGATCAAGCCTGATCTCGGAAGGATCGAGAAGTGGGTGC
TACGCAATGCCTTTGATGATGAAGAACGACCTTATCTACCAAAGCACATTCTATATAGACAGAAAG
AACAGTTTAGTGATGGAGTTGGGTATAGCTGGATAGATGGTCTGAAAGATCATGCAAATAAACATG
TCTCTGATACTATGCTGTCAAACGCAAGCTTTGTCTTCCCGGATAACACACCTCTGACAAAAGAAG
CGTACTACTACAGAACCATCTTCGAGAAGTTCTTCCCGAAGAGTGCTGCTAGAGCGACTGTACCAG
GAGGTCCAAGTATAGCTTGCAGTACCGCGAAAGCTGTAGAATGGGATGCAACTTGGTCAAAGAATC
TTGATCCGTCAGGCCGTGCGGCTCTTGGAGTTCATGTTGCAGCTTATGAAGAGGATAAAGCAGCTG
CTGCTGCTAAGGCTGGATCGGATTTAGTAGATCCTCTTCCTAAGAATGGAACATAAGAGAACAACA
CTACAGGCATTGAGGATATAAGCAAATGTTTTATTCTTCTACACAGAGAGATCGTTATCTTCTAGA
GGGATCAATGAATAAAAGCTTCGTCCATTTCTAGCTGGAGATTCCATGGATCTCCAGTTAGTGCAA
GTGATACACGTTGTCTACATTTGTACCTAAGTTTCTGCATTTTTTGTCGTTCTTTTGTGTTAGACA
AGTCTTGGACCCTAGATGATACTTCAGTTTCTTAGACGTTAAATTTGATGAATCCGAACTTGTTTG
ATTTCAAAGCCTGGCCTTTCTGC

SEQ ID NO: 110, Arabidopsis thaliana - AT3G47340.1#1

AGACATCAAAAACACGAATATCGATAGTACACTTCTACGTGCAATTTTCTCCTTTCTCTTCCTGGA
CATCTGTCTGTTTATTACATTTTCTTGTAATCTCTTTTTGGGGTTTTACAATATCTATCCCCTAAA
GTTTCGGAAAATTCTGTTTTTCTGTTCTCATTCTTCGTGATCTTTTTCACTTTCTTCAAAAAAAAA
ACATGTGTGGAATACTTGCCGTGTTAGGATGTTCCGATGATTCTCAGGCCAAGAGAGTTCGTGTTC
TTGAGCTTTCTCGCAGATTGAGGCACAGAGGACCTGACTGGAGTGGCTTATATCAGAACGGAGATA
ATTACTTGGCCCATCAACGTCTTGCCGTCATCGATCCTGCTTCCGGTGATCAACCTCTTTTCAACG
AGGACAAGACCATTGTTGTCACGGTGAACGGAGAGATTTATAACCATGAGGAGCTGAGAAAACGTC
TGAAGAATCACAAGTTCCGTACTGGTAGTGATTGTGAAGTCATTGCTCACTTGTACGAGGAGTATG
GTGTGGATTTTGTTGATATGTTGGATGGAATATTCTCCTTTGTGTTGCTCGACACACGAGATAACT
CTTTCATGGTGGCTCGTGATGCGATTGGTGTCACTTCGCTCTACATTGGTTGGGGACTAGACGGAT
CTGTGTGGATATCTTCAGAGATGAAAGGCCTAAACGATGACTGTGAGCATTTCGAAACGTTTCCTC
CAGGTCATTTTTATTCAAGCAAATTAGGAGGGTTTAAGCAATGGTATAATCCTCCTTGGTTCAATG
AATCTGTTCCTTCAACGCCTTATGAGCCTCTTGCGATAAGACGCGCCTTTGAAAACGCTGTGATTA
AGCGGTTGATGACTGATGTTCCATTTGGAGTTTTGCTCTCTGGTGGTCTTGATTCTTCCCTTGTTG
CCTCCATCACTGCACGTCACTTGGCCGGTACTAAGGCGGCTAAGCAATGGGGTCCTCAGCTCCATT
CCTTTTGCGTTGGTCTTGAGGGCTCACCGGACTTGAAGGCAGGGAAAGAGGTGGCGGAATATTTGG
GGACGGTGCACCACGAGTTCCACTTCTCGGTGCAGGACGGGATTGATGCGATAGAGGATGTGATTT
ACCATGTTGAGACCTATGATGTGACGACTATCAGAGCGAGCACACCGATGTTCTTGATGTCCCGGA
AAATCAAGTCTCTAGGGGTCAAGATGGTTCTCTCCGGCGAAGGTGCGGACGAGATCTTTGGAGGGT
ACCTCTATTTCCACAAGGCACCTAACAAGAAAGAGTTTCACCAAGAAACTTGTCGCAAGATCAAGG
CTCTTCACAAGTATGACTGTCTAAGAGCCAACAAATCTACCTCTGCCTTTGGACTAGAGGCACGTG
TTCCTTTCCTTGACAAAGACTTCATCAACACAGCTATGTCTCTCGACCCTGAATCCAAGATGATCA
AGCCAGAGGAAGGAAGGATCGAGAAATGGGTTCTAAGGAGAGCCTTTGACGACGAAGAACGTCCTT
ATCTACCAAAACACATTCTCTACAGACAGAAAGAACAGTTCAGTGATGGTGTTGGCTACAGTTGGA
TCGATGGCCTGAAAGATCACGCTGCTCAAAATGTCAATGACAAGATGATGTCGAACGCGGGGCATA
TCTTCCCTCACAACACTCCAAACACTAAAGAAGCTTACTACTACAGAATGATCTTTGAAAGGTTCT
TCCCGCAGAACTCTGCGAGACTAACGGTTCCTGGAGGTGCCACCGTGGCTTGTTCGACTGCAAAGG
CAGTGGAGTGGGATGCAAGCTGGTCCAACAATATGGATCCATCAGGAAGAGCCGCTATCGGAGTTC
ACCTTTCGGCCTACGATGGCAAGAACGTGGCATTGACCATACCACCACTTAAGGCAATTGACAACA
TGCCGATGATGATGGGTCAAGGAGTTGTGATTCAGTCATAACTTCGAAGGAGAAATGGATGAAATA

FIGURE 5 (continued)

TGTGTTATATCTTCCCAATGGGTGAAGTGTTTTGTATGATTTTAAAAATAAGAATGTGATCCTTTT
TTTTTCCTATGAAGATCTGAATGTATAATCTATCTTGTAAAAATTTGTTTCTTTGTAAGATTTGAA
TGTACCGCTTTTACGTAGATCGATGTACATCAATCTTATAAGTTTCAATTATGTATTATATTATGT
CGATTTGCCAAAAATAAATCTAAAACCTC

SEQ ID NO: 111, Arabidopsis thaliana - AT5G10240.1#1

TCCATTTCTCTGAAGCCGTTGTGTTCTCTTATTGCCGCCACCACCACCATGTGTGGGATTCTCGCT
GTGTTAGGCTGCGTCGATAACTCTCAAGCTAAACGTTCCCGTATCATCGAACTCTCTCGCAGATTG
AGGCATAGAGGTCCTGACTGGAGTGGTCTACATTGTTATGAGGATTGTTATTTGGCTCATGAGCGT
TTGGCTATCGTTGACCCCACTTCTGGAGATCAACCACTCTATAACGAAGATAAGACCATTGCTGTC
ACGGTCAATGGAGAGATTTACAACCACAAGGCTTTGCGTGAAAATTTGAAGTCTCACCAATTCCGT
ACTGGGAGTGATTGTGAAGTGATTGCCCATCTTTACGAAGAACATGGAGAGGAATTTGTCGACATG
TTGGATGGCATGTTTGCATTTGTGCTTCTTGATACCCGAGACAAAAGCTTTATTGCTGCAAGGGAT
GCCATTGGTATCACTCCACTCTACATCGGGTGGGGTCTCGATGGTTCTGTTTGGTTTGCTTCCGAG
ATGAAAGCACTTAGTGATGATTGTGAGCAGTTTATGTGCTTCCCCCCAGGCCACATCTATTCAAGT
AAACAAGGTGGGCTTAGGAGGTGGTACAACCCCCCGTGGTTCTCTGAGGTTGTTCCTTCAACCCCA
TATGATCCCCTAGTGGTGCGCAATACTTTTGAGAAGGCTGTTATAAAACGACTAATGACTGATGTG
CCTTTTGGTGTCCTCCTATCTGGTGGATTAGATTCATCCCTTGTTGCTTCAGTAGCATTACGCCAT
CTGGAAAAAATCAGAAGCTGCTTGTCAGTGGGGTTCAAAGTTGCACACATTTTGTATCGGTTTGAAG
GGATCCCCGGATCTTAAAGCTGGCAGAGAAGTCGCTGACTATTTAGGAACTCGCCACCACGAGTTA
CACTTTACAGTTCAGGACGGAATAGATGCCATAGAAGAAGTCATCTACCATGTTGAGACCTATGAT
GTGACTACTATTAGAGCCAGCACTCCAATGTTTCTTATGTCGCGAAAAATCAAATCGCTTGGTGTA
AAGATGGTTCTTTCTGGGGAAGGCTCTGATGAAATTTTTGGAGGATATTTGTACTTCCATAAAGCT
CCCAACAAGAAGGAATTTCATGAGGAAACATGTCGAAAGATCAAAGCTCTTCATCAATATGACTGC
TTGAGGGCTAACAAATCAACTTCTGCATGGGGTGTTGAGGCTCGTGTACCTTTCCTCGATAAAGAA
TTTATAAATGTCGCAATGAGCATCGATCCAGAGTGGAAGATGATTAGGCCTGATTTGGGAAGGATC
GAGAAATGGGTGTTACGCAATGCCTTTGATGATGAGAAAAATCCTTACCTACCAAAGCACATTCTA
TATAGGCAGAAAGAACAGTTCAGTGATGGAGTTGGATACAGCTGGATTGATGGTCTAAAAGATCAT
GCAAACAAACATGTCTCTGAGACAATGCTGATGAACGCAAGCTTTGTCTTCCCTGATAACACACCT
TTGACAAAAGAAGCTTACTACTACAGAACCATCTTTGAAAAGTTCTTCCCTAAGAGTGCTGCTAGA
GCAACTGTACCAGGAGGTCCAAGTGTGGCATGTAGCACAGCAAAAGCTGTGGAATGGGACGCAGCT
TGGTCACAGAATCTTGACCCATCAGGTCGTGCGGCTCTTGGAGTTCATGTTTCAGCTTATGGGGAA
GATAAAACCGAAGATTCTCGTCCCGAGAAGCTACAGAAACTAGCAGAGAAGACTCCAGCCATTGTT
TGAGGATAAACAAACAAGGTTTCAGCTAATGTTGAATCGTGCAATACTCTTATTGTCTCAAAGACA
ATAGATATCCTCTTCTATAGGTTCTAAAAAGGCTTTCTTTTTTTCTTGTTTTCTGGGGTTCTTTGG
ATGTGTACCTAATAAGTTCCTGGTGAATTTCTGTGTTTAGTGTTATTAGACAATCCATGAAAGCTT
GATACTTCAGATTATGAACGTTATTTTCATGAATCCGATTCTTTCTTTC

SEQ ID NO: 112, Triticum aestivum - TA71252 - 4565#1

GCGACGTGTAGCCCTGCTCTCCGCCATCTCCGGCCAGGCATCTATCTACCTACAAGTAGAGCCAAG
CCATTCCTGCACACCTCCATACAGAAACACAATTCAGATCGACTAGCTCGCTGCTGGCTGTAGAGG
ACGATCGACGACGATCCAGAGGAGCAGCATAACCGAGGAGAGCGGAGCATGTGCGGCATACTAGCG
GTGCTGGGGTGCGGCGACGAGTCGCAGGGGAAGAGGGTCCACGTGCTAGAGCTCTCGCGCAGGCTC
AAGCACCGGGGCCCGGACTGGAGCGGCCTGCACCAGGTCGCCGACAACTACCTCTGCCACCAGCGC
CTCGCCATCATCGACCCGGCCTCCGGCGACCAGCCGCTCTACAACGAGGACAAGTCCATCGCCGTT
GCCGTCAACGGGGAGGTCTACAACCATGAGGAGCTTCGGGCACGGCTCTCCGGACACAGGTTCCGG
ACCGGCAGTGACTGCGAGGTCATCGCCCATCTGTATGAGGAATACGGAGAAAGCTTCATTGACATG
TTGGATGGTGTTTTCTCCTTCGTGTTACTTGACGCACGAGATAACAGCTTCATTGCTGCTCGTGAT

FIGURE 5 (continued)

GCCATTGGTGTCACGCCTCTCTACATTGGCTGGGGAATTGATGGTTCAGTGTGGATATCTTCGGAG
ATGAAAGGACTAAACGATGATTGTGAGCACTTCGAGATCTTCCCGCCTGGTAATCTTTACTCCAGC
AAAGAGAAGTCCTTCAAGAGATGGTATAACCCTCCTTGGTTCTCTGAGGTCATCCCCTCGGTTCCC
TATGACCCACTGCGTCTCAGATCGGCATTTGAAAAGGCTGTTATCAAGAGGCTCATGACAGATGTT
CCATTTGGCGTCCTCCTCTCCGGTGGTCTCGACTCATCATTGGTGGCTGCTGTCGCAGCCCGTCAT
TTCGCTGGGACGAAGGCTGCAAAGCGCTGGGGAACTAGGCTCCACTCCTTCTGTGTGGGGCTTGAG
GGATCACCAGATCTCAAGGCTGCAAAGGAGGTCGCGGATCACCTGGGTACCGTGCACCACGAGTTC
AACTTCACAGTTCAGGATGGCATCGATGCAATTGAAGATGTGATATACCACATTGAAACATATGAT
GTGACGACGATCAGGGCAAGCACACTGATGTTCCAGATGTCACGCAAGATCAAGGCGCTTGGAGTC
AAGATGGTCATCTCCGGTGAGGGTGCCGATGAGATCTTCGGAGGGTACTTGTATTTCCACAAGGCC
CCTAACAAGGAGGAGTTCCACCAGGAAACATGTCGGAAGATAAAAGCTCTCCATCAGTACGATTGC
TTGAGGGCCAACAAAGCAACATCTGCATGGGGCCTTGAGGTTCGTGTGCCATTCTTGGACAAGGAG
TTCATCAATGAGGCTATGAGCATAGATCCCGAATGGAAGATGATCCGGCCTGATCTTGGAAGAATT
GAGAAATGGATACTGAGGAAAGCGTTCGATGATGAGGAGCGACCCTTCCTGCCGAAGCATATTCTG
TACAGGCAGAAGGAGCAGTTTAGTGATGGTGTTGGGTATAGCTGGATTGATGGCCTGAAGGATCAT
GCAGCCTCAAATGTGAGTGATAAGATGATGTCCAATGCAAAGTTCATCTACCCACACAACACCCCA
ACAACTAAAGAGGCCTACTATTACAGGATGATCTTTGAGAGGTACTTCCCCCAGAGCTCGGCGATC
CTCACGGTGCCAGGCGGGCCAAGCGTGGCGTGCAGCACAGCCAAGGCTATAGAGTGGGATGCCCAA
TGGTCTGGGAACCTGGACCCCTCTGGGAGAGCAGCGCTTGGAGTCCATCTCTCAGCCTACGAGCAG
GACACGGTCGCTGTGGGAGGTAGCAACAAGCCTGGGGTGATGAACACCGTGGTACCTGGTGTTGCC
ATTGAGACTTGATGAATGGTACATGTATCATATCGTGTCCTACTAAAGGCAAATAAGAACGGTTGT
GTGCATCGCTTCATGTAGAGGCCGGGCATACTCCTTTTCAAAAAAAAAAGAGAAAATAAGATGCAT
ATGTTCTTGTCAGCGTTGTAATAAGACGGGCCTATGTTTTGCTATTTAATTAAAGGGTTAATTATC
CTTTTGCCTTGAGTGATGTCTGTGTGCTC

SEQ ID NO: 113, Triticum aestivum - TA54599 - 4565#1

GCACGAGGCCCATCCTCCTTCAGAAGCACAGAGAGAGATCTTCTAGCTACATACTGTTGCCGTCGA
TCCAGGAAAATGTGCGGCATACTGGCGGTGCTGGGCTGCGCTGATGACACCCAGGGGAAGAGAGTG
CGCGTGCTCGAGCTCTCGCGCAGGCTCAAGCACCGCGGCCCCGACTGGAGCGGCATGCACCAGGTT
GGCGACTGCTACCTCTCCCACCAGCGCCTCGCCATTATCGACCCTGCCTCTGGCGACCAGCCGCTC
TACAACGAGGACAAGTCCATCGTCGTCACAGTGAATGGAGAGATCTACAACCATGAACAGCTCCGG
GCGCAGCTCTCCTCCCACACGTTCAGGACAGGCAGCGACTGCGAGGTCATCGCACACCTGTACGAG
GAGCATGGGGAGAACTTCATCGACATGCTGGATGGTGTCTTCTCCTTCGTCTTGCTCGATACACGC
GACAACAGCTTCATTGCTGCACGTGATGCCATTGGCGTCACACCCCTCTATATTGGCTGGGGAATT
GATGGGTCGGTATGGATATCATCAGAGATGAAGGGCCTGAATGATGATTGTGAGCACTTTGAGATC
TTTCCTCCTGGCCATCTCTACTCCAGCAAGCAGGGAGGCTTCAAGAGATGGTACAACCCACCTTGG
TTCTCCGAGGTCATTCCTTCAGTGCCATATGACCCACTTGCTCTCAGGAAGGCTTTCGAAAAGGCT
GTCATCAAGAGGCTTATGACGGACGTTCCATTCGGTGTTCTACTCTCTGGTGGCCTTGACTCATCA
TTGGTTGCAGCCGTTACAGTTCGTCACCTGGCAGGAACAAAGGCTGCAAAGCGCTGGGGGACTAAG
CTTCACTCTTTTTGTGTCGGACTTGAGGGGTCACCTGATCTGAAGGCTGCAAAGGAGGTAGCCAAT
TACCTGGGCACCATGCACCATGAGTTCACCTTCACTGTTCAGGACGGCATTGATGCAATTGAGGAT
GTGATTTATCACACCGAAACATATGATGTGACGACAATCAGGGCAAGCACGCCAATGTTCCTGATG
TCACGCAAGATCAAGTCACTTGGCGTCAAGATGGTCATCTCTGGTGAGGGTTCCGATGAGATTTTC
GGAGGGTACCTCTACTTCCACAAGGCACCCAACAAAGAGGAGCTCCACCGTGAGACATGTCAAAAG
ATCAAAGCTCTGCATCAGTACGATTGCTTGAGGGCCAACAAGGCAACATCTGCATGGGGCCTCGAA
GCACGTGTGCCATTCTTGGACAAGGAGTTTATCAATGAGGCAATGAGCATTGATCCTGAGTGGAAG
ATGATCCGGCCTGATCTTGGAAGAATTGAGAAATGGATGCTGAGGAAAGCATTTGATGACGAGGAG
CAACCATTCCTGCCGAAGCACATTCTGTACAGGCAGAAAGAGCAGTTCAGTGATGGTGTTGGCTAC

FIGURE 5 (continued)

AGCTGGATTGATGGCCTAAAGGCTCACGCAGAATCAAATGTGACAGATAAGATGATGTCAAATGCA
AAGTTCATCTACCCACACAACACCCCGACTACAAAAGAGGCCTACTGTTACAGGATGATATTTGAG
AGGTTCTTCCCCCAGAACTCGGCGATCCTGACGGTGCCAGGTGGGCCAAGCGTTGCATGCAGCACG
GCGAAGGCGGTAGAGTGGGATGCCCAGTGGTCAGGGAACCTGGATCCCTCAGGGAGAGCAGCACTT
GGAGTCCATCTCTCGGCCTATGAACAGGAGCATCTCCCAGCAACCATCATGGCAGGAACCAGCAAG
AAGCCAAGGATGATCGAGGTTGCGGCGCCTGGTGTCGCAATTGAGAGTTGATGGTGTCCTGTCCTG
CTTGCCGTTTCTGATAAGAAATAAGATGTACCTGGTCTTGCCATTAGAGTGGTGCAGACCTAAGGT
TTGAGTGAAGATTGTGCATTAATGTTTCTATTGTTCTTATGAAATCGGAGACCGGTGATTTCTAAT
CCTTTCTGGCAACTTCCATCAAAACATTATTACATGATGGTTATTATTTGAC

SEQ ID NO: 114, Vitis vinifera - GSVIVT00024074001#1
ATGTGCGGAATACTTGCAGTTCTGGGTTGTTCTGATGATTCCCAGGCCAAAAGGGTCCGATTGTTT
TACCATTGTTATTTATGCTTCTGTGATAGGTTGAAGCATCGTGGTCCTGACTGGAGTGGGCTATAC
CAACATGGAGATTGTTATTTAGCTCATCAGCGGCTAGCAATCATCGATCCAGCTTCTGGTGATCAG
CCTCTATATAATGAAAACCAAGCCATTGTTGTGACGGTGAATGGAGAAATTTATAACCATGAGGAG
TTGAGGAAGAGCATGCCAAATCACAAGTTCAGGACCGGGAGCGATTGCGATGTCATTGCCCATTTG
TACGAGGAGCATGGGGAAAATTTTGTGGACATGTTGGATGGAATGTTCTCATTTGTCCTGCTGGAT
ACCCGTGATGATAGCTTCATTGTTGCCCGAGATGCCATCGGAATCACCTCCCTCTATATTGGTTGG
GGACTTGATGGTAGCTCGGTATGGATTTCATCTGAGCTCAAAGGTTTGAATGATGACTGTGAACAT
TTTGAGAGCTTTCCACCTGGTCACATGTACTCTAGCAAAGAGGGTGGATTCAAAAGATGGTACAAC
CCCCCTTGGTTCTCTGAGGCTATTCCATCGGCACCATATGACCCTCTTGTTCTGAGGCGAGCTTTT
GAGAATGCCGTGATCAAGAGGTTAATGACCGATGTTCCTTTTGGGGTTCTGCTGTCAGGAGGTCTG
GATTCATCCTTAGTTGCCTCTATTACCGCTCGCCACTTAGCAGGCACAAAGGCTGCTAAGCAGTGG
GGAGCACAGCTCCATTCCTTCTGTGTTGGGCTAGAGGGCTCACCGGATCTGAAGGCTGCAAAAGAA
GTTGCAGACTATTTGGGCACCGTTCACCACGAGTTTCACTTCACCGTTCAGGATGGTATCGATGCC
ATTGAGGATGTTATTTACCATATTGAAACTTATGATGTGACAACGATCCGAGCAAGTACCCCTATG
TTTCTCATGTCGCGTAAGATTAAGTCACTAGGAGTGAAGATGGTGATATCCGGAGAGGGCTCTGAT
GAGATTTTTGGTGGGTACTTATACTTTCACAAGGCGCCCAACAAGGAAGAGTTCCATAGGGAAACA
TGTCGCAAGATAAAGGCACTCTACCAGTATGATTGCTTGAGAGCTAATAAATCAACATCTGCATGG
GGTTTGGAAGCCCGGGTCCCCTTTTTAGACAAGGAATTCATTAAAGTTGCAATGGATATTGACCCT
GAGTGGAAGATGATAAAGCCAGAACAAGGGCGAATTGAGAAATGGGTTCTGAGGAGGGCTTTTGAT
GATGAGGAACAACCCTATCTGCCAAAGCATATTCTCTACAGGCAAAAAGAGCAATTCAGTGATGGT
GTCGGCTACAGTTGGATTGATGGGCTCAAAGCCCATGCGTCACAACATGTGACCGATAAAATGATG
CTCAATGCTTCACATATCTTCCCACACAATACCCCTACCACAAAAGAAGCCTACTATTACCGAATG
ATCTTTGAGAGGTTCTTCCCACAGAACTCAGCTAGGCTGACTGTTCCGGGAGGAGCAAGCGTTGCA
TGCAGCACTGCCAAAGCAGTTGAATGGGATTCTGCGTGGTCAAATAACCTTGATCCTTCTGGCAGG
GCGGCATTAGGAGTCCATCTTTCAGCTTATGACCAGAAGTTAACCACAGTCAGTGCTGCAAATGTG
CCAACAAAGATCATTGATAATATGCCGCGGATTATGGAAGTAACCGCACCCTGA

SEQ ID NO: 115, Volvox carteri - 65699 - e - gw1.50.7.1#1
ATGTGCGGAATCCTAGCTGTGCTCAACTCCACGGATGATAGCCCGGCGATGAGGGCGAAGGTGCTG
GCGCTTAGTCGTCGCCAGAAGCATCGTGGCCCCGACTGGTCGGGGATGCACCAGTTTGGCAACAAC
TTCCTGGCGCATGAGCGGCTTGCGATTATGGATCCCAGCTCGGGCGATCAGCCGCTGTACAACGAG
GACAAGTCTATCGTCGTGACGGTGAACGGCGAGATCTACAATTATAAGGAACTGCGCAAGGAGATC
TCTGACAAGTGCCCTGGCAAGAAGTTCCGCACCAACAGCGACTGTGAGGTGATCAGCCACCTGTAC
GAACTGTACGGCGAGGCAGTTGCCAACAAGCTGGACGGCTTCTTTGCCTTTGTACTGCTGGACACT
CGCAACAACACCTTCTTCGCGGCGCGCGATCCGTTGGGCGTCACCTGCATGTACATTGGCTGGGGC
CGGGATGGCAGCGTGTGGCTGTCCTCCGAGATGAAATGTCTCAAGGACGACTGTGCGCGCTTCCAG

FIGURE 5 (continued)

CAATTCCCTCCCGGCCATTATTACTCGTCCAAGACAGGCGAGTTTGTGCGGTACTTCAACCCCCAG
TTTTACCTGGACTTTGAGGCAGAGCCGCAGGTTTTCCCCTCGGTGCCCTACGACCCCGTCACGTTG
CGCACGGCGTTTGAGGCGGCCGTGGAGAAGCGCATGATGAGCGACGTGCCCTTCGGTGTGCTGCTG
AGTGGCGGTCTGGACAGCAGCCTTGTGGCCTCTATCGCGGCCCGCAAAATCAAGCGGGAGGGCAGT
GTGTGGGGCAAGCTGCACAGCTTCTGCGTTGGTCTGGAGGGCAGCCCCGACCTCAAGGCAGGTGCC
GCTGTGGCTGAGTTTCTGGGCACCGACCACCACGAGTTCCACTTTACAGTGCAGGAGGGCATTGAC
GCCATCTCGGAGGTCATTTACCACATCGAGACGTTTGACGTGACCACGATCCGCGCCTCCACTCCC
ATGTTCCTGATGAGCCGCAAGATCAAGGCCCTGGGTGTCAAGATGGTGCTGTCCGGAGAAGGCTCG
GATGAGGTGTTCGGGGGTTACCTCTACTTCCATAAGGCTCCCAGCAAGGATGAGTTCCACAGCGAA
ACGGTTCGCAAGCTGAAGGACCTGTTCAAGTACGACTGCCTGCGAGCCAACAAGGCCACCATGGCC
TGGGGTGTAGAGGCGCGTGTGCCCTTCCTGGACCGGGCATTCCTGGATGTGGCCATGTCCATTGAC
CCGGCGGAGAAGATGATTGACAAGAGCAAGGGCCGGATCGAGAAATACATTCTCCGGAAAGCCTTC
GATACGCCCGAGGATCCATACCTGCCTAAGGAGGTACTGTGGCGCCAGAAGGAGCAGTTCAGCGAC
GGCGTGGGCTACAACTGGATTGATGGGCTCAAGGCGCATGCTGAGAGCCAAGTCAGCGATGAGATG
CTCAAGAACGCCGTGCACAGATTCCCGGACAACACCCCGCGCACCAAGGAGGCCTACTGGTACCGC
TCTATCTTTGAGAGCCACTTCCCGCAGCGTGCTGCTATGGAGACGGTGCCGGGTGGTCCCTCAGTG
GCTTGCTCCACCGCGACAGCCGCCCTGTGGGATGCAGCGTGGGCCGGGAAGGAGGACCCGTCGGGC
CGCGCCGTGGCGGGCGTTCATGACGCTGCTTACGAGGAAGGCGCGGAAGCCAATGGCGAGCCCGCA
TCCAAAAAGCAAAAGGTCTGA

SEQ ID NO: 116, Zea mays - TA174465 - 4577#1

GATCGTCTCGTCTCCCTCCCAAAAAAAAAAAAAAAAACTGCTCGGTTGCTGCTCCTGCTCCGCCGC
GCCGGCATCATGTGTGGCATCTTAGCCGTGCTCGGTTGCTCCGACTGGTCTCAGGCAAAGAGGGCT
CGCATCCTCGCCTGCTCCAGAAGGTTGAAGCACAGGGGCCCCGACTGGTCGGGCCTCTACCAGCAC
GAGGGCAACTTCCTGGCGCAGCAGCGGCTCGCCGTCGTCTCCCCGCTGTCCGGCGACCAGCCGCTG
TTCAACGAGGACCGCACCGTCGTGGTGGTGGCCAATGGAGAGATCTACAACCACAAGAACGTCCGG
AAGCAGTTCACCGGCACACACAACTTCAGCACGGGCAGTGACTGCGAGGTCATCATCCCCCTGTAC
GAGAAGTACGGCGAGAACTTCGTGGACATGCTGGACGGGGTGTTCGCGTTCGTGCTCTACGACACC
CGCGACAGGACCTACGTGGCGGCGCGCGACGCCATCGGCGTCAACCCGCTCTACATCGGCTGGGGC
AGTGACGGTTCCGTCTGGATCGCGTCCGAGATGAAGGCGCTGAACGAGGACTGCGTGCGCTTCGAG
ATCTTCCCGCCGGGCCACCTCTACTCCAGCGCCGGCGGCGGGTTCCGGCGGTGGTACACCCCGCAC
TGGTTCCAGGAGCAGGTGCCCCGGATGCCGTACCAGCCGCTCGTCCTCAGAGAGGCCTTCGAGAAG
GCGGTCATCAAGAGGCTCATGACTGACGTCCCGTTCGGGGTCCTCCTCTCCGGCGGCCTCGACTCC
TCGCTCGTCGCCTCCGTCACCAAGCGCCACCTCGTCGAGACCGAGGCCGCCGAGAAGTTCGGCACC
GAGCTCCACTCCTTTGTCGTCGGCCTCGAGGGCTCTCCTGACCTGAAGGCCGCACGAGAGGTCGCT
GACTACCTTGGAACCATCCATCACGAGTTCCACTTCACCGTACAGGACGGCATCGACGCGATCGAG
GAGGTGATCTACCACGACGAGACGTACGACGTGACGACGATCCGGGCCAGCACGCCCATGTTCCTG
ATGGCTCGCAAGATCAAGTCGCTGGGCGTGAAGATGGTGCTGTCCGGGGAGGGCTCCGACGAGCTC
CTGGGCGGCTACCTCTACTTCCACTTCGCCCCCAACAAGGAGGAGTTCCACAGGGAGACCTGCCGC
AAGGTGAAGGCCCTGCACCAGTACGACTGCCTGCGCGCCAACAAGGCCACGTCGGCGTGGGGCCTG
GAGGTCCGCGTGCCGTTCCTCGACAAGGAGTTCATCAACGTCGCGATGGGCATGGACCCCGAATGG
AAAATGTACGACAAGAACCTGGGCCGCATCGAGAAGTGGGTCATGAGGAAGGCGTTCGACGACGAC
GAGCACCCTTACCTGCCCAAGCATATTCTCTACAGGCAGAAAGAACAGTTCAGTGACGGCGTTGGC
TACAACTGGATCGATGGCCTCAAATCCTTCACTGAACAGCAGGTGACGGATGAGATGATGAACAAC
GCCGCCCAGATGTTCCCCTACAACACGCCCGTCAACAAGGAGGCCTACTACTACCGGATGATATTC
GAGAGGCTCTTCCCTCAGGACTCGGCGAGGGAGACGGTGCCGTGGGGCCCGAGCATCGCCTGCAGC
ACGCCCGCGGCCATCGAGTGGGTGGAGCAGTGGAAGGCCTCCAACGACCCCTCCGGCCGCTTCATC
TCCTCCCACGACTCCGCCGCCACCGACCACACCGGCGGTAAGCCGGCGGTGGCCAACGGCGGCGGC

FIGURE 5 (continued)

CACGGCGCCGCGAACGGCACGGTCAACGGCAAGGACGTCGCAGTCGCGATCGCGGTCTGACGAGAG
TACGTGCTCGCGCACCTCCCTGCTAGCTTCTACCGGGCTGCAGCCTGCAGCATGCACTGTGCGAGC
ACAGCCGATCAGCGCCAATAAACTGGAGGATAAGAACGACTGGTAGGTGTGTGTGTGTGTCGTGCG
TGCCCACCGGCCATATCCCGGTGCGGCAGCACGTGCTATTGTTACGTGTTGTACTGCCGCCAGCGT
ACGTGTCTGTGTGTCTCGATCATATCTGTACGTTTTTAGATTTAGAAGAAAAAAAAAAAGGCATGTC
CGTGTCTGTATGTCTGGATCATATCTGTACGTTCTTAGATTTAGAAGAAAGAAGAAAAACATTATA
TACGTACGTCCATGTCTCT

SEQ ID NO: 117, Zea mays - X82849#1

ATGTGTGGGATTCTGGCGGTGCTGGGCGTCGTTGAGGTCTCCCTCGCCAAGCGCTCCCGCATCATT
GAGCTCTCGCGCAGGTTACGGCACCGAGGGCCTGATTGGAGTGGTTTGCACTGTCATGAGGATTGT
TACCTTGCACACCAGCGGTTGGCTATTATCGATCCTACATCTGGAGACCAGCCTTTGTACAATGAG
GATAAAACAGTTGTTGTAACGGTGAACGGCGAAATTTACAATCATGAAGAATTGAAAGCTAAGTTG
AAAACTCATGAGTTCCAAACTGGCAGTGATTGTGAAGTTATAGCCCATCTTTACGAAGAATATGGC
GAAGAATTTGTGGATATGTTGGATGGAATGTTCTCCTTTGTTCTTCTTGATACACGTGATAAAAGC
TTCATCGCAGCTCGTGATGCTATTGGCATCTGCCCTTTATACATGGGATGGGGTCTTGATGGATCA
GTCTGGTTTTCTTCAGAGATGAAGGCATTGAGTGATGATTGTGAACGCTTCATAACATTTCCCCCA
GGGCATCTCTACTCCAGCAAGACAGGTGGTCTAAGGAGATGGTACAACCCACCATGGTTTTCAGAG
ACTGTCCCTTCAACCCCTTACAATGCTCTCTTCCTCCGGGAGATGTTTGAGAAGGCTGTTATTAAG
AGGCTGATGACTGATGTGCCATTTGGTGTGCTTTTATCTGGTGGACTCGACTCTTCTTTGGTTGCA
TCTGTTGCTTCGCGGCACTTAAACGAAACAAAGGTTGACAGGCAGTGGGGAAATAAATTGCATACT
TTCTGTATAGGCTTGAAGGGTTCTCCTGATCTTAAAGCTGCTAGAGAAGTTGCTGATTACCTCAGC
ACTGTACATCATGAGTTCCACTTCACAGTGCAGGAGGGGATTGATGCCTTGGAAGAAGTCATCTAC
CATATTGAGACATATGATGTTACAACAATCAGAGCAAGTACCCCAATGTTTTTGATGTCACGCAAA
ATCAAATCTTTGGGTGTGAAGATGGTTATTTCTGGCGAAGGTTCAGATGAAATTTTTGGTGGTTAC
CTTTATTTTCACAAGGCACCAAACAAGAAAGAATTCCTAGAGGAAACATGTCGGAAGATAAAAGCA
CTACATCTGTATGACTGCTTGAGAGCTAACAAAGCAACTTCTGCCTGGGGTGTTGAGGCTCGTGTT
CCATTCCTTGACAAAAGTTTCATCAGTGTAGCAATGGACATTGATCCTGAATGGAACATGATAAAA
CGTGACCTCGGTCGAATTGAGAAGTGGGTCATGAGGAAGGCGTTCGACGACGACGAGCACCCTTAC
CTGCCCAAGCATATTCTCTACAGGCAGAAAGAACAGTTCAGTGACGGCGTTGGCTACAACTGGATC
GATGGCCTCAAATCCTTCACTGAACAGCAGGTGACGGATGAGATGATGAACAACGCCGCCCAGATG
TTCCCCTACAACACGCCCGTCAACAAGGAGGCCTACTACTACCGGATGATATTCGAGAGGCTCTTC
CCTCAGGACTCGGCGAGGGAGACGGTGCCGTGGGGCCCGAGCATCGCCTGCAGCACGCCCGCGGCC
ATCGAGTGGGTGGAGCAGTGGAAGGCCTCCAACGACCCCTCCGGCCGCTTCATCTCCTCCCACGAC
TCCGCCGCCACCGACCACACGGCGGTAAGCCGGCGGTGGCCAACGGCGGCGGCACGGCCGGCGAAC
GGCACGGTCAACGGCAAGGACGTGCCAGTGCCGATCGCGGTCTGACGAGAGTACGTGCTCGCGCAC
CTCCCTGCTAGCTTCTACCGGGCTGCAGCCTGCAGCATGCACTGTGCGAGCACAGCCGATCAGCGC
CAATAAACTGGAGGATAAGAACGACTGGTAGCTGTGTGTGTGTGTGTCGTGCGTGCCCACCGGCCA
TATCCCGGTGCGGCAGCACGTGCTATTGTTACGTGTTGTACTGCCACCAGCGTACGTGTCTGTGTG
TCTCGATCATATCTGTACGTTTTTAGATTTAGAAGAGAAAAAAAAAGTATGCCCGTGTCTGTATGT
CTGGATCATATCTGTACGTTCTTAGATTTAGAAGA

SEQ ID NO: 118, Zea mays - TA182904 - 4577#1

GGAATTCCCCGGGATCAAGGAGCACCGTCTGCTGCTCGCTCTATAAAACGAACGGAGGCTGCAGAG
CAGAGCAGAGCAGAGCAAGAAGCTTTACAGTGAACGAGTGAGTATGTGCGGCATACTTGCTGTGCT
CGGGTGCGCCGACGAGGCCAAGGGCAGCAGCAAGAGGTCCCGGGTGCTGGAGCTGTCGCGGCGGCT
GAAGCACCGGGGCCCCGACTGGAGCGGCCTCCGGCAGGTGGGCGACTGCTACCTCTCTCACCAGCG
CCTCGCCATCATCGACCCGGCCTCTGGCGACCAGCCCCTCTACAACGAGGACCAGTCGGTGGTCGT

FIGURE 5 (continued)

CGCCGTCAACGGCGAGATCTACAACCACCTGGACCTCAGGAGCCGCCTCGCCGGCGCAGGCCACAG
CTTCAGGACCGGCAGCGACTGCGAGGTCATCGCGCACCTGTACGAGGAGCATGGAGAAGAGTTCGT
GGACATGCTGGACGGCGTCTTCTCCTTCGTGCTGCTGGACACTCGCCATGGCGACCGCGCGGGCAG
CAGCTTCTTCATGGCTGCTCGCGACGCCATCGGTGTGACGCCCCTCTACATCGGATGGGGAGTCGA
TGGGTCGGTGTGGATTTCGTCGGAGATGAAGGCCCTGCACGACGAGTGTGAGCACTTCGAGATCTT
CCCTCCGGGGCATCTCTACTCCAGCAACACCGGCGGATTCAGCAGGTGGTACAACCCTCCTTGGTA
CGACGACGACGACGACGAGGAGGCCGTCGTCACCCCCTCCGTCCCCTACGACCCGCTGGCGCTAAG
GAAGGCGTTCGAGAAGGCCGTGGTGAAGCGGCTGATGACAGACGTCCCGTTCGGCGTCCTGCTCTC
CGGCGGGCTGGACTCGTCGCTGGTGGCGACCGTCGCCGTGCGCCACCTCGCCCGGACAGAGGCCGC
CAGGCGCTGGGGCACCAAGCTCCACTCCTTCTGCGTGGGCCTGGAGGGGTCCCCTGACCTCAAGGC
GGCCAGGGAGGTGGCGGAGTACCTGGGCACCCTGCACCATGAGTTCCACTTCACTGTTCAGGACGG
CATCGACGCCATCGAGGACGTGATCTACCACACGGAGACGTACGACGTCACCACGATCAGGGCGAG
CACGCCCATGTTCCTCATGTCGCGCAAGATCAAGTCGCTCGGGGTCAAGATGGTCATCTCCGGCGA
GGGCTCCGACGAGCTCTTCGGAGGCTACCTCTACTTCCACAAGGCGCCCAACAAGGAGGAGTTGCA
CCGAGAGACGTGTAGGAAGGTTAAGGCTCTGCATCAGTACGACTGCCTGAGAGCCAACAAGGCGAC
ATCAGCTTGGGGCCTGGAGGCTCGCGTCCCGTTCCTGGACAAGGAGTTCATCAATGCGGCCATGAG
CATCGATCCTGAGTGGAAGATGGTCCAGCCTGATCTTGGAAGGATTGAGAAGTGGGTGCTGAGGAA
GGCATTCGACGACGAGGAGCAGCCATTCCTGCCCAAGCATATCCTCTACAGACAGAAGGAGCAGTT
CAGTGACGGCGTTGGGTACAGCTGGATCGATGGCCTGAAGGCTCATGCAACATCAAATGTGACTGA
CAAGATGCTGTCAAATGCAAAGTTCATCTTCCCACACAACACTCCGACCACCAAGGAGGCCTACTA
CTACAGGATGGTCTTCGAGAGGTTCTTCCCACAGAAATCTGCTATCCTGACGGTACCTGGTGGGCC
AAGTGTGGCGTGCAGCACAGCCAAGGCCATCGAGTGGGACGCACAATGGTCAGGAAATCTGGACCC
CTCGGGAAGGGCGGCACTGGGCGTCCATCTCGCCGCCTACGAACACCAACATGATCCCGAGCATGT
CCCGGCGGCCATTGCAGCAGGAAGCGGCAAGAAGCCAAGGACGATTAGGGTGGCACCGCCTGGCGT
TGCCATCGAGGGATAGACGACGACGCATATATAAGCTTCCTACTTTTGTTTCAATGCATGCATGCT
ATGTATCTGTGTCCACCGGCTGTCTAGCCTTATCATCATCACTGTCTGCAACAAATTAATAATCAA
GTGGTATGGGGTACCTACGTTTAATGTATACGGAGTATTGTATTGCTTGTGTGTGGTATGCTTAGG
TTGGCCGTGAGTAAGGGATTACAAGTATTCGATATCGGGTGTTTCTATAGGTTGAAGTGCTCATAA
AGGGCTCCCTATCCTCTATGGTCATGTTTGTAATAGTTTTTTTTCTTAAAGAGCTTTTCTATGAAT
TTGGATTCCTGTT

SEQ ID NO: 119, Zea mays - TA11549 - 4577999#1
CGAGCGCTCAGCGTCTCGTCTCCTCCTCCCCACAAAAAGCCGCTGAATTGCTCCGTCGGCGTCATG
TGTGGCATCTTAGCCGTGCTCGGATGCTCCGACTGCTCCCAGGCCAGGAGGGCTCGCATCCTCGCC
TGCTCCAGAAGGCTGAAGCACAGGGGCCCCGACTGGTCGGGCCTCTACCAGCACGAGGGCAACTTC
CTGGCGCAGCAGCGGCTCGCCATCGTCTCCCCGCTGTCCGGCGACCAGCCGCTGTTCAACGAGGAC
CGCACCGTCGTGGTGGTGGCCAATGGAGAGATCTACAACCACAAGAACGTCCGGAAGCAGTTCACC
GGCGCGCACAGCTTCAGCACCGGCAGTGACTGCGAGGTCATCATCCCCCTGTACGAGAAGTACGGC
GAGAACTTCGTGGACATGCTGGACGGAGTCTTCGCGTTCGTGCTCTACGACACGCGAGACAGGACC
TACGTGGCGGCACGCGACGCCATCGGCGTCAACCCGCTCTACATCGGCTGGGGCAGCGACGGTTCC
GTCTGGATGTCGTCCGAGATGAAGGCGCTGAACGAGGACTGCGTGCGCTTCGAGATCTTCCCGCCG
GGGCACCTCTACTCCAGCGCCGCCGGCGGGTTCCGCCGGTGGTACACCCCGCACTGGTTCCAGGAG
CAGGTGCCCCGGACGCCGTACCAGCCGCTCGTCCTTAGAGAGGCCTTCGAGAAGGCGGTTATCAAG
AGGCTCATGACCGACGTCCCGTTCGGGGTCCTCCTCTCCGGCGGCCTCGACTCCTCCCTCGTCGCC
TCCGTCACCAAGCGCCACCTCGTCAAGACCGACGCCGCCGGAAAGTTCGGCACAGAGCTCCACTCC
TTCGTCGTCGGCCTCGAGGGCTCCCCTGACCTGAAGGCCGCACGAGAGGTCGCTGACTACCTCGGA
ACCACCCATCACGAGTTCCATTTCACCGTACAGGACGGCATCGACGCGATCGAGGAGGTGATCTAC
CACGACGAGACGTACGACGTGACGACGATCCGGGCCAGCACGCCCATGTTCCTGATGGCTCGCAAG

FIGURE 5 (continued)

ATCAAGTCGCTGGGCGTGAAGATGGTGCTGTCCGGGGAGGGCTCCGACGAGCTCCTGGGCGGCTAC
CTCTACTTCCACTTCGCCCCCAACAGGGAGGAGCTCCACAGGGAGACCTGCCGCAAGGTGAAGGCC
CTGCACCAGTACGACTGCCTGCGCGCCAACAAGGCGACGTCGGCGTGGGGCCTGGAGGTCCGCGTG
CCGTTCCTCGACAAGGAGTTCGTCGACGTCGCGATGGGCATGGACCCCGAATGGAAAATGTACGAC
AAGAACCTGGGTCGCATCGAGAAGTGGGTCCTGAGGAAGGCGTTCGACGACGAGGAGCACCCTTAC
CTGCCCGAGCATATTCTGTACAGGCAGAAAGAACAGTTCAGTGACGGAGTGGGCTACAACTGGATC
GATGGACTCAAAGCCTTCACCGAACAGCAGGTTGATGGTCGTCGTCGAAGTTAGCTAACCAGCGCT
GACGTTCCCCCCCATGTCCAGGTGACGGATGAGATGATGAACAGCGCCGCCCAGATGTTCCCGTAC
AACACGCCCGTCAACAAGGAGGCCTACTACTACCGGATGATATTCGAGAGGCTCTTCCCTCAGGAC
TCGGCGAGGGAGACGGTGCCGTGGGGCCCGAGCATCGCCTGCAGCACGCCCGCGGCCATCGAGTGG
GTGGAGCAGTGGAAGGCCTCCAACGACCCCTCCGGCCGCTTCATCTCCTCCCACGACTCCGCCGCC
ACCGACCGCACCGGAGACAAGCTGGCGGTGGTCAACGGCGACGGGCACGGCGCGGCGAACGGCACG
GTCAACGGCAACGACGTCGCTGTCGCCATCGCGGTGTAACAGTAATGAACTGGAGGATAGGGACGA
ACGAACGACTGGTAGGTGTGGCGTACCTGCCGCGTGCCCACCGGCCGGCCATATATATCGAATCCC
GGCCCGGCGCGGCAGCACGTGCTATTGTTACGTGTCACCAGCGTACGTGTCTGTGTAGTGCCTCGA
TCGTATCTGTACGTCTTTAGGAAAAGGTGTGTCCGTGTGTATTGTATGTGTGTGAGCAAGCGTGCG
TGACGCGCTCTGCCTGTGTGACAAAGCAGAGCAGTACAAGCTCAGGCATTTTCTGTCCGAGCGATG
ATTTGAACTGGATCTATCATCTCTGAATTGAACTCGGCCGGACGACGACCTACCGCTAAAATTATT
CCCAGCTGGATTTCGGTACGTGTCCCCGTTGTTCGTTCTCGCGGCTGTGACTGTGACCGAACCTGC
TGCTACAAGTGCGCGTAAAGGATCTGGTTCCACGTGTCCGGCACGCCGGGCACGCACCAGTGGATG
CAGTCCGTGTACGTCTGCGGGTCGGCCCTCTGCGCGTCGGTGAGCAGCTCGCCGCCGGTCTCGGTG
TACACGGAGACGTGGGCGTCGACGCGGTGCTCCGTCAGCTGCGTGACGTTGAGCAGCGTCACGGGC
ACCCGCATCCGGCCCACCACGTCGGACATCACCTCCATCATCCGCCGGTCCGCGCCGCTGCCCCAG
TACCCCTTCCGCGTGACGGGCAACGTCTCGTTGTAGCACCGGATGCCGCCCTCCCGGCCCCAGTCC
TCGCTCCTCATGTGCGTGGTGGAGATCGACATGAAGAAGACCCTTGTGGCGTTGGGATCGATGTTG
GCGTCCACCCAGTTGGCCCATGTCTTGAGTCCAAGCCGGAACGCGACCCAGGCGTCCAGCTCCTCG
TACCCGTCGTCCCCGAACGACCCCCACACTGATTTGATCCTGCTGCCGGTCATCCACCACACGTAG
CTGTCGAAGACGAGGATGTCCACGCCCTTCCAGTGCCTAGCGTGCAGCTCGACGGCGTCGACGTGG
AGCACGCGGCCGTCGGCCCCCAGCCGGATGTTGCGGTCGGAGTTGGCCTCCACCAGGTACGGCGCC
CAGTAGAACTCGATCGTCGCGTTGTACTCCGTGGCGGTGAAGACGGACAGGGTGGTGCTGCGCTCC
ATGGACCGCGCGGTGTAGGGCACGGCGGAGTTGACGAGGCAGACCATGGAGAGCCACTGCCCCATC
TGCAGCGAGTCGCCAACGAACATCATCCGCTTCCCCCGCAGCGTCTCCAGCAACGCCACCGGGTCG
AACCTTGGGAGACTGCAGTCGTCTAGGTGCCAATCCCAGCGCAGGTAGTCGCTGTCCGGCCTGCCG
TTCCTCTGGCAAGAGACCTGCCTGTCGATGAACGGGCATGTTTGGTCCGTGTAAAGCAGCTCCTTG
GACCTGTTGTACGCCCAGTACCCCTCCGTCACGCTGCACCGGCTCGGGTCGAACGCTGCCTTCGCC
GGCTGCGGCGGCGGCGTTAGCGGCATCTTCCTCGTCGTCGTCCCGGCATGTAGAGACGTCGTCCTC
TTGTGCTCCTTCGCTTTCCCCTTCTTCTCCATGATCTCAGTGAGTGAGCGGAGGTCGTCGGTGAAG
ATGACGCCCGCTAGAGCCAGCCCGCCGATGATTGCCACCACCACTGACAGCGGGGCCCGCCCCTTC
ATCCGCTTCACCGCTGCTATCTGAATCTGAACCATGAAGCTCAGATGCTACGTGGATGCTGGCATG
CAGCAATGCTAGCTTGTTGCAGGCTCAAGGTGTGAGACGGCTTATCGATTTATTTGCAGCTGCTCT
TTGT

SEQ ID NO: 120, Zea mays - TA15078 - 4577999#1

CCGAGGCGGCGCTTTTGGGGTCGGAAGCGACACGGGCGCCGGGCGGGTCCGCGGGTGGTGGTGCTA
CTGCTAGCAAGCAGCAGCAGGCGACGCTAGGCGAGAGCCCCAGTCGGAGCAGGCCACCATGTGCGG
CATCCTCGCTGTCCTCGGCGTCGCTGAGGTCTCCCTCGCCAAGCGCTCCCGCATCATTGAGCTCTC
GCGCAGGTTACGGCACCGAGGGCCTGATTGGAGTGGTTTGCACTGTCATGAGGATTGTTACCTTGC
ACACCAGCGGTTGGCTATTATCGATCCTACATCTGGAGACCAGCCTTTGTACAATGAGGATAAAAC

FIGURE 5 (continued)

AGTTGTTGTAACGGTGAACGGAGAGATCTATAACCATGAAGAATTGAAAGCTAAGTTGAAAACTCA
TGAGTTCCAAACTGGCAGTGATTGTGAAGTTATAGCCCATCTTTACGAAGAATATGGCGAAGAATT
TGTGGATATGTTGGATGGAATGTTCTCCTTTGTTCTTCTTGATACACGTGATAAAAGCTTCATCGC
AGCTCGTGATGCTATTGGCATCTGCCCTTTATACATGGGATGGGGTCTTGATGGATCAGTCTGGTT
TTCTTCAGAGATGAAGGCATTGAGTGATGATTGTGAACGCTTCATAACATTTCCCCCAGGGCATCT
CTACTCCAGCAAGACAGGTGGTCTAAGGAGATGGTACAACCCACCATGGTTTTCAGAGACGGTCCC
TTCAACCCCTTACAATGCTCTCTTCCTCCGGGAGATGTTTGAGAAGGCTGTTATTAAGAGGCTGAT
GACTGATGTGCCATTTGGTGTGCTTTTATCTGGTGGACTCGACTCTTCTTTGGTTGCATCTGTTGC
TTCGCGGCACTTTAACGAAACAAAGGGTGACAGGCAGTGGGGAAATAAATTGCATACTTTCTGTAT
AGGCTTGAAGGGTTCTCCTGATCTTAAAGCTGCTAGAGAAGTTGCTGATTACCTCAGCACTGTACA
TCATGAGTTCCACTTCACAGTGCAGGAGGGCATTGATGCCTTGGAAGAAGTCATCTACCATATTGA
GACATATGATGTTACAACAATCAGAGCAAGTACCCCAATGTTTTTGATGTCACGCAAAATCAAATC
TTTGGGTGTGAAGATGGTTATTTCTGGCGAAGGTTCAGATGAAATTTTTGGTGGTTACCTTTATTT
TCACAAGGCACCAAACAAGAAAGAATTCCATGAGGAAACATGTCGGAAGATAAAAGCACTACATCT
GTATGACTGCTTGAGAGCTAACAAAGCAACTTCTGCCTGGGGTGTTGAGGCTCGTGTTCCATTCCT
TGACAAAAGTTTCATCAGTGTAGCAATGGACATTGATCCTGATTGGAAGATGATAAAACGTGACCT
CGGTCGAATTGAGAAATGGGTTATCCGTAATGCATTTGATGATGATGAGAGGCCCTATTTACCTAA
GCACATTCTCTACAGGCAAAAGGAACAGTTCAGTGATGGTGTTGGGTATAGTTGGATCGATGGATT
GAAGGACCATGCCAGCCAACATGTCTCCGATTCCATGATGATGAATGCTGGCTTTGTTTACCCAGA
GAACACACCCACAACAAAAGAAGGGTACTACTACAGAATGATATTCGAGAAATTCTTTCCCAAGCC
TGCAGCAAGGTCAACTGTTCCTGGAGGTCCTAGTGTGGCCTGCAGCACTGCCAAAGCTGTTGAATG
GGACGCATCCTGGTCCAAGAACCTTGATCCTTCTGGCCGTGCTGCTTTGGGTGTTCACGATGCTGC
GTATGAAGACACTGCAGGGAAAACTCCTGCCTCTGCTGATCCTGTCTCAGACAAGGGCCTTCGTCC
AGCTATTGGCGAAAGCCTAGGGACACCCGTTGCTTCAGCCACAGCTGTCTAACCTTATGTTTATCA
CCCAGCAATGCTTGAAACAGCAAAGGTTGTCCATTGCTTGTTTCAGTTTCCTTCCGATCATGTTTT
TAGTTCCATCAATCAAGCAATGGAGACATGCTTGTGCTTCATACTTGGCAGCATCGTGTTTGGGTT
TTCACTGGGCAGTACTGTTTAATTTTTATGGACTGAAAAGACTCAGTTTTGTAAATATTCGTCACT
GTGACCAATTCCTGTGGTGGTTTATGTGATTTGCAGATTGCAGTGGTTAGTGTATCTTCCNCAATT
TTCACTCCTTT

SEQ ID NO: 121, Brassica napus - P3_BPS4LI_BN06M@BN06MC14360 43814276@14314#1

GAATTCTCCGGGTCGACGATTTCGTACGAAATCGTCATTGCCGCCACCATCCATCAACCATGTGTG
GGATTCTCGCTGTTCTAGGCTGCGTCGATAACTCTCAAGCCACACGTTCTCGTATCATCAAACTCT
CTCGCAGATTGAGGCATAGAGGTCCTGATTGGAGCGGGCTTCATTGTTATGAGGATTGTTACTTGG
CTCATGAGCGTTTGGCCATCATTGACCCCATTTCTGGAGACCAGCCTCTCTACAGCGAAGATAAGA
CCGTCGTTGTCACGGTGAATGGAGAGATATACAATCACAAGGCATTGCGTGAAAGTGAAAGTCTGA
AGTCTCACAAGTACCATACCGGGAGTGATTGTGAAGTGCTTGCCCATCTTTATGAAGAACATGGAG
AGGAATTTATCAACATGTTGGACGGCATGTTTGCATTTGTCCTTCTTGATACTAAGGACAAAAGTT
ATATTGCTGTAAGGGATGCCATTGGTGTCATCCCACTCTACATTGGCTGGGGTCTCGATGGTTCTG
TCTGGTTTGCTTCTGAGATGAAAGCACTTAGTGATGATTGTGAACAGTTTATGGCTTTCCCACCAG
GCCACATCTATTCCAGTAAACAAGGTGGTCTTAGGAGGTGGTACAACCCTCCATGGTTCTCTGAGC
TCGTTCCTTCAACCCCTTATGATCCCTTAGTATTGCGAGATACTTTCGAGAAGGCTGTAATAAAGA
GACTAATGACCGATGTGCCTTTTGGTGTCCTACTCTCTGGAGGACTAGACTCATCTCTTGTTGCTT
CAGTGGCTATACGCCATTTGGAAAAGTCAGATGCTCGTCAGTGGGGTTCCAAGCTGCACACCTTTT
GCATTGGTTTAAAGGGATCTCCGGATCTTAAAGCTGGTAAAGAAGTTGCTGACTATCTAGGAACTC
GCCACCACGAGCTCCACTTTACAGTTCAGGAAGGGATAGACGCCATAGAAGAAGTTATATACCATG
TTGAGACCTATGACGTGACTACCATAAGAGCAAGCACTCCCATGTTTCTCATGTCGAGAAAAATCA

FIGURE 5 (continued)

AATCGCTTGGTGTGAAGATGGTTCTCTCTGGTGAAGGCTCTGATGAGATCTTTGGAGGGTATTTGT
ACTTCCACAAAGCACCTAACAAGAAGGAGTTACACGAGGAAACATGCCGAAAGATCAAAGCACTTT
ATCAATATGATTGCTTGAGGGCTAACAAATCAACTTCTGCGTGGGGTGTTGAGGCTCGTGTGCCTT
TCCTTGATAAAGCGTTTCTAGATGTAGCAATGGGCATTGATCCAGAGTGGAAGATGATCAGGCCTG
ACTTGGGAAGGATTGAGAAATGGGTGTTACGCAATGCCTTTGATGATGAGAAGAATCCTTATCTAC
CAAAGCACATTCTGTACAGGCAGAAGGAACAGTTCAGTGATGGAGTTGGATACAGCTGGATTGACG
GTCTGAAAGATCATGCAAACAAACATGTCTCTGACGCAATGCTGACGAACGCAAACTTTGTCTTCC
CGGAGAACACACCTTTGACAAAGGAGGCTTACTACTACAGAGCCATCTTTGAAAAGTTCTTCCCTA
AGAGCGCTGCTAGAGCAACTGTACCAGGAGGTCCAAGTGTAGCATGTAGTACTGCAAAAGCTGTGG
AGTGGGACGCAGCTTGGAAAGGGAACCTTGACCCGTCGGGTCGTGCGGCTCTTGGAGTTCATGTTG
CAGCTTATGAAGGAGATAAAGCTGAAGATCCTCGTCCTGAGAAGGTACAGAAGCTGGCAGAGAAAA
CTGCAGAAGCCATTGTTTGAGGATGAAACGAATGTTTGAGTCGTGCGTTTCTTTTATTTTCTCATA
AGACAATACGTTATTATCATCTTCCGTAGGATCAATAAGTACAATAAGTTGTCTCTCTTTAACTGA
ATTGAGGTGGGAGTGTCTGAGGTTGTACCTAAGTTGTTGGTGATTTTCTGGTTCTTTCATTTGTCA
CAAAGTTTTCAGCGTTTCTTTTATGTATGATGTATCGTTCACCCCTGTTAATCTAGATTTGGTTCA
GTTCAAAAAAAAAAAAAAAAAAAAGCGGACGCTCTAGA

SEQ ID NO: 122, Triticum aestivum – BPS_Hyseq_TA Wheat@c54713691@13255#1

GGCCTGGCCCGCTACGAACCCCAAACGCGCATCTCTCCTAGCCCCCTCCCTGCTGCTCTACCACCA
CCGTGCCGCCGTAGAACGCCGTACCTGACCCCCCACCACCACCTGCGCCTGCGTCGCCGCCGGCGC
CGTCGCCGTCGCCCGTCCGTACTAGTCGGGGCATCGCCGGTGATTAGTCAAATCACCTTCGGAGCT
CGCGACCACCCAAATCACCCGCGGAGTCTCGCCAACGAGCAGGGACCGCCCGCCGGCCGCCACCAT
GTGCGGCATCCTCGCCGTCCTCGGCGTCGGCGACGTCTCCCTCGCCAAGCGCTCCCGCATCATCGA
GCTCTCCCGCCGATTACGGCACAGAGGCCCTGATTGGAGTGGTATACACAGCTTTGAGGATTGCTA
TCTTGCACACCAGCGGTTGGCTATTGTTGATCCTACATCTGGAGACCAGCCATTGTACAACGAGGA
CAAAACAGTTGTTGTGACGGTGAATGGAGAGATCTATAATCATGAAGAACTGAAAGCTAAGCTGAA
ATCTCATCAATTCCAAACTGGTAGTGATTGTGAAGTTATTGCTCACCTATATGAGGAATACGGGGA
GGAATTTGTGGATATGCTGGATGGCATGTTCTCGTTTGTGCTTCTTGACACACGTGATAAAAGCTT
CATTGCTGCCCGTGATGCTATTGGCATCTGTCCTTTGTACATGGGCTGGGGTCTTGATGGGTCAGT
TTGGTTTTCTTCAGAGATGAAGGCATTGAGTGATGATTGCGAGCGCTTCATATCGTTCCCTCCTGG
ACACTTGTACTCAAGCAAAACAGGTGGCCTAAGGAGGTGGTACAACCCCCCATGGTTTTCAGAAAG
CATTCCCTCAGCCCCCTATGATCCTCTCCTCATCCGAGAGAGTATTGAGAAGGCTGCTATTAAGAG
GCTAATGACTGATGTGACATTTGGCGTTCTCTTGTCTGGTGGGCTTGACTCTTCTTTGGTGGCTTC
TGTTGTTTCACGCTACTTGGCAGAAACAAAAGTTGCTAGGCAGTGGCGAAACAAACTGCACACCTT
TTGCATCGGCATGAAGGGTTCTCCTGATCTTAAAGCTGCTAAGGAAGTTGCTGACTACCTTGGCAC
AGTCCATCATGAATTACACTTCACAGTGCAGGAGGGCATTGATGCTTTGGAAGAAGTTATATATCA
CATCGAGACGTATGATGTCACGACCATTAGAGCAAGTACCCCAATGTTTCTAATGTCTCGGAAAAT
CAAATCGTTGGGTGTGAAGATGGTTCTTTCGGGAGAAGGCTCCGATGAAATATTTGGTGGTTATCT
TTATTTTCACAAGGCACCAAACAAAAAGGAACTACATGAGGAAACATGTAGGAAGATAAAAGCTCT
CCATTTATATGATTGTTTGAGAGCGAACAAAGCAACTTCTGCCTGGGGTCTCGAGGCTCGTGTTCC
ATTCCTCGACAAAAACTTCATCAATGTAGCAATGGACCTGGATCCGGAATGTAAGATGATAAGACG
TGATCTTGGCCGGATCGAGAAATGGGTTCTGCGTAATGCATTTGATGATGAGGAGAAGCCCTATTT
ACCCAAGCACATTCTTTACAGGCAAAAAGAACAATTCAGCGATGGGGTTGGGTACAGTTGGATTGA
TGGATTGAAGGACCATGCTAAAGCACATGTGTCGGATTCCATGATGACGAACGCCAGCTTTGTTTA
CCCTGAAAACACACCCACAACAAAAGAGGCCTACTATTACAGGACCGTATTCGAGAAGTTCTATCC
CAAGAATGCTGCTAGGCTAACGGTGCCAGGAGGTCCCAGCATCGCATGCAGCACCGCTAAAGCTGT
CGAATGGGACGCCGCCTGGTCCAAGCTCCTCGACCCGTCTGGCCGCGCCGCTCTTGGCGTGCACGA

FIGURE 5 (continued)

TGCGGCGTACAAAGAAAAGGCTCCTGCATCGGTCGATCCTGCCGTGGATAACGTCTCACGTTCACC
TGCACATGACGTCAAAAGACTCAAAACCGCCATTTCAGCAGCTGCTGTATAACCTTCCATTCCATG
GTTCCAAAAATGCCGTCGCTTAGTTTTAATCCTAGCAATCCTGTCTGTAGTTCATTCAGTCATGCA
GTGCAGAAATCGCTTTGCTCTACTTTTTCGTTCATGTTGTGCTTTCGCATGTATGTACCAAGTTAG
TTTGTTTATGCAGCGAGCGTTTGCGTCGTAAATAAATATTTCACCGTGGTTGATATCCTTGTGTTG
CTCAGTGTTTGGTTTGCAAGCTGCAAATTGCACTAATAAATTCC

FIGURE 5 (continued)

```
SEQID63 MCGILAVLGVADVSLAKRSRIIELSRRLRHRGPDWSGIHCYQDCYLAHQRLAIVDPTSGDQPLYNEDKSVVVTVNGEIYN 80
SEQID67 MCGILAVLGVADVSLAKRSRIIELSRRLRHRGPDWSGIHCYQDCYLAHQRLAIVDPTSGDQPLYNEDKSVVVTVNGEIYN 80

SEQID63 HEELKANLKSHKFQTASDCEVIAHLYEEYGEEFVDMLDGMFAFVLLDTRDKSFIAARDAIGICPLYMGWGLDGSVWFSSE 160
SEQID67 HEELKANLKSHKFQTASDCEVIAHLYEEYGEEFVDMLDGMFAFVLLDTRDKSFIAARDAIGICPLYMGWGLDGSVWFSSE 160

SEQID63 MKALGDDCERFISFPPGHLYSSKTGGLRRWYNPPWFSESIPSTPYNPLLLRQSFEKAIIKRLMTDVPFGVLLSGGLDSSL 240
SEQID67 MKALSDDCERFISFPPGHLYSSKTGGLRRWYNPPWFSESIPSTPYNPLLLRQSFEKAIIKRLMTDVPFGVLLSGGLDSSL 240

SEQID63 VASVVSRHLAEAKVAAQWGNKLHTFCIGLKGSPDLRAAKEVADYLGTVHHELHFTVQEGIDALEEVIYHVETYDVTTIRA 320
SEQID67 VASVVSRHLAEAKVAAQWGNKLHTFCIGLKGSPDLRAAKEVADYLGTVHHELHFTVQEGIDALEEVIYHVETYDVTTIRA 320

SEQID63 STPMFLMSRKIKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLGANKSTSAWGVEARVPFLD 400
SEQID67 STPMFLMSRKIKSLGVKMVLSGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLRANKSTSAWGVEARVPFLD 400

SEQID63 KNFINVAMDIDPEWKMIKRDLGRIEKWVLRNAFDDEEKPYLPKHILYRQKEQFSDGVGYSWIDGLKDHANEHVSDSMMMN 480
SEQID67 KNFINVAMDIDPEWKMIKRDLGRIEKWVLRNAFDDEEKPYLPKHILYRQKEQFSDGVGYSWIDGLKDHANEHVSDSMMMN 480

SEQID63 ASFVYPENTPVTKEAYYYRTIFEKFFPKNAARLTVPGGPSVACSTAKAVEWDAAWSKNLDPSGRAALGVHDAAYEDTLQK 560
SEQID67 ASFVYPENTPVTKEAYYYRTIFEKFFPKNAARLTVPGGPSVACSTAKAVEWDAAWSKNLDPSGRAALGVHDAAYEDTLQK 560

SEQID63 SPASANPVLDNGFGPALGESMVKTVASATAV 591
SEQID67 SPASANPVLDNGFGPALGESMVKTVASATAV 591
```

FIGURE 6

```
CLUSTAL 2.0.3 multiple sequence alignment

Pt829702            MCGILAVLGCS-DDSQAKRFR-----VLELSRRLKHRGPDWSGLFQHGD--FYLAHQRLA
Vv00024074001       MCGILAVLGCS-DDSQAKRVRLFYHCYLCFCDRLKHRGPDWSGLYQHGD--CYLAHQRLA
GmU77679            MCGILAVLGCS-DSSQAKRVR-----VLELSRRLKHRGPDWSGLHQYGD--NYLAHQRLA
BoTA5921            MCGILALLGCS-DDSQAKRVR-----VLELSRRLRHRGPDWSGIYQNGF--NYLAHQRLA
BN06MC14360         MCGILAVLGCV-DNSQATRSR-----IIKLSRRLRHRGPDWSGLHCYED--CYLAHERLA
AT5G65010           MCGILAVLGCI-DNSQAKRSR-----IIELSRRLRHRGPDWSGLHCYED--CYLAHERLA
Os06g0265000#1      MCGILAVLGVA-DVSLAKRSR-----IIELSRRLRHRGPDWSGIHCYQD--CYLAHQRLA
ZmTA15078           MCGILAVLGVA-EVSLAKRSR-----IIELSRRLRHRGPDWSGLHCHED--CYLAHQRLA
Ta13255             MCGILAVLGVG-DVSLAKRSR-----IIELSRRLRHRGPDWSGIHSFED--CYLAHQRLA
GmTA51197           MCGILAVLGCV-DNSQTKRAR-----IIELSRRLRHRGPDWSGIHCYED--CYLAHQRLA
Pp180723            MCGILAILGADGAVPSAGRDR-----ALALSRRLRHRGPDWSGLFEGKDSWCYLAHERLA
Vc65699             MCGILAVLNST-DDSPAMRAK-----VLALSRRQKHRGPDWSGMHQFGN--NFLAHERLA
Cr140252            MCGILAVLNTT-DDSQAMRSR-----VLALSRRQRHRGPDWSGMHQFGN--NFLAHERLA
                    ******:*.      . : * :       : :. * :********:.       :***:***

Pt829702            IIDPASGDQPLFNEDQAIVVTVNGEIYNHEELR----KRLPNHKFRTGSDCDVISHLYEE
Vv00024074001       IIDPASGDQPLYNENQAIVVTVNGEIYNHEELR----KSMPNHKFRTGSDCDVIAHLYEE
GmU77679            IVDPASGDQPLFNEDKTVVVTVNGEIYNHEELR----KQLPNHTFRTGSDCDVIAHLYEE
BoTA5921            IIDPDSGDQPLFNEDKSIVVTVNGEIYNHEELR----KGLKNHKFHTGSDCDVIAHLYEE
BN06MC14360         IIDPISGDQPLYSEDKTVVVTVNGEIYNHKALRE--SESLKSHKYHTGSDCEVLAHLYEE
AT5G65010           IIDPTSGDQPLYNEDKTVAVTVNGEIYNHKILR----EKLKSHQFRTGSDCEVIAHLYEE
Os06g0265000#1      IVDPTSGDQPLYNEDKSVVVTVNGEIYNHEELK----ANLKSHKFQTASDCEVIAHLYEE
ZmTA15078           IIDPTSGDQPLYNEDKTVVVTVNGEIYNHEELK----AKLKTHEFQTGSDCEVIAHLYEE
Ta13255             IVDPTSGDQPLYNEDKTVVVTVNGEIYNHEELK----AKLKSHQFQTGSDCEVIAHLYEE
GmTA51197           IVDPTSGDQPLYNEDKTIIVTVNGEIYNHKQLR----QKLSSHQFRTGSDCEVIAHLYEE
Pp180723            IIDPASGDQPLYNGTKDIVVAANGEIYNHELLK----KNMKPHEYHTQSDCEVIAHLYED
Vc65699             IMDPSSGDQPLYNEDKSIVVTVNGEIYNYKELRKEISDKCPGKKFRTNSDCEVISHLYEL
Cr140252            IMDPASGDQPLFNEDRTIVVTVNGEIYNYKELRQQITDACPGKKFATNSDCEVISHLYEL
                    *:** ******:.  : : *:.******:: *:          : : * ***:*::****

Pt829702            YGENFVDMLDGMFSFVLLDTRDNSFIVARDAIGITSLYIGWGLDG-SVWISSELKGLNDD
Vv00024074001       HGENFVDMLDGMFSFVLLDTRDDSFIVARDAIGITSLYIGWGLDGSSVWISSELKGLNDD
GmU77679            HGENFMDMLDGISSFVLLDTRDNSFIVARDAIGVTSLYIGWGLDG-SVWISSELKGLNDD
BoTA5921            HGENFVDMLDGIFSFVLLDTRDNSFMVARDAVGVTSLYIGWGLDG-SLWVSSEMKGLHED
BN06MC14360         HGEEFINMLDGMFAFVLLDTKDKSYIAVRDAIGVIPLYIGWGLDG-SVWFASEMKALSDD
AT5G65010           HGEEFIDMLDGMFAFVLLDTRDKSFIAARDAIGITPLYIGWGLDG-SVWFASEMKALSDD
Os06g0265000#1      YGEEFVDMLDGMFAFVLLDTRDKSFIAARDAIGICPLYMGWGLDG-SVWFSSEMKALSDD
ZmTA15078           YGEEFVDMLDGMFSFVLLDTRDKSFIAARDAIGICPLYMGWGLDG-SVWFSSEMKALSDD
Ta13255             YGEEFVDMLDGMFSFVLLDTRDKSFIAARDAIGICPLYMGWGLDG-SVWFSSEMKALSDD
GmTA51197           HGEEFVNMLDGMFAFILLDTRDKSFIAARDAIGITPLYLGWGHDG-STWFASEMKALSDD
Pp180723            VGEEVVNMLDGMWSFVLVDSRDNSFIAARDPIGITPLYLGWGADGRTVWFASEMKALKDD
Vc65699             YGEAVANKLDGFFAFVLLDTRNNTFFAARDPLGVTCMYIGWGRDG-SVWLSSEMKCLKDD
Cr140252            HGEKVASMLDGFFAFVVLDTRNNTFYAARDPIGITCMYIGWGRDG-SVWLSSEMKCLKDD
                     **  .  . ***: :*:::*:::.:: ..**.:*:  :*:*** ** : *.:**:* * :*

Pt829702            CEHFKCFPPGHIYSSKSGG-LRRWYNPLWFSEA------IPSTPYDPLALRRAFEKAVIK
Vv00024074001       CEHFESFPPGHMYSSKEGG-FKRWYNPPWFSEA------IPSAPYDPLVLRRAFENAVIK
GmU77679            CEHFESFPPGHLYSSKERA-FRRWYNPPWLSLA------IPSAPYDPLALRHAFEKLWIK
BoTA5921            CEHFEAFPPGHLYSSKSGGGFKQWYNPPWFNES------VPSTPYEPLAIRSAFEDAVIK
BN06MC14360         CEQFMAFPPGHIYSSKQGG-LRRWYNPPWFSEL------VPSTPYDPLVLRDTFEKAVIK
AT5G65010           CEQFMSFPPGHIYSSKQGG-LRRWYNPPWYNEQ------VPSTPYDPLVLRNAFEKAVIK
Os06g0265000#1      CERFISFPPGHLYSSKTGG-LRRWYNPPWFSES------IPSTPYNPLLLRQSFEKAIIK
ZmTA15078           CERFITFPPGHLYSSKTGG-LRRWYNPPWFSET------VPSTPYNALFLREMFEKAVIK
Ta13255             CERFISFPPGHLYSSKTGG-LRRWYNPPWFSES------IPSAPYDPLLIRESIEKAAIK
GmTA51197           CERFISFPPGHIYSSKQGG-LRRWYNPPWFSED------IPSTPYDPTLLRETFERAVVK
Pp180723            CERLEVFPPGHIYSSKAGG-LRRYYNPQWFSETF-----VPETPYQPLELRSAFEKAVVK
Vc65699             CARFQQFPPGHYYSSKTGE-FVRYFNPQFYLDFEAEPQVFPSVPYDPVTLRTAFEAAVEK
Cr140252            CTRFQQFPPGHFYNSKTGE-FTRYYNPKYFLDFEAKPQRFPSAPYDPVALRQAFEQSVEK
                    * ::  ***** *.**     : :::** :          .*..**:.  :*  :*    *
```

FIGURE 7

```
Pt829702             RLMTDVPFGVLLSGGLDSSLVAAVTARHLAGTQAARQWGAHLHSFCVGLENSPDLKAARE
Vv00024074001        RLMTDVPFGVLLSGGLDSSLVASITARHLAGTKAAKQWGAQLHSFCVGLEGSPDLKAAKE
GmU77679             RLMTDVPFGVLLSGGLDSSLVAAVTARYLAGTKAAKQWGTKLHSFCVGLEGAPDLKATKE
BoTA5921             RLMTDVPFGVLLSGGLDSSLVASITARHLAGTKAAKRWGPQLHSFCVGLEGSPDLKAGKE
BN06MC14360          RLMTDVPFGVLLSGGLDSSLVASVAIRHLEKS-DARQWGSKLHTFCIGLKGSPDLKAGKE
AT5G65010            RLMTDVPFGVLLSGGLDSSLVAAVALRHLEKSEAARQWGSQLHTFCIGLQGSPDLKAGRE
Os06g0265000#1       RLMTDVPFGVLLSGGLDSSLVASVVSRHLAEAKVAAQWGNKLHTFCIGLKGSPDLRAAKE
ZmTA15078            RLMTDVPFGVLLSGGLDSSLVASVASRHFNETKGDRQWGNKLHTFCIGLKGSPDLKAARE
Ta13255              RLMTDVTFGVLLSGGLDSSLVASVVSRYLAETKVARQWRNKLHTFCIGMKGSPDLKAAKE
GmTA51197            RMMTDVPFGVLLSGGLDSSLVAAVVNRYLAESESARQWGSQLHTFCIGLKGSPDLKAAKE
Pp180723             RLMTDVPFGVLLSGGLDSSLVASVAARHLAETKAVRIWGNELHSFCVGLEGSPDLKAARE
Vc65699              RMMSDVPFGVLLSGGLDSSLVASIAARKIKRE--GSVWG-KLHSFCVGLEGSPDLKAGAA
Cr140252             RMMSDVPFGVLLSGGLDSSLVASIAARKIKRE--GSVWG-KLHSFCVGLPGSPDLKAGAQ
                     *:*:**.***************::. * :        *   .**:**:*: .:***:*

Pt829702             VADYLGTIHHEFHFTVQDGIDAIEDVIYHVETYDVTTIRASTPMFLLARKIKALGVKMVI
Vv00024074001        VADYLGTVHHEFHFTVQDGIDAIEDVIYHIETYDVTTIRASTPMFLMSRKIKSLGVKMVI
GmU77679             VAEYIGTVHHEFHYTVQDGIDAIEDVIYHIETYDVTTIRASIPMFLMSRKIKSLGVKWVI
BoTA5921             VAEYLGTVHHEFHFTVQDGIDAIEDVIYHVETYDVTTIRASTPMFLMSRKIKSLGVKMVL
BN06MC14360          VADYLGTRHHELHFTVQEGIDAIEEVIYHVETYDVTTIRASTPMFLMSRKIKSLGVKMVL
AT5G65010            VADYLGTRHHEFQFTVQDGIDAIEEVIYHIETYDVTTIRASTPMFLMSRKIKSLGVKMVL
Os06g0265000#1       VADYLGTVHHELHFTVQEGIDALEEVIYHVETYDVTTIRASTPMFLMSRKIKSLGVKMVL
ZmTA15078            VADYLSTVHHEFHFTVQEGIDALEEVIYHIETYDVTTIRASTPMFLMSRKIKSLGVKMVI
Ta13255              VADYLGTVHHELHFTVQEGIDALEEVIYHIETYDVTTIRASTPMFLMSRKIKSLGVKMVL
GmTA51197            VADYLGTRHHELYFTVQEGIDALEEVIYHIETYDVTTIRASTAMFLMSRKIKALGVKMVL
Pp180723             VAKYIGTRHHEFNFTVQEGLDALSDVIYHVETYDVTTIRASTPMFLMTRKIKALGVKMVL
Vc65699              VAEFLGTDHHEFHFTVQEGIDAISEVIYHIETFDVTTIRASTPMFLMSRKIKALGVKMVL
Cr140252             VAEFLGTDHHEFHFTVQEGIDAISEVIYHIETFDVTTIRASTPMFLMSRKIKALGVKMVL
                     **.::.* ***: :***:*:**:.:****:**:********  .***::****:**** *:

Pt829702             SGEGSDEIFGGYLYFHKAPNKEELHGETCRKIKALHQYDCLRANKATSAWGLEARVPFLD
Vv00024074001        SGEGSDEIFGGYLYFHKAPNKEEFHRETCRKIKALYQYDCLRANKSTSAWGLEARVPFLD
GmU77679             SGEGSDVFFGGYLYFHKAPNKEEFHQETCRTIIVLHRYDCSRANKSTFVWGLEARVPFLD
BoTA5921             SGEGSDEIFGGYLYFHKAPNKQEFHQETCRKIKALHKYDCLRANKATSAFGLEARVPFLD
BN06MC14360          SGEGSDEIFGGYLYFHKAPNKKELHEETCRKIKALYQYDCLRANKSTSAWGVEARVPFLD
AT5G65010            SGEGSDEILGGYLYFHKAPNKKEFHEETCRKIKALHQFDCLRANKSTSAWGVEARVPFLD
Os06g0265000#1       SGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLRANKSTSAWGVEARVPFLD
ZmTA15078            SGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLRANKATSAWGVEARVPFLD
Ta13255              SGEGSDEIFGGYLYFHKAPNKKELHEETCRKIKALHLYDCLRANKATSAWGLEARVPFLD
GmTA51197            SGEGSDEIFGGYLYFHKAPNKKEFHEETCRKIKALHLYDCLRANKSTAAWGVEARVPFLD
Pp180723             SGEGSDEIFGGYLYFHKAPNREEFHHELVRKIKALHLYDCQRANKSTSAWGLEARVPFLD
Vc65699              SGEGSDEVFGGYLYFHKAPSKDEFHSETVRKLKDLFKYDCLRANKATMAWGVEARVPFLD
Cr140252             SGEGSDEVFGGYLYFHKAPNKEEFQSETVRKIQDLYKYDCLRANKSTMAWGVEARVPFLD
                     ****** .:**********.:.*:: *  *.:  *. :** ****:* .:*:********

Pt829702             KDFINVAMAIDPEWKMIKP--GRIEKWVLRKAFDDEEHPYLPKHILYRQKEQFSDGVGYS
Vv00024074001        KEFIKVAMDIDPEWKMIKPEQGRIEKWVLRRAFDDEEQPYLPKHILYRQKEQFSDGVGYS
GmU77679             KEFIRVAMNIDPECKMIKKEEGRIEKWALRRAFDDEEHPYLPKHILYRQKEQFSDGVGYG
BoTA5921             KEFINTAMSLDPESKMIKPEEGRIEKWVLRRAFDDEERPYLPKHILYRQKEQFSDGVGYS
BN06MC14360          KAFLDVAMGIDPEWKMIRPDLGRIEKWVLRNAFDDEKNPYLPKHILYRQKEQFSDGVGYS
AT5G65010            KEFLNVAMSIDPEWKLIKPDLGRIEKWVLRNAFDDEERPYLPKHILYRQKEQFSDGVGYS
Os06g0265000#1       KNFINVAMDIDPEWKMIKRDLGRIEKWVLRNAFDDEEKPYLPKHILYRQKEQFSDGVGYS
ZmTA15078            KSFISVAMDIDPDWKMIKRDLGRIEKWVIRNAFDDDERPYLPKHILYRQKEQFSDGVGYS
Ta13255              KNFINVAMDLDPECKMIRRDLGRIEKWVLRNAFDDEEKPYLPKHILYRQKEQFSDGVGYS
GmTA51197            KEFINVAMSIDPEWKMIRPDLGRIEKWVLRNAFDDDKNPYLPKHILYRQKEQFSDGVGYS
Pp180723             KEFMDVAMMIDPSEKMIRKDLGRIEKWVLRKAFDDEERPYLPKHILYRQKEQFSDGVGYS
Vc65699              RAFLDVAMSIDPAEKMIDKSKGRIEKYILRKAFDTPEDPYLPKEVLWRQKEQFSDGVGYN
Cr140252             RHFLDVAMEIDPAEKMIDKSKGRIEKYILRKAFDTPEDPYLPNEVLWRQKEQFSDGVGYN
                     : *: .** :**  *:*    *****: :*.***  : ****:.:*:***********.
```

FIGURE 7 (continued)

```
Pt829702         WIDGLKAHAELHVHDKMMQNAEHIFPHNTPTTKEAYYYRMIFERFFPQNSARLTVPGGAS
Vv00024074001    WIDGLKAHASQHVTDKMMLNASHIFPHNTPTTKEAYYYRMIFERFFPQNSARLTVPGGAS
GmU77679         WIDGLKAHAEKHVTDRMMLNAANIFPFNTPTTKEAYHYRMIFERFFPQNSCRLTVPGGTS
BoTA5921         WIDGLKAHAAENVNDKMMSKAAFIFPHNTPLTKEAYYYRMIFERFFPQNSARLTVPGGAT
BN06MC14360      WIDGLKDHANKHVSDAMLTNANFVFPENTPLTKEAYYYRAIFEKFFPKSAARATVPGGPS
AT5G65010        WIDGLKDHANKHVSDTMLSNASFVFPDNTPLTKEAYYYRTIFEKFFPKSAARATVPGGPS
Os06g0265000#1   WIDGLKDHANEHVSDSMMMNASFVYPENTPVTKEAYYYRTIFEKFFPKNAARLTVPGGPS
ZmTA15078        WIDGLKDHASQHVSDSMMMNAGFVYPENTPTTKEGYYYRMIFEKFFPKPAARSTVPGGPS
Ta13255          WIDGLKDHAKAHVSDSMMTNASFVYPENTPTTKEAYYYRTVFEKFYPKNAARLTVPGGPS
GmTA51197        WIDGLKDHANKQVTDATMMAANFIYPENTPTTKEGYLYRTIFEKFFPKNAAKATVPGGPS
Pp180723         WIDGLKEYAESHVTDQMMKHAKHVYPFNTPNTKEGYYYRMIFEKHFPQQSARMTVPGGPS
Vc65699          WIDGLKAHAESQVSDEMLKNAVHRFPDNTPRTKEAYWYRSIFESHFPQRAAMETVPGGPS
Cr140252         WIDGLKAHADSQVSDDMMKTAAHRYPDNTPRTKEAYWYRSIFETHFPQRAAVETVPGGPS
                 ****** :*  :* *  :  *   :* *** ***.* ** :** .:*: :.  *****.:

Pt829702         VACSTAKAVEWDASWSNNLDPSGRAALGVHLSAYE---------QQAALASAGVVPPEII
Vv00024074001    VACSTAKAVEWDSAWSNNLDPSGRAALGVHLSAYD---------QKLTTVSAANVPTKII
GmU77679         VACSTAKAVEWDAAWSNNLDPSGRAALGVHASAYG---------NQVKAVE----PEKII
BoTA5921         VACSTAKAVEWDASWSNNMDPSGRAAIGVHLSAYD---------GSKVALPLP-APHKAI
BN06MC14360      VACSTAKAVEWDAAWKGNLDPSGRAALGVHVAAYE-------------GDKAEDPRPEKV
AT5G65010        IACSTAKAVEWDATWSKNLDPSGRAALGVHVAAYE-------------EDKAAAAAKAGS
Os06g0265000#1   VACSTAKAVEWDAAWSKNLDPSGRAALGVHDAAYEDTLQKSPASANPVLDNGFG-PALGE
ZmTA15078        VACSTAKAVEWDASWSKNLDPSGRAALGVHDAAYEDTAGKTPASADPVSDKGLR-PAIGE
Ta13255          IACSTAKAVEWDAAWSKLLDPSGRAALGVHDAAYK---EKAPASVDPAVDNVSRSPAHDV
GmTA51197        VACSTAKAVEWDAAWSKNLDPSGRAALGIHDAAYD------------AVDTKIDEPKNGT
Pp180723         VACSTATAVAWDEAWANNLDPSGRAALGCHDSAYTDKHSEKAAPAAEANGTASHENGHTF
Vc65699          VACSTATAALWDAAWAGKEDPSGRAVAGVHDAAYE------------EGAEANGEPASKK
Cr140252         VACSTATAALWDATWAGKEDPSGRAVAGVHDSAYD------------AAAAANGEPAAKK
                 :*****.*. ** :*    ******. * * :**

Pt829702         DNLPRMMKVGAPGVAIQS
Vv00024074001    DNMPRIMEVTAP------
GmU77679         P----KMEVSPLGVAI--
BoTA5921         DDIPMMMGQEVVIQT---
BN06MC14360      QKLAEKTAEAIV------
AT5G65010        DLVDPLPKNGT-------
Os06g0265000#1   SMVKTVASATAV------
ZmTA15078        SLGTPVASATAV------
Ta13255          KRLKTAISAAAV------
GmTA51197        L-----------------
Pp180723         SKPKSTLDATILKTQAVH
Vc65699          QKV---------------
Cr140252         AKK---------------
```

FIGURE 7 (continued)