



(86) Date de dépôt PCT/PCT Filing Date: 2012/11/12
(87) Date publication PCT/PCT Publication Date: 2013/05/23
(85) Entrée phase nationale/National Entry: 2014/04/14
(86) N° demande PCT/PCT Application No.: IB 2012/056346
(87) N° publication PCT/PCT Publication No.: 2013/072832
(30) Priorités/Priorities: 2011/11/14 (EP11189031.5);
2011/11/14 (US61/559,202)

(51) Cl.Int./Int.Cl. *C12N 15/29* (2006.01),
A01H 5/00 (2006.01), *C12N 15/82* (2006.01),
C12N 5/14 (2006.01)
(71) Demandeur/Applicant:
BASF PLANT SCIENCE COMPANY GMBH, DE
(72) Inventeurs/Inventors:
VAN CAMP, WIM, BE;
VANDENABEELE, STEVEN, BE;
LOUWERS, MARIEKE, BE;
MARTIN, JEROME, BE
(74) Agent: ROBIC

(54) Titre : PLANTES AYANT DES CARACTERES AMELIORES ASSOCIES AU RENDEMENT ET PROCEDE DE
FABRICATION ASSOCIE
(54) Title: PLANTS HAVING ENHANCED YIELD-RELATED TRAITS AND A METHOD FOR MAKING THE SAME

(57) **Abrégé/Abstract:**

A method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a GRP (Growth Related Protein) polypeptide is provided. Plants having modulated expression of a nucleic acid encoding a GRP polypeptide, which plants have enhanced yield-related traits relative to control plants, and constructs comprising the GRP-encoding nucleic acid, useful in performing the methods for enhancing yield-related traits in plants are also provided.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)
(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 May 2013 (23.05.2013)

(10) International Publication Number
WO 2013/072832 A1

- (51) **International Patent Classification:**
C12N 15/29 (2006.01) *A01H 5/00* (2006.01)
C12N 15/82 (2006.01) *C12N 5/14* (2006.01)
- (21) **International Application Number:**
PCT/IB2012/056346
- (22) **International Filing Date:**
12 November 2012 (12.11.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/559,202 14 November 2011 (14.11.2011) US
11189031.5 14 November 2011 (14.11.2011) EP
- (71) **Applicant (for all designated States except US):** **BASF PLANT SCIENCE COMPANY GMBH** [DE/DE]; 67056 Ludwigshafen (DE).
- (71) **Applicant (for MN only):** **BASF (CHINA) COMPANY LIMITED** [CN/CN]; 300 Jiangxinsha Road, Shanghai, 200137 (CN).
- (72) **Inventors; and**
- (71) **Applicants (for US only):** **Van Camp, Wim** [BE/BE]; Witbakkerstraat 14, B-9051 Sint-Denijs-Westrem (BE). **Vandenabeele, Steven** [BE/BE]; Meersport 23, B-9700 Oudenaarde (BE). **Louwers, Marieke** [BE/BE]; Prinses Clementinalaan 9, B-9000 Gent (BE). **Martin, Jerome** [FR/BE]; Weldadigheidstraat 16, B-9050 Gent (BE).
- (74) **Agent:** **BASF SE**; Carl-Bosch-Strasse 38, 67056 Ludwigshafen (DE).

- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

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

(57) **Abstract:** A method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a GRP (Growth Related Protein) polypeptide is provided. Plants having modulated expression of a nucleic acid encoding a GRP polypeptide, which plants have enhanced yield-related traits relative to control plants, and constructs comprising the GRP-encoding nucleic acid, useful in performing the methods for enhancing yield-related traits in plants are also provided.

WO 2013/072832 A1

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

Background

The present invention relates generally to the field of molecular biology and concerns a method for enhancing yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a GRP (Growth Related Protein) polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid encoding a GRP polypeptide, which plants have enhanced yield-related traits relative to corresponding wild type plants or other control plants. The invention also provides constructs useful in the methods of the invention.

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

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

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

endosperm, in particular, assimilates the metabolic precursors of carbohydrates, oils and proteins and synthesizes them into storage macromolecules to fill out the grain.

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

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

Crop yield may therefore be increased by optimising one of the above-mentioned factors.

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

It has now been found that various yield-related traits may be improved in plants by modulating expression in a plant of a nucleic acid encoding a GRP (Growth Related Protein) polypeptide in a plant.

Detailed description of the invention

The present invention shows that modulating expression in a plant of a nucleic acid encoding a GRP 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 GRP polypeptide and optionally selecting for plants having enhanced yield-related traits. According to another embodiment, the present invention provides a method for producing plants having enhanced yield-related traits relative to control plants, wherein said method comprises the steps of modulating expression in said plant of a nucleic acid encoding a GRP polypeptide as described herein and optionally selecting for plants having enhanced yield-related traits.

In one embodiment, the preferred method for modulating expression of a nucleic acid encoding a GRP polypeptide is increasing expression by introducing and expressing in a plant a nucleic acid encoding a GRP polypeptide, preferably a recombinant nucleic acid encoding a GRP polypeptide. In another embodiment, the preferred method of modulating expression is by introducing a genetic construct designed for decreasing expression of a nucleic acid encoding a GRP.

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

A “GRP polypeptide” as defined herein refers to a polypeptide selected from Table A in the Examples section, and to homologues and derivatives thereof.

In a particular embodiment, preferred GRP proteins for use in the methods of the invention include Zeama_GRMZM2G180172, Zeama_GRMZM2G012224_T01, Zeama_GRMZM2G055585, Zeama_GRMZM2G049398_T01, Zeama_GRMZM2G127850_T01, Zm protein kinase-related (ZM1s60983186), Orysa_LOC_Os03g46490, Zeama_GRMZM2G126566, Zeama_GRMZM2G031724, Zeama_GRMZM2G047684, Zm indole-3-acetic acid amido synthetase-like 1, Zm indole-3-acetic acid amido synthetase-like 2, Zeama_GRMZM2G031983_T01, Zeama_GRMZM2G347043, Orysa_LOC_Os06g01440, Zeama_GRMZM2G108716, Zm embryo-abundant protein-related, Zm ATP-binding protein, Zeama_unknown protein b.

Homologues of polypeptides from Table A preferably have 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 one of the sequences of the proteins listed in Table A. The overall

sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG Wisconsin Package, Accelrys), preferably with default parameters and preferably with sequences of mature proteins (i.e. without taking into account secretion signals or transit peptides). In one embodiment the sequence identity level is determined by comparison of the polypeptide sequences over the entire length of the sequence of the proteins listed in Table A.

Alternatively or additionally, homologues of polypeptides from Table A can be identified by screening for the presence of domains or motifs in common with one of the polypeptides from Table A or by screening for the same biological activity as that of one of the polypeptides from Table A. Tools for identifying domains are known in the art and comprise querying databases like InterPro (Hunter et al., Nucleic Acids Res. 37 (Database Issue):D224-228, 2009) with a protein sequence from Table A or of homologous sequences, as exemplified in Example 4. Also the identification of motifs is known in the art, for example by using the MEME algorithm (Bailey and Elkan, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994). At each position within a MEME motif, the residues are shown that are present in the query set of sequences with a frequency higher than 0.2. Residues within square brackets represent alternatives.

Alternatively or additionally, the sequence identity level is determined by comparison of one or more conserved domains or motifs in a given protein sequence provided in Table A with corresponding conserved domains or motifs in other GRP polypeptides. Compared to overall sequence identity, the sequence identity will generally be higher when only conserved domains or motifs are considered. Preferably the motifs in a GRP polypeptide have, in increasing order of preference, at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one or more of the motifs present in a protein sequence of Table A.

The terms “domain”, “signature” and “motif” are defined in the “definitions” section herein. Preferably the homologue of a given protein from Table A has the same domain and motif organisation as in that particular protein from Table A.

According one embodiment, there is provided a method for improving yield-related traits as provided herein in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a GRP polypeptide as defined herein, preferably a nucleic acid encoding a GRP polypeptide listed in Table A.

Preferably, GRP polypeptides (at least in their native form) typically have a biological function similar or identical to that of one of the polypeptides of Table A. Tools and techniques for measuring biological functions are known in the art. In addition, nucleic acids

encoding GRP polypeptides, when expressed in rice according to the methods of the present invention as outlined in Examples 7 and 9, give plants having increased yield related traits, such as those provided in Example 10. Another function of the nucleic acid sequences encoding GRP polypeptides is to confer information for synthesis of the GRP protein that increases yield or yield related traits as described herein, when such a nucleic acid sequence of the invention is transcribed and translated in a living plant cell.

The present invention is illustrated by transforming plants with a nucleic acid sequence selected from Table A, encoding the GRP polypeptides. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any GRP-encoding nucleic acid or GRP polypeptide as defined herein. The term "GRP" or "GRP polypeptide" as used herein also intends to include homologues as defined hereunder of the proteins given in Table A. Preferably the homologue is an orthologue or a paralogue of a protein given in Table A, the terms "orthologues" and "paralogues" being as defined herein. Orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search as described in the definitions section; where the query sequence is one of the protein or DNA sequences listed in Table A, the second BLAST (back-BLAST) would be against maize, rice or Arabidopsis sequences.

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

Further nucleic acid variants useful in practising the methods of the invention include portions of nucleic acids encoding GRP polypeptides, nucleic acids hybridising to nucleic acids encoding GRP polypeptides, splice variants of nucleic acids encoding GRP polypeptides, allelic variants of nucleic acids encoding GRP polypeptides and variants of nucleic acids encoding GRP polypeptides obtained by gene shuffling. The terms hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

Nucleic acids encoding GRP polypeptides need not be full-length nucleic acids, since performance of the methods of the invention does not rely on the use of full-length nucleic acid sequences. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a

portion of any one of the nucleic acid sequences given in Table A of the Examples section, or a portion of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A of the Examples section.

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

Portions useful in the methods of the invention, encode a GRP polypeptide as defined herein or at least part thereof, and have substantially the same biological activity as the amino acid sequences given in Table A of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A of the Examples section, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A of the Examples section. Preferably the portion is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A of the Examples section, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A of the Examples section. Most preferably the portion is a portion of a nucleic acid listed in Table A. Preferably, the portion encodes a fragment of an amino acid sequence which has at least 25% sequence identity to one of the proteins listed in Table A.

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, more preferably under conditions of high stringency, with a nucleic acid encoding a GRP polypeptide as defined herein, or with a portion as defined herein. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a nucleic acid capable of hybridizing to the complement of a nucleic acid encoding any one of the proteins given in Table A of the Examples section, or to a nucleic acid encoding an orthologue, paralogue or homologue of any one of the proteins given in Table A.

Hybridising sequences useful in the methods of the invention encode a GRP polypeptide as defined herein, having substantially the same biological activity as any one of the amino acid sequences given in Table A of the Examples section. Preferably, the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding any one of the proteins given in Table A of the Examples section, or to a portion of any of these sequences, a portion being as defined herein, or the hybridising sequence is capable of

hybridising to the complement of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A of the Examples section. Most preferably, the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding the polypeptide as represented by one of the sequences given in Table A or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined herein.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence which has at least 25% sequence identity to one of the proteins listed in Table A.

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

Preferred splice variants are splice variants of a nucleic acid provided in Table A, or a splice variant of a nucleic acid encoding an orthologue or paralogue of one of the proteins listed in Table A. Preferably, the amino acid sequence encoded by the splice variant has at least 25% sequence identity to one of the proteins listed in Table A.

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

The polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as one of the GRP polypeptides listed in Table A of the Examples section. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of one of the nucleic acids provided in Table A or an allelic variant of a nucleic acid encoding an orthologue or paralogue of a protein listed in Table A. Preferably, the amino acid sequence encoded by the allelic variant has at least 25% sequence identity to one of the proteins listed in Table A.

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

homologue of any of the amino acid sequences given in Table A of the Examples section, which variant nucleic acid is obtained by gene shuffling.

Preferably, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling has at least 25% sequence identity to one of the proteins listed in Table A.

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.). GRP polypeptides differing from the sequence of a protein listed in Table A by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined herein) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention.

Nucleic acids encoding GRP 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 GRP polypeptide-encoding nucleic acid is from a plant, further preferably from a monocotyledonous plant, more preferably from the family Poaceae, most preferably the nucleic acid is from *Oryza sativa* or *Zea mays*. In another embodiment, the GRP encoding nucleic acid is from a dicotyledonous plant, more preferably from the family Brassicaceae, most preferably the nucleic acid is from *Arabidopsis thaliana*. In a particular embodiment, the expression of the endogenous GRP encoding nucleic acid is downregulated. For optimal performance, the gene silencing techniques used for the reduction or substantial elimination of endogenous GRP gene expression require the use of GRP sequences that are substantially homologous to the endogenous GRP gene to be downregulated. For example, a GRP nucleic acid from rice (be it a full length GRP sequence or a fragment) is transformed into a rice plant. The GRP nucleic acid need not be introduced into the same plant variety. Most preferably, the GRP nucleic acid from rice is a sufficient length of substantially contiguous nucleotides of any one of the SEQ ID NOs listed in Table A or a sufficient length of substantially contiguous nucleotides of a nucleic acid sequence encoding an orthologue or paralogue of a GRP listed in Table A. A person skilled in the art would be well aware of what would constitute a sufficient length of substantially contiguous nucleotides to perform any of the gene silencing methods defined hereinabove, this may be as little as 20 or fewer substantially contiguous nucleotides in some cases.

In another embodiment the present invention extends to recombinant chromosomal DNA comprising a nucleic acid sequence useful in the methods of the invention, wherein said nucleic acid is present in the chromosomal DNA as a result of recombinant methods, but is not in its natural genetic environment. In a further embodiment the recombinant chromosomal DNA of the invention is comprised in a plant cell.

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

The present invention thus provides a method for increasing yield-related traits, especially one or more of early vigour, root biomass, shoot biomass, and seed yield of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding a GRP polypeptide as defined herein.

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

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

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

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

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

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding GRP polypeptides. The gene constructs may be inserted into vectors, which may be commercially available, suitable for transforming into plants or host cells and suitable for expression of the gene of interest in the transformed cells. The invention also provides use of a gene construct as defined herein in the methods of the invention.

More specifically, the present invention provides a construct comprising:

- (a) a nucleic acid encoding a GRP 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 GRP polypeptide is as defined above. The term “control sequence” and “termination sequence” are as defined herein.

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

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

The skilled artisan is well aware of the genetic elements that must be present on the genetic construct in order to successfully transform, select and propagate host cells containing the sequence of interest. The sequence of interest is operably linked to one or more control sequences (at least to a promoter).

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

In one embodiment, the constitutive promoter is a ubiquitous constitutive promoter of medium strength. Preferably it is a plant derived promoter, e.g. a promoter of plant chromosomal origin, such as a GOS2 promoter or a promoter of substantially the same strength and having substantially the same expression pattern (a functionally equivalent promoter), more preferably the promoter is the promoter GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 643, most preferably the constitutive promoter is as represented by SEQ ID NO: 643. In another embodiment, the constitutive promoter is a CaMV-35S promoter, preferably the 35S promoter is substantially similar to SEQ ID NO: 644. See the "Definitions" section herein for further examples of constitutive promoters.

According to another preferred embodiment of the invention, the nucleic acid encoding a GRP polypeptide is operably linked to a root-specific promoter. The root-specific promoter is preferably a pericycle specific UAS promoter (Brand and Perrimon Development 118(2): 401-415, 1993; Laplaze et al., The Plant cell, 19, 3889-38900, 2007) or a promoter of substantially the same strength and having substantially the same expression pattern (a functionally equivalent promoter). Examples of other root-specific promoters which may also be used to perform the methods of the invention are shown in Table 2b in the "Definitions" section.

It should be clear that the applicability of the present invention is not restricted to the GRP polypeptide-encoding nucleic acid represented by those listed in Table A, nor is the applicability of the invention restricted to the rice GOS2 promoter or to the 35S promoter when expression of a GRP polypeptide-encoding nucleic acid is driven by a constitutive promoter.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Those skilled in the art will be aware of terminator sequences that may be suitable for use in performing the invention. In one embodiment, the construct comprises an expression cassette comprising a GOS2 promoter, substantially similar to SEQ ID NO: 643, operably linked to the nucleic acid encoding the GRP polypeptide. More preferably, the construct furthermore comprises a zein terminator (t-zein) linked to the 3' end of the GRP coding sequence. In another embodiment, the construct comprises an expression cassette comprising a CaMV-35S promoter substantially similar to SEQ ID NO: 644, operably linked to the nucleic acid encoding the GRP polypeptide.

Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

According to one embodiment 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. In another embodiment, the modulated expression is decreased expression. Methods for decreasing expression of nucleic acids or genes, or gene products, are well documented in the art and examples are provided in the definitions section.

As mentioned above, a preferred method for modulating expression of a nucleic acid encoding a GRP polypeptide is by introducing and expressing in a plant a nucleic acid encoding a GRP polypeptide comprised in a genetic construct designed for either increased expression or for decreased expression; however the effects of performing the method, i.e. enhancing yield-related traits may also be achieved using other well known techniques, including but not limited to T-DNA activation tagging, TILLING, homologous recombination. A description of these techniques is provided in the definitions section.

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

More specifically, the present invention provides a method for the production of transgenic plants having enhanced yield-related traits, which method comprises:

- (i) introducing and expressing in a plant or plant cell a GRP polypeptide-encoding nucleic acid or a genetic construct comprising a GRP 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 GRP polypeptide as defined herein.

Cultivating the plant cell under conditions promoting plant growth and development, may or may not include regeneration and/or growth to maturity. Accordingly, in a particular embodiment of the invention, the plant cell transformed by the method according to the invention is regenerable into a transformed plant. In another particular embodiment, the plant cell transformed by the method according to the invention is not regenerable into a transformed plant, i.e. cells that are not capable to regenerate into a plant using cell culture techniques known in the art. While plants cells generally have the characteristic of totipotency, some plant cells can not be used to regenerate or propagate intact plants from said cells. In one embodiment of the invention the plant cells of the invention are such cells. In another embodiment the plant cells of the invention are plant cells that do not sustain themselves in an autotrophic way.

The nucleic acid may be introduced directly into a plant cell or into the plant itself (including introduction into a tissue, organ or any other part of a plant). According to a preferred

feature of the present invention, the nucleic acid is preferably introduced into a plant or plant cell by transformation. The term "transformation" is described in more detail in the "definitions" section herein.

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

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

In a further embodiment the invention extends to seeds comprising the expression cassettes of the invention, the genetic constructs of the invention, or the nucleic acids encoding the GRP and/or the GRP polypeptides as described above.

The invention also includes host cells containing an isolated nucleic acid encoding a GRP polypeptide as defined above. In one embodiment host cells according to the invention are plant cells, yeasts, bacteria or fungi. Host plants for the nucleic acids, construct, expression cassette or the vector used in the method according to the invention are, in principle, advantageously all plants which are capable of synthesizing the polypeptides used in the inventive method. In a particular embodiment the plant cells of the invention overexpress the nucleic acid molecule of the invention.

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

group consisting of maize, wheat, rice, soybean, cotton, oilseed rape including canola, sugarcane, sugar beet and alfalfa. Advantageously the methods of the invention are more efficient than the known methods, because the plants of the invention have increased yield and/or tolerance to an environmental stress compared to control plants used in comparable methods.

The invention also extends to plants, plant parts or harvestable parts of a plant such as, but not limited to seeds, leaves, fruits, flowers, stems, roots, rhizomes, tubers and bulbs, which plants or plant parts or harvestable parts comprise a recombinant nucleic acid encoding a GRP polypeptide. The invention furthermore relates to products derived or produced, preferably directly derived or produced, from a harvestable part of such a plant, such as dry pellets, meal or powders, oil, fat and fatty acids, starch or proteins.

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

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

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

The present invention also encompasses use of nucleic acids encoding GRP polypeptides as described herein and use of these GRP polypeptides in enhancing any of the aforementioned yield-related traits in plants. For example, nucleic acids encoding GRP polypeptide described herein, or the GRP polypeptides themselves, may find use in breeding programmes in which a DNA marker is identified which may be genetically linked

to a GRP polypeptide-encoding gene. The nucleic acids/genes, or the GRP polypeptides themselves may be used to define a molecular marker. This DNA or protein marker may then be used in breeding programmes to select plants having enhanced yield-related traits as defined herein in the methods of the invention. Furthermore, allelic variants of a GRP polypeptide-encoding nucleic acid/gene may find use in marker-assisted breeding programmes. Nucleic acids encoding GRP 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.

Moreover, the present invention relates to the following specific embodiments:

1. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a GRP polypeptide listed in Table A.
2. Method according to embodiment 1, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said GRP polypeptide.
3. Method according to embodiment 1 or 2, wherein said enhanced yield-related traits comprise increased yield and/or increased early vigour relative to control plants, wherein said increased yield preferably comprises increased biomass and/or increased seed yield relative to control plants.
4. Method according to any one of embodiments 1 to 3, wherein said enhanced yield-related traits are obtained under non-stress conditions.
5. Method according to any one of embodiments 1 to 3, wherein said enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.
6. Method according to any one of embodiments 1 to 5, wherein said nucleic acid encoding a GRP is of plant origin, preferably from a monocotyledonous or a dicotyledonous plant.
7. Method according to any one of embodiments 1 to 6, wherein said nucleic acid encoding a GRP listed in Table A is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
8. Method according to any one of embodiments 1 to 7, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A.

9. Method according to any one of embodiments 1 to 8, wherein said nucleic acid is operably linked to a constitutive promoter of plant origin, preferably to a medium strength constitutive promoter of plant origin, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
10. Method according to any one of embodiments 1 to 8, wherein said nucleic acid is operably linked to a strong constitutive promoter, preferably the CaMV-35S promoter.
11. Plant, or part thereof, or plant cell, obtainable by a method according to any one of embodiments 1 to 10, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding a GRP polypeptide as defined in any of embodiments 1 and 6 to 8.
12. Construct comprising:
 - (i) nucleic acid encoding an GRP as defined in any of embodiments 1 and 6 to 8;
 - (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - (iii) a transcription termination sequence.
13. Construct according to embodiment 12, wherein one of said control sequences is a constitutive promoter of plant origin, preferably to a medium strength constitutive promoter of plant origin, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
14. Construct according to embodiment 12, wherein one of said control sequences is a strong constitutive promoter, preferably the CaMV-35S promoter.
15. Use of a construct according to any one of embodiments 12 to 14 in a method for making plants having enhanced yield-related traits, preferably increased yield and/or early vigour relative to control plants.
16. Plant, plant part or plant cell transformed with a construct according to any one of embodiments 12 to 14.
17. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield and/or increased early vigour relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants, comprising:
 - (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an GRP polypeptide as defined in any of embodiments 1 and 6 to 10; and

- (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.
- 18. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield and/or early vigour relative to control plants, resulting from modulated expression of a nucleic acid encoding an GRP polypeptide as defined in any of embodiments 1 and 6 to 8 or a transgenic plant cell derived from said transgenic plant.
- 19. Transgenic plant according to embodiment 11, 16 or 18, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.
- 20. Harvestable parts of a plant according to embodiment 19, wherein said harvestable parts are preferably shoot biomass and/or root biomass and/or seeds.
- 21. Products derived from a plant according to embodiment 19 and/or from harvestable parts of a plant according to embodiment 20.
- 22. Use of a nucleic acid encoding an GRP polypeptide as defined in any of embodiments 1 and 6 to 8 for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield and/or early vigour, relative to control plants.
- 23. A method for manufacturing a product comprising the steps of growing the plants according to embodiment 12, 16, 19 or 20 and producing said product from or by said plants; or parts thereof, including seeds.
- 24. Products produced from a plant according to embodiment 19 and/or from harvestable parts of a plant according to embodiment 20.
- 25. Construct according to any one of embodiments 12 to 14 comprised in a plant cell.
- 26. Recombinant chromosomal DNA comprising the construct according to any one of embodiments 12 to 14.

Definitions

The following definitions will be used throughout the present application. The section captions and headings in this application are for convenience and reference purpose only and should not affect in any way the meaning or interpretation of this application. The technical terms and expressions used within the scope of this application are generally to

be given the meaning commonly applied to them in the pertinent art of plant biology, molecular biology, bioinformatics and plant breeding. All of the following term definitions apply to the complete content of this application. The term “essentially”, “about”, “approximately” and the like in connection with an attribute or a value, particularly also define exactly the attribute or exactly the value, respectively. The term “about” in the context of a given numeric value or range relates in particular to a value or range that is within 20%, within 10%, or within 5% of the value or range given.

Peptide(s)/Protein(s)

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

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

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

Homologue(s)

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

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

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

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

epitope, c-myc epitope, FLAG®-epitope, lacZ, CMP (calmodulin-binding peptide), HA epitope, protein C epitope and VSV epitope.

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

Table 1: Examples of conserved amino acid substitutions

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

Amino acid substitutions, deletions and/or insertions may readily be made using peptide synthetic techniques known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulation. Methods for the manipulation of DNA sequences to produce substitution, insertion or deletion variants of a protein are well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA are well known to those skilled in the art and include M13 mutagenesis, T7-Gen in vitro mutagenesis (USB, Cleveland, OH), QuickChange Site Directed mutagenesis (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols (see *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989 and yearly updates)).

Derivatives

“Derivatives” include peptides, oligopeptides, polypeptides which may, compared to the amino acid sequence of the naturally-occurring form of the protein, such as the protein of interest, comprise substitutions of amino acids with non-naturally occurring amino acid residues, or additions of non-naturally occurring amino acid residues. “Derivatives” of a protein also encompass peptides, oligopeptides, polypeptides which comprise naturally

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

Domain, Motif/Consensus sequence/Signature

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

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

Specialist databases exist for the identification of domains, for example, SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864; Letunic et al. (2002) Nucleic Acids Res 30, 242-244), InterPro (Mulder et al., (2003) Nucl. Acids. Res. 31, 315-318), Prosite (Bucher and Bairoch (1994), A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation. (In) ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., pp53-61, AAAI Press, Menlo Park; Hulo et al., Nucl. Acids. Res. 32:D134-D137, (2004)), or Pfam (Bateman et al., Nucleic Acids Research 30(1): 276-280 (2002)). A set of tools for *in silico* analysis of protein sequences is available on the ExPASy proteomics server (Swiss Institute of Bioinformatics (Gasteiger et al., ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31:3784-3788(2003)). Domains or motifs may also be identified using routine techniques, such as by sequence alignment.

Methods for the alignment of sequences for comparison are well known in the art, such methods include GAP, BESTFIT, BLAST, FASTA and TFASTA. GAP uses the algorithm of

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

Reciprocal BLAST

Typically, this involves a first BLAST involving BLASTing a query sequence (for example using any of the sequences listed in Table A of the Examples section) against any sequence database, such as the publicly available NCBI database. BLASTN or TBLASTX (using standard default values) are generally used when starting from a nucleotide sequence, and BLASTP or TBLASTN (using standard default values) when starting from a protein sequence. The BLAST results may optionally be filtered. The full-length sequences of either the filtered results or non-filtered results are then BLASTed back (second BLAST) against sequences from the organism from which the query sequence is derived. The results of the first and second BLASTs are then compared. A paralogue is identified if a high-ranking hit from the first blast is from the same species as from which the query sequence is derived, a BLAST back then ideally results in the query sequence amongst the highest hits; an orthologue is identified if a high-ranking hit in the first BLAST is not from the same species as from which the query sequence is derived, and preferably results upon BLAST back in the query sequence being among the highest hits.

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

be used, followed by a neighbour joining tree, to help visualize clustering of related genes and to identify orthologues and paralogues.

Hybridisation

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

The term "stringency" refers to the conditions under which a hybridisation takes place. The stringency of hybridisation is influenced by conditions such as temperature, salt concentration, ionic strength and hybridisation buffer composition. Generally, low stringency conditions are selected to be about 30°C lower than the thermal melting point (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 T_m is the temperature under defined ionic strength and pH, at which 50% of the target sequence hybridises to a perfectly matched probe. The T_m is dependent upon the solution conditions and the base composition and length of the probe. For example, longer sequences hybridise specifically at higher temperatures. The maximum rate of hybridisation is obtained from about 16°C up to 32°C below T_m . The presence of monovalent cations in the hybridisation solution reduce the electrostatic repulsion between the two nucleic acid strands thereby promoting hybrid formation; this effect is visible for sodium concentrations of up to 0.4M (for higher concentrations, this effect may be ignored). Formamide reduces the melting temperature of DNA-DNA and DNA-RNA duplexes with 0.6 to 0.7°C for each percent formamide, and addition of 50% formamide allows hybridisation to be performed at 30 to 45°C, though the rate of hybridisation will be lowered. Base pair mismatches reduce the hybridisation rate and the thermal stability of the duplexes. On average and for large

probes, the T_m decreases about 1°C per % base mismatch. The T_m may be calculated using the following equations, depending on the types of hybrids:

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

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

2) DNA-RNA or RNA-RNA hybrids:

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

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

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

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

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

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

^c L = length of duplex in base pairs.

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

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

Besides the hybridisation conditions, specificity of hybridisation typically also depends on the function of post-hybridisation washes. To remove background resulting from non-specific hybridisation, samples are washed with dilute salt solutions. Critical factors of such washes include the ionic strength and temperature of the final wash solution: the lower the salt concentration and the higher the wash temperature, the higher the stringency of the wash. Wash conditions are typically performed at or below hybridisation stringency. A positive hybridisation gives a signal that is at least twice of that of the background. Generally, suitable stringent conditions for nucleic acid hybridisation assays or gene amplification detection procedures are as set forth above. More or less stringent conditions may also be selected. The skilled artisan is aware of various parameters which may be altered during washing and which will either maintain or change the stringency conditions.

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.

Endogenous gene

Reference herein to an “endogenous” gene not only refers to the gene in question as found in a plant in its natural form (i.e., without there being any human intervention), but also refers to that same gene (or a substantially homologous nucleic acid/gene) in an isolated form subsequently (re)introduced into a plant (a transgene). For example, a transgenic plant containing such a transgene may encounter a substantial reduction of the transgene expression and/or substantial reduction of expression of the endogenous gene. The isolated gene may be isolated from an organism or may be manmade, for example by chemical synthesis.

Gene shuffling/Directed evolution

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

Construct

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

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

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

Regulatory element/Control sequence/Promoter

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 1/500,000 transcripts per cell. Conversely, a “strong promoter” drives expression of a coding sequence at high level, or at about 1/10 transcripts to about 1/100 transcripts to about 1/1000 transcripts per cell. Generally, by “medium strength promoter” is intended a promoter that drives expression of a coding sequence at a lower level than a strong promoter, in particular at a level that is in all instances below that obtained when under the control of a 35S CaMV promoter.

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	Koyama et al. J Biosci Bioeng. 2005 Jan;99(1):38-42.; Mudge et al. (2002, Plant J. 31:341)
Medicago phosphate transporter	Xiao et al., 2006, Plant Biol (Stuttg). 2006 Jul;8(4):439-49
Arabidopsis Pyk10	Nitz et al. (2001) Plant Sci 161(2): 337-346
root-expressible genes	Tingey et al., EMBO J. 6: 1, 1987.
tobacco auxin-inducible gene	Van der Zaal et al., Plant Mol. Biol. 16, 983, 1991.
β -tubulin	Oppenheimer, et al., Gene 63: 87, 1988.
tobacco root-specific genes	Conkling, et al., Plant Physiol. 93: 1203, 1990.
B. napus G1-3b gene	United States Patent No. 5, 401, 836
SbPRP1	Suzuki et al., Plant Mol. Biol. 21: 109-119, 1993.
LRX1	Baumberger et al. 2001, Genes & Dev. 15:1128
BTG-26 Brassica napus	US 20050044585
LeAMT1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
The LeNRT1-1 (tomato)	Lauter et al. (1996, PNAS 3:8139)
class I patatin gene (potato)	Liu et al., Plant Mol. Biol. 17 (6): 1139-1154

KDC1 (<i>Daucus carota</i>)	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 (<i>Arabidopsis</i>)	Diener et al. (2001, Plant Cell 13:1625)
NRT2;1Np (<i>N. plumbaginifolia</i>)	Quesada et al. (1997, Plant Mol. Biol. 34:265)

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

Table 2c: Examples of seed-specific promoters

Gene source	Reference
seed-specific genes	Simon et al., Plant Mol. Biol. 5: 191, 1985;
	Scofield et al., J. Biol. Chem. 262: 12202, 1987.;
	Baszczynski et al., Plant Mol. Biol. 14: 633, 1990.
Brazil Nut albumin	Pearson et al., Plant Mol. Biol. 18: 235-245, 1992.
legumin	Ellis et al., Plant Mol. Biol. 10: 203-214, 1988.
glutelin (rice)	Takaiwa et al., Mol. Gen. Genet. 208: 15-22, 1986;
	Takaiwa et al., FEBS Letts. 221: 43-47, 1987.
zein	Matzke et al Plant Mol Biol, 14(3):323-32 1990
napA	Stalberg et al, Planta 199: 515-519, 1996.
wheat LMW and HMW glutenin-1	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
wheat SPA	Albani et al, Plant Cell, 9: 171-184, 1997
wheat α , β , γ -gliadins	EMBO J. 3:1409-15, 1984
barley ltr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	Mena et al, The Plant Journal, 116(1): 53-62, 1998
blz2	EP99106056.7
synthetic promoter	Vicente-Carbajosa et al., Plant J. 13: 629-640, 1998.
rice prolamin NRP33	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice α -globulin Glb-1	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122,

	1996
rice α -globulin REB/OHP-1	Nakase et al. Plant Mol. Biol. 33: 513-522, 1997
rice ADP-glucose pyrophosphorylase	Trans Res 6:157-68, 1997
maize ESR gene family	Plant J 12:235-46, 1997
sorghum α -kafirin	DeRose et al., Plant Mol. Biol 32:1029-35, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
rice oleosin	Wu et al, J. Biochem. 123:386, 1998
sunflower oleosin	Cummins et al., Plant Mol. Biol. 19: 873-876, 1992
PRO0117, putative rice 40S ribosomal protein	WO 2004/070039
PRO0136, rice alanine aminotransferase	unpublished
PRO0147, trypsin inhibitor ITR1 (barley)	unpublished
PRO0151, rice WSI18	WO 2004/070039
PRO0175, rice RAB21	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039
α -amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al, Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β -like gene	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Barley Ltp2	Kalla et al., Plant J. 6:849-60, 1994
Chi26	Leah et al., Plant J. 4:579-89, 1994
Maize B-Peru	Selinger et al., Genetics 149;1125-38,1998

Table 2d: examples of endosperm-specific promoters

Gene source	Reference
glutelin (rice)	Takaiwa et al. (1986) Mol Gen Genet 208:15-22; Takaiwa et al. (1987) FEBS Letts. 221:43-47
zein	Matzke et al., (1990) Plant Mol Biol 14(3): 323-32
wheat LMW and HMW glutenin-1	Colot et al. (1989) Mol Gen Genet 216:81-90, Anderson et al. (1989) NAR 17:461-2
wheat SPA	Albani et al. (1997) Plant Cell 9:171-184
wheat gliadins	Rafalski et al. (1984) EMBO 3:1409-15
barley ltr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Cho et al. (1999) Theor Appl Genet 98:1253-62; Muller et al. (1993) Plant J 4:343-55; Sorenson et al. (1996) Mol Gen Genet 250:750-60
barley DOF	Mena et al, (1998) Plant J 116(1): 53-62
blz2	Onate et al. (1999) J Biol Chem 274(14):9175-82

synthetic promoter	Vicente-Carbajosa et al. (1998) Plant J 13:629-640
rice prolamin NRP33	Wu et al, (1998) Plant Cell Physiol 39(8) 885-889
rice globulin Glb-1	Wu et al. (1998) Plant Cell Physiol 39(8) 885-889
rice globulin REB/OHP-1	Nakase et al. (1997) Plant Molec Biol 33: 513-522
rice ADP-glucose pyrophosphorylase	Russell et al. (1997) Trans Res 6:157-68
maize ESR gene family	Opsahl-Ferstad et al. (1997) Plant J 12:235-46
sorghum kafirin	DeRose et al. (1996) Plant Mol Biol 32:1029-35

Table 2e: Examples of embryo specific promoters:

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

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., Plant Physiol. 2001 Nov;127(3):1136-46
Maize Phosphoenolpyruvate carboxylase	Leaf specific	Kausch et al., Plant Mol Biol. 2001 Jan;45(1):1-15
Rice Phosphoenolpyruvate carboxylase	Leaf specific	Lin et al., 2004 DNA Seq. 2004 Aug;15(4):269-76
Rice small subunit Rubisco	Leaf specific	Nomura et al., Plant Mol Biol. 2000 Sep;44(1):99-106
rice beta expansin EXBP9	Shoot specific	WO 2004/070039

Pigeonpea small subunit Rubisco	Leaf specific	Panguluri et al., Indian J Exp Biol. 2005 Apr;43(4):369-72
Pea RBCS3A	Leaf specific	

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 <i>et al.</i> (1996) Proc. Natl. Acad. Sci. USA, 93: 8117-8122
Rice metallothionein	Meristem specific	BAD87835.1
WAK1 & WAK 2	Shoot and root apical meristems, and in expanding leaves and sepals	Wagner & Kohorn (2001) Plant Cell 13(2): 303–318

Terminator

The term “terminator” encompasses a control sequence which is a DNA sequence at the end of a transcriptional unit which signals 3' processing and polyadenylation of a primary transcript and termination of transcription. The terminator can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The terminator to be added may be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

Selectable marker (gene)/Reporter gene

"Selectable marker", "selectable marker gene" or "reporter gene" includes any gene that confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells that are transfected or transformed with a nucleic acid construct of the invention. These marker genes enable the identification of a successful transfer of the nucleic acid molecules via a series of different principles. Suitable markers may be selected from markers that confer antibiotic or herbicide resistance, that introduce a new metabolic trait or that allow visual selection. Examples of selectable marker genes include genes conferring resistance to antibiotics (such as 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/luciferase system) or fluorescence (Green Fluorescent Protein, GFP, and derivatives thereof). This list represents only a small number of possible markers. The skilled worker is familiar with such markers. Different markers are preferred, depending on 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 present in, or originating from, the genome of said plant, or are present in the genome of said plant but not at their natural locus in the genome of said plant, it being possible for the nucleic acids to be expressed homologously or heterologously. However, as mentioned, transgenic also means that, while the nucleic acids according to the invention or used in the inventive method are at their natural position in the genome of a plant, the sequence has been modified with regard to the natural sequence, and/or that the regulatory sequences of the natural sequences have been modified. Transgenic is preferably understood as meaning the expression of the nucleic acids according to the invention at an unnatural locus in the genome, i.e. homologous or, preferably, heterologous expression of the nucleic acids takes place. Preferred transgenic plants are mentioned herein.

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

Modulation

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

Expression

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

Increased expression/overexpression

The term “increased expression” or “overexpression” as used herein means any form of expression that is additional to the original wild-type expression level. For the purposes of

this invention, the original wild-type expression level might also be zero, i.e. absence of expression or immeasurable expression.

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

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

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

Decreased expression

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

For the reduction or substantial elimination of expression an endogenous gene in a plant, a sufficient length of substantially contiguous nucleotides of a nucleic acid sequence is

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

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

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

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

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

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

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

Antisense nucleic acid sequences can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid sequence may be complementary to the entire nucleic acid sequence (in this case a stretch of substantially contiguous nucleotides derived from the gene of interest, or from any nucleic acid capable of encoding an orthologue, paralogue or homologue of the protein of interest), but may also be an oligonucleotide that is antisense to only a part of the nucleic acid sequence (including the mRNA 5' and 3' UTR). For example, the antisense oligonucleotide sequence may be

complementary to the region surrounding the translation start site of an mRNA transcript encoding a polypeptide. The length of a suitable antisense oligonucleotide sequence is known in the art and may start from about 50, 45, 40, 35, 30, 25, 20, 15 or 10 nucleotides in length or less. An antisense nucleic acid sequence according to the invention may be constructed using chemical synthesis and enzymatic ligation reactions using methods known in the art. For example, an antisense nucleic acid sequence (e.g., an antisense oligonucleotide sequence) may be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acid sequences, e.g., phosphorothioate derivatives and acridine substituted nucleotides may be used. Examples of modified nucleotides that may be used to generate the antisense nucleic acid sequences are well known in the art. Known nucleotide modifications include methylation, cyclization and 'caps' and substitution of one or more of the naturally occurring nucleotides with an analogue such as inosine. Other modifications of nucleotides are well known in the art.

The antisense nucleic acid sequence can be produced biologically using an expression vector into which a nucleic acid sequence has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest). Preferably, production of antisense nucleic acid sequences in plants occurs by means of a stably integrated nucleic acid construct comprising a promoter, an operably linked antisense oligonucleotide, and a terminator.

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.

Transformation

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

The transfer of foreign genes into the genome of a plant is called transformation. Transformation of plant species is now a fairly routine technique. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable ancestor cell. The methods described for the transformation and regeneration of plants from plant tissues or plant cells may be utilized for transient or for stable transformation. Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts (Krens, F.A. et al., (1982) Nature 296, 72-74; Negrutiu I et al. (1987) Plant Mol Biol 8: 363-373); electroporation of protoplasts (Shillito R.D. et al. (1985) Bio/Technol 3, 1099-1102); microinjection into plant material (Crossway A et al., (1986) Mol. Gen Genet 202: 179-185); DNA or RNA-coated particle bombardment (Klein TM et al., (1987) Nature 327: 70) infection with (non-integrative) viruses and the like. Transgenic plants, including transgenic crop plants, are preferably produced via *Agrobacterium*-mediated transformation. An advantageous transformation method is the transformation *in planta*. To this end, it is possible, for example, to allow the agrobacteria to act on plant seeds or to inoculate the plant meristem with agrobacteria. It has proved particularly expedient in accordance with

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

In addition to the transformation of somatic cells, which then have to be regenerated into intact plants, it is also possible to transform the cells of plant meristems and in particular those cells which develop into gametes. In this case, the transformed gametes follow the natural plant development, giving rise to transgenic plants. Thus, for example, seeds of *Arabidopsis* are treated with agrobacteria and seeds are obtained from the developing plants of which a certain proportion is transformed and thus transgenic [Feldman, KA and Marks MD (1987). *Mol Gen Genet* 208:1-9; Feldmann K (1992). In: C Koncz, N-H Chua and J Shell, eds, *Methods in Arabidopsis Research*. Word Scientific, Singapore, pp. 274-289]. Alternative methods are based on the repeated removal of the inflorescences and incubation of the excision site in the center of the rosette with transformed agrobacteria, whereby transformed seeds can likewise be obtained at a later point in time (Chang (1994). *Plant J.* 5: 551-558; Katavic (1994). *Mol Gen Genet*, 245: 363-370). However, an especially effective method is the vacuum infiltration method with its modifications such as the “floral dip” method. In the case of vacuum infiltration of *Arabidopsis*, intact plants under reduced pressure are treated with an agrobacterial suspension [Bechthold, N (1993). *C R Acad Sci*

Paris Life Sci, 316: 1194-1199], while in the case of the "floral dip" method the developing floral tissue is incubated briefly with a surfactant-treated agrobacterial suspension [Clough, SJ and Bent AF (1998) *The Plant J.* 16, 735-743]. A certain proportion of transgenic seeds are harvested in both cases, and these seeds can be distinguished from non-transgenic seeds by growing under the above-described selective conditions. In addition the stable transformation of plastids is of advantages because plastids are inherited maternally in most crops reducing or eliminating the risk of transgene flow through pollen. The transformation of the chloroplast genome is generally achieved by a process which has been schematically displayed in Klaus et al., 2004 [*Nature Biotechnology* 22 (2), 225-229]. Briefly the sequences to be transformed are cloned together with a selectable marker gene between flanking sequences homologous to the chloroplast genome. These homologous flanking sequences direct site specific integration into the plastome. Plastidal transformation has been described for many different plant species and an overview is given in Bock (2001) *Transgenic plastids in basic research and plant biotechnology. J Mol Biol.* 2001 Sep 21; 312 (3):425-38 or Maliga, P (2003) *Progress towards commercialization of plastid transformation technology. Trends Biotechnol.* 21, 20-28. Further biotechnological progress has recently been reported in form of marker free plastid transformants, which can be produced by a transient co-integrated marker gene (Klaus et al., 2004, *Nature Biotechnology* 22(2), 225-229).

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

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

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

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

T-DNA activation tagging

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

TILLING

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

Homologous recombination

“Homologous recombination” allows introduction in a genome of a selected nucleic acid at a defined selected position. Homologous recombination is a standard technology used routinely in biological sciences for lower organisms such as yeast or the moss *Physcomitrella*. Methods for performing homologous recombination in plants have been described not only for model plants (Offringa et al. (1990) EMBO J 9(10): 3077-84) but also for crop plants, for example rice (Terada et al. (2002) Nat Biotech 20(10): 1030-4; Iida and Terada (2004) Curr Opin Biotech 15(2): 132-8), and approaches exist that are generally applicable regardless of the target organism (Miller et al, Nature Biotechnol. 25, 778-785, 2007).

Yield related Trait(s)

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

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

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

Flowers in maize are unisexual; male inflorescences (tassels) originate from the apical stem and female inflorescences (ears) arise from axillary bud apices. The female inflorescence produces pairs of spikelets on the surface of a central axis (cob). Each of the female spikelets encloses two fertile florets, one of them will usually mature into a maize kernel once fertilized. Hence a yield increase in maize may be manifested as one or more of the following: increase in the number of plants established per square meter, an increase in the number of ears per plant, an increase in the number of rows, number of kernels per row,

kernel weight, thousand kernel weight, ear length/diameter, increase in the seed filling rate, which is the number of filled florets (i.e. florets containing seed) divided by the total number of florets and multiplied by 100), among others.

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

Early flowering time

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

Early vigour

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

Increased growth rate

The increased growth rate may be specific to one or more parts of a plant (including seeds), or may be throughout substantially the whole plant. Plants having an increased growth rate may have a shorter life cycle. The life cycle of a plant may be taken to mean the time needed to grow from a mature seed up to the stage where the plant has produced mature seeds, similar to the starting material. This life cycle may be influenced by factors such as speed of germination, early vigour, growth rate, greenness index, flowering time and speed of seed maturation. The increase in growth rate may take place at one or more stages in

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

Stress resistance

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

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

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

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

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

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

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

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

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

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

Increase/Improve/Enhance

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

Seed yield

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

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

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

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

Greenness Index

The “greenness index” as used herein is calculated from digital images of plants. For each pixel belonging to the plant object on the image, the ratio of the green value versus the red value (in the RGB model for encoding color) is calculated. The greenness index is expressed as the percentage of pixels for which the green-to-red ratio exceeds a given threshold. Under normal growth conditions, under salt stress growth conditions, and under reduced nutrient availability growth conditions, the greenness index of plants is measured in

the last imaging before flowering. In contrast, under drought stress growth conditions, the greenness index of plants is measured in the first imaging after drought.

Biomass

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

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

Marker assisted breeding

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

Use as probes in (gene mapping)

Use of nucleic acids encoding the protein of interest for genetically and physically mapping the genes requires only a nucleic acid sequence of at least 15 nucleotides in length. These nucleic acids may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning, A Laboratory Manual) of restriction-digested plant genomic DNA may be probed with the nucleic acids encoding the protein of interest. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) Genomics 1: 174-181) in order to construct a genetic map. In addition, the nucleic

acids may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the nucleic acid encoding the protein of interest in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

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

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

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

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

Plant

The term "plant" as used herein encompasses whole plants, ancestors and progeny of the plants and plant parts, including seeds, shoots, stems, leaves, roots (including tubers), flowers, and tissues and organs, wherein each of the aforementioned comprise the gene/nucleic acid of interest. The term "plant" also encompasses plant cells, suspension

cultures, callus tissue, embryos, meristematic regions, gametophytes, sporophytes, pollen and microspores, again wherein each of the aforementioned comprises the gene/nucleic acid of interest.

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

durum, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum*, *Triticum monococcum* or *Triticum vulgare*), *Tropaeolum minus*, *Tropaeolum majus*, *Vaccinium* spp., *Vicia* spp., *Vigna* spp., *Viola odorata*, *Vitis* spp., *Zea mays*, *Zizania palustris*, *Ziziphus* spp., amongst others.

Control plant(s)

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

Examples

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

DNA manipulation: unless otherwise stated, recombinant DNA techniques are performed 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: GRP Sequences and homologues thereof

Table A provides a list of nucleic acid sequences and protein sequences used for generating transgenic plants described in Example 7.

Plant Source	ZM array number	Description	Nucleic acid SEQ ID NO:	Protein SEQ ID NO:
Oryza sativa	ZM4s07788	Orysa_auxin-repressed protein	1	2
Oryza sativa	ZM1s68450033	Orysa_CBS domain containing membrane protein	3	4
Oryza sativa	ZM1s60334585	Orysa_dehydration-induced protein ERD15	5	6
Oryza sativa	ZM1s57410125	Orysa_LOC_Os01g06640	7	8
Oryza sativa	ZM1s58705134	Orysa_LOC_Os01g63690	9	10
Oryza sativa	ZM1s60103573	Orysa_LOC_Os01g68650	11	12
Oryza sativa	ZM1s57443595	Orysa_LOC_Os01g70860.1	13	14
Oryza sativa	ZM1s60213342	Orysa_LOC_Os02g04120	15	16
Oryza sativa	ZM4s05301	Orysa_LOC_Os02g04230	17	18
Oryza sativa	ZM1s60940818	Orysa_LOC_Os02g06720 for downregulation	19	20
Oryza sativa	ZM1s60940818	Orysa_LOC_Os02g06725	21	22
Oryza sativa	ZM1s61299604	Orysa_LOC_Os02g07240	23	24
Oryza sativa	ZM4s31512	Orysa_LOC_Os02g15800	25	26
Oryza sativa	ZM1s65223113	Orysa_LOC_Os02g36490	27	28
Oryza sativa	ZM1s62030237	Orysa_LOC_Os02g45170	29	30
Oryza sativa	ZM1s57223022	Orysa_LOC_Os02g48990	31	32
Oryza sativa	ZM1s60187950	Orysa_LOC_Os02g52560	33	34
Oryza sativa	ZM1s62114467	Orysa_LOC_Os02g58730	35	36
Oryza sativa	ZM1s62145107	Orysa_LOC_Os03g09220	37	38
Oryza sativa	ZM1s59517939	Orysa_LOC_Os03g10210	39	40
Oryza sativa	ZM1s59034758	Orysa_LOC_Os03g11770	41	42

Oryza sativa	ZM1s62059350	Orysa_LOC_Os03g17790	43	44
Oryza sativa	ZM1s62187860	Orysa_LOC_Os03g17800	45	46
Oryza sativa	ZM4s07964	Orysa_LOC_Os03g46490	47	48
Oryza sativa	ZM1s62219636	Orysa_LOC_Os03g55670	49	50
Oryza sativa	ZM1s66928877.1	Orysa_LOC_Os03g64219	51	52
Oryza sativa	ZM4s14939	Orysa_LOC_Os04g35790	53	54
Oryza sativa	ZM4s22201	Orysa_LOC_Os05g07070	55	56
Oryza sativa	ZM1s62216784	Orysa_LOC_Os05g40770	57	58
Oryza sativa	ZM1a61978583	Orysa_LOC_Os05g50510	59	60
Oryza sativa	ZM1s62075387	Orysa_LOC_Os06g01440	61	62
Oryza sativa	ZM4s29073	Orysa_LOC_Os06g05120	63	64
Oryza sativa	ZM1s57336820	Orysa_LOC_Os06g06250.2	65	66
Oryza sativa	ZM1s68504043	Orysa_LOC_Os06g11840.1	67	68
Oryza sativa	ZM3s01411	Orysa_LOC_Os07g35004	69	70
Oryza sativa	ZM1s67010779	Orysa_LOC_Os08g01390 for downregulation	71	72
Oryza sativa	ZM1s57853588	Orysa_LOC_Os08g01760	73	74
Oryza sativa	ZM1s67094909	Orysa_LOC_Os08g26870	75	76
Oryza sativa	ZM1s60496018	Orysa_LOC_Os08g31250	77	78
Oryza sativa	ZM1s57505415	Orysa_LOC_Os08g37070	79	80
Oryza sativa	ZM4s14446	Orysa_LOC_Os08g40030	81	82
Oryza sativa	ZM4s23596	Orysa_LOC_Os08g42400	83	84
Oryza sativa	ZM1s58474578	Orysa_LOC_Os08g44270	85	86
Oryza sativa	ZM3s01404	Orysa_LOC_Os09g11440	87	88

Oryza sativa	ZM1s58561763	Orysa_LOC_Os09g20284	89	90
Oryza sativa	ZM1s62127538	Orysa_LOC_Os10g39130	91	92
Oryza sativa	ZM1a61994414	Orysa_LOC_Os10g40710	93	94
Oryza sativa	ZM1s57236502	Orysa_LOC_Os11g27400	95	96
Oryza sativa	ZM4s20980	Orysa_LOC_Os12g32970	97	98
Oryza sativa	ZM4s14399	Orysa_LOC_Os12g32992	99	100
Oryza sativa	ZM4s27634	Orysa_Os04g42470	101	102
Oryza sativa	ZM1a62119712	Orysa_Os04g54300	103	104
Oryza sativa	ZM1s62025034	Orysa_putative serine/threonine protein kinase	105	106
Oryza sativa	ZM1s62132790	Orysa_WRKY7 for downregulation	107	108
Zea mays	ZM1s59458241	Zeama_AC214266.3_FGT002	109	110
Zea mays	ZM1s62033152	Zeama_AC214350.3_FGT007	111	112
Zea mays	ZM1s62096975	Zeama_ATACP5; protein serine/threonine phosphatase	113	114
Zea mays	ZM1s60481108	Zeama_calmodulin binding	115	116
Zea mays	ZM4s37206	Zeama_GRMZM2G000645	117	118
Zea mays	ZM1s57508357	Zeama_GRMZM2G000825	119	120
Zea mays	ZM4s16403	Zeama_GRMZM2G001221	121	122
Zea mays	ZM4s22803	Zeama_GRMZM2G004459	123	124
Zea mays	ZM1s57740507	Zeama_GRMZM2G005040_T01	125	126
Zea mays	ZM4s13892	Zeama_GRMZM2G005260	127	128
Zea mays	ZM1s65250927	Zeama_GRMZM2G005435	129	130
Zea mays	ZM4S19342	Zeama_GRMZM2G007466	131	132
Zea mays	ZM1S62132058	Zeama_GRMZM2G007729	133	134

Zea mays	ZM1s57554822	Zeama_GRMZM2G007899	135	136
Zea mays	ZM1s57259830	Zeama_GRMZM2G008657	137	138
Zea mays	ZM1S61984928	Zeama_GRMZM2G009412	139	140
Zea mays	ZM4s11447	Zeama_GRMZM2G009837_T01	141	142
Zea mays	ZM1s65113928	Zeama_GRMZM2G010356_T01	143	144
Zea mays	ZM1S62273254	Zeama_GRMZM2G011110	145	146
Zea mays	ZM1s57807604	Zeama_GRMZM2G011357_T01	147	148
Zea mays	ZM1s62063823	Zeama_GRMZM2G012224_T01	149	150
Zea mays	ZM1S58972630	Zeama_GRMZM2G013627	151	152
Zea mays	ZM4S02093	Zeama_GRMZM2G013657	153	154
Zea mays	ZM1s58375824	Zeama_GRMZM2G014750	155	156
Zea mays	ZM1S57638002	Zeama_GRMZM2G015433	157	158
Zea mays	ZM1S60466311	Zeama_GRMZM2G018689	159	160
Zea mays	ZM1s57738137	Zeama_GRMZM2G018771	161	162
Zea mays	ZM1s57429621	Zeama_GRMZM2G020150	163	164
Zea mays	ZM4s41243	Zeama_GRMZM2G020281_T03	165	166
Zea mays	ZM1s61998920	Zeama_GRMZM2G021270	167	168
Zea mays	ZM1s62262119	Zeama_GRMZM2G023242	169	170
Zea mays	ZM1S57850754	Zeama_GRMZM2G025812	171	172
Zea mays	ZM1s57263625	Zeama_GRMZM2G027052	173	174
Zea mays	ZM4s45997	Zeama_GRMZM2G028218	175	176
Zea mays	ZM1s60841542	Zeama_GRMZM2G029692	177	178
Zea mays	ZM1s62202989	Zeama_GRMZM2G030284	179	180

Zea mays	ZM1s59258289	Zeama_GRMZM2G031724	181	182
Zea mays	ZM1s55817549	Zeama_GRMZM2G031983_T01	183	184
Zea mays	ZM1s58561763	Zeama_GRMZM2G034152_T01	185	186
Zea mays	ZM3s21693	Zeama_GRMZM2G034157	187	188
Zea mays	ZM1s68299591	Zeama_GRMZM2G036120	189	190
Zea mays	ZM4S19630	Zeama_GRMZM2G036134	191	192
Zea mays	ZM1s60105986	Zeama_GRMZM2G036351	193	194
Zea mays	ZM1s57431009	Zeama_GRMZM2G037001_T01	195	196
Zea mays	ZM4s19587	Zeama_GRMZM2G038066	197	198
Zea mays	ZM1s61992039	Zeama_GRMZM2G040517_T01	199	200
Zea mays	ZM1s60429690	Zeama_GRMZM2G041774	201	202
Zea mays	ZM1s60312671	Zeama_GRMZM2G042895	203	204
Zea mays	ZM1S68765176	Zeama_GRMZM2G043147	205	206
Zea mays	ZM4s39700	Zeama_GRMZM2G043275_T01	207	208
Zea mays	ZM4s13777	Zeama_GRMZM2G044027_T01	209	210
Zea mays	ZM1s58678671	Zeama_GRMZM2G044576_T01	211	212
Zea mays	ZM1s62146990	Zeama_GRMZM2G044851_T01	213	214
Zea mays	ZM1a68352777	Zeama_GRMZM2G047684	215	216
Zea mays	ZM1S55843902	Zeama_GRMZM2G049091	217	218
Zea mays	ZM4s45895	Zeama_GRMZM2G049398_T01	219	220
Zea mays	ZM4s46105	Zeama_GRMZM2G052078	221	222
Zea mays	ZM1s58862414	Zeama_GRMZM2G052509	223	224
Zea mays	ZM1s62012391	Zeama_GRMZM2G053766	225	226

Zea mays	ZM4s14974	Zeama_GRMZM2G055458	227	228
Zea mays	ZM1S60958648	Zeama_GRMZM2G055585	229	230
Zea mays	ZM1s67282786	Zeama_GRMZM2G056038	231	232
Zea mays	ZM1s58335548	Zeama_GRMZM2G056407_T01	233	234
Zea mays	ZM3S00100	Zeama_GRMZM2G056600	235	236
Zea mays	ZM4s03858	Zeama_GRMZM2G057283_T01	237	238
Zea mays	ZM1s62139486	Zeama_GRMZM2G057789	239	240
Zea mays	ZM1s58751265	Zeama_GRMZM2G058690_T02	241	242
Zea mays	ZM4s42869	Zeama_GRMZM2G059580	243	244
Zea mays	ZM1S58075442	Zeama_GRMZM2G059799	245	246
Zea mays	ZM1s62207107	Zeama_GRMZM2G060163	247	248
Zea mays	ZM1s59183793	Zeama_GRMZM2G061745	249	250
Zea mays	ZM1s62167066	Zeama_GRMZM2G061996_T01	251	252
Zea mays	ZM1s57461756	Zeama_GRMZM2G062160	253	254
Zea mays	ZM4s10805	Zeama_GRMZM2G062531	255	256
Zea mays	ZM1s59012895	Zeama_GRMZM2G064096_T01	257	258
Zea mays	ZM4s15692	Zeama_GRMZM2G064960	259	260
Zea mays	ZM4a18634	Zeama_GRMZM2G066528	261	262
Zea mays	ZM4s03279	Zeama_GRMZM2G073755	263	264
Zea mays	ZM4s17107	Zeama_GRMZM2G075942	265	266
Zea mays	ZM1s62055151	Zeama_GRMZM2G077206	267	268
Zea mays	ZM4s24055	Zeama_GRMZM2G077458	269	270
Zea mays	ZM4s13911	Zeama_GRMZM2G078198	271	272

Zea mays	ZM1s57963623	Zeama_GRMZM2G079805	273	274
Zea mays	ZM1s60747775	Zeama_GRMZM2G080295_T01	275	276
Zea mays	ZM1s62026868	Zeama_GRMZM2G082032	277	278
Zea mays	ZM1s62161812	Zeama_GRMZM2G082853	279	280
Zea mays	ZM1s65109710	Zeama_GRMZM2G083068	281	282
Zea mays	ZM4s33018	Zeama_GRMZM2G083935	283	284
Zea mays	ZM1S58460807	Zeama_GRMZM2G084264	285	286
Zea mays	ZM1S60657408	Zeama_GRMZM2G084587	287	288
Zea mays	ZM1s62162828	Zeama_GRMZM2G087103	289	290
Zea mays	ZM1S57495842	Zeama_GRMZM2G087323	291	292
Zea mays	ZM1s58524553	Zeama_GRMZM2G088162	293	294
Zea mays	ZM1S57862466	Zeama_GRMZM2G089259	295	296
Zea mays	ZM1s66890980.1	Zeama_GRMZM2G090487_T01	297	298
Zea mays	ZM1s59343090	Zeama_GRMZM2G090595_T04	299	300
Zea mays	ZM1S59395629	Zeama_GRMZM2G091228	301	302
Zea mays	ZM4s22079	Zeama_GRMZM2G091540	303	304
Zea mays	ZM1s57288195	Zeama_GRMZM2G092137	305	306
Zea mays	ZM1s55864617	Zeama_GRMZM2G093217	307	308
Zea mays	ZM4s17135	Zeama_GRMZM2G093900	309	310
Zea mays	ZM1S57836157	Zeama_GRMZM2G095392	311	312
Zea mays	ZM1s58775973	Zeama_GRMZM2G098046	313	314
Zea mays	ZM1s68295211	Zeama_GRMZM2G100120	315	316
Zea mays	ZM1s61258515	Zeama_GRMZM2G100260_T01	317	318

Zea mays	ZM3s00430	Zeama_GRMZM2G100467_T01	319	320
Zea mays	ZM4s18637	Zeama_GRMZM2G101062	321	322
Zea mays	ZM1s62058690	Zeama_GRMZM2G101125	323	324
Zea mays	ZM1s62258567	Zeama_GRMZM2G101390	325	326
Zea mays	ZM1S65226266	Zeama_GRMZM2G101545	327	328
Zea mays	ZM1s61974300	Zeama_GRMZM2G102347	329	330
Zea mays	ZM3S29285	Zeama_GRMZM2G102356	331	332
Zea mays	ZM4s05881	Zeama_GRMZM2G102786_T01	333	334
Zea mays	ZM4s37616	Zeama_GRMZM2G104610_T01	335	336
Zea mays	ZM1A57699438	Zeama_GRMZM2G105331	337	338
Zea mays	ZM4s21938	Zeama_GRMZM2G105348	339	340
Zea mays	ZM1s58374993	Zeama_GRMZM2G107737_T01	341	342
Zea mays	ZM3s01330	Zeama_GRMZM2G107771_T01	343	344
Zea mays	ZM1S62039207	Zeama_GRMZM2G107839	345	346
Zea mays	ZM4S17089	Zeama_GRMZM2G108115	347	348
Zea mays	ZM4s14568	Zeama_GRMZM2G108716	349	350
Zea mays	ZM4s26880	Zeama_GRMZM2G109130	351	352
Zea mays	ZM1S67201464	Zeama_GRMZM2G110289	353	354
Zea mays	ZM4s18047	Zeama_GRMZM2G111451_T01	355	356
Zea mays	ZM4a14205	Zeama_GRMZM2G112210	357	358
Zea mays	ZM1S58101138	Zeama_GRMZM2G112903	359	360
Zea mays	ZM1s59421232	Zeama_GRMZM2G114048	361	362
Zea mays	ZM1s57921389	Zeama_GRMZM2G114552	363	364

Zea mays	ZM1s62190425	Zeama_GRMZM2G114873	365	366
Zea mays	ZM1s57679770	Zeama_GRMZM2G115975	367	368
Zea mays	ZM1s62204322	Zeama_GRMZM2G116526	369	370
Zea mays	ZM1s57370356	Zeama_GRMZM2G116632_T01	371	372
Zea mays	ZM1S65126433	Zeama_GRMZM2G117989	373	374
Zea mays	ZM4S19741	Zeama_GRMZM2G119906	375	376
Zea mays	ZM1s62126803	Zeama_GRMZM2G121074	377	378
Zea mays	ZM4s11447	Zeama_GRMZM2G122228_T01	379	380
Zea mays	ZM1s61995217	Zeama_GRMZM2G123212	381	382
Zea mays	ZM1s60485681	Zeama_GRMZM2G124863	383	384
Zea mays	ZM4s22440	Zeama_GRMZM2G125923	385	386
Zea mays	ZM4S45457	Zeama_GRMZM2G126397	387	388
Zea mays	ZM1S65239103	Zeama_GRMZM2G126566	389	390
Zea mays	ZM1s65368508	Zeama_GRMZM2G127609	391	392
Zea mays	ZM4s27549	Zeama_GRMZM2G127695	393	394
Zea mays	ZM1s68704853	Zeama_GRMZM2G127850_T01	395	396
Zea mays	ZM4s06440	Zeama_GRMZM2G127893	397	398
Zea mays	ZM1s60510343	Zeama_GRMZM2G132854	399	400
Zea mays	ZM1s57861573	Zeama_GRMZM2G137596_T01	401	402
Zea mays	ZM1S57327061	Zeama_GRMZM2G138131	403	404
Zea mays	ZM1s65129831	Zeama_GRMZM2G139082_T01	405	406
Zea mays	ZM1s58505914	Zeama_GRMZM2G142735	407	408
Zea mays	ZM4s07707	Zeama_GRMZM2G143253_T01	409	410

Zea mays	ZM1s59390494	Zeama_GRMZM2G147623	411	412
Zea mays	ZM1s62038184	Zeama_GRMZM2G152354	413	414
Zea mays	ZM4s04639	Zeama_GRMZM2G152561_T02	415	416
Zea mays	ZM4s17046	Zeama_GRMZM2G152908	417	418
Zea mays	ZM4s29494	Zeama_GRMZM2G154648	419	420
Zea mays	ZM1s57348884	Zeama_GRMZM2G156110_T01	421	422
Zea mays	ZM4s10874	Zeama_GRMZM2G158232	423	424
Zea mays	ZM1s57457322	Zeama_GRMZM2G158313_T01	425	426
Zea mays	ZM1s62093950	Zeama_GRMZM2G159034	427	428
Zea mays	ZM4s19175	Zeama_GRMZM2G161295	429	430
Zea mays	ZM1s57320274	Zeama_GRMZM2G163406	431	432
Zea mays	ZM1s62027902	Zeama_GRMZM2G163409	433	434
Zea mays	ZM1S58592636	Zeama_GRMZM2G163641	435	436
Zea mays	ZM4S06145	Zeama_GRMZM2G164781	437	438
Zea mays	ZM1s57368153	Zeama_GRMZM2G164868	439	440
Zea mays	ZM4s11036	Zeama_GRMZM2G165461	441	442
Zea mays	ZM1s62097506	Zeama_GRMZM2G166946	443	444
Zea mays	ZM1s57441890	Zeama_GRMZM2G168365	445	446
Zea mays	ZM1s58699795	Zeama_GRMZM2G168603_T01	447	448
Zea mays	ZM1s61048230	Zeama_GRMZM2G170400	449	450
Zea mays	ZM4s16359	Zeama_GRMZM2G172657	451	452
Zea mays	ZM1a62154780	Zeama_GRMZM2G174558	453	454
Zea mays	ZM1s62106945	Zeama_GRMZM2G174975	455	456

Zea mays	ZM1s57848822	Zeama_GRMZM2G175280_T02	457	458
Zea mays	ZM1s57305612	Zeama_GRMZM2G176217	459	460
Zea mays	ZM1S68614544	Zeama_GRMZM2G177867	461	462
Zea mays	ZM1s67253381	Zeama_GRMZM2G178815_T01	463	464
Zea mays	ZM4s30153	Zeama_GRMZM2G179679	465	466
Zea mays	ZM1S62018816	Zeama_GRMZM2G180172	467	468
Zea mays	ZM1s66894694	Zeama_GRMZM2G302245_T01	469	470
Zea mays	ZM1a61141890	Zeama_GRMZM2G304362	471	472
Zea mays	ZM4s06217	Zeama_GRMZM2G307119_T01	473	474
Zea mays	ZM1s62016000	Zeama_GRMZM2G310368	475	476
Zea mays	ZM1s60135266	Zeama_GRMZM2G312510	477	478
Zea mays	ZM1S60857688.1	Zeama_GRMZM2G320920	479	480
Zea mays	ZM1s58409566	Zeama_GRMZM2G321023	481	482
Zea mays	ZM1s62083330	Zeama_GRMZM2G327032	483	484
Zea mays	ZM1s61980732	Zeama_GRMZM2G334181_T01	485	486
Zea mays	ZM1s62099803	Zeama_GRMZM2G340130	487	488
Zea mays	ZM1s68469961	Zeama_GRMZM2G341404	489	490
Zea mays	ZM1A57511883	Zeama_GRMZM2G345840	491	492
Zea mays	ZM1s62064657	Zeama_GRMZM2G347043	493	494
Zea mays	ZM1s57766027	Zeama_GRMZM2G347808	495	496
Zea mays	ZM1s57319204	Zeama_GRMZM2G355752	497	498
Zea mays	ZM4s16535	Zeama_GRMZM2G359952	499	500
Zea mays	ZM1s57952044	Zeama_GRMZM2G368698_T02	501	502

Zea mays	ZM4S03593	Zeama_GRMZM2G394410	503	504
Zea mays	ZM1S57374543.1	Zeama_GRMZM2G401521	505	506
Zea mays	ZM1s61979139	Zeama_GRMZM2G402417	507	508
Zea mays	ZM1s62009031	Zeama_GRMZM2G407223	509	510
Zea mays	ZM1S62216882	Zeama_GRMZM2G408598	511	512
Zea mays	ZM1a62058338	Zeama_GRMZM2G411032	513	514
Zea mays	ZM1A57765201	Zeama_GRMZM2G412601	515	516
Zea mays	ZM1s60200790	Zeama_GRMZM2G413337	517	518
Zea mays	ZM1S66879359	Zeama_GRMZM2G416388	519	520
Zea mays	ZM1S58357601	Zeama_GRMZM2G421212	521	522
Zea mays	ZM1s67275443	Zeama_GRMZM2G422240	523	524
Zea mays	ZM1S57790868	Zeama_GRMZM2G438195	525	526
Zea mays	ZM1s62144084	Zeama_GRMZM2G445634	527	528
Zea mays	ZM1s62046845	Zeama_GRMZM2G447984	529	530
Zea mays	ZM1s55843675	Zeama_GRMZM2G450153	531	532
Zea mays	ZM1s62083561	Zeama_GRMZM2G451254	533	534
Zea mays	ZM4s06441	Zeama_GRMZM2G458164	535	536
Zea mays	ZM1s58973247	Zeama_GRMZM2G468756_T01	537	538
Zea mays	ZM1s62187177	Zeama_GRMZM2G476069	539	540
Zea mays	ZM1s57375517	Zeama_GRMZM2G477503	541	542
Zea mays	ZM4s03412	Zeama_GRMZM2G479260	543	544
Zea mays	ZM4s11878	Zeama_GRMZM2G481888_T01	545	546
Zea mays	ZM4s26962	Zeama_GRMZM2G701362	547	548

Zea mays	ZM1s60573546	zeama_heat shock 70 protein	549	550
Zea mays	ZM3s00628	Zeama_Methyl transferase like	551	552
Zea mays	ZM1s61220142	Zeama_nucleic acid binding / zinc ion binding	553	554
Zea mays	ZM4s15474	Zeama_Pleiotropic Drug Resistance 12 like	555	556
Zea mays	ZM1s65094362	Zeama_putative nuclear RNA binding protein	557	558
Zea mays	ZM4a37469	Zeama_putative zinc-finger protein	559	560
Zea mays	ZM4s40513	Zeama_SYTE like	561	562
Zea mays	ZM1a58376939	Zeama_transferase protein like	563	564
Zea mays	ZM1s61987040	Zeama_unknown protein a	565	566
Zea mays	ZM1s62016945	Zeama_unknown protein b	567	568
Zea mays	ZM1s62111342	Zeama_unknown protein c	569	570
Zea mays	ZM1s61053034	Zeama_Unknown protein d	571	572
Zea mays	ZM1s58196890	Zeama_Unknown protein e	573	574
Zea mays	ZM1s57483885	Zeama_unknown protein f	575	576
Zea mays	ZM4s16739	Zeama_unknown protein g	577	578
Zea mays	ZM1a58376939	Zeama_unknown protein h	579	580
Zea mays	ZM1s55866371	Zeama_ZM1s55866371	581	582
Zea mays	ZM1s55873260	Zeama_ZM1s55873260	583	584
Zea mays	ZM1s58530935	Zeama_ZM1s58530935	585	586
Zea mays	ZM1s59031313	Zeama_ZM1s59031313	587	588
Zea mays	ZM1s61996727	Zeama_ZM1s61996727	589	590
Zea mays	ZM1s65116286	Zeama_ZM1s65116286	591	592
Zea mays	ZM1s62111307	Zm AAA-type ATPase family protein-like	593	594

Zea mays	ZM1s57394235	Zm ATP-binding protein	595	596
Zea mays	ZM1s58425424	Zm AUXIN RESPONSE FACTOR 16	597	598
Zea mays	ZM1s62214832	Zm Cyclin A2;4-like	599	600
Zea mays	ZM1s68382282	Zm embryo-abundant protein-related	601	602
Zea mays	ZM1s57390634	Zm histidine kinase 1	603	604
Zea mays	ZM4s13278	Zm indole-3-acetic acid amido synthetase-like 1	605	606
Zea mays	ZM1s61040508	Zm indole-3-acetic acid amido synthetase-like 2	607	608
Zea mays	ZM1s62107983	Zm indoleacetic acid-induced protein 13-like 1	609	610
Zea mays	ZM1s59209789	Zm indoleacetic acid-induced protein 13-like 2	611	612
Zea mays	ZM1s58194594	Zm indoleacetic acid-induced protein 18	613	614
Zea mays	ZM1s57509684	Zm indoleacetic acid-induced protein 20-like	615	616
Zea mays	ZM3s01845	Zm MYB78 like	617	618
Zea mays	ZM1s62004697	Zm PHYTOCHROME-ASSOCIATED PROTEIN 1-like	619	620
Zea mays	ZM4s18277	Zm Polygalacturonase inhibiting protein 1-like	621	622
Zea mays	ZM1s57260204	Zm Proliferating Cellular Nuclear Antigen 1	623	624
Zea mays	ZM1s60983186	Zm protein kinase-related	625	626
Zea mays	ZM1s60157931	Zm U-box domain-containing protein	627	628
Zea mays	ZM4s09532	Zm unknown protein 1	629	630
Zea mays	ZM1s62122111	Zm unknown protein 2	631	632
Zea mays	ZM1s62075438	Zm unknown protein 3	633	634
Zea mays	ZM1s57691797	Zm unknown protein 4	635	636
Zea mays	ZM1s62132790	Zm WRKY (WRKY DNA-binding protein 23-like)	637	638
Zea mays	ZM1s60455877	Zm WRKY 40-like	639	640

Sequences (full length cDNA, ESTs or genomic) related to those listed in Table A above can be 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, polypeptides encoded by the nucleic acids listed in Table A can be used with the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis can be viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflect the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons may also be 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.

Sequences have been tentatively assembled and publicly disclosed by research institutions, such as The Institute for Genomic Research (TIGR; beginning with TA). For instance, the Eukaryotic Gene Orthologs (EGO) database may be used to identify such related sequences, either by keyword search or by using the BLAST algorithm with the nucleic acid sequence or polypeptide sequence of interest. Special nucleic acid sequence databases have been created for particular organisms, e.g. for certain prokaryotic organisms, such as by the Joint Genome Institute. Furthermore, access to proprietary databases, has allowed the identification of novel nucleic acid and polypeptide sequences.

Example 2: Alignment of GRP polypeptide sequences

Alignment of the polypeptide sequences representing a GRP protein of Table A and its homologues may be performed using the ClustalW 2.0 algorithm of progressive alignment (Thompson et al. (1997) Nucleic Acids Res 25:4876-4882; Chenna et al. (2003). Nucleic Acids Res 31:3497-3500) with standard setting (slow alignment, similarity matrix: Gonnet, gap opening penalty 10, gap extension penalty: 0.2). Minor manual editing may be done to further optimise the alignment. Generation of multiple alignments is well known in the art.

A phylogenetic tree of a GRP polypeptide of Table A and its homologues may be constructed by aligning GRP sequences using, for example, MAFFT (Kato and Toh (2008) - Briefings in Bioinformatics 9:286-298) with default settings. A neighbour-joining tree is then calculated using, for example, Quick-Tree (Howe et al. (2002), Bioinformatics 18(11): 1546-7), with 100 bootstrap repetitions. The dendrogram is drawn using, for example,

Dendroscope (Huson et al. (2007), BMC Bioinformatics 8(1):460). Confidence levels for 100 bootstrap repetitions are usually indicated for major branchings.

Example 3: Calculation of global percentage identity between polypeptide sequences

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention are determined using MatGAT (Matrix Global Alignment Tool) software (BMC Bioinformatics. 2003 4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. Campanella JJ, Bitincka L, Smalley J; software hosted by Ledion Bitincka). MatGAT generates similarity/identity matrices for DNA or protein sequences without needing pre-alignment of the data. The program performs a series of pair-wise alignments using the Myers and Miller global alignment algorithm, calculates similarity and identity, and then places the results in a distance matrix. 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 protein analysis are typically: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2.

Like for full length sequences, a MATGAT table based on subsequences of a specific domain, may be generated. Based on a multiple alignment of GRP polypeptides, a skilled person may select conserved sequences and submit as input for a MaTGAT analysis. This approach is useful where overall sequence conservation among GRP proteins is rather low.

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.

Results of the InterPro scan (InterPro database, release 34.0) of a selection of GRP polypeptide sequence from Table A are listed hereunder.

Sequence "ZM1S55817549" crc64 checksum: 1DB9E7773AC36AAF length: 361 aa.

InterPro IPR000679 Zinc finger, GATA-type
Molecular Function: transcription factor activity (GO:0003700), Cellular Component: nucleus (GO:0005634),
Biological Process: regulation of transcription, DNA-dependent (GO:0006355), Molecular Function: zinc ion
binding (GO:0008270), Molecular Function: sequence-specific DNA binding (GO:0043565)
method AccNumber shortName location
HMMPfam PF00320 GATA T[191-226] 1.1e-18
HMMSmart SM00401 no description T[185-237] 3.3e-19
ProfileScan PS50114 GATA_ZN_FINGER_2 T[189-221] 12.466
ScanRegExp PS00344 GATA_ZN_FINGER_1 ?[191-216] NA

InterPro IPR013088 Zinc finger, NHR/GATA-type
Molecular Function: transcription factor activity (GO:0003700), Biological Process: regulation of
transcription, DNA-dependent (GO:0006355), Molecular Function: zinc ion binding (GO:0008270)
method AccNumber shortName location
Gene3D G3DSA:3.30.50.10no description T[185-229] 2.4e-05

InterPro NULL NULL
method AccNumber shortName location
HMPanther PTHR10071 TRANSCRIPTION FACTOR GATA (GATA BINDING FACTOR)T[11-18] 1.7e-05 T[188-221]
1.7e-05
Seg seg seg
NA ?[322-334] NA ?[134-139] NA ?[219-236] NA ?[287-302]
superfamily SSF57716 Glucocorticoid receptor-like (DNA-binding domain)T[180-246] 3.7e-14

Sequence "ZM1S60958648" crc64 checksum: 10C1558B1D15EA57 length: 426 aa.

InterPro IPR013781 Glycoside hydrolase, subgroup, catalytic core
Molecular Function: catalytic activity (GO:0003824), Biological Process: carbohydrate metabolic process
(GO:0005975), Molecular Function: cation binding (GO:0043169)
method AccNumber shortName location
Gene3D G3DSA:3.20.20.80no description T[28-410] 2.9e-132

InterPro IPR017853 Glycoside hydrolase, catalytic core
method AccNumber shortName location
superfamily SSF51445 (Trans)glycosidases T[28-400] 7.5e-89

InterPro NULL NULL

method	AccNumber	shortName	location
Seg	seg	seg	?[2-14] NA ?[70-83] NA ?[91-96] NA
?[409-425] NA			
SignalPHMM	signalp	signal-peptide	?[1-22] NA
Sequence "ZM1S60983186" crc64 checksum: E8BAF56AF1838CBC length: 491 aa.			
InterPro	IPR000719	Protein kinase, core	
Molecular Function:	protein kinase activity (GO:0004672), Molecular Function: ATP binding (GO:0005524),		
Biological Process:	protein amino acid phosphorylation (GO:0006468)		
method	AccNumber	shortName	location
BlastProDom	PD000001	Q94GE0_ORYSA_Q94GE0;	T[69-313] 3e-134
ProfileScan	PS50011	PROTEIN_KINASE_DOM	T[55-322] 20.160
InterPro	IPR001245	Tyrosine protein kinase	
Molecular Function:	protein-tyrosine kinase activity (GO:0004713), Molecular Function: ATP binding (GO:0005524), Biological Process: protein amino acid phosphorylation (GO:0006468)		
method	AccNumber	shortName	location
HMMPfam	PF07714	Pkinase_Tyr	T[68-134] 2.1e-06
HMMSmart	SM00219	no description	T[55-307] 2.1e-08
InterPro	IPR002290	Serine/threonine protein kinase	
Molecular Function:	protein serine/threonine kinase activity (GO:0004674), Molecular Function: ATP binding (GO:0005524), Biological Process: protein amino acid phosphorylation (GO:0006468)		
method	AccNumber	shortName	location
HMMSmart	SM00220	no description	T[55-316] 7.4e-08
InterPro	IPR011009	Protein kinase-like	
method	AccNumber	shortName	location
superfamily	SSF56112	Protein kinase-like (PK-like)	T[37-327] 2.2e-43
InterPro	IPR011990	Tetratricopeptide-like helical	
Molecular Function:	binding (GO:0005488)		
method	AccNumber	shortName	location
Gene3D	G3DSA:1.25.40.10	no description	T[378-481] 1.5e-09
InterPro	NULL	NULL	
method	AccNumber	shortName	location
Gene3D	G3DSA:1.10.510.10	no description	T[131-335] 7.2e-23

Gene3D	G3DSA:3.30.200.20no description	T[33-131]	4.8e-11
HMPanther	PTHR23258	SERINE-THREONINE PROTEIN KINASE, PLANT-TYPET	[29-484] 4.6e-262
HMPanther	PTHR23258:SF457PROTEIN KINASE-LIKE PROTEIN	T[29-484]	4.6e-262
superfamily	SSF48452	TPR-like	T[382-480] 9.5e-10
Sequence "ZM1S62018816" crc64 checksum: DCA5D5DC3B026BD4 length: 426 aa.			
InterPro	NULL	location	
method	AccNumber	shortName	
Seg	seg	seg	?[112-137] NA ?[248-263] NA
TMHMM	tmhmm	transmembrane_regions	?[315-337] NA
Sequence "ZM1S62063823" crc64 checksum: 329D087C5DB89A45 length: 390 aa.			
InterPro	IPR006946	Protein of unknown function DUF642	
method	AccNumber	shortName	
HMPFfam	PF04862	DUF642	location
			T[64-389] 2.4e-193
InterPro	IPR008979	Galactose-binding like	
method	AccNumber	shortName	
superfamily	SSF49785	Galactose-binding domain-like	location
			T[51-209] 4.2e-07 T[219-388] 0.0019
InterPro	NULL	location	
method	AccNumber	shortName	
Seg	seg	seg	?[8-42] NA
SignalPHMM	signalp	signal-peptide	?[1-25] NA
Sequence "ZM4S45895" crc64 checksum: 9F0BD4B998057088 length: 119 aa.			
InterPro	IPR007741	Mitochondrial ribosome	
method	AccNumber	shortName	
HMPFfam	PF05047	L51_S25_CI-B8	location
			T[18-91] 3.2e-24
InterPro	NULL	location	
method	AccNumber	shortName	
HMPanther	PTHR21396	39S RIBOSOMAL PROTEIN L43	T[4-114] 1.2e-48
HMPanther	PTHR21396:SF2	60S RIBOSOMAL PROTEIN L51	T[4-114] 1.2e-48
Sequence "ZM1S68704853" crc64 checksum: C1B601532AE00FF3 length: 305 aa.			
InterPro	IPR001680	WD40 repeat	
method	AccNumber	shortName	
			location

BlastProDom	PD000018	Q8W403_ORYSA_Q8W403;	T[50-84]	4e-13	T[209-244]	4e-05
FPrintScan	PR00320	GPROTEINBRPT	T[72-86]	8.9e-06	T[182-196]	8.9e-06
T[232-246]	8.9e-06					
HMMPfam	PF00400	WD40	T[1-39]	0.0003	T[45-85]	2.3e-10
133]	1.3e-07	T[140-195]	4.3e-09	T[203-245]	1.5e-08	
HMMSmart	SM00320	no description	T[2-39]	0.41	T[44-85]	2.2e-08
133]	2.7e-05	T[139-195]	0.033	T[202-245]	T[254-290]	0.55
ProfileScan	PS50082	WD_REPEATS_2	T[7-37]	8.537	T[51-85]	10.408
204]	8.704	T[209-244]	10.943			
ProfileScan	PS50294	WD_REPEATS_REGION	T[7-299]	27.948		
InterPro	IPR011046	WD40 repeat-like				
method	AccNumber	shortName	location			
superfamily	SSF50978	WD40 repeat-like	T[1-290]	1.9e-50		
InterPro	IPR015943	WD40/YVTN repeat-like				
method	AccNumber	shortName	location			
Gene3D	G3DSA:2.130.10.10	no description	T[8-294]	7.3e-61		
InterPro	NULL	NULL				
method	AccNumber	shortName	location			
HMMPanther	PTHR11024	PROTEIN TRANSPORT PROTEIN SEC13-RELATED	T[7-305]	1.7e-175		
HMMPanther	PTHR11024:SF2	PROTEIN TRANSPORT PROTEIN SEC13	T[7-305]	1.7e-175		
Sequence "ZM4S07964" crc64 checksum: 5F668E8D83CF5645 length: 82 aa.						
InterPro	IPR001931	Ribosomal protein S21e				
Molecular Function: structural constituent of ribosome (GO:0003735), Cellular Component: intracellular (GO:0005622), Cellular Component: ribosome (GO:0005840), Biological Process: translation (GO:0006412)						
method	AccNumber	shortName	location			
BlastProDom	PD006584	RS21_ORYSA_P35687;	T[26-74]	8e-23		
HMMPiR	PIRSF002148	Ribosomal protein S21e	T[1-82]	1e-50		
HMMPanther	PTHR10442	40S RIBOSOMAL PROTEIN S21E	T[2-82]	4.9e-47		
HMMPfam	PF01249	Ribosomal_S21e	T[1-81]	8.5e-60		
ScanRegExp	PS00996	RIBOSOMAL_S21E	T[11-19]	8e-5		
Sequence "ZM1S65239103" crc64 checksum: D9F7188DB3E52343 length: 300 aa.						
InterPro	IPR001005	SANT, DNA-binding				
Molecular Function: DNA binding (GO:0003677)						
method	AccNumber	shortName	location			

HMMSmart	SM00717	no description	T[13-63]	7.2e-12	T[66-114]	5.6e-15
ProfileScan	PS50090	MYB_3	T[9-61]	15.442	T[62-112]	15.512
ScanRegExp	PS00037	MYB_1	?[17-25]	NA		
ScanRegExp	PS00334	MYB_2	T[89-112]	8e-5		
InterPro	IPR009057	Homeodomain-like	location			
method	AccNumber	shortName	T[14-84]	9.3e-19	T[86-146]	1e-13
superfamily	SSF46689	Homeodomain-like				
InterPro	IPR012287	Homeodomain-related				
Molecular Function: DNA binding (GO:0003677), Biological Process: regulation of transcription (GO:0045449)						
method	AccNumber	shortName	location			
Gene3D	G3DSA:1.10.10.60	no description	T[12-64]	2.2e-13	T[68-116]	1e-16
InterPro	IPR014778	Myb, DNA-binding				
method	AccNumber	shortName	location			
HMPfam	PF00249	Myb_DNA-binding	T[14-61]	4.7e-10	T[67-112]	5.2e-10
InterPro	IPR015495	Myb transcription factor				
method	AccNumber	shortName	location			
HMPanther	PTHR10641	MYB-RELATED	T[3-142]	2.5e-97		
InterPro	NULL	NULL				
method	AccNumber	shortName	location			
HMPanther	PTHR10641:SF30	R2R3-MYB TRANSCRIPTION FACTOR	T[3-142]	2.5e-97		
Seg	seg	seg	?[131-143]	NA	?[191-236]	NA
Sequence "ZM1S59258289" crc64 checksum: E21693C982A6C0C9 length: 338 aa.						
InterPro	IPR005123	2OG-Fe(II) oxygenase				
method	AccNumber	shortName	location			
HMPfam	PF03171	2OG-FeII_Oxy	T[177-287]	1.5e-33		
InterPro	NULL	NULL				
method	AccNumber	shortName	location			
Gene3D	G3DSA:2.60.120.33	no description	T[29-335]	6.2e-76		
HMPanther	PTHR10209	IRON/ASCORBATE-DEPENDENT OXIDOREDUCTASE	FAMILY MEMBERT	[39-326]	6.3e-133	
HMPanther	PTHR10209:SF8	GIBBERELLIN 2-BETA-DIOXYGENASE 2	T[39-326]	6.3e-133		
Seg	seg	seg	?[113-125]	NA	?[142-152]	NA

superfamily	SSF511197	Clavaminate synthase-like	T[29-329]	1.4e-78
Sequence "ZM1A68352777" crc64 checksum: 0ED63F3B6A2E49FA length: 563 aa.				
InterPro	IPR001128	Cytochrome P450		
Molecular Function: monooxygenase activity (GO:0004497), Molecular Function: iron ion binding (GO:0005506), Biological Process: electron transport (GO:0006118), Molecular Function: heme binding (GO:0020037)				
method	AccNumber	shortName	location	
FPrintScan	PR00385	P450	T[354-371]	6.5e-14 T[407-418] 6.5e-14
T[484-493]	6.5e-14 T[493-504]	6.5e-14		
Gene3D	G3DSA:1.10.630.10no description		T[94-548]	1.6e-104
HMMPanther	PTHR19383	CYTOCHROME P450	T[27-73]	5.4e-185 T[94-514] 5.4e-185
HMMPfam	PF00067	p450	T[86-542]	1.4e-59
ScanRegExp	PS00086	CYTOCHROME_P450	?	[486-495] NA
superfamily	SSF48264	Cytochrome P450	T[57-544]	2.7e-109
Sequence "ZM1A68352777" crc64 checksum: 0ED63F3B6A2E49FA length: 563 aa.				
InterPro	IPR002401	Cytochrome P450, E-class, group I		
Molecular Function: monooxygenase activity (GO:0004497), Molecular Function: iron ion binding (GO:0005506), Biological Process: electron transport (GO:0006118), Molecular Function: heme binding (GO:0020037)				
method	AccNumber	shortName	location	
FPrintScan	PR00463	EP450I	T[104-123]	2.3e-24 T[128-149] 2.3e-24
T[343-360]	2.3e-24 T[363-389]	2.3e-24 T[483-493] 2.3e-24 T[493-516]	2.3e-24	
Sequence "ZM1A68352777" crc64 checksum: 0ED63F3B6A2E49FA length: 563 aa.				
InterPro	NULL	NULL		
method	AccNumber	shortName	location	
HMMPanther	PTHR19383:SF174CYTOCHROME P450, SUBFAMILY 72		T[27-73]	5.4e-185 T[94-514] 5.4e-185
Seg	seg	seg	?	[18-30] NA ?[327-335] NA
SignalPHMM	signalp	signal-peptide	?	[1-35] NA
TMHMM	tmhmm	transmembrane_regions	?	[13-35] NA
Sequence "ZM4S13278" crc64 checksum: 8FA559338699CF78 length: 610 aa.				
InterPro	IPR004993	GH3 auxin-responsive promoter		
method	AccNumber	shortName	location	
HMMPfam	PF03321	GH3	T[35-584]	0
Sequence "ZM4S13278" crc64 checksum: 8FA559338699CF78 length: 610 aa.				
InterPro	NULL	NULL		
method	AccNumber	shortName	location	
Seg	seg	seg	?	[13-29] NA

Sequence "ZM1S61040508" crc64 checksum: 131843319F9E7FD8 length: 614 aa.
InterPro IPR004993 GH3 auxin-responsive promoter location
method AccNumber shortName T[33-588] 0
HMMPfam PF03321 GH3

InterPro NULL
method AccNumber shortName
Seg seg ?[72-83] NA ?[386-396] NA

Sequence "ZM1S62064657" crc64 checksum: 0B1A10A3C6485A7F length: 312 aa.
InterPro IPR003441 No apical meristem (NAM) protein
Molecular Function: DNA binding (GO:0003677), Biological Process: regulation of transcription (GO:0045449)
method AccNumber shortName location
HMMPfam PF02365 NAM T[17-143] 5.6e-86
ProfileScan PS51005 NAC T[17-169] 60.395
superfamily SSF101941 NAC domain T[1-170] 1.3e-68

InterPro NULL
method AccNumber shortName location
Seg seg ?[141-158] NA ?[198-211] NA

Sequence "ZM4S14568" crc64 checksum: B6E028AA73574C57 length: 368 aa.
InterPro IPR001222 Zinc finger, TFIIS-type
Molecular Function: nucleic acid binding (GO:0003676), Molecular Function: zinc ion binding (GO:0008270),
Molecular Function: transcription regulator activity (GO:0030528), Biological Process: regulation of
transcription (GO:0045449)
method AccNumber shortName location
HMMPfam PF01096 TFIIS_C T[328-366] 7.6e-23
HMMSmart SM00440 no description T[328-367] 3.5e-22
ProfileScan PS51133 ZF_TFIIS_2 T[326-366] 14.437
ScanRegExp PS00466 ZF_TFIIS_1 T[330-365] 8e-5

InterPro IPR003617 Transcription elongation factor, TFIIS/CRSP70, N-terminal, sub-type
Cellular Component: nucleus (GO:0005634), Biological Process: transcription (GO:0006350)
method AccNumber shortName location
HMMSmart SM00509 no description T[8-85] 2.3e-12

InterPro IPR003618 Transcription elongation factor S-II, central region

Biological Process: transcription (GO:0006350)

method	AccNumber	shortName	location
HMMPfam	PF07500	TFIIS_M	T[189-317] 2.1e-48
HMMSmart	SM00510	no description	T[191-306] 3.2e-23
superfamily	SSF46942	Elongation factor TFIIS domain 2	T[172-299] 4.5e-24

InterPro IPR006289 Transcription elongation factor, TFIIS
Molecular Function: nucleic acid binding (GO:0003676), Molecular Function: RNA polymerase II transcription factor activity (GO:0003702), Molecular Function: translation elongation factor activity (GO:0003746), Molecular Function: protein binding (GO:0005515), Cellular Component: nucleus (GO:0005634), Biological Process: RNA elongation (GO:0006354), Biological Process: regulation of transcription, DNA-dependent (GO:0006355)

method	AccNumber	shortName	location
HMMTigr	TIGR01385	TFSII: transcription elongation factor S-IIT[11-368]	4.4e-101

InterPro IPR010990 Transcription elongation factor, TFIIS/elongin A/CRSP70, N-terminal
Molecular Function: DNA binding (GO:0003677), Molecular Function: protein binding (GO:0005515), Cellular Component: nucleus (GO:0005634), Biological Process: transcription (GO:0006350), Molecular Function: transcription regulator activity (GO:0030528)

method	AccNumber	shortName	location
HMMPfam	PF08711	TFIIS	T[1-93] 5.5e-06
superfamily	SSF47676	Conserved domain common to transcription factors TFIIS, elongin A, CRSP70T[9-84] 2.5e-10	

InterPro IPR014765 Transcription elongation factor, TFIIS/CRSP70, N-terminal
Molecular Function: DNA binding (GO:0003677), Molecular Function: protein binding (GO:0005515), Cellular Component: nucleus (GO:0005634), Biological Process: transcription (GO:0006350), Molecular Function: transcription regulator activity (GO:0030528)

method	AccNumber	shortName	location
Gene3D	G3DSA:1.25.40.260	no description	T[24-100] 3e-07

InterPro IPR016492 Transcription elongation factor, IIS
Molecular Function: transcription elongation regulator activity (GO:0003711), Biological Process: regulation of transcription from RNA polymerase II promoter (GO:0006357)

method	AccNumber	shortName	location
HMMPIR	PIRSF006704	Transcription elongation factor, IIS typeT[2-368]	1.7e-125

InterPro NULL NULL

method	Gene3D	AccNumber	shortName	location
		G3DSA:2.20.25.10	no description	T[299-368] 4.4e-20
	HMPanther	PTHR11477	TRANSCRIPTION ELONGATION FACTOR S-II	T[1-97] 8.4e-51 T[129-367] 8.4e-51
	Seg	seg	seg	?[10-23] NA
	superfamily	SSF57783	Zinc beta-ribbon	T[305-367] 3.1e-21

Sequence "ZM1S62075387" crc64 checksum: 912650C1CC470182 length: 752 aa.

InterPro	IPR008217	Protein of unknown function DUF125, transmembrane
method	AccNumber	shortName
HMPfam	PF01988	DUF125
		location
		T[565-732] 7.099998842111971E-11

Example 5: Topology prediction of the GRP polypeptide sequences

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

A number of parameters must be selected before analysing a sequence, such as organism group (non-plant or plant), cutoff sets (none, predefined set of cutoffs, or user-specified set of cutoffs), and the calculation of prediction of cleavage sites (yes or no).

Results of TargetP 1.1 analysis of a selection of GRP polypeptide sequence from Table A are listed hereunder in Table B.

The “plant” organism group has been selected, no cutoffs defined, and the predicted length of the transit peptide requested.

Table B: TargetP 1.1 analysis of some polypeptide sequences listed in Table A, represented by their identifier. Abbreviations: Len, Length; cTP, Chloroplastic transit peptide; mTP, Mitochondrial transit peptide, SP, Secretory pathway signal peptide, other, Other subcellular targeting, Loc, Predicted Location; RC, Reliability class; TPlen, Predicted transit peptide length.

Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
ZM1s55817549	361	0.056	0.334	0.056	0.728	—	4	—
cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
ZM1S60958648	426	0.006	0.403	0.716	0.018	S	4	22
cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
ZM1s60983186	491	0.009	0.431	0.084	0.775	—	4	—
cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
ZM1S62018816	426	0.501	0.796	0.002	0.042	M	4	12

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM1s62063823	390	0.192	0.035	0.646	0.012	S	3	25

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM4s45895	119	0.026	0.527	0.062	0.138	M	4	102

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM1s68704853	305	0.079	0.093	0.052	0.919	—	1	—

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM4s07964	82	0.108	0.285	0.064	0.769	—	3	—

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM1S65239103	300	0.105	0.058	0.040	0.882	—	2	—

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM1s59258289	338	0.325	0.231	0.184	0.270	C	5	27

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM1a68352777	563	0.038	0.249	0.146	0.370	—	5	—

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM4s13278	610	0.136	0.210	0.181	0.436	—	4	—

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM1s61040508	614	0.296	0.147	0.165	0.435	—	5	—

cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen

ZM1s62064657	312	0.072	0.219	0.049	0.886	—	2	—
cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
ZM4s14568	368	0.131	0.047	0.052	0.913	—	2	—
cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
ZM1s62075387	752	0.160	0.089	0.076	0.895	—	2	—
cutoff		0.000	0.000	0.000	0.000			

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

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

Example 6: Cloning of the GRP encoding nucleic acid sequence

The nucleic acid sequences was amplified by PCR using as template a custom-made *Zea mays*, *Oryza sativa* or *Arabidopsis thaliana* seedlings cDNA libraries. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were designed using standard methods for gene specific amplification and included the AttB sites for Gateway recombination. The amplified PCR fragment was purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombined in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an “entry clone”. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising the coding sequences provided in Table A 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. In one set of experiments, the rice GOS2 promoter (SEQ ID NO: 643) for constitutive

expression was located upstream of this Gateway cassette. In one embodiment, the Gateway cassette comprising the nucleic acid sequence of interest was designed for overexpressing the nucleic acid sequence of interest; in another embodiment, the Gateway cassette comprising the nucleic acid sequence of interest was designed for decreasing expression of the nucleic acid sequence of interest (as indicated in Table C). In another set of experiments, the CaMV-35S promoter for constitutive expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector 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 to 60 minutes, preferably 30 minutes in sodium hypochlorite solution (depending on the grade of contamination), followed by a 3 to 6 times, preferably 4 time wash with sterile distilled water. The sterile seeds were then germinated on a medium containing 2,4-D (callus induction medium). After incubation in light for 6 days scutellum-derived calli is transformed with *Agrobacterium* as described herein below.

Agrobacterium strain LBA4404 containing the expression vector was used for co-cultivation. *Agrobacterium* was inoculated on AB medium with the appropriate antibiotics and cultured for 3 days at 28°C. The bacteria were then collected and suspended in liquid co-cultivation medium to a density (OD₆₀₀) of about 1. The calli were immersed in the suspension for 1 to 15 minutes. The callus tissues were then blotted dry on a filter paper and transferred to solidified, co-cultivation medium and incubated for 3 days in the dark at 25°C. After washing away the *Agrobacterium*, the calli were grown on 2,4-D-containing medium for 10 to 14 days (growth time for indica: 3 weeks) under light at 28°C - 32°C in the presence of a selection agent. During this period, rapidly growing resistant callus developed. After transfer of this material to regeneration media, the embryogenic potential was released and shoots developed in the next four to six weeks. Shoots were excised from the calli and incubated for 2 to 3 weeks on an auxin-containing medium from which they were transferred to soil. Hardened shoots were grown under high humidity and short days in a greenhouse.

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

35 to 90 independent T0 rice transformants were generated for one construct. The primary transformants were transferred from a tissue culture chamber to a greenhouse. After a quantitative PCR analysis to verify copy number of the T-DNA insert, only single copy

transgenic plants that exhibit tolerance to the selection agent were kept for harvest of T1 seed. Seeds were then harvested three to five months after transplanting. The method yielded single locus transformants at a rate of over 50 % (Aldemita and Hodges1996, Chan et al. 1993, Hiei et al. 1994).

Example 8: Transformation of other crops

Corn transformation

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

Wheat transformation

Transformation of wheat is performed with the method described by Ishida et al. (1996) Nature Biotech 14(6): 745-50. The cultivar Bobwhite (available from CIMMYT, Mexico) is commonly used in transformation. Immature embryos are co-cultivated with *Agrobacterium tumefaciens* containing the expression vector, and transgenic plants are recovered through organogenesis. After incubation with *Agrobacterium*, the embryos are grown in vitro on callus induction medium, then regeneration medium, containing the selection agent (for example imidazolinone but various selection markers can be used). The Petri plates are incubated in the light at 25 °C for 2-3 weeks, or until shoots develop. The green shoots are transferred from each embryo to rooting medium and incubated at 25 °C for 2-3 weeks, until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Soybean transformation

Soybean is transformed according to a modification of the method described in the Texas A&M patent US 5,164,310. Several commercial soybean varieties are amenable to transformation by this method. The cultivar Jack (available from the Illinois Seed

foundation) is commonly used for transformation. Soybean seeds are sterilised for *in vitro* sowing. The hypocotyl, the radicle and one cotyledon are excised from seven-day old young seedlings. The epicotyl and the remaining cotyledon are further grown to develop axillary nodes. These axillary nodes are excised and incubated with *Agrobacterium tumefaciens* containing the expression vector. After the cocultivation treatment, the explants are washed and transferred to selection media. Regenerated shoots are excised and placed on a shoot elongation medium. Shoots no longer than 1 cm are placed on rooting medium until roots develop. The rooted shoots are transplanted to soil in the greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Rapeseed/canola transformation

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

Alfalfa transformation

A regenerating clone of alfalfa (*Medicago sativa*) is transformed using the method of (McKersie et al., 1999 Plant Physiol 119: 839–847). Regeneration and transformation of alfalfa is genotype dependent and therefore a regenerating plant is required. Methods to obtain regenerating plants have been described. For example, these can be selected from the cultivar Rangelander (Agriculture Canada) or any other commercial alfalfa variety as described by Brown DCW and A Atanassov (1985. Plant Cell Tissue Organ Culture 4: 111-112). Alternatively, the RA3 variety (University of Wisconsin) has been selected for use in tissue culture (Walker et al., 1978 Am J Bot 65:654-659). Petiole explants are cocultivated with an overnight culture of *Agrobacterium tumefaciens* C58C1 pMP90 (McKersie et al., 1999 Plant Physiol 119: 839–847) or LBA4404 containing the expression vector. The explants are cocultivated for 3 d in the dark on SH induction medium containing 288 mg/ L

Pro, 53 mg/ L thioproline, 4.35 g/ L K₂SO₄, and 100 µm acetosyringinone. The explants are washed in half-strength Murashige-Skoog medium (Murashige and Skoog, 1962) and plated on the same SH induction medium without acetosyringinone but with a suitable selection agent and suitable antibiotic to inhibit *Agrobacterium* growth. After several weeks, somatic embryos are transferred to BOi2Y development medium containing no growth regulators, no antibiotics, and 50 g/ L sucrose. Somatic embryos are subsequently germinated on half-strength Murashige-Skoog medium. Rooted seedlings were transplanted into pots and grown in a greenhouse. T1 seeds are produced from plants that exhibit tolerance to the selection agent and that contain a single copy of the T-DNA insert.

Cotton transformation

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

Sugarbeet transformation

Seeds of sugarbeet (*Beta vulgaris* L.) are sterilized in 70% ethanol for one minute followed by 20 min. shaking in 20% Hypochlorite bleach e.g. Clorox® regular bleach (commercially available from Clorox, 1221 Broadway, Oakland, CA 94612, USA). Seeds are rinsed with sterile water and air dried followed by plating onto germinating medium (Murashige and Skoog (MS) based medium (Murashige, T., and Skoog, ., 1962. Physiol. Plant, vol. 15, 473-497) including B5 vitamins (Gamborg et al.; Exp. Cell Res., vol. 50, 151-8.) supplemented with 10 g/l sucrose and 0,8% agar). Hypocotyl tissue is used essentially for the initiation of shoot cultures according to Hussey and Hepher (Hussey, G., and Hepher, A., 1978. Annals

of Botany, 42, 477-9) and are maintained on MS based medium supplemented with 30g/l sucrose plus 0,25mg/l benzylamino purine and 0,75% agar, pH 5,8 at 23-25°C with a 16-hour photoperiod. *Agrobacterium tumefaciens* strain carrying a binary plasmid harbouring a selectable marker gene, for example nptII, is used in transformation experiments. One day before transformation, a liquid LB culture including antibiotics is grown on a shaker (28°C, 150rpm) until an optical density (O.D.) at 600 nm of ~1 is reached. Overnight-grown bacterial cultures are centrifuged and resuspended in inoculation medium (O.D. ~1) including Acetosyringone, pH 5,5. Shoot base tissue is cut into slices (1.0 cm x 1.0 cm x 2.0 mm approximately). Tissue is immersed for 30s in liquid bacterial inoculation medium. Excess liquid is removed by filter paper blotting. Co-cultivation occurred for 24-72 hours on MS based medium incl. 30g/l sucrose followed by a non-selective period including MS based medium, 30g/l sucrose with 1 mg/l BAP to induce shoot development and cefotaxim for eliminating the *Agrobacterium*. After 3-10 days explants are transferred to similar selective medium harbouring for example kanamycin or G418 (50-100 mg/l genotype dependent). Tissues are transferred to fresh medium every 2-3 weeks to maintain selection pressure. The very rapid initiation of shoots (after 3-4 days) indicates regeneration of existing meristems rather than organogenesis of newly developed transgenic meristems. Small shoots are transferred after several rounds of subculture to root induction medium containing 5 mg/l NAA and kanamycin or G418. Additional steps are taken to reduce the potential of generating transformed plants that are chimeric (partially transgenic). Tissue samples from regenerated shoots are used for DNA analysis. Other transformation methods for sugarbeet are known in the art, for example those by Linsey & Gallois (Linsey, K., and Gallois, P., 1990. *Journal of Experimental Botany*; vol. 41, No. 226; 529-36) or the methods published in the international application published as WO9623891A.

Sugarcane transformation

Spindles are isolated from 6-month-old field grown sugarcane plants (Arencibia et al., 1998. *Transgenic Research*, vol. 7, 213-22; Enriquez-Obregon et al., 1998. *Planta*, vol. 206, 20-27). Material is sterilized by immersion in a 20% Hypochlorite bleach e.g. Clorox® regular bleach (commercially available from Clorox, 1221 Broadway, Oakland, CA 94612, USA) for 20 minutes. Transverse sections around 0,5cm are placed on the medium in the top-up direction. Plant material is cultivated for 4 weeks on MS (Murashige, T., and Skoog, ., 1962. *Physiol. Plant*, vol. 15, 473-497) based medium incl. B5 vitamins (Gamborg, O., et al., 1968. *Exp. Cell Res.*, vol. 50, 151-8) supplemented with 20g/l sucrose, 500 mg/l casein hydrolysate, 0,8% agar and 5mg/l 2,4-D at 23°C in the dark. Cultures are transferred after 4 weeks onto identical fresh medium. *Agrobacterium tumefaciens* strain carrying a binary plasmid harbouring a selectable marker gene, for example hpt, is used in transformation experiments. One day before transformation, a liquid LB culture including antibiotics is grown on a shaker (28°C, 150rpm) until an optical density (O.D.) at 600 nm of ~0,6 is reached. Overnight-grown bacterial cultures are centrifuged and resuspended in MS based inoculation medium (O.D. ~0,4) including acetosyringone, pH 5,5. Sugarcane embryogenic callus pieces (2-4 mm) are isolated based on morphological characteristics as compact

structure and yellow colour and dried for 20 min. in the flow hood followed by immersion in a liquid bacterial inoculation medium for 10-20 minutes. Excess liquid is removed by filter paper blotting. Co-cultivation occurred for 3-5 days in the dark on filter paper which is placed on top of MS based medium incl. B5 vitamins containing 1 mg/l 2,4-D. After co-cultivation calli are washed with sterile water followed by a non-selective cultivation period on similar medium containing 500 mg/l cefotaxime for eliminating remaining *Agrobacterium* cells. After 3-10 days explants are transferred to MS based selective medium incl. B5 vitamins containing 1 mg/l 2,4-D for another 3 weeks harbouring 25 mg/l of hygromycin (genotype dependent). All treatments are made at 23°C under dark conditions. Resistant calli are further cultivated on medium lacking 2,4-D including 1 mg/l BA and 25 mg/l hygromycin under 16 h light photoperiod resulting in the development of shoot structures. Shoots are isolated and cultivated on selective rooting medium (MS based including, 20g/l sucrose, 20 mg/l hygromycin and 500 mg/l cefotaxime). Tissue samples from regenerated shoots are used for DNA analysis. Other transformation methods for sugarcane are known in the art, for example from the international application published as WO2010/151634A and the granted European patent EP1831378.

Example 9: Phenotypic evaluation procedure

9.1 Evaluation setup

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

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

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

Drought screen

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

Nitrogen use efficiency screen

T1 or T2 plants 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

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

9.2 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.

9.3 Parameters measured

From the stage of sowing until the stage of maturity the plants were passed several times through a digital imaging cabinet. At each time point digital images (2048x1536 pixels, 16 million colours) were taken of each plant from at least 6 different angles as described in WO2010/031780. These measurements were used to determine different parameters.

Biomass-related parameter measurement

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

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

Parameters related to development time

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

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

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

Seed-related parameter measurements

The mature primary panicles were harvested, counted, bagged, barcode-labelled and then dried for three days in an oven at 37°C. The panicles were then threshed and all the seeds were collected and counted. The seeds are usually covered by a dry outer covering, the husk. The filled husks (herein also named filled florets) were separated from the empty ones using an air-blowing device. The empty husks were discarded and the remaining fraction was counted again. The filled husks were weighed on an analytical balance.

The total number of seeds was determined by counting the number of filled husks that remained after the separation step. The total seed weight (Seed Yield in Table C) was measured by weighing all filled husks harvested from a plant.

The total number of seeds (or florets) per plant was determined by counting the number of husks (whether filled or not) harvested from a plant.

Thousand Kernel Weight (TKW) is extrapolated from the number of seeds counted and their total weight.

The Harvest Index (HI) in the present invention is defined as the ratio between the total seed weight and the above ground area (mm²), multiplied by a factor 10⁶.

The number of flowers per panicle as defined in the present invention is the ratio between the total number of seeds over the number of mature primary panicles.

The “seed fill rate” or “seed filling rate” as defined in the present invention is the proportion (expressed as a %) of the number of filled seeds (i.e. florets containing seeds) over the total number of seeds (i.e. total number of florets). In other words, the seed filling rate is the percentage of florets that are filled with seed.

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

The results of the evaluation of transgenic rice plants in the T1 generation and expressing a nucleic acid encoding a GRP polypeptide as given in Table A are presented below in Table C. In a number of cases the gene of interest was downregulated, indicated in the column “expression”. In the Yield screen, plants were grown under optimal conditions, in the Nitrogen screen, plants were grown under reduced nutrient availability as described above, in the Drought screen, plants were subjected to reduced water availability, as described above.

Table C: Data summary for transgenic T1 rice plants; for each parameter, the phenotype is indicated as follows: ++++: at least 4 positive events with a positive difference >5%, p<0.2 for any of the parameters indicated, +++: 3 positive events with a positive difference >5%, p<0.1, ++: 2 positive events with a positive difference >10%, p<0.1.

Identifier	Protein SEQ ID NO:	Phenotype ++++	Phenotype +++	Phenotype ++	Screen	Expression
Orysa_CBS domain containing membrane protein	4			Flowers per panicle	Yield	downregulation
Orysa_LOC_Os01g06640	8		Seed yield	Biomass, Number of flowers, Fillrate, Harvest index; Seed size (TKW)	Yield	downregulation
Orysa_LOC_Os01g63690	10		Seed yield, Fillrate, Harvest index, Number of filled seeds		Yield	downregulation
Orysa_LOC_Os02g36490	28			Harvest index	Yield	downregulation
Orysa_LOC_Os02g52560	34			Fillrate, Harvest index	Yield	downregulation
Orysa_LOC_Os03g10210	40		Fillrate		Yield	downregulation
Orysa_LOC_Os03g17790	44			Biomass	Yield	downregulation
Orysa_LOC_Os03g17800	46		Seed yield, number of filled seeds	Fillrate, Harvest Index, Flowers per panicle	Yield	downregulation
Orysa_LOC_Os03g46490	48	Flowers per panicle		Seed Yield, Number of flowers	Yield	downregulation
Orysa_LOC_Os03g55670	50			Fillrate	Yield	downregulation
Orysa_LOC_Os04g35790	54			Fillrate, Harvest index	Yield	downregulation

Orysa_LOC_Os05g07070	56				Seed yield, Fillrate, number of filled seeds	Yield	downregulation
Orysa_LOC_Os05g40770	58		Harvest Index		Seed yield, Fillrate, Seed size (TKW), Number of filled seeds	yield	downregulation
Orysa_LOC_Os06g01440	62	Seed Yield, Fillrate, Harvest Index, Number of filled seeds				yield	downregulation
Orysa_LOC_Os06g06250.2	66				Fillrate	Yield	downregulation
Orysa_LOC_Os08g01760	74				Seed yield, Number of flowers, Fillrate, Harvest index, Number of panicles	Yield	downregulation
Orysa_LOC_Os08g26870	76				Fillrate, Harvest index	Yield	downregulation
Orysa_LOC_Os10g39130	92				Number of flowers	Nitrogen	downregulation
Orysa_WRKY7	108				Seed yield, Fillrate, Harvest index, Number of panicles, Number of filled seeds	Drought	downregulation
Zeama_GRMZM2G005260	128				Fillrate	Drought	normal
Zeama_GRMZM2G005435	130				Seed siz (TKW)	Drought	normal
Zeama_GRMZM2G010356_ T01	144				Fillrate	Yield	normal
Zeama_GRMZM2G012224_ T01	150	Fillrate				Nitrogen	normal

Zeama_GRMZM2G013627	152				Seed yield, Fillrate, Harvest index, Number of of panicles, Number of filled seeds	Drought	normal
Zeama_GRMZM2G014750	156				Seed yield, Number of flowers, Fillrate, Harvest index, Flowers per panicle	Drought	normal
Zeama_GRMZM2G015433	158			Biomass		Nitrogen	normal
Zeama_GRMZM2G020281_ T03	166				Seed yield, Fillrate, Harvest index, Seed size (TKW), Number of filled seeds	yield	normal
Zeama_GRMZM2G028218	176				Seed size (TKW)	Yield	normal
Zeama_GRMZM2G031724	182		Root Shoot Index			Drought	normal
Zeama_GRMZM2G031983_ T01	184		Seed size (TKW)	Fillrate	Harvest Index	Yield	normal
Zeama_GRMZM2G037001_ T01	196				Harvest index, Seed size (TKW), Number of Panicles	Yield	normal
Zeama_GRMZM2G040517_ T01	200				Seed yield, Harvest index, Number of filled seeds	Nitrogen	normal
Zeama_GRMZM2G043147	206				Seed yield, Fillrate, Harvest index, Number of filled seeds	Drought	normal
Zeama_GRMZM2G044027_	210			Biomass,	Seed yield, number of	Yield	normal

T01				Number of flowers	panicles, flowers per panicle		
Zeama_GRMZM2G047684	216	Root Shoot Index			Nitrogen	normal	
Zeama_GRMZM2G049398_ T01	220	Fillrate, Harvest index			Seed Yield, number of filled seeds	normal	
Zeama_GRMZM2G053766	226			Number of Panicles	Drought	normal	
Zeama_GRMZM2G055585	230	Fillrate		Harvest Index	Drought	normal	
Zeama_GRMZM2G058690_ T02	242			Biomass	Yield	normal	
Zeama_GRMZM2G059580	244				Fillrate	normal	
Zeama_GRMZM2G059799	246				Flowers per panicle	normal	
Zeama_GRMZM2G061745	250				Seed yield, Fillrate, Number of filled seeds	normal	
Zeama_GRMZM2G061996_ T01	252				Seed yield, Fillrate, Harvest index, Number of filled seeds	normal	
Zeama_GRMZM2G062160	254			Fillrate, Number of filled seeds	Seed yield, Harvest index	normal	
Zeama_GRMZM2G062531	256			Flowers per panicle	Drought	normal	
Zeama_GRMZM2G077458	270				Seed yield, Fillrate, number of filled seeds	normal	
Zeama_GRMZM2G087323	292				Seed yield, Number of filled seeds	normal	
Zeama_GRMZM2G092137	306				Seed size (TKW)	normal	
Zeama_GRMZM2G101390	326			Biomass	Nitrogen	normal	

Zeama_GRMZM2G102356	332				Fillrate, Number of panicles	Drought	normal
Zeama_GRMZM2G104610_T01	336				Fillrate	Yield	normal
Zeama_GRMZM2G105331	338				Number of panicles	Drought	normal
Zeama_GRMZM2G105348	340			Number of filled seeds	Seed yield, fillrate, harvest index	Drought	normal
Zeama_GRMZM2G107839	346				Seed yield, Fillrate, Harvest index	Drought	normal
Zeama_GRMZM2G108716	350	Seed Yield, Fillrate, Harvest Index, Number of filled seeds				Drought	normal
Zeama_GRMZM2G110289	354				Biomass, Seed yield	Nitrogen	normal
Zeama_GRMZM2G112210	358				Number of panicles	Drought	normal
Zeama_GRMZM2G115975	368				Seed yield, Fillrate, Harvest index	Drought	normal
Zeama_GRMZM2G116632_T01	372				Fillrate	Yield	normal
Zeama_GRMZM2G119906	376				Seed size (TKW)	Nitrogen	normal
Zeama_GRMZM2G126397	388				Seed yield, Fillrate	Drought	normal
Zeama_GRMZM2G126566	390	Flowers per panicle				Drought	normal
Zeama_GRMZM2G127850_T01	396	fillrate, harvest index, number of filled seeds			Seed size (TKW), seed yield	yield	normal
Zeama_GRMZM2G127893	398				Seed yield, Fillrate,	Drought	normal

					Harvest index, number of filled seeds		
Zeama_GRMZM2G137596_ T01	402				Emergence Vigor, Flowers per panicle	Yield	normal
Zeama_GRMZM2G156110_ T01	422				Emergence Vigor, Seed yield, Fillrate, Harvest index, Seed size (TKW), Number of filled seeds	Yield	normal
Zeama_GRMZM2G159034	428				Seed yield, Fillrate, Harvest index, Number of filled seeds	Drought	normal
Zeama_GRMZM2G161295	430				Flowers per panicle	Nitrogen	normal
Zeama_GRMZM2G163406	432				Fillrate, Number of panicles	Drought	normal
Zeama_GRMZM2G174558	454			Flowers per panicle	Root Shoot Index	Nitrogen	normal
Zeama_GRMZM2G178815_ T01	464			Fillrate	Seed yield, Harvest index	Nitrogen	normal
Zeama_GRMZM2G180172	468		Biomass			Nitrogen	normal
Zeama_GRMZM2G302245_ T01	470			Seed size (TKW)		Yield	normal
Zeama_GRMZM2G334181_ T01	486			Fillrate	Seed yield, Harvest index	yield	normal
Zeama_GRMZM2G340130	488			Seed size (TKW)		Nitrogen	normal
Zeama_GRMZM2G347043	494		Seed size (TKW)			Nitrogen	normal
Zeama_GRMZM2G355752	498				Fillrate	Drought	normal

Zeama_GRMZM2G359952	500		Flowers per panicle			Nitrogen	normal
Zeama_GRMZM2G401521	506		Biomass		Seed yield, Fillrate, number of filled seeds	Nitrogen	normal
Zeama_GRMZM2G413337	518				Flowers per panicle	Nitrogen	normal
Zeama_GRMZM2G447984	530				Harvest index, Number of panicles	Nitrogen	normal
Zeama_GRMZM2G479260	544				Seed yield, Fillrate	Drought	normal
Zeama_GRMZM2G481888_T01	546				Biomass, RootMax, Fillrate	Yield	normal
Zeama_nucleic acid binding / zinc ion binding	554		Fillrate		Harvest Index	Yield	normal
Zeama_putative zinc-finger protein	560				Biomass, Fillrate, Number of filled seeds	Yield	normal
Zeama_SYTE like	562				Harvest index	Drought	normal
Zeama_unknown protein a	566				Seed size (TKW)	Drought	normal
Zeama_unknown protein b	568				Emergence Vigor, Fillrate, Seed size (TKW)	Drought	normal
Zeama_Unknown protein e	574				Seed yield, Fillrate	Drought	normal
Zeama_unknown protein g	578				Flowers per panicle	Drought	normal
Zeama_ZM1s61996727	590		Biomass		RootMax	Nitrogen	normal
Zm ATP-binding protein	596				Seed yield, Fillrate, Harvest index, Number of filled seeds	Drought	normal
Zm embryo-abundant protein-related	602		Seed size (TKW)		Seed yield, Fillrate, Harvest index, Number	Drought	normal

					of filled seeds			
Zm indole-3-acetic acid amido synthetase-like 1	606	Root Shoot Index				Drought	normal	
Zm indole-3-acetic acid amido synthetase-like 2	608	Root Shoot Index				Drought	normal	
Zm indoleacetic acid-induced protein 13-like 1	610				Seed size (TKW)	Drought	normal	
Zm indoleacetic acid-induced protein 13-like 2	612				Seed yield, Fillrate, Harvest index, Number of filled seeds	Drought	normal	
Zm indoleacetic acid-induced protein 18	614				Fillrate, Seed size (TKW)	Drought	normal	
Zm indoleacetic acid-induced protein 20-like	616				Seed yield, Harvest index	Drought	normal	
Zm PHYTOCHROME-ASSOCIATED PROTEIN 1-like	620				Biomass, seed yield, seed size (TKW)	Drought	normal	
Zm protein kinase-related	626	Fillrate, Seed size (TKW)	Seed Yield		Harvest Index	Yield	normal	
Zm protein kinase-related	626	Seed size (TKW)	Seed Yield		Fillrate, Harvest Index, number of filled seeds	Drought	normal	

Example 11: Phenotypical analysis of Arabidopsis transformed with GRP encoding sequences.

In order to analyse the potential role of selected genes in root development, a detailed analysis of the root system architecture was performed on transgenic Arabidopsis lines expressing the maize transcripts.

Entry clones were prepared by cloning the *Zea mays* coding sequence in pDONRTM221 or PDONRTM/Zeo as described in Example 7. Ectopic over-expression in Arabidopsis was obtained by cloning the coding sequence in pK7WG2 destination vectors, under control of the 35S promoter (Karimi et al., Trends in plant science 7(5): 193-195, 2002), or in a destination vector with the UAS promoter, which was transformed by floral dip (Clough and Bent, The Plant journal 16(6): 735-743, 1998) in Arabidopsis thaliana (L.) Heyhn. plants, ecotype Col0. In subsequent generations single-insert and homozygous lines were selected by growth on selective medium. For xylem-pole pericycle specific expression, vectors were prepared by cloning the required fragments into pH7m24GW or pB7m24GW (Karimi et al., Plant physiology, 145, 1144-1154, 2007; Hilson, Trends in plant science, 11, 133-141, 2006) which were transformed into plants from the J0121 GAL4 enhancer-trap line (Laplaze et al., 2005).

For phenotypic analyses Arabidopsis thaliana seeds (7 to 15 plants) were sown on solid 0.5xMS medium in square Petri dishes in near-vertical position and after 48h stratification at 4°C, incubated in a growth chamber at 22°C under continuous light conditions (110 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ photosynthetically active radiation, supplied by cool-white fluorescent tungsten tubes; Osram) till 10 days after germination (DAG). Lines were not grouped per construct but analysed randomly, depending on seed availability and capacity. In each experiment a Col0 was included as a reference, which was grown for one generation under the same conditions as the transgenic lines.

The number of emerged lateral roots (LR) was counted and a high resolution scan was made from the plates with a flatbed scanner (Epson Perfection V700 Photo). Scans were analysed with ImageJ (<http://rsbweb.nih.gov/>) and WinRHIZO (Regent Instruments) software, which was used to measure the total length of the root system. Through image analyses, primary root length (PR) was measured (expressed in cm) and size of the naked zone (in cm), which is the distance between the root tip and the youngest emerged lateral. The lateral root density (LRD) was calculated by dividing the number of laterals by the length of the primary root minus the naked zone. Average values were calculated per line and are represented as percentages compared to the non-transformed Col0 control. Data are presented in Table D and Table E.

Table D: Data summary for transgenic Arabidopsis plants screened for lateral root initiation on the primary root; for each parameter the increase is shown as: + = 0-10%, ++ = 10-20%, +++ = 20-30%, ++++ = >30%; for the scores between brackets the p-value was ≤ 0.05 , otherwise the p-value was ≤ 0.01 .

Identifier	Expression	Line nr	PR	LR	LRD	Total Root Length
ZM1s60983186	35S-up	2	++	+++	(+)	
ZM1s68382282	35S-up	1	(++)	++		
	35S-up	2	(++)	(+++)		
	35S-up	3			(+)	
	35S-up	4			+	
ZM1s57394235	35S-up	1			(++)	
	35S-up	3	+			
	35S-up	4	++			
ZM1s62016945	35S-up	1	++	++++	(++++)	
	35S-up	2		(++++)	(++++)	
	35S-up	4			++	
	35S-up	5	+++	+++		
ZM1s58194594	UAS-up	1		++		
ZM1s62122111	35S-up	4				(+++)
	35S-up	5		(++)		(++++)
ZM4s18277	UAS-up	1	(++)		++	
ZM1s61987040	35S-up	2			++	
	35S-up	3			+++	
ZM1s62111342	35S-up	3				(++)
	35S-up	5				(++)
ZM3s00628	35S-up	1	+++	++++	++	(++++)
	35S-up					

In addition, the Arabidopsis knock out line SALK_022198 (downregulation of the Ara orthologue of ZM1s62132790 (At5g64810, SEQ ID NO: 641) showed improved root growth: the primary root length was increased with 17,5%, the lateral root number was increased with 16,0 % and the total root length was 26,8% increased.

Table E: Data summary for transgenic Arabidopsis plants screened for lateral root initiation on the adventitious root; for each parameter the increase is shown as: + = 0-10%, ++ = 10-20%, +++ = 20-30%, ++++ = >30%; for the scores between brackets the p-value was ≤ 0.05 , otherwise the p-value was ≤ 0.01 . For the adventitious roots, calculations for the LRD were made with or without taking the naked zone into account (respectively indicated by +NZ or -NZ).

Identifier	Expression	Line nr	PR	LR	LRD -NZ	LRD +NZ	Total Root Length
ZM4s18047	35S-up	1	++	+++			
	35S-up	4	(+)	(++)			
ZM1s61992039	35S-up	1		(++)		(++)	
	35S-up	4		+++	++	++	

	35S-up	5	++	+++		(++)	
ZM4s07707	35S-up	2		(++)	(+)	(+)	
	35S-up	3	++	+++	++	++	
	35S-up	4	++	+++		(+)	
	35S-up	5	++	+++	(+)	(+)	
ZM1s57368153	35S-up	3	++				
	35S-up	4	++	(++)			(++++)
ZM1s62202989	35S-up	1			(+)		
	35S-up	2	++	++			
	35S-up	4	++	(++)			
ZM1s57508357	35S-up	1			+++	+++	
	35S-up	2			++	+++	
	35S-up	3			++	++	
	35S-up	4			(+)	+	
ZM1s62258567	35S-up	1	+				
	35S-up	4	++	++	(+)		+++
	35S-up	5			(+)		
ZM1a62154780	35S-up	1	+++	++++	+++	++	(++++)
	35S-up	2	++++	++++	++	+++	(++++)
	35S-up	3	(+)	+++	++	++	
	35S-up	4	++	++++	+++	+++	
	35S-up	5	+++	++++	+++	+++	++++
	35S-up	6		(++)	(+)	(++)	
	35S-up	7	++	+++	(++)	++	
	35S-up	9	++	++++	++	++	
	35S-up	10		(++)	(+)	(+)	
ZM3s01330	35S-up	1		(+)			
	35S-up	3	(+)				
	35S-up	5	(+)	(++)			(+++)
ZM4s29494	35S-up	3			+		
	35S-up	4			+		
ZM1s58530935	35S-up	5	++	+++			
ZM1s62083330	35S-up	5	(++)	(++)			(++++)
ZM4s16359	35S-up	1			(+)		
	35S-up	2	++	++			
	35S-up	5	(+)				++
ZM1s61996727	35S-up	3			++	(++)	
	35S-up	4			++	++	
	35S-up	5			(++)	++	

Claims

1. A method for the production of a transgenic plant having enhanced yield relative to a control plant, comprising introduction and expression in a plant cell or plant a nucleic acid encoding a GRP polypeptide listed in Table A or a homologue thereof.
2. Method according to claim 1, wherein said nucleic acid encoding a GRP polypeptide
 - (i) encodes the GRP polypeptide represented by SEQ ID NO: 626, or
 - (ii) encodes a homologue of SEQ ID NO: 626 having at least 75% overall sequence identity to SEQ ID NO : 626, or
 - (iii) is a nucleic acid hybridising to a nucleic acid of (i) or (ii);
3. The method of claim 1 or 2, further comprising cultivation of said plant cell or plant under conditions promoting plant growth and development.
4. The method of claim 3, further comprising selection of plants with enhanced yield.
5. Method according to claim 1 or 2, wherein said increased yield comprises increased seed yield and/or increased biomass.
6. Method according to claim 5, wherein said increased seed yield comprises at least one parameter selected from the group comprising increased total seed weight, increased seed size, increased harvest index, increased fill rate, and increased number of filled seeds.
7. Method according to claim 5, wherein said increased biomass is root biomass
8. Method according to any one of claims 1 to 7, wherein said nucleic acid is operably linked to a constitutive promoter,
9. Method according to claim 8, wherein said constitutive promoter is a constitutive promoter of medium strength, preferably a GOS2 promoter.
10. Method according to claim 8, wherein said constitutive promoter is a CaMV-35S promoter.
11. Plant, or part thereof, or plant cell, obtainable by a method according to any one of claims 1 to 10, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding a GRP polypeptide as defined in claim 1 or 2.
12. Method according to any of claims 1 to 10, wherein said increased yield is obtained under non-stress conditions.

13. Method according to any of claims 1 to 10, wherein said increased yield is obtained under drought stress conditions.
14. Construct comprising:
 - (i) nucleic acid encoding a GRP polypeptide as defined in claim 1 or 2,
 - (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - (iii) a transcription termination sequence.
15. Construct of claim 14, wherein said one or more control sequences is a constitutive promoter of medium strength, preferably a GOS2 promoter.
16. Construct of claim 14, wherein said one or more control sequences is a CaMV-35S promoter.
17. Use of a construct according to any one of claims 14 to 16 in a method for making plants having enhanced yield, relative to control plants.
18. Plant, plant part or plant cell transformed with a construct according to any one of claims 14 to 16.
19. Construct according to any one of embodiments 14 to 16 comprised in a plant cell.
20. Transgenic plant having enhanced yield relative to control plants, resulting from introduction and expression of a recombinant nucleic acid encoding a GRP polypeptide as defined in claim 1 or 2 in said plant, or a transgenic plant cell derived from said transgenic plant.
21. Transgenic plant according to claim 11 or 20, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.
22. Harvestable parts of a plant according to claim 20 or 21, wherein said harvestable parts are preferably shoot biomass and/or root biomass and/or seeds.
23. Products derived from a plant according to claim 20 or 21 and/or from harvestable parts of a plant according to claim 22.
24. Use of a nucleic acid encoding a GRP polypeptide as defined in claim 1 or 2 for enhancing yield as defined in any of claims 4 to 6 in a transgenic plant relative to a control plant.