



(51) International Patent Classification:

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

(21) International Application Number:

PCT/IB2012/051938

(22) International Filing Date:

18 April 2012 (18.04.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/478,067 22 April 2011 (22.04.2011) US
61/485,131 12 May 2011 (12.05.2011) US

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

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SANZ MOLINERO, Ana Isabel** [ES/ES]; Isla Malaita 1, 6B, E-28035 Madrid (ES). **FRANKARD, Valerie** [BE/BE]; Avenue des Croix du Feu 84, E-1410 Waterloo (BE). **VANDENABEELE, Steven** [BE/BE]; Meersport 23, B-9700 Oudenaarde (BE).

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

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

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

Published:

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

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

(57) Abstract: Provided is a method for enhancing yield-related traits in plants by modulating expression of a nucleic acid encoding a bZIP-like polypeptide or a BCAT4-like polypeptide in a plant. Also provided are plants having modulated expression of a nucleic acid encoding a bZIP-like polypeptide or a BCAT4-like polypeptide, which plants have enhanced yield-related traits compared with control plants. Also provided are constructs comprising bZIP-like polypeptide-encoding nucleic acids or BCAT4-like polypeptide-encoding nucleic acids, useful in enhancing yield-related traits in plants.



PLANTS HAVING ENHANCED YIELD-RELATED TRAITS AND 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 bZIP-like (basic Leucine Zipper) polypeptide, or a BCAT4-like (Branched-Chain AminoTransferase 4-like) polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid encoding a bZIP-like 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.

With respect to bZIP-like polypeptides, bZIP proteins are a group of transcription factors, containing a conserved domain (ZIP) that participates in the formation of homo or heterodimers with other bZIP proteins and a DNA binding domain. These transcription factors form a large family and are present across fungi, animals and plants.

In plants, up to 13 groups of bZIP transcription factors have been identified in plants using evolutionary studies on nucleotide sequences. (Corrêa et al. 2008).

A subgroup within the group G of bZIP transcription factors contains a G-BOX motif binding domain. The genes of this group are mostly related to morphogenic responses to light maturation and LEA genes repression, ABA regulation, Adh activation and photomorphogenesis (Corrêa et al. 2008).

With respect to BCAT4-like polypeptides, BCAT or Branched Chain AminoTransferase, sometimes also named Branched Chain AminoTransaminase, catalyses the last step of synthesis, the transamination of the branched-chain amino acids leucine, isoleucine and valine to their respective alpha-keto acids and/or the initial step of degradation of these amino acids.

Diebold et al. (Plant Physiol, 2002, 129, 540-550) have identified in Arabidopsis seven putative BCAT genes. Maloney et al. (Plant Physiol., 2010, 153, 925-936) identified six BCAT genes from the cultivated tomato *Solanum Lycopersicum*.

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 bZIP-like polypeptide, or a BCAT4-like polypeptide.

Detailed description of the invention

With respect to bZIP-like polypeptides, the present invention shows that modulating expression in a plant of a nucleic acid encoding a bZIP-like polypeptide gives plants having enhanced yield-related traits relative to control plants.

With respect to BCAT4-like polypeptides, the present invention shows that modulating expression of a nucleic acid encoding a BCAT4-like polypeptide as defined herein gives plants having enhanced yield-related traits, in particular increased yield or more particular increased seed yield 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 bZIP-like polypeptide, or a BCAT4-like 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 bZIP-like polypeptide, or a BCAT4-like polypeptide, as described herein and optionally selecting for plants having enhanced yield-related traits.

A preferred method for modulating (preferably, increasing) expression of a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide, is by introducing and expressing in a plant a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide.

Any reference hereinafter to a "protein useful in the methods of the invention" is taken to

mean a bZIP-like polypeptide, or a BCAT4-like polypeptide, as defined herein. Any reference hereinafter to a “nucleic acid useful in the methods of the invention” is taken to mean a nucleic acid capable of encoding such a bZIP-like polypeptide, or a BCAT4-like polypeptide. The nucleic acid to be introduced into a plant (and therefore useful in performing the methods of the invention) is any nucleic acid encoding the type of protein which will now be described, hereafter also named “*bZIP-like* nucleic acid”, or “*BCAT4-like* nucleic acid”, or “*bZIP-like* gene”, or “*BCAT4-like* gene”.

A “bZIP-like polypeptide” as defined herein refers to any polypeptide comprising a Basic Leucine Zipper domain (PF00170, SM00338, PS50217) and a G-box binding domain of the MFMR type (PF07777).

The Basic Leucine Zipper Domain (bZIP domain, PF00170) is found in many DNA binding eukaryotic proteins. One part of the domain contains a region that mediates sequence specific DNA binding properties and the leucine zipper that is required for the dimerisation of two DNA binding regions. The DNA binding region comprises a number of basic amino acids such as arginine and lysine.

The G-box binding protein MFMR domain (PF07777) is found at the N-terminus of the PF00170 bZIP domain. It typically ranges in length between 150 and 200 amino acids, but may be shorter, such as in SEQ ID NO: 2. The N-terminal half is rather rich in proline residues and has been termed the PRD (proline rich domain) whereas the C-terminal half is more polar and has been named the MFMR (multifunctional mosaic region). It has been suggested that some of these motifs may be involved in mediating protein-protein interactions.

Additionally or alternatively, the bZIP-like polypeptide comprises one or more of the following motif(s):

Motif 1 (SEQ ID NO: 119): ELKR[EQ][KR]RKQSNRESARRSRLRKQAE[CTA]EEL

Motif 2 (SEQ ID NO: 120): [AQ][RH][KR]VE[SAV]L[TS][HAT]ENx[SAT]L[RKQ][SD]E[LI][QNS][RQ][LF]

Motif 3 (SEQ ID NO: 121): [HPL][AN][PI][HPG][PM][YD][MLV]W

In one embodiment the bZIP-like polypeptide comprises additionally or alternatively one or more of motif(s):

Motif 4 (SEQ ID NO: 122): RELKRQKRKQSNRESARRSRLRKQAECEELQ

Motif 5 (SEQ ID NO: 123): [AG]TNLN[IM]GMD[LV]WN

Motif 6 (SEQ ID NO: 124): MPPYGTPVPYPA[LIM]YPP

In another embodiment, the bZIP-like polypeptide comprises additionally or alternatively one or more of motif(s):

Motif 7 (SEQ ID NO: 125): NE[RL]ELKRE[RK]RKQSNRESARRSRLRKQAE[TA]EELA[RH]

[KR]V[EDQ][SAV]LT[AT]EN[LM][TAS]L[KRQ]

Motif 8 (SEQ ID NO: 126): [IA][ED][TS]P[TA]KSSGNTD[RQ]GL[MLV][NK]KLK[GE]FDGL
[AT]MSIGN

Motif 9 (SEQ ID NO: 127): [PL]PQ[PH]MMPPYG[APT]PY

Motif 10 (SEQ ID NO: 128): NSGAKL[HR]QLLD[AT][SN]PR[AT]DAVAAG

Motif 11 (SEQ ID NO: 129): EI[NS][RKQ][LF]TE[NK]SEK[LM][KR][LM][EQ]N[AS][ATK]
L[MRT][EV][KH]

Preferably, motif 7 extends to WLQNE[RL]ELKRE[RK]RKQSNRESARRSRLRKQAE[AT]EE
LA[RIH][KR]V[EQ][VSA]LT[AST]EN[ML][ATS]L[QKR] (Motif 12, SEQ ID NO: 130).

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

Motifs 1 to 12 were derived using the MEME algorithm (Bailey and Elkan, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994). At each position within a MEME motif, the residues are shown that are present in the query set of sequences with a frequency higher than 0.2. Residues within square brackets represent alternatives.

More preferably, the bZIP-like polypeptide comprises in increasing order of preference, 1, 2, 3 or more of motifs 1 to 12, preferably 4 or more, more preferably 5 or more of motifs 1 to 12, most preferably 6 or more of motifs 1 to 12.

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

In one embodiment the sequence identity level is determined by comparison of the polypeptide sequences over the entire length of the sequence of SEQ ID NO: 2 or SEQ ID NO: 4.

Compared to overall sequence identity, the sequence identity will generally be higher when

only conserved domains or motifs are considered. Preferably the motifs in a bZIP-like polypeptide have, in increasing order of preference, at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one or more of the motifs represented by SEQ ID NO: 119 to SEQ ID NO: 130 (Motifs 1 to 12).

In other words, in another embodiment a method is provided wherein said bZIP-like polypeptide comprises a conserved domain (or motif) with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the conserved domain starting with amino acid E186 up to amino acid E240 in SEQ ID NO: 2 (which corresponds to the conserved domain starting with amino acid E262 up to amino acid N322 in SEQ ID NO: 4). Additionally or alternatively, the bZIP-like polypeptide useful in the methods of the present invention comprises a conserved domain (or motif) with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the conserved domain starting with amino acid M1 up to amino acid N99 in SEQ ID NO: 2 (which corresponds to the conserved domain starting with amino acid M1 up to amino acid R175 in SEQ ID NO: 4). Preferably said bZIP-like polypeptide comprises both conserved domains.

A "BCAT4-like polypeptide" as defined herein refers to any polypeptide comprising the signature sequence represented by AN(KREN)(RKH)W(VIT)PP(PTAQWFRH)GKG (SEQ ID NO: 216).

The term "BCAT4-like" or "BCAT4-like polypeptide" as used herein also intends to include homologues as defined hereunder of "BCAT4-like polypeptide".

Preferably, the BCAT4-like polypeptide comprises one or more of the following motifs:

- (i) Motif 13 represented by ANKRWVPP[PT]GKGSLYIRP (SEQ ID NO: 213),
- (ii) Motif 14 represented by RP[ED]ENA[ML]RM[IQK]xGA[ED]R[ML]CM (SEQ ID NO: 214),
- (iii) Motif 15 represented by LNYGQGLFEGLKAYR[KT]ED (SEQ ID NO: 215).

Motifs 13 to 15 were derived using the MEME algorithm (Bailey and Elkan, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994). At each position within a MEME motif, the residues are shown that are present in the query set of sequences with a frequency higher than 0.2. Residues within square brackets represent alternatives.

More preferably, the BCAT4-like polypeptide comprises in increasing order of preference, at least 2, or all 3 motifs.

Additionally or alternatively, the homologue of a BCAT4-like protein has in increasing order of preference at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% overall sequence identity to the amino acid sequence represented by SEQ ID NO: 142, provided that the homologous protein comprises any one or more of the conserved motifs as outlined above. The overall sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG Wisconsin Package, Accelrys), preferably with default parameters and preferably with sequences of mature proteins (i.e. without taking into account secretion signals or transit peptides). In one embodiment the sequence identity level is determined by comparison of the polypeptide sequences over the entire length of the sequence of SEQ ID NO: 142.

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 BCAT4-like polypeptide have, in increasing order of preference, at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one or more of the motifs represented by SEQ ID NO: 213 to SEQ ID NO: 215 (Motifs 13 to 15).

In other words, in another embodiment a method is provided wherein said BCAT4-like polypeptide comprises a conserved domain (or motif) with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the conserved domain starting with amino acid 185 up to amino acid 202 in SEQ ID NO: 142, or to the conserved domain starting with amino acid 150 up to amino acid 167 in SEQ ID NO: 142, or to the conserved domain starting with amino acid 126 up to amino acid 143 in SEQ ID NO: 142.

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

With respect to bZIP-like polypeptides, the bZIP-like polypeptide sequence, when used in the construction of the phylogenetic tree as provided in Correa et al. (2008), preferably clusters with the group G of bZIP-like polypeptides rather than with any other group of bZIP-like polypeptides.

Furthermore, bZIP-like polypeptides (at least in their native form) typically have DNA binding activity. More particularly, bZIP-like polypeptides typically bind to the G-box sequence (ABRE oligonucleotide in Example 6, SEQ ID NO: 139). Tools and techniques for

measuring DNA binding activity are well known in the art, see for example the DNA binding assay in Liao et al. (Planta, 228, 225-240, 2008). Further details are provided in Example 6.

In addition, bZIP-like polypeptides, when expressed in rice according to the methods of the present invention as outlined in Examples 7 and 8, give plants having increased yield related traits, in particular increased seed yield when grown under conditions of nitrogen limitation or drought stress.

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

With respect to BCAT4-like polypeptides, the BCAT4-like polypeptide is preferably encoded by a nucleic acid molecule comprising a nucleic acid molecule selected from the group consisting of:

- (i) a nucleic acid represented by any one of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211;
- (ii) the complement of a nucleic acid represented by any one of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211;
- (iii) a nucleic acid encoding the polypeptide as represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid being deducible from a polypeptide sequence as represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212, and further preferably conferring enhanced yield-related traits relative to control plants;
- (iv) a nucleic acid having, in increasing order of preference at least 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with any of the nucleic acid sequences of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211, and further preferably conferring enhanced yield-related traits relative to control plants;

- (v) a first nucleic acid molecule which hybridizes with a second nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;
- (vi) a nucleic acid encoding said polypeptide having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212, and preferably conferring enhanced yield-related traits relative to control plants; or
- (vii) a nucleic acid comprising any combination(s) of features of (i) to (vi) above.

Preferably, the polypeptide sequence which when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, clusters with the group of BCAT4-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 142 rather than with any other group.

Furthermore, BCAT4-like polypeptides, at least in their native form, typically have transamination activity.

In addition, BCAT4-like polypeptides, when expressed in rice according to the methods of the present invention as outlined in Examples 7 and 8, give plants having increased yield related traits, in particular increased seed yield, more in particular increase in total weight of the seeds, increase in fillrate, increase in harvestindex and increased number of seeds.

In one embodiment of the present invention the function of the nucleic acid sequences of the invention is to confer information for synthesis of the BCAT4-like polypeptide that increases yield or yield related traits, when such a nucleic acid sequence of the invention is transcribed and translated in a living plant cell.

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

Examples of nucleic acids encoding bZIP-like polypeptides are given in Table A1 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table A1 of the Examples section are

example sequences of orthologues and paralogues of the bZIP-like polypeptide represented by SEQ ID NO: 2, the terms “orthologues” and “paralogues” being as defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search as described in the definitions section; where the query sequence is SEQ ID NO: 1 or SEQ ID NO: 2, the second BLAST (back-BLAST) would be against *Solanum lycopersicum* sequences. Where the query sequence is SEQ ID NO: 3 or SEQ ID NO: 4, the second BLAST (back-BLAST) would be against *Populus trichocarpa* sequences.

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

With respect to BCAT4-like polypeptides, the present invention is illustrated by transforming plants with the nucleic acid sequence represented by SEQ ID NO: 141, encoding the polypeptide sequence of SEQ ID NO: 142. However, performance of the invention is not restricted to these sequences; the methods of the invention may advantageously be performed using any BCAT4-like-encoding nucleic acid or BCAT4-like polypeptide as defined herein.

Examples of nucleic acids encoding BCAT4-like polypeptides are given in Table A2 of the Examples section herein. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table A2 of the Examples section are example sequences of orthologues and paralogues of the BCAT4-like polypeptide represented by SEQ ID NO: 142, the terms “orthologues” and “paralogues” being as defined herein. Further orthologues and paralogues may readily be identified by performing a so-called reciprocal blast search as described in the definitions section; where the query sequence is SEQ ID NO: 141 or SEQ ID NO: 142, the second BLAST (back-BLAST) would be against *Populus trichocarpa* 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 A1 and A2 of the Examples section, the terms “homologue” and “derivative” being as defined herein. Also useful in the methods of the invention are nucleic acids encoding homologues and derivatives of orthologues or paralogues of any one of the amino acid sequences given in Table A1 and A2 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 bZIP-like polypeptides, or BCAT4-like polypeptides, nucleic acids hybridising to nucleic acids encoding bZIP-like polypeptides, or BCAT4-like polypeptides, splice variants of nucleic acids encoding bZIP-like polypeptides, or BCAT4-like polypeptides, allelic variants of nucleic acids encoding bZIP-like polypeptides, or BCAT4-like polypeptides, and variants of nucleic acids encoding bZIP-like polypeptides, or BCAT4-like polypeptides, obtained by gene shuffling. The terms hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

Nucleic acids encoding bZIP-like polypeptides, or BCAT4-like polypeptides, need not be full-length nucleic acids, since performance of the methods of the invention does not rely on the use of full-length nucleic acid sequences. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a portion of any one of the nucleic acid sequences given in Table A1 and A2 of the Examples section, or a portion of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 and A2 of the Examples section.

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

With respect to bZIP-like polypeptides, portions useful in the methods of the invention, encode a bZIP-like polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A1 of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A1 of the Examples section, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A1 of the Examples section. Preferably the portion is at least 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A1 of the Examples section, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A1 of the Examples section. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 1 or SEQ ID NO: 3. Preferably, the portion encodes a fragment of an amino acid sequence which, when used in the construction of the phylogenetic tree as provided in Correa et al. (2008), clusters with the group G of bZIP-like polypeptides rather than with any other group of bZIP-like polypeptides, and/or comprises one or more of the motifs 1 to 12 as described above, and/or has DNA binding activity, and/or has at least 17% sequence identity to SEQ ID NO: 2, or at least 23% sequence identity to SEQ ID NO: 4.

With respect to BCAT4-like polypeptides, portions useful in the methods of the invention, encode a BCAT4-like polypeptide as defined herein, and have substantially the same biological activity as the amino acid sequences given in Table A2 of the Examples section. Preferably, the portion is a portion of any one of the nucleic acids given in Table A2 of the Examples section, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A2 of the Examples section. Preferably the portion is at least 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table A2 of the Examples section, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A2 of the Examples section. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 141. Preferably, the portion encodes a fragment of an amino acid sequence which, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, clusters with the group of BCAT4-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 142 rather than with any other group, and/or comprises the signature peptide as represented by SEQ ID NO: 216, and/or comprises at least one of the motifs 13 to 15 as represented in SEQ ID NO: 213 to 215, respectively, and/or has aminotransferase biological activity, and/or has at least 80% sequence identity to SEQ ID NO: 142.

Another nucleic acid variant useful in the methods of the invention is a nucleic acid capable of hybridising, under reduced stringency conditions, preferably under stringent conditions, with a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide, as defined herein, or with a portion as defined herein. According to the present invention, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a nucleic acid capable of hybridizing to the complement of a nucleic acid encoding any one of the proteins given in Table A1 and A2 of the Examples section, or to the complement of a nucleic acid encoding an orthologue, paralogue or homologue of any one of the proteins given in Table A1 and A2.

Hybridising sequences useful in the methods of the invention encode a POI polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in Table A1 and A2 of the Examples section. Preferably, the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding any one of the proteins given in Table A1 and A2 of the Examples section, or to a portion of any of these sequences, a portion being as defined herein, or the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table A1 and A2 of the Examples section.

With respect to bZIP-like polypeptides, the hybridising sequence is most preferably capable of hybridising to the complement of a nucleic acid as represented by SEQ ID NO: 1 or SEQ

ID NO: 3, or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined above.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence which, when full-length and when used in the construction of the phylogenetic tree as provided in Correa et al. (2008), clusters with the group G of bZIP-like polypeptides rather than with any other group of bZIP-like polypeptides, and/or comprises one or more of the motifs 1 to 12 as described above, and/or has DNA binding activity, and/or has at least 17% sequence identity to SEQ ID NO: 2, or at least 23% sequence identity to SEQ ID NO: 4.

With respect to BCAT4-like polypeptides, the hybridising sequence is most preferably capable of hybridising to the complement of a nucleic acid as represented by SEQ ID NO: 141 or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined above.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence which, when full-length and used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, clusters with the group of BCAT4-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 142 rather than with any other group, and/or comprises the signature peptide as represented by SEQ ID NO: 216, and/or comprises at least one of the motifs 13 to 15 as represented in SEQ ID NO: 213 to 215, respectively, and/or has aminotransferase biological activity, and/or has at least 80% sequence identity to SEQ ID NO: 142.

Another nucleic acid variant useful in the methods of the invention is a splice variant encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide, as defined hereinabove, a splice variant being as defined herein.

In another embodiment, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a splice variant of a nucleic acid encoding any one of the proteins given in Table A1 and A2 of the Examples section, or a splice variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 and A2 of the Examples section.

With respect to bZIP-like polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 1 or SEQ ID NO: 3, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2 or SEQ ID NO: 4. Preferably, the amino acid sequence encoded by the splice variant, when used in the construction of the phylogenetic tree as provided in Correa et al. (2008), clusters with the group G of bZIP-like polypeptides rather than with any other group of bZIP-like polypeptides, and/or comprises one or more of the motifs 1 to 12 as described above, and/or has DNA binding activity, and/or has at least 17% sequence identity to SEQ ID NO: 2, or at least 23%

sequence identity to SEQ ID NO: 4.

With respect to BCAT4-like polypeptides, preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 141, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 142. Preferably, the amino acid sequence encoded by the splice variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, clusters with the group of BCAT4-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 142 rather than with any other group, and/or comprises the signature peptide as represented by SEQ ID NO: 216, and/or comprises at least one of the motifs 13 to 15 as represented in SEQ ID NO: 213 to 215, respectively, and/or has aminotransferase biological activity, and/or has at least 80% sequence identity to SEQ ID NO: 142.

Another nucleic acid variant useful in performing the methods of the invention is an allelic variant of a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide, as defined hereinabove, an allelic variant being as defined herein.

In yet another embodiment, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant an allelic variant of a nucleic acid encoding any one of the proteins given in Table A1 and A2 of the Examples section, or comprising introducing and expressing in a plant an allelic variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 and A2 of the Examples section.

With respect to bZIP-like polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the bZIP-like polypeptide of SEQ ID NO: 2 and any of the amino acids depicted in Table A1 of the Examples section. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 1 or SEQ ID NO: 3, or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2 or SEQ ID NO: 4. Preferably, the amino acid sequence encoded by the allelic variant, when used in the construction of the phylogenetic tree as provided in Correa et al. (2008), clusters with the group G of bZIP-like polypeptides rather than with any other group of bZIP-like polypeptides, and/or comprises one or more of the motifs 1 to 12 as described above, and/or has DNA binding activity, and/or has at least 17% sequence identity to SEQ ID NO: 2, or at least 23% sequence identity to SEQ ID NO: 4.

With respect to BCAT4-like polypeptides, the polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the BCAT4-like polypeptide of SEQ ID NO: 142 and any of the amino acids depicted in Table A2 of the Examples section. Allelic variants exist in nature, and encompassed within

the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 141 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 142. Preferably, the amino acid sequence encoded by the allelic variant, when used in the construction of a phylogenetic tree, such as the one depicted in Figure 7, clusters with the group of BCAT4-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 142 rather than with any other group, and/or comprises the signature peptide as represented by SEQ ID NO: 216, and/or comprises at least one of the motifs 13 to 15 as represented in SEQ ID NO: 213 to 215, respectively, and/or has aminotransferase biological activity, and/or has at least 80% sequence identity to SEQ ID NO: 142.

Gene shuffling or directed evolution may also be used to generate variants of nucleic acids encoding bZIP-like polypeptides, or BCAT4-like polypeptides, as defined herein; the term "gene shuffling" being as defined herein.

In yet another embodiment, there is provided a method for enhancing yield-related traits in plants, comprising introducing and expressing in a plant a variant of a nucleic acid encoding any one of the proteins given in Table A1 and A2 of the Examples section, or comprising introducing and expressing in a plant a variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table A1 and A2 of the Examples section, which variant nucleic acid is obtained by gene shuffling.

With respect to bZIP-like polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling, when used in the construction of the phylogenetic tree as provided in Correa et al. (2008), preferably clusters with the group G of bZIP-like polypeptides rather than with any other group of bZIP-like polypeptides, and/or comprises one or more of the motifs 1 to 12 as described above, and/or has DNA binding activity, and/or has at least 17% sequence identity to SEQ ID NO: 2, or at least 23% sequence identity to SEQ ID NO: 4.

With respect to BCAT4-like polypeptides, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling, when used in the construction of a phylogenetic tree such as the one depicted in Figure 7, preferably clusters with the group of BCAT4-like polypeptides comprising the amino acid sequence represented by SEQ ID NO: 142 rather than with any other group, and/or comprises the signature peptide as represented by SEQ ID NO: 216, and/or comprises at least one of the motifs 13 to 15 as represented in SEQ ID NO: 213 to 215, respectively, and/or has aminotransferase biological activity, and/or has at least 80% sequence identity to SEQ ID NO: 142.

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.).

bZIP-like polypeptides differing from the sequence of SEQ ID NO: 2 or of SEQ ID NO: 4 by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined above) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention.

BCAT4-like polypeptides differing from the sequence of SEQ ID NO: 142 by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined above) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention.

Concerning bZIP-like polypeptides, nucleic acids encoding bZIP-like polypeptides may be derived from any natural or artificial source. The nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. Preferably the bZIP-like polypeptide-encoding nucleic acid is from a plant, further preferably from a dicotyledonous plant. In one embodiment, the bZIP-like polypeptide-encoding nucleic acid is from the family Solanaceae, preferably from *Solanum lycopersicum*. In another embodiment, the bZIP-like polypeptide-encoding nucleic acid is from the family Salicaceae, preferably from *Populus trichocarpa*.

Nucleic acids encoding BCAT4-like polypeptides may be derived from any natural or artificial source. The nucleic acid may be modified from its native form in composition and/or genomic environment through deliberate human manipulation. Preferably the BCAT4-like polypeptide-encoding nucleic acid is from a plant, further preferably from a dicotyledonous plant, more preferably from the family Salicaceae, most preferably the nucleic acid is from *Populus trichocarpa*.

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 yield, especially increased seed yield and/or increased biomass relative to control plants. It should be noted that the increased biomass was obtained without effect on flowering time, in particular, no delay in flowering time was observed. The terms "yield" and "seed yield" are described in more detail in the "definitions" section herein.

Reference herein to enhanced yield-related traits is taken to mean an increase early vigour and/or in biomass (weight) of one or more parts of a plant, which may include (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 biomass and/or seeds, and performance of the methods of the invention results in plants having increased biomass and/or increased seed yield relative to the seed yield of control plants. In one preferred embodiment, the increased yield is increased biomass, in another preferred embodiment, the increased yield is increased seed yield, in yet another embodiment, the increased yield is increased biomass and increased seed yield.

The present invention provides a method for increasing yield-related traits, in particular yield, especially seed yield and/or biomass of plants, relative to control plants, which method comprises modulating expression in a plant of a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like 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 bZIP-like polypeptide, or a BCAT4-like polypeptide, as defined herein.

Performance of the methods of the invention gives plants grown under non-stress conditions or under mild drought conditions increased yield relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield in plants grown under non-stress conditions or under mild drought conditions, which method comprises modulating expression in a plant of a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide.

Performance of the methods of the invention gives plants grown under conditions of drought, increased yield relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield in plants grown under conditions of drought which method comprises modulating expression in a plant of a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide.

Performance of the methods of the invention gives plants grown under conditions of nutrient deficiency, particularly under conditions of nitrogen deficiency, increased yield relative to control plants grown under comparable conditions. Therefore, according to the present invention, there is provided a method for increasing yield in plants grown under conditions of nutrient deficiency, which method comprises modulating expression in a plant of a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like 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 bZIP-like polypeptide, or a BCAT4-like polypeptide.

The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding bZIP-like polypeptides, or BCAT4-like 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 bZIP-like polypeptide, or a BCAT4-like polypeptide, as defined above;
- (b) one or more control sequences capable of driving expression of the nucleic acid sequence of (a); and optionally
- (c) a transcription termination sequence.

Preferably, the nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like 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 POI 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 POI 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.

The constitutive promoter is preferably a ubiquitous constitutive promoter of medium strength. More preferably it is a plant derived promoter, e.g. a promoter of plant chromosomal origin, such as a GOS2 promoter or a promoter of substantially the same strength and having substantially the same expression pattern (a functionally equivalent promoter), more preferably the promoter is the promoter GOS2 promoter from rice. Further preferably the constitutive promoter is represented by a nucleic acid sequence substantially similar to SEQ ID NO: 131 or SEQ ID NO: 218, most preferably the constitutive promoter is as represented by SEQ ID NO: 131 or SEQ ID NO: 218. See the “Definitions” section herein for further examples of constitutive promoters.

With respect to bZIP-like polypeptides, it should be clear that the applicability of the present invention is not restricted to the bZIP-like polypeptide-encoding nucleic acid represented by SEQ ID NO: 1 or SEQ ID NO: 3, nor is the applicability of the invention restricted to the rice GOS2 promoter when expression of a bZIP-like polypeptide-encoding nucleic acid is driven by a constitutive promoter.

With respect to BCAT4-like polypeptides, it should be clear that the applicability of the present invention is not restricted to the BCAT4-like polypeptide-encoding nucleic acid represented by SEQ ID NO: 141, nor is the applicability of the invention restricted to expression of a BCAT4-like polypeptide-encoding nucleic acid when driven by a constitutive promoter.

With respect to bZIP-like polypeptides, optionally, one or more terminator sequences may be used in the construct introduced into a plant. Preferably, the construct comprises an expression cassette comprising a GOS2 promoter, substantially similar to SEQ ID NO: 131, operably linked to the nucleic acid encoding the bZIP-like polypeptide. More preferably, the construct comprises a zein terminator (t-zein) linked to the 3' end of the *bZIP-like* coding sequence. Most preferably, the expression cassette comprises a sequence having in increasing order of preference at least 95%, at least 96%, at least 97%, at least 98%, at least 99% identity to the sequence represented by SEQ ID NO: 132 (Le expression cassette) or SEQ ID NO: 133 (Pt expression cassette). Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

With respect to BCAT4-like polypeptides, optionally, one or more terminator sequences may be used in the construct introduced into a plant. Preferably, the construct comprises an

expression cassette comprising a GOS2 promoter, substantially similar to SEQ ID NO: 218, operably linked to the nucleic acid encoding the BCAT4-like polypeptide. More preferably, the construct comprises a zein terminator (t-zein) linked to the 3' end of the BCAT4-like polypeptide coding sequence. Most preferably, the expression cassette comprises a sequence having in increasing order of preference at least 95%, at least 96%, at least 97%, at least 98%, at least 99% identity to the sequence represented by SEQ ID NO: 217 (pGOS2::BCAT4-like::t-zein sequence). Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

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

As mentioned above, a preferred method for modulating expression of a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide, is by introducing and expressing in a plant a nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide; however the effects of performing the method, i.e. enhancing yield-related traits may also be achieved using other well known techniques, including but not limited to T-DNA activation tagging, TILLING, homologous recombination. A description of these techniques is provided in the definitions section.

The invention also provides a method for the production of transgenic plants having enhanced yield-related traits relative to control plants, comprising introduction and expression in a plant of any nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide, as defined herein.

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

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

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

Preferably, the nucleic acid encoding bZIP-like polypeptide as defined herein and to be introduced into the plant is an isolated nucleic acid or is comprised in a genetic construct.

With respect to BCAT4-like polypeptides, the present invention more specifically provides a method for the production of transgenic plants having enhanced yield-related traits, particularly increased yield, more in particular increased seed yield, which method comprises:

- (i) introducing and expressing in a plant or plant cell a BCAT4-like polypeptide-encoding nucleic acid or a genetic construct comprising a BCAT4-like polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development.

The nucleic acid of (i) may be any of the nucleic acids capable of encoding a BCAT4-like polypeptide as defined herein.

Preferably, the nucleic acid encoding BCAT4-like polypeptide as defined herein and to be introduced into the plant is an isolated nucleic acid or is comprised in a genetic construct.

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 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 bZIP-like polypeptide, or a BCAT4-like 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 the bZIP-like polypeptide, or the BCAT4-like polypeptide, and/or the bZIP-like polypeptides, or BCAT4-like polypeptides, as described above.

The invention also includes host cells containing an isolated nucleic acid encoding a bZIP-like polypeptide, or a BCAT4-like 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 harvestable parts of a plant such as, but not limited to seeds, leaves, fruits, flowers, stems, roots, rhizomes, tubers and bulbs, which harvestable parts comprise a recombinant nucleic acid encoding a POI 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 bZIP-like polypeptides, or BCAT4-like polypeptides, as described herein and use of these bZIP-like polypeptides, or BCAT4-like polypeptides, in enhancing any of the aforementioned yield-related traits in plants. For example, nucleic acids encoding POI polypeptide described herein, or the bZIP-like polypeptides, or the BCAT4-like polypeptides, themselves, may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to a gene encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide. The nucleic acids/genes, or the bZIP-like polypeptides, or the BCAT4-like polypeptides, themselves may be used to define a molecular marker. This DNA or protein marker may then be used in breeding programmes to select plants having enhanced yield-related traits as defined herein in the methods of the invention. Furthermore, allelic variants of a nucleic acid/gene encoding a bZIP-like polypeptide, or a BCAT4-like polypeptide, may find use in marker-assisted breeding programmes. Nucleic acids encoding bZIP-like polypeptides, or BCAT4-like polypeptides, may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired

phenotypes.

Moreover, concerning the bZIP-like polypeptides, 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 bZIP-like polypeptide, wherein said bZIP-like polypeptide comprises Basic Leucine Zipper Domain (PF00170) and a G-box binding domain of the MFMR type (PF07777) and one or more of motifs 1 to 3, represented by SEQ ID NO: 119, SEQ ID NO: 120 or SEQ ID NO: 121.
2. Method according to embodiment 1, wherein said bZIP-like polypeptide comprises one or more of motifs 4 to 6.
3. Method according to embodiment 1, wherein said bZIP-like polypeptide comprises one or more of motifs 7 to 12.
4. Method according to any one of embodiments 1 to 3, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said bZIP-like polypeptide.
5. Method according to embodiment 1 to 4, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.
6. Method according to embodiment 1 to 5, wherein said enhanced yield-related traits are obtained without effect on flowering time of the plant.
7. Method according to any one of embodiments 1 to 3, wherein said enhanced yield-related traits are obtained under non-stress conditions.
8. Method according to any one of embodiments 1 to 3, wherein said enhanced yield-related traits are obtained under conditions of drought stress or nitrogen deficiency.
9. Method according to any one of embodiments 1 to 8, wherein said nucleic acid encoding a bZIP-like is of plant origin, preferably from a dicotyledonous plant.
10. Method according to any one of embodiments 1 to 8, wherein said nucleic acid encoding a bZIP-like encodes any one of the polypeptides listed in Table A1 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
11. Method according to any one of embodiments 1 to 8, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A1.
12. Method according to any one of embodiments 1 to 11, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 2 or SEQ ID NO: 4.
13. Method according to any one of embodiments 1 to 12, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
14. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method ac-

according to any one of embodiments 1 to 13, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding a bZIP-like polypeptide as defined in any of embodiments 1 to 3 and 9 to 12.

15. Construct comprising:
 - (i) nucleic acid encoding a bZIP-like as defined in any of embodiments 1 to 3 and 9 to 12;
 - (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - (iii) a transcription termination sequence.
16. Construct according to embodiment 15, wherein one of said control sequences is a constitutive promoter, preferably a medium strength constitutive promoter, preferably a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
17. Use of a construct according to embodiment 15 or 16 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
18. Plant, plant part or plant cell transformed with a construct according to embodiment 15 or 16.
19. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants, comprising:
 - (i) introducing and expressing in a plant cell or plant a nucleic acid encoding a bZIP-like polypeptide as defined in any of embodiments 1 to 3 and 9 to 12; and
 - (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.
20. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding a bZIP-like polypeptide as defined in any of embodiments 1 to 3 and 9 to 12 or a transgenic plant cell derived from said transgenic plant.
21. Transgenic plant according to embodiment 14, 18 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 embodiment 21, wherein said harvestable parts are preferably shoot biomass and/or seeds.
23. Products derived from a plant according to embodiment 21 and/or from harvestable parts of a plant according to embodiment 22.
24. Use of a nucleic acid encoding a bZIP-like polypeptide as defined in any of

embodiments 1 to 3 and 9 to 12, for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.

25. A method for manufacturing a product, comprising the steps of growing the plants according to embodiment 14, 18, 20 or 21 and producing said product from or by said plants; or parts thereof, including seeds.

Moreover, concerning the BCAT4-like polypeptides, 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 an BCAT4-like polypeptide, wherein said BCAT4-like polypeptide comprises the signature sequence represented by SEQ ID NO: 216.
2. Method according to embodiment 1, wherein said polypeptide is encoded by a nucleic acid molecule comprising a nucleic acid molecule selected from the group consisting of:
 - (i) a nucleic acid represented by any one of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211;
 - (ii) the complement of a nucleic acid represented by any one of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211;
 - (iii) a nucleic acid encoding the polypeptide as represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid being deducible from a polypeptide sequence as represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212 and further preferably conferring enhanced yield-related traits relative to control plants;
 - (iv) a nucleic acid having, in increasing order of preference at least 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with any of the nucleic acid sequences of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211, and further preferably

- conferring enhanced yield-related traits relative to control plants;
- (v) a first nucleic acid molecule which hybridizes with a second nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;
 - (vi) a nucleic acid encoding said polypeptide having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212 and preferably conferring enhanced yield-related traits relative to control plants; or
 - (vii) a nucleic acid comprising any combination(s) of features of (i) to (vi) above.
3. Method according to embodiment 1 or 2, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said BCAT4-like polypeptide.
 4. Method according to any one of embodiment 1 to 3, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.
 5. Method according to any one of embodiments 1 to 4, wherein said enhanced yield-related traits are obtained under conditions of drought stress.
 6. Method according to any of embodiments 1 to 5, wherein said BCAT4-like polypeptide comprises one or more of the following motifs:
 - (i) Motif 13 represented by SEQ ID NO: 213,
 - (ii) Motif 14 represented by SEQ ID NO: 214,
 - (iii) Motif 15 represented by SEQ ID NO: 215.
 7. Method according to any one of embodiments 1 to 6, wherein said nucleic acid encoding an BCAT4-like is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Salicaceae, more preferably from the genus *Populus*, most preferably from *Populus trichocarpa*.
 8. Method according to any one of embodiments 1 to 7, wherein said nucleic acid encoding a BCAT4-like polypeptide encodes any one of the polypeptides listed in Table A2 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
 9. Method according to any one of embodiments 1 to 8, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A2.
 10. Method according to any one of embodiments 1 to 9, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 142.
 11. Method according to any one of embodiments 1 to 10, wherein said nucleic acid is

- operably linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
12. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method according to any one of embodiments 1 to 11, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an BCAT4-like polypeptide as defined in any of embodiments 1, 2 and 6 to 11.
 13. Construct comprising:
 - (i) nucleic acid encoding an BCAT4-like as defined in any of embodiments 1, 2 and 6 to 11;
 - (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - (iii) a transcription termination sequence.
 14. Construct according to embodiment 13, wherein one of said control sequences is a constitutive promoter, preferably a medium strength constitutive promoter, preferably a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
 15. Use of a construct according to embodiment 13 or 14 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
 16. Plant, plant part or plant cell transformed with a construct according to embodiment 13 or 14.
 17. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants, comprising:
 - (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an BCAT4-like polypeptide as defined in any of embodiments 1, 2 and 6 to 11; 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 relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding an BCAT4-like polypeptide as defined in any of embodiments 1, 2 and 6 to 11 or a transgenic plant cell derived from said transgenic plant.
 19. Transgenic plant according to embodiment 12, 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 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 BCAT4-like polypeptide as defined in any of embodiments 1, 2 and 6 to 11 for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.
 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. Construct according to embodiment 13 or 14 comprised in a plant cell.
 25. Recombinant chromosomal DNA comprising the construct according to embodiment 13 or 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. As used herein, the term “comprising” also encompasses the term “consisting of”.

Peptide(s)/Protein(s)

The terms “peptides”, “oligopeptides”, “polypeptide” and “protein” are used interchangeably 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, <i>Plant Cell</i> , 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 (Daucus carota)	Downey et al. (2000, J. Biol. Chem. 275:39420)
TobRB7 gene	W Song (1997) PhD Thesis, North Carolina State University, Raleigh, NC USA
OsRAB5a (rice)	Wang et al. 2002, Plant Sci. 163:273
ALF5 (Arabidopsis)	Diener et al. (2001, Plant Cell 13:1625)
NRT2;1Np (N. plumbaginifolia)	Quesada et al. (1997, Plant Mol. Biol. 34:265)

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

Table 2c: Examples of seed-specific promoters

Gene source	Reference
seed-specific genes	Simon et al., Plant Mol. Biol. 5: 191, 1985;

	Scofield et al., J. Biol. Chem. 262: 12202, 1987.;
	Baszczynski et al., Plant Mol. Biol. 14: 633, 1990.
Brazil Nut albumin	Pearson et al., Plant Mol. Biol. 18: 235-245, 1992.
legumin	Ellis et al., Plant Mol. Biol. 10: 203-214, 1988.
glutelin (rice)	Takaiwa et al., Mol. Gen. Genet. 208: 15-22, 1986;
	Takaiwa et al., FEBS Letts. 221: 43-47, 1987.
zein	Matzke et al Plant Mol Biol, 14(3):323-32 1990
napA	Stalberg et al, Planta 199: 515-519, 1996.
wheat LMW and HMW glutenin-1	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
wheat SPA	Albani et al, Plant Cell, 9: 171-184, 1997
wheat α , β , γ -gliadins	EMBO J. 3:1409-15, 1984
barley Itr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	Mena et al, The Plant Journal, 116(1): 53-62, 1998
blz2	EP99106056.7
synthetic promoter	Vicente-Carbajosa et al., Plant J. 13: 629-640, 1998.
rice prolamin NRP33	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice a-globulin Glb-1	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
rice α -globulin REB/OHP-1	Nakase et al. Plant Mol. Biol. 33: 513-522, 1997
rice ADP-glucose pyrophos- phorylase	Trans Res 6:157-68, 1997
maize ESR gene family	Plant J 12:235-46, 1997
sorghum α -kafirin	DeRose et al., Plant Mol. Biol 32:1029-35, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39:257-71, 1999
rice oleosin	Wu et al, J. Biochem. 123:386, 1998
sunflower oleosin	Cummins et al., Plant Mol. Biol. 19: 873-876, 1992
PRO0117, putative rice 40S ribosomal protein	WO 2004/070039
PRO0136, rice alanine aminotransferase	unpublished
PRO0147, trypsin inhibitor ITR1 (barley)	unpublished
PRO0151, rice WSI18	WO 2004/070039
PRO0175, rice RAB21	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039
α -amylase (Amy32b)	Lanahan et al, Plant Cell 4:203-211, 1992; Skriver et al,

	Proc Natl Acad Sci USA 88:7266-7270, 1991
cathepsin β -like gene	Cejudo et al, Plant Mol Biol 20:849-856, 1992
Barley Ltp2	Kalla et al., Plant J. 6:849-60, 1994
Chi26	Leah et al., Plant J. 4:579-89, 1994
Maize B-Peru	Selinger et al., Genetics 149;1125-38,1998

Table 2d: examples of endosperm-specific promoters

Gene source	Reference
glutelin (rice)	Takaiwa et al. (1986) Mol Gen Genet 208:15-22; Takaiwa et al. (1987) FEBS Letts. 221:43-47
zein	Matzke et al., (1990) Plant Mol Biol 14(3): 323-32
wheat LMW and HMW glutenin-1	Colot et al. (1989) Mol Gen Genet 216:81-90, Anderson et al. (1989) NAR 17:461-2
wheat SPA	Albani et al. (1997) Plant Cell 9:171-184
wheat gliadins	Rafalski et al. (1984) EMBO 3:1409-15
barley 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. World Scientific, Singapore, pp. 274-289]. Alternative methods are based on the repeated removal of the inflorescences and incubation of the excision site in the center of the rosette with transformed agrobacteria, whereby transformed seeds can likewise be obtained at a later point in time (Chang (1994). Plant J. 5: 551-558; Katavic (1994). Mol Gen Genet, 245: 363-370). However, an especially effective method is the vacuum infiltration method with its modifications such as the "floral dip" method. In the case of vacuum infiltration of *Arabidopsis*, intact plants under reduced pressure are treated with an agrobacterial suspension [Bechthold, N (1993). C R Acad Sci Paris Life Sci, 316: 1194-1199], while in the case of the "floral dip" method the developing floral tissue is incubated briefly with a surfactant-treated agrobacterial suspension [Clough, SJ and Bent AF (1998) The Plant J. 16, 735-743]. A certain proportion of transgenic seeds are harvested in both cases, and these seeds can be distinguished from non-transgenic seeds by growing under the above-described selective conditions. In addition the stable transformation of plastids is of advantages because plastids are inherited maternally is most crops reducing or eliminating the risk of transgene flow through pollen. The transformation of the chloroplast genome is generally achieved by a process which has been schematically displayed in Klaus et al., 2004 [Nature Biotechnology 22 (2), 225-229]. Briefly the sequences to be transformed are cloned together with a selectable marker gene between flanking sequences homologous to the chloroplast genome. These homologous flanking sequences direct site specific integration into the plastome. Plastidal transformation has been described for many different plant species and an overview is given in Bock (2001) Transgenic plastids in basic research and plant biotechnology. J Mol Biol. 2001 Sep 21; 312 (3):425-38 or Maliga, P (2003) Progress towards commercialization of plastid transformation technology. Trends Biotechnol. 21, 20-28. Further biotechnological progress has recently been reported in form of marker free plastid transformants, which can be produced by a transient co-integrated 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 excelsea*, *Beta vulgaris*, *Brassica* spp. (e.g. *Brassica napus*, *Brassica rapa* ssp. [canola, oilseed rape, turnip rape]), *Cadaba farinosa*, *Camellia sinensis*, *Canna indica*, *Cannabis sativa*, *Capsicum* spp., *Carex elata*, *Carica papaya*, *Carissa macrocarpa*, *Carya* spp., *Carthamus tinctorius*, *Castanea* spp., *Ceiba pentandra*, *Cichorium endivia*, *Cinnamomum* spp., *Citrullus lanatus*, *Citrus* spp., *Cocos* spp., *Coffea* spp., *Colocasia esculenta*, *Cola* spp., *Corchorus* sp., *Coriandrum sativum*, *Corylus* spp., *Crataegus* spp., *Crocus sativus*, *Cucurbita* spp., *Cucumis* spp., *Cynara* spp., *Daucus carota*, *Desmodium* spp., *Dimocarpus longan*, *Dioscorea* spp., *Diospyros* spp., *Echinochloa* spp., *Elaeis* (e.g. *Elaeis guineensis*, *Elaeis oleifera*), *Eleusine coracana*, *Eragrostis tef*, *Erianthus* sp., *Eriobotrya japonica*, *Eucalyptus* sp., *Eugenia uniflora*, *Fagopyrum* spp., *Fagus* spp.,

Festuca arundinacea, *Ficus carica*, *Fortunella* spp., *Fragaria* spp., *Ginkgo biloba*, *Glycine* spp. (e.g. *Glycine max*, *Soja hispida* or *Soja max*), *Gossypium hirsutum*, *Helianthus* spp. (e.g. *Helianthus annuus*), *Hemerocallis fulva*, *Hibiscus* spp., *Hordeum* spp. (e.g. *Hordeum vulgare*), *Ipomoea batatas*, *Juglans* spp., *Lactuca sativa*, *Lathyrus* spp., *Lens culinaris*, *Linum usitatissimum*, *Litchi chinensis*, *Lotus* spp., *Luffa acutangula*, *Lupinus* spp., *Luzula sylvatica*, *Lycopersicon* spp. (e.g. *Lycopersicon esculentum*, *Lycopersicon lycopersicum*, *Lycopersicon pyriforme*), *Macrotyloma* spp., *Malus* spp., *Malpighia emarginata*, *Mammea americana*, *Mangifera indica*, *Manihot* spp., *Manilkara zapota*, *Medicago sativa*, *Melilotus* spp., *Mentha* spp., *Miscanthus sinensis*, *Momordica* spp., *Morus nigra*, *Musa* spp., *Nicotiana* spp., *Olea* spp., *Opuntia* spp., *Ornithopus* spp., *Oryza* spp. (e.g. *Oryza sativa*, *Oryza latifolia*), *Panicum miliaceum*, *Panicum virgatum*, *Passiflora edulis*, *Pastinaca sativa*, *Pennisetum* sp., *Persea* spp., *Petroselinum crispum*, *Phalaris arundinacea*, *Phaseolus* spp., *Phleum pratense*, *Phoenix* spp., *Phragmites australis*, *Physalis* spp., *Pinus* spp., *Pistacia vera*, *Pisum* spp., *Poa* spp., *Populus* spp., *Prosopis* spp., *Prunus* spp., *Psidium* spp., *Punica granatum*, *Pyrus communis*, *Quercus* spp., *Raphanus sativus*, *Rheum rhabarbarum*, *Ribes* spp., *Ricinus communis*, *Rubus* spp., *Saccharum* spp., *Salix* sp., *Sambucus* spp., *Secale cereale*, *Sesamum* spp., *Sinapis* sp., *Solanum* spp. (e.g. *Solanum tuberosum*, *Solanum integrifolium* or *Solanum lycopersicum*), *Sorghum bicolor*, *Spinacia* spp., *Syzygium* spp., *Tagetes* spp., *Tamarindus indica*, *Theobroma cacao*, *Trifolium* spp., *Tripsacum dactyloides*, *Triticosecale rimpai*, *Triticum* spp. (e.g. *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum*, *Triticum monococcum* or *Triticum vulgare*), *Tropaeolum minus*, *Tropaeolum majus*, *Vaccinium* spp., *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.

Description of figures

The present invention will now be described with reference to the following figures in which: **Figure 1** represents the domain structure of SEQ ID NO: 2 and SEQ ID NO: 4 with the conserved motifs 1 to 12 indicated.

Figure 2 represents a multiple alignment of various bZIP-like polypeptides. The consensus sequence line at the bottom of the alignment gives an indication of the conserved regions

within the group of bZIP-like proteins. These alignments can be used for defining further motifs or signature sequences, when using conserved amino acids. Panel A shows an alignment of sequences comprising SEQ ID NO: 2, panel B shows an alignment of sequences comprising SEQ ID NO: 4.

Figure 3 shows the MATGAT tables of Example 3. Panel A shows a table for sequences comprising SEQ ID NO: 2, panel B shows a table for sequences comprising SEQ ID NO: 4.

Figure 4 represents the binary vector used for increased expression in *Oryza sativa* of a bZIP-like-encoding nucleic acid (such as SEQ ID NO: 1 or SEQ ID NO: 3) under the control of a rice GOS2 promoter (pGOS2).

Figure 5 represents the domain structure of SEQ ID NO: 142 with conserved motifs.

Figure 6 represents a multiple alignment of various BCAT4-like polypeptides. The asterisks indicate identical amino acids among the various protein sequences, colons represent highly conserved amino acid substitutions, and the dots represent less conserved amino acid substitution; on other positions there is no sequence conservation. These alignments can be used for defining further motifs or signature sequences, when using conserved amino acids. The corresponding SEQ ID NOs for the aligned polypeptide sequences shown in Figure 6 are:

- SEQ ID NO: 160 for G.max_Glyma06g05280
- SEQ ID NO: 168 for H.annuus_TC54245
- SEQ ID NO: 146 for A.thaliana_AT1G10070
- SEQ ID NO: 198 for P.trichocarpa_scaff_II.1054
- SEQ ID NO: 144 for A.thaliana_AT1G10060
- SEQ ID NO: 154 for A.thaliana_AT3G49680
- SEQ ID NO: 156 for A.thaliana_AT5G65780
- SEQ ID NO: 142 for P.trichocarpa_BCAT4-like
- SEQ ID NO: 158 for G.max_Glyma01g40420
- SEQ ID NO: 166 for G.max_Glyma11g04870
- SEQ ID NO: 182 for M.truncatula_TC114768
- SEQ ID NO: 202 for S.lycopersicum_TC213629
- SEQ ID NO: 170 for H.vulgare_TC165564
- SEQ ID NO: 188 for O.sativa_LOC_Os05g48450
- SEQ ID NO: 208 for Z.mays_TA12434_4577999
- SEQ ID NO: 174 for Hordeum_vulgare_PUT-169a-Horde
- SEQ ID NO: 176 for Hordeum_vulgare_subsp_vulgare_
- SEQ ID NO: 190 for O.sativa_Os03g0106400
- SEQ ID NO: 186 for O.sativa_LOC_Os04g47190
- SEQ ID NO: 204 for T.aestivum_TC320973
- SEQ ID NO: 210 for Zea_mays_GRMZM2G047347_T03
- SEQ ID NO: 192 for P.patens_TC31354
- SEQ ID NO: 164 for P.patens_TC33668
- SEQ ID NO: 172 for H.vulgare_TC186077
- SEQ ID NO: 206 for T.aestivum_TC325793

SEQ ID NO: 184 for O.sativa_LOC_Os03g12890
SEQ ID NO: 212 for Zea_mays_GRMZM2G153536_T03
SEQ ID NO: 148 for A.thaliana_AT1G50090
SEQ ID NO: 150 for A.thaliana_AT1G50110
SEQ ID NO: 152 for A.thaliana_AT3G19710
SEQ ID NO: 162 for G.max_Glyma07g30510
SEQ ID NO: 164 for G.max_Glyma08g06750
SEQ ID NO: 178 for M.truncatula_AC159872_36
SEQ ID NO: 196 for P.trichocarpa_804339
SEQ ID NO: 200 for P.trichocarpa_scaff_IX.827
SEQ ID NO: 180 for M.truncatula_AC159872_55

Figure 7 shows phylogenetic tree of BCAT4-like polypeptides, as described in example 2.

Figure 8 shows the MATGAT table of Example 3.

Figure 9 represents the binary vector used for increased expression in *Oryza sativa* of a BCAT4-like-encoding nucleic acid under the control of a rice GOS2 promoter (pGOS2).

Examples

The present invention will now be described with reference to the following examples, which are by way of illustration 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: Identification of sequences related to the nucleic acid sequence used in the methods of intervention

1. bZIP-like polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 1 or 3 and SEQ ID NO: 2 or 4 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402). The program is used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. For example, the polypeptide encoded by the nucleic acid of SEQ ID NO: 1 or SEQ ID NO: 3 was used for the TBLASTN algorithm, with default settings and

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

Table A1 provides a list of nucleic acid sequences and amino acid sequences related to SEQ ID NO: 1/2 and SEQ ID NO: 3/4 respectively.

Table A1: Examples of bZIP-like nucleic acids and polypeptides:

Plant source	Nuclei acid SEQ ID NO:	Protein SEQ ID NO:
<i>S. lycopersicum</i> bZIP-like	1	2
<i>P. tricarpa</i> bZIP-like	3	4
<i>A.lyrata</i> _322281	5	6
<i>A.lyrata</i> _895903	7	8
<i>A.lyrata</i> _944204	9	10
<i>A.thaliana</i> _AT2G46270.2	11	12
<i>A.thaliana</i> _AT4G01120.1	13	14
<i>A.thaliana</i> _AT4G36730.2	15	16
<i>Aquilegia</i> _sp_TC21139	17	18
<i>B.napus</i> _TC63500	19	20
<i>B.napus</i> _TC63534	21	22
<i>B.napus</i> _TC63535	23	24
<i>B.napus</i> _TC63581	25	26
<i>B.napus</i> _TC89867	27	28
<i>C.canephora</i> _TC1061	29	30
<i>C.canephora</i> _TC5243	31	32
<i>C.clementina</i> _TC35556	33	34
<i>C.endivia</i> _EL359780	35	36
<i>C.maculosa</i> _EH744483	37	38
<i>C.maculosa</i> _TA2192_215693	39	40
<i>C.roseus</i> _AF084971	41	42

C.roseus_AY027510	43	44
C.sinensis_TC24527	45	46
C.sinensis_TC9189	47	48
C.tinctorius_TA1437_4222	49	50
E.esula_TC3985	51	52
F.vesca_TA11395_57918	53	54
G.hirsutum_TC137570	55	56
G.hirsutum_TC148476	57	58
G.max_Glyma03g41590.4	59	60
G.max_Glyma07g06620.1	61	62
G.max_Glyma11g06960.1	63	64
G.max_Glyma16g03190.1	65	66
G.max_Glyma16g25600.2	67	68
G.max_Glyma19g44190.1	69	70
H.annuus_BU027457	71	72
H.annuus_TC42219	73	74
H.petiolaris_TA3844_4234	75	76
H.tuberosus_TA2407_4233	77	78
H.tuberosus_TA2966_4233	79	80
M.truncatula_AC137602_8.5	81	82
M.truncatula_AC148484_8.5	83	84
N.tabacum_BP133908	85	86
N.tabacum_NP917548	87	88
N.tabacum_TC40642	89	90
N.tabacum_TC76189	91	92
P.taeda_TA23200_3352	93	94
P.trichocarpa_719452	95	96
P.trifoliata_TA6299_37690	97	98
P.vulgaris_TC11785	99	100
S.lycopersicum_TC211600	101	102
S.lycopersicum_TC213303	103	104
S.tuberosum_TC185019	105	106
S.tuberosum_TC186959	107	108
S.tuberosum_TC187892	109	110
T.pratense_TA1890_57577	111	112

Medicago truncatula GBF3-like	113	114
V.vinifera_GSVIVT00014657001	115	116
V.vinifera_GSVIVT00024984001	117	118

2. BCAT4-like polypeptides

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 141 and SEQ ID NO: 142 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215:403-410; and Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The program is used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. For example, the polypeptide encoded by the nucleic acid of SEQ ID NO: 141 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflect the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search. For example the E-value may be increased to show less stringent matches. This way, short nearly exact matches may be identified.

Table A2 provides a list of nucleic acid sequences related to SEQ ID NO: 141 and SEQ ID NO: 142.

Table A2: Examples of BCAT4-like nucleic acids and polypeptides:

Plant Source	Nucleic acid SEQ ID NO:	Protein SEQ ID NO:
A.thaliana_AT1G10060.1	143	144
A.thaliana_AT1G10070.1	145	146
A.thaliana_AT1G50090.1	147	148
A.thaliana_AT1G50110.1	149	150
A.thaliana_AT3G19710.1	151	152
A.thaliana_AT3G49680.1	153	154
A.thaliana_AT5G65780.1	155	156
G.max_Glyma01g40420.1	157	158
G.max_Glyma06g05280.1	159	160

G.max_Glyma07g30510.1	161	162
G.max_Glyma08g06750.1	163	164
G.max_Glyma11g04870.1	165	166
H.annuus_TC54245	167	168
H.vulgare_TC165564	169	170
H.vulgare_TC186077	171	172
Hordeum_vulgare_PUT-169a- Hordeum_vulgare-79158	173	174
Hordeum_vulgare_subsp_ vulgare_AK251931	175	176
M.truncatula_AC159872_36	177	178
M.truncatula_AC159872_55	179	180
M.truncatula_TC114768	181	182
O.sativa_LOC_Os03g12890	183	184
O.sativa_LOC_Os04g47190	185	186
O.sativa_LOC_Os05g48450	187	188
O.sativa_Os03g0106400	189	190
P.patens_TC31354	191	192
P.patens_TC33668	193	194
P.trichocarpa_804339	195	196
P.trichocarpa_scaff_II.1054	197	198
P.trichocarpa_scaff_IX.827	199	200
S.lycopersicum_TC213629	201	202
T.aestivum_TC320973	203	204
T.aestivum_TC325793	205	206
Z.mays_TA12434_4577999	207	208
Zea_mays_GRMZM2G047347_T03	209	210
Zea_mays_GRMZM2G153536_T03	211	212

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 sequences to the polypeptide sequences used in the methods of the invention

1. bZIP-like polypeptides

Alignment of the polypeptide sequences was performed using the ClustalW algorithm (Thompson et al. (1997) *Nucleic Acids Res* 25:4876-4882; Chenna et al. (2003) *Nucleic Acids Res* 31:3497-3500) as present in VectorNTI (Invitrogen), with standard setting (gap opening penalty: 10, gap extension penalty: 0.05, gap separation penalty range: 8). Minor manual editing was done to further optimise the alignment. The bZIP-like polypeptides are aligned in Figure 2 A and B.

2. BCAT4-like polypeptides

Alignment of the polypeptide sequences was performed using the ClustalW 2.0.11 algorithm of progressive alignment (Thompson et al. (1997) *Nucleic Acids Res* 25:4876-4882; Chenna et al. (2003). *Nucleic Acids Res* 31:3497-3500) with standard setting (slow alignment, similarity matrix: Gonnet, gap opening penalty 10, gap extension penalty: 0.2). Minor manual editing was done to further optimise the alignment. The BCAT4-like polypeptides are aligned in Figure 6.

A phylogenetic tree of BCAT4-like polypeptides (Figure 7) was constructed by aligning BCAT4-like sequences using MAFFT (Katoh and Toh (2008) - *Briefings in Bioinformatics* 9:286-298). A neighbour-joining tree was calculated using Quick-Tree (Howe et al. (2002), *Bioinformatics* 18(11): 1546-7), 100 bootstrap repetitions. The tree was drawn using Dendroscope (Huson et al. (2007), *BMC Bioinformatics* 8(1):460). Confidence levels for 100 bootstrap repetitions are indicated for major branchings.

Example 3: Calculation of global percentage identity between polypeptide sequences

Global percentages of similarity and identity between full length polypeptide sequences useful in performing the methods of the invention were determined using one of the methods available in the art, the MatGAT (Matrix Global Alignment Tool) software (*BMC Bioinformatics*. 2003 4:29. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. Campanella JJ, Bitincka L, Smalley J; software hosted by Ledion Bitincka). MatGAT generates similarity/identity matrices for DNA or protein sequences without needing pre-alignment of the data. The program performs a series of pair-wise alignments using the Myers and Miller global alignment algorithm (with a gap opening penalty of 12, and a gap extension penalty of 2), calculates similarity and identity using for example Blosum 62 (for polypeptides), and then places the results in a distance matrix.

1. bZIP-like polypeptides

Results of the MatGAT analysis are shown in Figure 3 with global similarity and identity percentages over the full length of the polypeptide sequences. Sequence similarity is shown

in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2. The sequence identity (in%) between the bZIP-like polypeptide sequences useful in performing the methods of the invention can be as low as 18% (but is generally higher than 30%) compared to SEQ ID NO: 2. For SEQ ID NO: 4, the sequence identity with other bZIP-like polypeptide sequences useful in performing the methods of the invention can be as low as 23% but is on average higher than 48%)

2. BCAT4-like polypeptides

Results of the MatGAT analysis are shown in Figure 8 with global similarity and identity percentages over the full length of the polypeptide sequences. Sequence similarity is shown in the bottom half of the dividing line and sequence identity is shown in the top half of the diagonal dividing line. Parameters used in the analysis were: Scoring matrix: Blosum62, First Gap: 12, Extending Gap: 2. The sequence identity (in%) between the BCAT4-like polypeptide sequences useful in performing the methods of the invention can be as low as 37.9% (is generally higher than 37.9%) compared to SEQ ID NO: 142.

Table B: Description of proteins in Figure 8:

1. A.thaliana_AT1G10060
2. A.thaliana_AT1G10070
3. A.thaliana_AT1G50090
4. A.thaliana_AT1G50110
5. A.thaliana_AT3G19710
6. A.thaliana_AT3G49680
7. A.thaliana_AT5G65780
8. G.max_Glyma01g40420
9. G.max_Glyma06g05280
10. G.max_Glyma07g30510
11. G.max_Glyma08g06750
12. G.max_Glyma11g04870
13. H.annuus_TC54245
14. H.vulgare_TC165564
15. H.vulgare_TC186077
16. Hordeum_vulgare_PUT-169a-Hordeum_vulgare-79158
17. Hordeum_vulgare_subsp_vulgare_AK251931
18. M.truncatula_AC159872_36
19. M.truncatula_AC159872_55
20. M.truncatula_TC114768
21. O.sativa_LOC_Os03g12890
22. O.sativa_LOC_Os04g47190
23. O.sativa_LOC_Os05g48450
24. O.sativa_Os03g0106400

25. P.patens_TC31354
26. P.patens_TC33668
27. P.trichocarpa_804339
28. P.trichocarpa_scaff_II.1054
29. P.trichocarpa_scaff_IX.827
30. S.lycopersicum_TC213629
31. P.trichocarpa_BCAT4-like
32. T.aestivum_TC320973
33. T.aestivum_TC325793
34. Z.mays_TA12434_4577999
35. Zea_mays_GRMZM2G047347_T03
36. Zea_mays_GRMZM2G153536_T03

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

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

1.bZIP-like polypeptides

The results of the InterPro scan (InterPro database, release 31.0) of the polypeptide sequence as represented by SEQ ID NO: 2 and 4 are presented in Table C.

Table C1: InterPro scan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 2.

<i>InterPro ID</i>	<i>Domain name</i>	<i>Short name</i>	<i>Description</i>	<i>Location</i>
IPR004827	SM00338	no description	Basic-leucine zipper (bZIP) transcription factor	186-250
	PS50217	BZIP	Basic-leucine zipper (bZIP) transcription factor	188-251
	PS00036	BZIP_BASIC	Basic-leucine zipper (bZIP) transcription factor	193-208
IPR011616	PF00170	bZIP_1	bZIP transcription factor, bZIP-1	186-240
IPR012900	PF07777	MFMR	G-box binding, MFMR	001-099

No IPR ID	G3DSA:1.20.5.170	no description	NULL	180-246
		PTHR22952:SF5	CYCLIC-AMP-DEPENDENT TRANSCRIPTION FACTOR ATF-6 BETA	158-232
		PTHR22952	CAMP-RESPONSE ELEMENT BINDING PROTEIN-RELATED	158-232

Table C2: InterPro scan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 4.

<i>InterPro ID</i>	<i>Domain name</i>	<i>Short name</i>	<i>Description</i>	<i>Location</i>
NULL	G3DSA:1.20.5.170	no description	NULL	254-319
IPR004827	PS00036	BZIP_BASIC	Basic-leucine zipper (bZIP) transcription factor	267-282
IPR012900	PF07777	MFMR	G-box binding, MFMR	1-175
IPR011616	PF00170	bZIP_1	bZIP transcription factor, bZIP-1	262-322
	PTHR22952:SF73	GBF2 (G-BOX BINDING FACTOR 2); DNA BINDING / TRANSCRIPTION FACTOR	NULL	12-333
	PTHR22952	FAMILY NOT NAMED	NULL	12-333
IPR004827	SM00338	no description	Basic-leucine zipper (bZIP) transcription factor	260-324
IPR004827	PS50217	BZIP	Basic-leucine zipper (bZIP) transcription factor	262-325

In an embodiment a bZIP-like polypeptide comprises a conserved domain (or motif) with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the conserved domain starting with amino acid M1 up to amino acid N99 in SEQ ID NO: 2 (which corresponds to the conserved domain starting with amino acid M1 up to amino acid R175 in SEQ ID NO: 4).

2. BCAT4-like polypeptides

SEQ ID NO: 142 was checked against the NCBI database for conserved domains and it was found that SEQ ID NO: 142 is part of the BCAT_beta_family and part of the multidomains PLN02782.

In an embodiment a BCAT4-like polypeptide comprises a conserved domain (or motif) with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or

99% sequence identity to the conserved domain starting with amino acid 185 up to amino acid 202 in SEQ ID NO: 142, or to the conserved domain starting with amino acid 150 up to amino acid 167 in SEQ ID NO: 142, or to the conserved domain starting with amino acid 126 up to amino acid 143 in SEQ ID NO: 142.

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

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

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

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

- 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).

1.bZIP-like polypeptides

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

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

The results of TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID

NO: 2 are presented Table D. The “plant” organism group has been selected, no cutoffs defined, and the predicted length of the transit peptide requested. The subcellular localization of the polypeptide sequence as represented by SEQ ID NO: 2 may be the cytoplasm or nucleus, no transit peptide is predicted.

Table D: TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 2 and 4. 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
SEQ ID NO: 2	284	0.144	0.279	0.067	0.468	—	5	—
cutoff		0.000	0.000	0.000	0.000			
Name	Len	cTP	mTP	SP	other	Loc	RC	TPlen
SEQ ID NO: 4	399	0.238	0.053	0.045	0.940	—	2	—
cutoff		0.000	0.000	0.000	0.000			

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

Gel-shift analysis of GmbZIP DNA-binding ability(Liao et al., 2008)

Three copies of GLM (GTGAGTCAT) (Onate et al., J Biol Chem, 274:9175–9182, 1999), ABRE (CCACGTGG) (Jakoby et al., Trends Plant Sci 7:106–111, 2002), and PB-like (TGAAA) (Onate et al. 1999) elements are synthesised using standard techniques. Two complementary single-stranded oligonucleotides are mixed in 50 mM NaCl, heated at 70°C for 5 min and then cooled slowly to room temperature. Each annealed element is labeled with [γ - 32 P]ATP (about 110 TBq/mmol, Amersham) using T4 polynucleotide kinase and used as a probe. Gelshift analysis is performed with the 32 P-labeled probe and 2 μ g of recombinant protein. The competitive experiment is performed by adding an excess of 50X unlabeled element in addition to the 32 P-labeled probe. After incubation for 30 min at 25°C, the mixtures are subjected to electrophoresis in 6% (w/v) polyacrylamide gel on ice. The gel is placed on a sheet of Whatman 3MM filter paper, covered with plastic wrap and exposed to X-ray film overnight at -70°C with an intensifying screen.

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

1.bZIP-like polypeptides

The nucleic acid sequence of SEQ ID NO: 1 was amplified by PCR using as template a custom-made *Solanum lycopersicum* seedlings cDNA library. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 μ l PCR mix. The primers used were prm9943 (SEQ ID NO: 134; sense, start codon in bold): 5'-ggggacaagtttgataaaaaagcaggcttaaacaa**atg**cctccttatgggactc-3' and prm9944 (SEQ ID NO: 135; reverse, complementary): 5'-ggggaccacttttgataagaaagctg

ggtgctttccacttctcctaac-3', which include the AttB sites for Gateway recombination. Similarly, the nucleic acid sequence of SEQ ID NO: 3 was amplified by PCR using as template a custom-made *Populus trichocarpa* seedlings cDNA library using primers prm17402: 5'-ggggacaagttgtacaaaaaagcaggcttaacaatgggaaacattgaagaggg-3' (SEQ ID NO: 136) and prm17403: 5'ggggaccactttgtacaagaaagctgggttgaaccagtgtcatcaaccag-3' (SEQ ID NO: 137).

The amplified PCR fragments were purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment (comprising either SEQ ID NO: 1 or SEQ ID NO: 3) recombined in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone", pbZIP-like. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 1 or SEQ ID NO: 3 was then used in an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector contained as functional elements within the T-DNA borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the nucleic acid sequence of interest already cloned in the entry clone. A rice GOS2 promoter (SEQ ID NO: 131) for constitutive expression was located upstream of this Gateway cassette.

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

2. BCAT4-like polypeptides

The nucleic acid sequence was amplified by PCR using as template a custom-made *Populus trichocarpa* seedlings cDNA library. PCR was performed using a commercially available proofreading Taq DNA polymerase in standard conditions, using 200 ng of template in a 50 µl PCR mix. The primers used were prm15099 (SEQ ID NO: 219; sense): 5'-ggggacaa gtttgtacaaaaaagcaggcttaacaatggagagaagcgccgt-3' and prm15100 (SEQ ID NO: 220; reverse, complementary): 5'-ggggaccactttgtacaagaaagctgggttactgcagtacgcctaac-3', which include the AttB sites for Gateway recombination. The amplified PCR fragment was purified also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombined in vivo with the pDONR201 plasmid to produce, according to the Gateway terminology, an "entry clone", pBCAT4-like. Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

The entry clone comprising SEQ ID NO: 141 was then used in an LR reaction with a destination vector used for *Oryza sativa* transformation. This vector contained as functional

elements within the T-DNA borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the nucleic acid sequence of interest already cloned in the entry clone. A rice GOS2 promoter (SEQ ID NO: 218) for constitutive expression was located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector pGOS2::BCAT4-like (Figure 9) was transformed into *Agrobacterium* strain LBA4044 according to methods well known in the art.

Example 8: Plant transformation

Rice transformation

The *Agrobacterium* containing the expression vector was used to transform *Oryza sativa* plants. Mature dry seeds of the rice japonica cultivar Nipponbare were dehusked. Sterilization was carried out by incubating for one minute in 70% ethanol, followed by 30 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 Hodges 1996, Chan et al. 1993, Hiei et al. 1994).

Example 9: Transformation of other crops

Corn transformation

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

Wheat transformation

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

Soybean transformation

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

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

Rapeseed/canola transformation

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

Alfalfa transformation

A regenerating clone of alfalfa (*Medicago sativa*) is transformed using the method of (McKersie et al., 1999 Plant Physiol 119: 839–847). Regeneration and transformation of alfalfa is genotype dependent and therefore a regenerating plant is required. Methods to obtain regenerating plants have been described. For example, these can be selected from the cultivar Rangelander (Agriculture Canada) or any other commercial alfalfa variety as described by Brown DCW and A Atanassov (1985. Plant Cell Tissue Organ Culture 4: 111-112). Alternatively, the RA3 variety (University of Wisconsin) has been selected for use in tissue culture (Walker et al., 1978 Am J Bot 65:654-659). Petiole explants are cocultivated with an overnight culture of *Agrobacterium tumefaciens* C58C1 pMP90 (McKersie et al., 1999 Plant Physiol 119: 839–847) or LBA4404 containing the expression vector. The explants are cocultivated for 3 d in the dark on SH induction medium containing 288 mg/ L Pro, 53 mg/ L thioproline, 4.35 g/ L 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 acetosyringone 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 Hopher (Hussey, G., and Hopher, A., 1978. Annals of Botany, 42, 477-9) and are maintained on MS based medium supplemented with 30g/l

sucrose plus 0,25mg/l benzylamino purine and 0,75% agar, pH 5,8 at 23-25°C with a 16-hour photoperiod. *Agrobacterium tumefaciens* strain carrying a binary plasmid harbouring a 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 10: Phenotypic evaluation procedure

10.1 Evaluation setup

35 to 90 independent T0 rice transformants were generated. The primary transformants were transferred from a tissue culture chamber to a greenhouse for growing and harvest of T1 seed. Six events, of which the T1 progeny segregated 3:1 for presence/absence of the transgene, were retained. For each of these events, approximately 10 T1 seedlings containing the transgene (hetero- and homo-zygotes) and approximately 10 T1 seedlings lacking the transgene (nullizygotes) were selected by monitoring visual marker expression. The transgenic plants and the corresponding nullizygotes were grown side-by-side at random positions. Greenhouse conditions were of 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 (bZIP-like polypeptides)

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.

10.2 Statistical analysis: F test

A two factor ANOVA (analysis of variants) was used as a statistical model for the overall evaluation of plant phenotypic characteristics. An F test was carried out on all the parameters measured of all the plants of all the events transformed with the gene of the present invention. The F test was carried out to check for an effect of the gene over all the transformation events and to verify for an overall effect of the gene, also known as a global gene effect. The threshold for significance for a true global gene effect was set at a 5% probability level for the F test. A significant F test value points to a gene effect, meaning that it is not only the mere presence or position of the gene that is causing the differences in phenotype.

10.3 Parameters measured

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

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 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

1. bZIP-like polypeptides

The results of the evaluation of transgenic rice plants in the T1 generation and expressing a nucleic acid encoding the bZIP-like polypeptide of SEQ ID NO: 2 under nutrient-stress conditions are presented below in Table E1. When grown under conditions of nitrogen deficiency, an increase of at least 5% was observed for aboveground (or green) biomass (AreaMax), for root biomass (RootMax & RootThickMax) and for seed yield (total seed weight (totalwgseeds), number of filled seeds (nrfilledseed) and total number of seeds (nrtotalseed)). No negative effect on flowering time (delayed flowering) was observed.

In addition, plants expressing the *bZIP-like* nucleic acid of SEQ ID NO: 1 showed an increase in Thousand Kernel Weight, fillrate, and number of first panicles in at least one of the tested lines. One of the tested lines also showed an earlier flowering time.

Table E1: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for the T1 generation, for each parameter the p-value is <0.05.

Parameter	Overall increase
AreaMax	14.7
RootMax	8.8
totalwgseeds	18.1
nrtotalseed	10.0
nrfilledseed	16.1
RootThickMax	5.3

A similar trend was observed for plants grown under non-stress conditions: aboveground biomass, root biomass and seed yield (total seed weight, total number of seeds, fillrate, and number of filled seeds) were increased in at least 2 of the lines tested, compared to control plants

The results of the evaluation of transgenic rice plants in the T1 generation and expressing a nucleic acid encoding the bZIP-like polypeptide of SEQ ID NO: 4 under drought-stress conditions are presented below in Table E2. When grown under conditions of drought, an

increase of at least 5% was observed for aboveground (or green) biomass (AreaMax), for root biomass (RootThickMax) and for seed yield (total seed weight (totalwgseeds), fillrate, harvest index, and number of filled seeds (nrfilledseed)). No negative effect on flowering time (delayed flowering) was observed.

Table E2: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown for the T1 generation, for each parameter the p-value is <0.05.

Parameter	Overall increase
AreaMax	8.6
totalwgseeds	42.5
fillrate	46.6
harvestindex	33.5
nrfilledseed	44.0
RootThickMax	8.4

2. BCAT4-like polypeptides

The results of the evaluation of transgenic rice plants expressing a *BCAT4-like* nucleic acid under drought-stress conditions are presented hereunder. An increase of at least 5% was observed for total seed weight, fill rate, harvest index and number of seeds (Table E3).

The results of the evaluation of transgenic rice plants in the T1 generation and expressing a nucleic acid encoding the BCAT4-like polypeptide of SEQ ID NO: 142 under drought stress conditions are presented below in Table D. When grown under drought stress conditions, an increase of at least 5% was observed for seed yield, including total weight of seeds (totalwgseeds), fill rate, harvest index and number of seeds (nrfilledseed). In addition, 2 lines were clearly positive for GravityYMax, which is the height of the gravity center of the leafy biomass of the plant.

Table E3: Data summary for transgenic rice plants; for each parameter, the overall percent increase is shown (T1 generation), for each parameter the p-value is <0.05.

Parameter	Overall increase
totalwgseeds	34.5
fillrate	36.0
harvestindex	29.8
nrfilledseed	32.1

Claims

1. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding a bZIP-like polypeptide, wherein said bZIP-like polypeptide comprises Basic Leucine Zipper Domain (PF00170) and a G-box binding domain of the MFMR type (PF07777) and one or more of motifs 1 to 3, represented by SEQ ID NO: 119, SEQ ID NO: 120 or SEQ ID NO: 121.
2. Method according to claim 1, wherein said bZIP-like polypeptide comprises one or more of motifs 4 to 6.
3. Method according to claim 1, wherein said bZIP-like polypeptide comprises one or more of motifs 7 to 12.
4. Method according to any one of claims 1 to 3, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said bZIP-like polypeptide.
5. Method according to claim 1 to 4, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.
6. Method according to claim 1 to 5, wherein said enhanced yield-related traits are obtained without effect on flowering time of the plant.
7. Method according to any one of claims 1 to 3, wherein said enhanced yield-related traits are obtained under non-stress conditions.
8. Method according to any one of claims 1 to 3, wherein said enhanced yield-related traits are obtained under conditions of drought stress or nitrogen deficiency.
9. Method according to any one of claims 1 to 8, wherein said nucleic acid encoding a bZIP-like is of plant origin, preferably from a dicotyledonous plant.
10. Method according to any one of claims 1 to 8, wherein said nucleic acid encoding a bZIP-like encodes any one of the polypeptides listed in Table A1 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
11. Method according to any one of claims 1 to 8, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A1.
12. Method according to any one of claims 1 to 11, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 2 or SEQ ID NO: 4.
13. Method according to any one of claims 1 to 12, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
14. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method according to any one of claims 1 to 13, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding a bZIP-like polypeptide as defined in any of claims 1 to 3 and 9 to 12.
15. Construct comprising:
 - (i) nucleic acid encoding a bZIP-like as defined in any of claims 1 to 3 and 9 to 12;

- (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - (iii) a transcription termination sequence.
16. Construct according to claim 15, wherein one of said control sequences is a constitutive promoter, preferably a medium strength constitutive promoter, preferably a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
 17. Use of a construct according to claim 15 or 16 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
 18. Plant, plant part or plant cell transformed with a construct according to claim 15 or 16.
 19. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants, comprising:
 - (i) introducing and expressing in a plant cell or plant a nucleic acid encoding a bZIP-like polypeptide as defined in any of claims 1 to 3 and 9 to 12; and
 - (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.
 20. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding a bZIP-like polypeptide as defined in any of claims 1 to 3 and 9 to 12 or a transgenic plant cell derived from said transgenic plant.
 21. Transgenic plant according to claim 14, 18 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 21, wherein said harvestable parts are preferably shoot biomass and/or seeds.
 23. Products derived from a plant according to claim 21 and/or from harvestable parts of a plant according to claim 22.
 24. Use of a nucleic acid encoding a bZIP-like polypeptide as defined in any of claims 1 to 3 and 9 to 12, for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.
 25. A method for manufacturing a product, comprising the steps of growing the plants according to claim 14, 18, 20 or 21 and producing said product from or by said plants; or parts thereof, including seeds.
 26. A method for enhancing yield-related traits in plants relative to control plants, comprising modulating expression in a plant of a nucleic acid encoding an BCAT4-like

polypeptide, wherein said BCAT4-like polypeptide comprises the signature sequence represented by SEQ ID NO: 216.

27. Method according to claim 26, wherein said polypeptide is encoded by a nucleic acid molecule comprising a nucleic acid molecule selected from the group consisting of:
- (i) a nucleic acid represented by any one of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211;
 - (ii) the complement of a nucleic acid represented by any one of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211;
 - (iii) a nucleic acid encoding the polypeptide as represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid being deducible from a polypeptide sequence as represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, or 212 and further preferably conferring enhanced yield-related traits relative to control plants;
 - (iv) a nucleic acid having, in increasing order of preference at least 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with any of the nucleic acid sequences of SEQ ID NO: 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, or 211, and further preferably conferring enhanced yield-related traits relative to control plants;
 - (v) a first nucleic acid molecule which hybridizes with a second nucleic acid molecule of (i) to (iv) under stringent hybridization conditions and preferably confers enhanced yield-related traits relative to control plants;
 - (vi) a nucleic acid encoding said polypeptide having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by any one of SEQ ID NO: 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200,

- 202, 204, 206, 208, 210, or 212 and preferably conferring enhanced yield-related traits relative to control plants; or
- (vii) a nucleic acid comprising any combination(s) of features of (i) to (vi) above.
28. Method according to claim 26 or 27, wherein said modulated expression is effected by introducing and expressing in a plant said nucleic acid encoding said BCAT4-like polypeptide.
29. Method according to any one of claim 26 to 28, wherein said enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control plants.
30. Method according to any one of claims 26 to 29, wherein said enhanced yield-related traits are obtained under conditions of drought stress.
31. Method according to any of claims 26 to 30, wherein said BCAT4-like polypeptide comprises one or more of the following motifs:
- (i) Motif 13 represented by SEQ ID NO: 213,
 - (ii) Motif 14 represented by SEQ ID NO: 214,
 - (iii) Motif 15 represented by SEQ ID NO: 215.
32. Method according to any one of claims 26 to 31, wherein said nucleic acid encoding an BCAT4-like is of plant origin, preferably from a dicotyledonous plant, further preferably from the family Salicaceae, more preferably from the genus *Populus*, most preferably from *Populus trichocarpa*.
33. Method according to any one of claims 26 to 32, wherein said nucleic acid encoding a BCAT4-like polypeptide encodes any one of the polypeptides listed in Table A2 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with such a nucleic acid.
34. Method according to any one of claims 26 to 33, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table A2.
35. Method according to any one of claims 26 to 34, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 142.
36. Method according to any one of claims 26 to 35, wherein said nucleic acid is operably linked to a constitutive promoter, preferably to a medium strength constitutive promoter, preferably to a plant promoter, more preferably to a GOS2 promoter, most preferably to a GOS2 promoter from rice.
37. Plant, plant part thereof, including seeds, or plant cell, obtainable by a method according to any one of claims 26 to 36, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an BCAT4-like polypeptide as defined in any of claims 26, 27 and 31 to 36.
38. Construct comprising:
- (i) nucleic acid encoding an BCAT4-like as defined in any of claims 26, 27 and 31 to 36;
 - (ii) one or more control sequences capable of driving expression of the nucleic acid sequence of (i); and optionally
 - (iii) a transcription termination sequence.

39. Construct according to claim 38, wherein one of said control sequences is a constitutive promoter, preferably a medium strength constitutive promoter, preferably a plant promoter, more preferably a GOS2 promoter, most preferably a GOS2 promoter from rice.
40. Use of a construct according to claim 38 or 39 in a method for making plants having enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants.
41. Plant, plant part or plant cell transformed with a construct according to claim 38 or 39.
42. Method for the production of a transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass relative to control plants, comprising:
 - (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an BCAT4-like polypeptide as defined in any of claims 26, 27 and 31 to 36; and
 - (ii) cultivating said plant cell or plant under conditions promoting plant growth and development.
43. Transgenic plant having enhanced yield-related traits relative to control plants, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass, resulting from modulated expression of a nucleic acid encoding an BCAT4-like polypeptide as defined in any of claims 26, 27 and 31 to 36 or a transgenic plant cell derived from said transgenic plant.
44. Transgenic plant according to claim 37, 41 or 43, 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.
45. Harvestable parts of a plant according to claim 44, wherein said harvestable parts are preferably shoot biomass and/or seeds.
46. Products derived from a plant according to claim 44 and/or from harvestable parts of a plant according to claim 45.
47. Use of a nucleic acid encoding an BCAT4-like polypeptide as defined in any of claims 26, 27 and 31 to 36 for enhancing yield-related traits in plants relative to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass in plants relative to control plants.
48. A method for manufacturing a product comprising the steps of growing the plants according to claim 37, 41, 43 or 45 and producing said product from or by said plants; or parts thereof, including seeds.
49. Construct according to claim 38 or 39 comprised in a plant cell.
50. Recombinant chromosomal DNA comprising the construct according to claim 38 or 39.

51	100		
		A.lyrata_944204#1	FFPSVGSPPHAYMWGAQHMMMP--YGTVPYPAMYPPGAVYSHPSMPM
		A.thaliana_AT4G36730.2#1	FFPSVGSPPHAYMWGAQHMMMP--YGTVPYPAMYPPGAVYSHPSMPM
		Aquilegia_sp_TC21139#1	FFPTNVASPPPYMVG--QMVSP--YGTPIPYPAIYPHGGLYPHPNLAT
		B.napus_TC63535#1	FFPSVGSPPHAYMWGAQHMMMP--YGTVPYPAMYPPGAVYSHPSMPM
		B.napus_TC63581#1	FFPSVGSPPHAYMWGAQHMMMP--YGTVPYPAMYPPGAVYSHPSMPM
		B.napus_TC89867#1	FFPSVGSPPHAYMWGAQHMMMP--YGTVPYPAMYPPGAVYSHPSMPM
		C.canephora_TC1061#1	FFASTVASPPHAYLWGNQHPMLMP--YGTVPYPALYP--GGVYAHPNMAM
		C.clementina_TC35556#1	-----MPP--IGHPRFHPQVYFFPGGVCPSWPGGL
		C.endivia_EL359780#1	-----MGLRFHTPLHYPPAGVYAHPSMPM
		C.roseus_AY027510#1	-----MWGGQHPMLMP--YGTVPYPALYPPGGVYAHPTMAT
		C.sinensis_TC9189#1	FFASTVASPTPHPYLWGSQHPMLMP--YGTVPYQAIYPPGGVYAHPSMAT
		C.tinctorius_TA1437_4222#1	FFASTVASPTPHPYIWWGGQHPMSP--YGTVPYPALYPPAGVYAHPSMPM
		E.esula_TC3985#1	FFASTVPSPTPHPYLWGGQHPMPPPYGTVPYPALYPPAGVYSHPTMTT
		F.vesca_TA11395_57918#1	-----MMPP--YGTVPYPALYPPGGVYAHPSMVT
		G.hirsutum_TC137570#1	LFASNVASPTPHPYIWWGGQHPMLMP--YGTVPYPAVYPPRGVYAHPNMAP
		G.max_Glyma11g06960.1#1	FFASNIASPTPHSYMWGSQHPLI--YSTVPYPALYPPGNVYAHPSMAM
		G.max_Glyma16g25600.2#1	FFATTVASPTPHPYLWGGQHPMLMP--YGTVPYPALYPPGSIYAHPSMAV
		H.annuus_BU027457#1	-----
		H.petiolaris_TA3844_4234#1	FFPSIVASPPPHPYMWGGQHPMMP--YGAPVYPTLYSPPEVYAHPTPM
		H.tuberosus_TA2966_4233#1	-----MMP--YGTVPYAAAMYPPAGVYGHFPMMPM
		M.truncatula_AC148484_8.5#1	YYASTVASPTPHPYMWGGQHPMMP--YGTVPYPAMFPPGNIYAHPSMVV
		N.tabacum_BP133908#1	-----
		N.tabacum_TC76189#1	FFASPVASPASHPYLWGGQHPMLMP--YGAPVYPALYPPAGVYAHPNMAT
		P.taeda_TA23200_3352#1	-----
		P.trichocarpa_719452#1	-----MPP--YGTVPYQAIYPPGGVYAHPSMAT
		P.trifoliata_TA6299_37690#1	-----MPP--YGTVPYPALYPPAGVYAHPNMAT
		S.lycopersicum_like_bZIP	FFASPVASPASHPYMWGGQHPMLMP--YGTVPYPALYPPAGVYAHPNMAT
		S.lycopersicum_TC211600#1	FFPSTVASPTPHPYMWGGQHPMLMP--YGAPVYPALYPPAGVYAHPNMPPM
		S.tuberosum_TC186959#1	-----P-----ALYPPAGVYAHPNMPPM
		S.tuberosum_TC187892#1	YYASTVQPTPHPYMWGSQHPMLMP--YGTVPYPAMYPPGNIYAHPSMVV
		T.pratense_TA1890_57577#1	FFPSPVAPSPPHPYLWGGQHPMMP--YGTPLYPALYPRGALYAHPSMAT
		V.vinifera_GSVIVT00014657001#1	FFAPSVASPTPHPYLWGSQHPLI--YGTVPYSALYPPGGVYAHPNLAT
		V.vinifera_GSVIVT00024984001#1	FF S VASPTPHPYMWGGQHP M MPP YGTVPYPALYPPGGVYAH P M M
		Consensus	

FIGURE 2 A (continued)

	101		
A.lyrata_944204#1	(96)	PPNSGPTNKE-----TVKDQASGKKSXGSSKKKG-----EGGDKALS	150
A.thaliana_AT4G36730.2#1	(96)	PPNSGPTNKE-----PAKDQASGKKSXGSSKKKA-----EGGDKALS	
Aquilegia_sp_TC21139#1	(95)	AQGAAMP1TQTEENNSPVKKIKSSGN--IGVVGGKLIK-----ESGKAAS	
B.napus_TC63535#1	(93)	PPSSAPTNE-----TVKEQAPGKKSXGSLKRKG-----EGGEKAPS	
B.napus_TC63581#1	(96)	PQASGPTNTE-----TVKAQAPGKPKGNLKRKS-----GGSEKAPS	
B.napus_TC89867#1	(93)	PPASAPTNKE-----TVKEQAPGKKSXGSLKRKG-----EGGEKAPS	
C.canephora_TC1061#1	(61)	APGAVQAPIESDAKAP-DGKDRNTNKKLKGPSGNPGLIAVKAGESGKAAS	
C.clementina_TC35556#1	(28)	EFKPGPNQN-RAGRKGPGAKDRGSFKNSR-----GLRQVRLGRLFRQLL	
C.endivia_EL359780#1	(25)	TPSPAPANTEMEAKAY-EGKERATNKKSKGTSNGN-VNVRTGDSGIAAS	
C.roseus_AY027510#1	(36)	TPGTTQANAESDAVKVSEKDRPTSKRSRGASGNHGLVAAKVAESGKAAS	
C.sinensis_TC9189#1	(97)	-TPTAAPTNTPEEGKPEAKDRASAKKSKGTPGGKA-----GEIVKATS	
C.tinctorius_TA1437_4222#1	(69)	TPSTAPPNAEME VKAY-EGKERAANKKSKGTSNGNAAVVRTGESGKAAS	
E.esula_TC3985#1	(100)	-TPNSAPVNAEFEGKGPDKDRASAKKSGASAGKG-----GETGKATS	
F.vesca_TA11395_57918#1	(29)	TPASVPTNPESEKSTDGKERASAKKPKGAA--GLVSGKAGDGGKATS	
G.hirsutum_TC137570#1	(95)	-MPSSARNNG-----ADGKDRGVTKKPKGSSGSKV-----GESAKATS	
G.max_Glyma11g06960.1#1	(97)	-TLSTQNGTEFVKGSGDEKDRVSAKSKAVSANNGSK--AGDNGKASS	
G.max_Glyma16g25600.2#1	(96)	-TPSAVQOQNTIEGKGAEGKYRDSKKLKGPSANTASK--AGESGKAGS	
H.annuus_BU027457#1	(1)	-----MSTILV-LGLERVG-RQLRVQGMTPHKAACNVS-----	
H.petiolaris_TA3844_4234#1	(94)	VVR-----	
H.tuberosus_TA2966_4233#1	(29)	TPDVTQPTTEMESKAP-NGKDKVN-KKSKGSSGNVNASGAKTRESGKAAS	
M.truncatula_AC148484_8.5#1	(99)	-TPSAMHQTFEFGKGPDKDKDSSKKPKGTSANTSAK---AGEGGKAGS	
N.tabacum_BP133908#1	(1)	-----	
N.tabacum_TC76189#1	(95)	HTPNAVQANLESNRKDPKDRSTNKKLKASSGGKA-----GDSGKVAS	
P.taeda_TA23200_3352#1	(1)	-----	
P.trichocarpa_719452#1	(1)	-----	
P.trifoliata_TA6299_37690#1	(28)	TPTVAPTNT-ELEGKPEAKDRASAKKSK-----GTPGGKAGEIVKATS	
S.lycopersicum_like_bZIP	(28)	PAPNSVPANPEADGKPEGKDRNSSKKLKVCSGGKA-----GDNGKVTS	
S.lycopersicum_TC211600#1	(95)	PAPNSVPANPEADGKPEGKDRNSSKKLKVCSGGKA-----GDNGKVTS	
S.tuberosum_TC186959#1	(94)	-TPNTLQANPESDSKAPDGKDQNTSKLKGCSGGKA-----GEIGKAAS	
S.tuberosum_TC187892#1	(45)	-TPNTLQANPESDSKAPDGKDQNTSKLKGCSGGKA-----GEIGKAAS	
T.pratense_TA1890_57577#1	(94)	-APSAMHQTFEFGKGPDKDKDSSKKPKGTSANTGAK---AGESGKAGS	
V.vinifera_GSVIVT00014657001#1	(97)	-AQQVALTNTDMEVKTPDGKDPASIKKSKAASGNMGLISGKSGESGKAAS	
V.vinifera_GSVIVT00024984001#1	(97)	-APSAHLNPELEGKPEGKDKASAKKSKGTSNTVTK---GGESGKAAS	
Consensus	(101)	K DGKDR S KK KG SG G GE GKA S	

FIGURE 2 A (continued)

151	151	200
A.lyrata_944204#1	(133)	GSGNDG-VSHSDDSVTAGSSDENDDNANQEQGSVVRKPSFGQMLADASSQ
A.thaliana_AT4G36730.2#1	(133)	GSGNDG-ASHSDESVTAGSSDENENANQ--QGSIRKPSFGQMLADASSQ
Aquilegia_sp_TC21139#1	(137)	GSRNDGVSRAE-SGSEGSASDEFNHKDDS-ENKNKSFQDMLADGANA
B.napus_TC63535#1	(129)	GSGNDG-VSHSDESVTAGSSDENENANHQEHGSVRKPSPFGQMLADASSQ
B.napus_TC63581#1	(133)	GSGNDA-VSQSEESVTAGSSDENDDNANHQEQGSVVRKPSFGQMLADASSQ
B.napus_TC89867#1	(130)	GSGNDG-VSHSDESVTAGSSDENENANHQEQGSVVRKPSFGQMLADASSQ
C.canephora_TC1061#1	(110)	GSG-NDGATQSAESGSEGSDDGSDEN-NNHELSTATKKGSFDQMLADGATA
C.clementina_TC35556#1	(71)	VSG-NDGVFQSGGSDGSDGSDENCNQEFVGRKKGSFDKMLADANTE
C.endivia_EL359780#1	(73)	SSGNDGGATQSAESGSDGSDGSDEN-DQNEFSGGKKGSDFNQMLADANAQ
C.roseus_AY027510#1	(86)	ESG-NDGATQSAESGSEGSDDGSDEN-NNHELSTGTTKKGSEQMLADGATA
C.sinensis_TC9189#1	(140)	GSGNDGVSQSAE-SGSDGSSDASDEN-----GNRAGVC-----
C.tinctorius_TA1437_4222#1	(118)	SSG-NDGATQSAESGSDGSDGSDSEEN-DQHEYSGGKKGSDFNQMLADANAQ
E.esula_TC3985#1	(143)	GSGNDGASQSGE-SGSDGSDGSDEN---QEYGANKKGSFDQMLADANAQ
F.vesca_TA11395_57918#1	(76)	GSG-NDGASQSAESGSEGSDDGSEENGHNHQEYGANKKGSFDKMLADGANA
G.hirsutum_TC137570#1	(132)	GSGNDGGSQSGE-SGSEGTSDRSDESNQ-QEVNAGKKGSEFEQMLADANAQ
G.max_Glyma11g06960.1#1	(143)	GPRNDGTSQSAE-SGSEGSDDASDENTNQESATNKKKGSFDQMLVDGANA
G.max_Glyma16g25600.2#1	(142)	GSGNDGISQSGE-SGSEGSNASDENTNQESAAANKKGSFDLMLVDGANA
H.annuus_BU027457#1	(32)	-----SAESGNGSSDANDED-TQQEHSGSKKGSFHQMLADANAR
H.petiolaris_TA3844_4234#1	(102)	SSGNDGNTTQSAESDENDGSSDANLQNDKSG-NLVVHTPATAKNVAVEMQS
H.tuberosus_TA2966_4233#1	(77)	SSG-NDGGTQSAESGNGSSDASDED-NQDYSGGKKGSEFHQMLADANAR
M.truncatula_AC148484_8.5#1	(145)	GSGNDGFHSGD-SGSEGSNASDE--NQQESARNKKGSEFDLMLVDGANA
N.tabacum_BP133908#1	(1)	-----
N.tabacum_TC76189#1	(139)	GSGNDGATQSAE-TRSEGTSDTNDENDN-HEFFAASKKGSFDQMLADGASA
P.taeda_TA23200_3352#1	(1)	-----
P.trichocarpa_719452#1	(1)	-----MAHQEYGASKKGSFNQMLADANAQ
P.trifoliata_TA6299_37690#1	(71)	GSG-NDGVQSAESGSDGSSDASDENDGNQEFAGKKGSEFDQMLADANTE
S.lycopersicum_like_bZIP	(72)	GSGNDGATQSAE-SRSEGTSDTNDENDN-NEFAANKKGSFDQMLADGASA
S.lycopersicum_TC211600#1	(139)	GSGNDGATQSAE-SRSEGTSDTNDENDN-NEFAANKKGSFDQMLRDGASA
S.tuberosum_TC186959#1	(137)	GSGNDGGATRSAESGSEGSDDENDENDN-HEFSADKNRSFDLMLANGANA
S.tuberosum_TC187892#1	(88)	GSGNDWWCHKKC-----
T.pratense_TA1890_57577#1	(140)	GSGNDGFSQSGE-SGSEGSNGSDE--NQQESARNKKGSEFDLMLVNGANV
V.vinifera_GSVIVT00014657001#1	(146)	VSGNDGASQSGE-SGSEASSDATDENAN-QASSAVKKRS--FNLADGSNA
V.vinifera_GSVIVT00024984001#1	(142)	GSGNDGASPSAE-SGSEGSDDASDENTNQEFASSKKGSEFNQMLADANAQ
Consensus	(151)	GSGNDG S SE SGSEGSDDASDEN N QE AA KKGSE QMLADAA

FIGURE 2 A (continued)

201	250
A.lyrata_944204#1	-ST-----TG-EIQ-----GSVPMKPVAPGTN-LNIGMDL
A.thaliana_AT4G36730.2#1	-ST-----TG-EIQ-----GSVPMKPVAPGTN-LNIGMDL
Aquilegia_sp_TC21139#1	QNTSAHHSSTAVGSSLNGNGEPSANFPVPLPGNPVGDIAATN-LNIGMDL
B.napus_TC63535#1	SNT-----TGEMIQ-----GSVPMKPLAPGTN-LNMGMMDL
B.napus_TC63581#1	SNT-----TG-EIQ-----GSMMPKPVAPGTN-LNMGMMDL
B.napus_TC89867#1	SNT-----TG-EIQ-----GSMMPKPVAPGTN-LNMGMMDL
C.canephora_TC1061#1	QNNT-----SVANFQNS-----VPGNPVVSVPATN-LNIGMDL
C.clementina_TC35556#1	NNTA---EA--VPGS-----VPGKPVVSMPATN-LNIGMDL
C.endivia_EL359780#1	NNN-----FHTPV-----VPVNPVTSIPGTN-LIMRMDL
C.roseus_AY027510#1	QNST---AIANFPNS-----VPGN-PVAMPATN-LNIGMDL
C.sinensis_TC9189#1	-----
C.tinctorius_TA1437_4222#1	NNNS---GP-NIQTS-----VPGNPVVSIPGTN-LNMGMMDL
E.esula_TC3985#1	NN-----GIQGS-----VPGKPVASMAGAN-LXXGMDL
F.vesca_TA11395_57918#1	QNNT---GS---VPG-----KP---VVSMPATS-LNMGMMDL
G.hirsutum_TC137570#1	GK-----AAGAL-----VPAEPIVSMPGNY-FEYRN--
G.max_Glyma11g06960.1#1	RNN---SVSIIPQ-----P-GNPAVSMSPS-LNIGMNL
G.max_Glyma16g25600.2#1	QNN---SAGAISQSS-----VPGKPVVPMPATN-LNIGMDL
H.annuus_BU027457#1	NN-----NHVASMPATTNLYMGMMN
H.petiolaris_TA3844_4234#1	-----NP-----SDAPQTAVPP-----
H.tuberosus_TA2966_4233#1	NNS---GP-NVQMP-----VPGNPVVSMPATN-LNMGMMDL
M.truncatula_AC148484_8.5#1	QNN---TTGPISQSS-----VPGNPVVSIPATN-LNIGMDL
N.tabacum_BP133908#1	-----MFMGGLTCLLAPTF
N.tabacum_TC76189#1	QNN---PTTANYQTS-----MHAN-PVTVHATN-LNIGMDV
P.taeda_TA23200_3352#1	-----MALVPPAP-----VPGMSTVSLPTTN-LNIGMGV
P.trichocarpa_719452#1	STSA---GAN-IQAS-----VPGKPVASMPATN-LNIGMDL
P.trifoliata_TA6299_37690#1	NNTV---EA--VPGS-----VPGKPVVSMPATN-LNIGMDL
S.lycopersicum_like_bZIP	QNN---PAKENHPTS-----IHGN-PVTMPATN-LNIGMDV
S.lycopersicum_TC211600#1	QNN---PAKENHPTS-----IHG---ICTMPATN-LNIGMDV
S.tuberosum_TC186959#1	QT-----NPATG-----NP---VAMPATN-LNIGMDL
S.tuberosum_TC187892#1	-----
T.pratense_TA1890_57577#1	QNN---NTGPISQSP-----VPGNPVVSIPATN-LNIGMDL
V.vinifera_GSVIVT00014657001#1	KGNSAAQYTGGNHSAS-----VPGKPVVPMPPTS-LNIGMDL
V.vinifera_GSVIVT00024984001#1	NNI---SG-TSVQAS-----VPGKPVISMPATN-LNIGMDL
Consensus	QNN VPG VVSMPATN LNIGMDL

FIGURE 2 A (continued)

251	300
A.lyrata_944204#1	WSSQAGVPKDE--RELK-----R
A.thaliana_AT4G36730.2#1	WSSQAGVPKDE--RELK-----R
Aquilegia_sp_TC21139#1	WNASPVGSVPLR--ARSNASGVPAVAPVKRDGHEGIVPEHLWGQDERE
B.napus_TC63535#1	WSSQAGVPKDE--RELK-----R
B.napus_TC63581#1	WSSQTGVAVKDE--RELK-----R
B.napus_TC89867#1	WSSQTGVAVKDE--RELK-----R
C.canephora_TC1061#1	WNPSSGASGAMK--MRPNP--GVSPAVAPGM-----MTDQWIQDERE
C.clementina_TC35556#1	WNTSPAAAGAAK--MRTNPFASPAVAPAGI-----MPDQWIQDERE
C.endivia_EL359780#1	RNPS-TGNAAMK--MRTNHSKSRGEVPPMKPES-----
C.roseus_AY027510#1	WNASSAAPGAMK--MRPSH--GVPSAVAPGM-----VNDQWIQDERE
C.sinensis_TC9189#1	-----
C.tinctorius_TA1437_4222#1	WNPS-TSGGTMK--IRSNPSGVARAAVPPPMIGREG-MMPDQWVQDERE
E.esula_TC3985#1	WNPSAAAAGTAK--IRPN-----ASGAPSGITPA-----
F.vesca_TA11395_57918#1	WNPSPAGAGTAK--MRGN-----QSGAPSAV-----GGDHWIQDERE
G.hirsutum_TC137570#1	-GMECFPCCHR--SSKN-----QT-----
G.max_Glyma11g06960.1#1	WNASPAGDEAAK--MRQN--QSSGAVTPTIMGREVALGEHWIQDER-E
G.max_Glyma16g25600.2#1	WNASSGGAEAAK--MRHN--QSGAPGVA-----LGDQWVQDER-E
H.annuus_BU027457#1	WTPP-TGS-----VSQVPTPPVMADR-----WVQDERE
H.petiolaris_TA3844_4234#1	-PIMGQWRQDE--RELK-----
H.tuberosus_TA2966_4233#1	WNPS-AGSGSMK--MQPNHSGVSQPGVPPIMADQ-----WGQDERE
M.truncatula_AC148484_8.5#1	WNASSAGAEAAK--MRHN--QPGAPGAG-----ALGEQWMMQDDRE
N.tabacum_BP133908#1	WNPKCTQQFPLG-----
N.tabacum_TC76189#1	WNASSAGPGAI I--IQPN-----VNGPVIHEGRMNDQWVQDER-E
P.taeda_TA23200_3352#1	YNVPAQGSVTPVKGRQGTNSVSTIVPAASQLIPGHDGVPSELWVQDERE
P.trichocarpa_719452#1	WNAS-SAAGATK--MRPNPSCATSGVVPAG-----LPEQWIQDERE
P.trifoliata_TA6299_37690#1	WNTSPAAAGAAK--MRTNPSGASPAVAPAGI-----IPDQWIQDERE
S.lycopersicum_like_bZIP	WNASAAAGPAIK--IQQN-----ATGPVIGHEGRMNDQWIQEER-E
S.lycopersicum_TC211600#1	WNASAAAGPAIK--IQQN-----ATGPVIGHEGRMNDQWIQEER-E
S.tuberosum_TC186959#1	WNATAWRVPE-----
S.tuberosum_TC187892#1	-----
T.pratense_TA1890_57577#1	WNASPANAPEATK--LRHN--QSSAPGA-----GEQWMMQDDRE
V.vinifera_GSVIVT00014657001#1	WNASPAGGTPMK--TRPQSSGASPVASATIVGREGMLQDHQWIQDERE
V.vinifera_GSVIVT00024984001#1	WSASPGSGGATK--LRPNPSGIISSVAPAAMVREGVMPDQWIQDER-E
Consensus	WNAS AG K M N D E

FIGURE 2 A (continued)

301			
	A.lyrata_944204#1	(226)	QKR--KQSNRESARRSRLRKQAECEQEQQRVESLSNENQSLRDELQRLS
	A.thaliana_AT4G36730.2#1	(224)	QKR--KQSNRESARRSRLRKQAECEQEQQRVESLSNENQSLRDELQRLS
	Aquilegia_sp_TC21139#1	(281)	LKRQRKLSNRESARRSRLRKQAECEELQVKVDTLTDENDNLRKELELRLA
	B.napus_TC63535#1	(224)	QKR--KQSNRESARRSRLRKQAECEQEQQRVESLTSNQSLRDELQRLS
	B.napus_TC63581#1	(227)	QKR--KQSNRESARRSRLRKQAECEQEQQRVESLTSNQSLRDELQRLS
	B.napus_TC89867#1	(224)	QKR--KQSNRESARRSRLRKQAECEQEQQRVESLTSNQSLRDELQRLS
	C.canephora_TC1061#1	(228)	LKRQKRKQSNRESARRSRLRKQAECEELQQRVESLNSNRALRDELQKVS
	C.clementina_TC35556#1	(190)	LKRQKRKQFNRESARRSRLRKQAECEGELQARVGTLSNENRNLRDELQRLF
	C.endivia_EL359780#1	(182)	-----
	C.roseus_AY027510#1	(203)	LKRQKRKQSNRESARRSRLRKQAECEELQQRVETLSNENRALRDELQRLS
	C.sinensis_TC9189#1	(172)	-----
	C.tinctorius_TA1437_4222#1	(242)	LKRQKRKQSNRESARRSRLRKQAECEEPQARVEALSNNHSLRDELQRLS
	E.esula_TC3985#1	(243)	-----
	F.vesca_TA11395_57918#1	(186)	LKRQKRKQSNRESARRSRLRKQAECEELQKSVHGLTNNHGLKDELQRLS
	G.hirsutum_TC137570#1	(222)	-----
	G.max_Glyma11g06960.1#1	(265)	LKKQKRKQSNRESARRSRLRKQAECEELQKRVESLGSNQTLREELQRVS
	G.max_Glyma16g25600.2#1	(258)	LKRQKRKQSNRESARRSRLRKQAECEELQKRVESLGGENQTLREELQRLS
	H.annuus_BU027457#1	(119)	LKRQKRKQDNRESARRSRLRKQAECEALQATVETLNNHSLRDELQRLS
	H.petiolaris_TA3844_4234#1	(179)	QKR--KQSNRESARRSRLRKQAECEELQARVEVLSNENHTLKDELQRLS
	H.tuberosus_TA2966_4233#1	(194)	LKRLKQSNRESARRSRLRKQTECEELQARVETLNNHSLRDELQRLS
	M.truncatula_AC148484_8.5#1	(261)	LKRQKRKQSNRESARRSRLRKQAECEELQKRVEALGGENRTLREELQKLS
	N.tabacum_BP133908#1	(27)	-----FSGSHVVF-CFLFLQAECEELQRRVEALSSENHSLKDELQRLS
	N.tabacum_TC76189#1	(256)	LKRQKRKQSNRESARRSRLRKQAECEELQRRVEALSSENHSLKDELQRLS
	P.taeda_TA23200_3352#1	(79)	LKRQKRKQSNRESAKRSRMKKQMECEELSAKVETLTSENMALRNEINLIA
	P.trichocarpa_719452#1	(94)	LKRQKRKQSNRESARRSRLRKQAECEELQARVQNLSSDNSLRLNELQSLS
	P.trifoliata_TA6299_37690#1	(190)	LKRQKRKQSNRESARRSRLRKQAECEELQARVESLNSNENRNLRDELQRLS
	S.lycopersicum_like_bZIP	(189)	LKRQKRKQSNRESARRSRLRKQAECEELQRRVEALSSENHSLKDELQRLS
	S.lycopersicum_TC211600#1	(255)	LKRQKRKQSNRESARRSRLRKQAECEELQRRVEALSSENHSLKDELQRLS
	S.tuberosum_TC186959#1	(219)	-----
	S.tuberosum_TC187892#1	(100)	-----
	T.pratense_TA1890_57577#1	(253)	LKRQKRKQSNRESARRSRLRKQAECEELQKRVEALGGENRTLREELQKLS
	V.vinifera_GSVIVT00014657001#1	(275)	LKRQKRKQSNRESARRSRLRKQAECEELQSKVEILSNENHVLRREELHRLA
	V.vinifera_GSVIVT00024984001#1	(268)	LKRQKRKQSNRESARRSRLRKQAECEELQAKVETLSTENTALRDELQRLS
	Consensus	(301)	LKRQKRKQSNRESARRSRLRKQAECEELQ RVE LS EN SLRDELQRLS

FIGURE 2 A (continued)

351			
	122		
	123		
	124		
	125		
	126		
	127		
	128		
	129		
	130		
	131		
	132		
	133		
	134		
	135		
	136		
	137		
	138		
	139		
	140		
	141		
	142		
	143		
	144		
	145		
	146		
	147		
	148		
	149		
	150		
	151		
	152		
	153		
	154		
	155		
	156		
	157		
	158		
	159		
	160		
	161		
	162		
	163		
	164		
	165		
	166		
	167		
	168		
	169		
	170		
	171		
	172		
	173		
	174		
	175		
	176		
	177		
	178		
	179		
	180		
	181		
	182		
	183		
	184		
	185		
	186		
	187		
	188		
	189		
	190		
	191		
	192		
	193		
	194		
	195		
	196		
	197		
	198		
	199		
	200		
	201		
	202		
	203		
	204		
	205		
	206		
	207		
	208		
	209		
	210		
	211		
	212		
	213		
	214		
	215		
	216		
	217		
	218		
	219		
	220		
	221		
	222		
	223		
	224		
	225		
	226		
	227		
	228		
	229		
	230		
	231		
	232		
	233		
	234		
	235		
	236		
	237		
	238		
	239		
	240		
	241		
	242		
	243		
	244		
	245		
	246		
	247		
	248		
	249		
	250		
	251		
	252		
	253		
	254		
	255		
	256		
	257		
	258		
	259		
	260		
	261		
	262		
	263		
	264		
	265		
	266		
	267		
	268		
	269		
	270		
	271		
	272		
	273		
	274		
	275		
	276		
	277		
	278		
	279		
	280		
	281		
	282		
	283		
	284		
	285		
	286		
	287		
	288		
	289		
	290		
	291		
	292		
	293		
	294		
	295		
	296		
	297		
	298		
	299		
	300		
	301		
	302		
	303		
	304		
	305		
	306		
	307		
	308		
	309		
	310		
	311		
	312		
	313		
	314		
	315		
	316		
	317		
	318		
	319		
	320		
	321		
	322		
	323		
	324		
	325		
	326		
	327		
	328		
	329		
	330		
	331		
	332		
	333		
	334		
	335		
	336		
	337		
	338		
	339		
	340		
	341		
	342		
	343		
	344		
	345		
	346		
	347		
	348		
	349		
	350		
	351		
	352		
	353		
	354		
	355		
	356		
	357		
	358		
	359		
	360		
	361		
	362		
	363		
	364		
	365		
	366		
	367		
	368		
	369		
	370		
	371		
	372		
	373		
	374		
	375		
	376		
	377		
	378		
	379		
	380		
	381		
	382		
	383		
	384		
	385		
	386		
	387		
	388		
	389		
	390		
	391		
	392		
	393		
	394		
	395		
	396		
	397		
	398		
	399		
	400		

FIGURE 2 A (continued)

	401	-----	446
A.lyrata_944204#1	(313)	-----	-----
A.thaliana_AT4G36730.2#1	(314)	-----	-----
Aquilegia_sp_TC21139#1	(381)	GNNLSLSE	-----
B.napus_TC63535#1	(314)	-----	-----
B.napus_TC63581#1	(314)	-----	-----
B.napus_TC89867#1	(311)	-----	-----
C.canephora_TC1061#1	(324)	-----	-----
C.clementina_TC35556#1	(284)	-----	-----
C.endivia_EL359780#1	(182)	-----	-----
C.roseus_AY027510#1	(301)	-----	-----
C.sinensis_TC9189#1	(172)	-----	-----
C.tinctorius_TA1437_4222#1	(334)	-----	-----
E.esula_TC3985#1	(243)	-----	-----
F.vesca_TA11395_57918#1	(276)	-----	-----
G.hirsutum_TC137570#1	(222)	-----	-----
G.max_Glyma11g06960.1#1	(345)	-----	-----
G.max_Glyma16g25600.2#1	(338)	-----	-----
H.annuus_BU027457#1	(186)	-----	-----
H.petiolaris_TA3844_4234#1	(249)	-----	-----
H.tuberosus_TA2966_4233#1	(286)	-----	-----
M.truncatula_AC148484_8.5#1	(341)	-----	-----
N.tabacum_BP133908#1	(119)	YLCTL	-----
N.tabacum_TC76189#1	(347)	-----	-----
P.taeda_TA23200_3352#1	(179)	QVLSKISNLSSQRDEQRESEMSDSTGKCHAVLEANVRSDRVAAG	-----
P.trichocarpa_719452#1	(191)	-----	-----
P.trifoliata_TA6299_37690#1	(284)	-----	-----
S.lycopersicum_like_bZIP	(285)	-----	-----
S.lycopersicum_TC211600#1	(351)	-----	-----
S.tuberosum_TC186959#1	(219)	-----	-----
S.tuberosum_TC187892#1	(100)	-----	-----
T.pratense_TA1890_57577#1	(344)	-----	-----
V.vinifera_GSVIVT00014657001#1	(375)	DASPLNSSSTSDQNGKFSSNGKI	-----
V.vinifera_GSVIVT00024984001#1	(365)	-----	-----
Consensus	(401)	-----	-----

FIGURE 2 A (continued)

12/43

	51				100
A.lyrata_322281#1	(40)	GPRVAMPPYYNSALAAS-GHPPPPYMWNP-QHMMSPYGAPYAAVYPHGGG			
A.thaliana_AT2G46270.2#1	(41)	GPRVAMPPYYNSAMAAS-GHPPPPYMWNP-QHMMSPYGAPYAAVYPHGGG			
B.napus_TC63534#1	(40)	GPRVAIPYYNSAMAAAAGHPPPPYMWNP-QHMMSPYGTYYAAVYPHGGG			
A.lyrata_895903#1	(41)	GPRVGIPQYYNSNVAPGH--APPPYMWASPSMMAPIGAPYPPFCPPGG-			
A.thaliana_AT4G01120.1#1	(42)	GPRVGIPQYYNSNLAPGH--APPPYMWASPSMMAPIGAPYPPFCPPGG-			
B.napus_TC63500#1	(45)	GPRVAITPQYYNSNGHAP--PPPPYIWGSPSMMAPYGTYPFFCPPGG-			
C.canephora_TC5243#1	(1)	-----			
C.roseus_AF084971#1	(39)	GPRVAVPPYFSSAVASG--HPPHPYMWGPPQPMMPYGTYYAAIYAHGG-			
N.tabacum_NP917548#1	(39)	GPRVAVPPYVNSPVAFG--QAPHPCMWGLQPMMPPIGIPYAGIYAHGG-			
S.tuberosum_TC185019#1	(1)	-----			
N.tabacum_TC40642#1	(1)	-----			
S.lycopersicum_TC213303#1	(1)	-----			
C.maculosa_EH744483#1	(1)	-----			
C.maculosa_TA2192_215693#1	(41)	GPRMAMPYFNSAVASG--HAPPYMWGPPQHMP----PYAAMYPHGG-			
H.annuus_TC42219#1	(24)	GPRMAMPYFSSPVASG--HAPPYMWGPMHPMMP----PYAAIYPHGG-			
H.tuberosum_TA2407_4233#1	(36)	GPRMAMPYFNSPVASG--HAPPYMWGPMHPMMP----PYAAIYPHGG-			
C.sinensis_TC24527#1	(40)	GPRVAIPYYNSPIASG--HAPQPYMWGPAQPMMPYGYAPYAA-IYSTGG			
G.hirsutum_TC148476#1	(6)	GPHVNMPPYYSSAVASG--HAPPYMWGTPQPMMPYGYAPYAA-IYSHGG			
G.max_Glyma03g41590.4#1	(1)	-----MPPYYNSAVASG--HAPHPYMWGPPQPMMPYGYAPYAA-IYPHGG			
G.max_Glyma19g44190.1#1	(46)	GPRVTMPYYNSAVASG--HAPHPYMWGPPQPMMPYGYAPYAA-IYPHGG			
M.truncatula_AC137602_8.5#1	(43)	GPRVAMPPYYNSPVASG--HTHPYMWGPPQPMMPYGYAPYAA-MYPHGG			
G.max_Glyma07g06620.1#1	(44)	-PRVNIIPPYFNSAVASG--HAPHPYMWGPPQPMMPYGYAPYAA-FYSHGG			
G.max_Glyma16g03190.1#1	(44)	-PRVNIIPPYFNSAVASG--HAPHPYMWGPPQPMMPYGYAPYAA-FYSHGG			
P.vulgaris_TC11785#1	(43)	-PRVNIIPPYFNSAVASG--HAPHPYMWGPPQPMMPYGYAPYAA-FYSPGG			
Similar_to_legume_bZIP	(1)	-----MWGPPQPMMPYGYAPYAPFFYSHGG			
P.trichocarpa_bZIP_like	(39)	GPRVALPPYYNSAVASG--HAPHPYMWGLPQPMMPYGYAPYAT-VYSHG-			
Consensus	(51)	GPRVAIPYYNSAVASG HAP PYMWGPPQPMMPYGYAPYAA G			

FIGURE 2 B (continued)

13/43

	101		150
A.lyrata_322281#1	(88)	VYAHPIPMGSQPQKTPPLATP-----GTHLSIDTPTTKSTGNTDNGLM	
A.thaliana_AT2G46270.2#1	(89)	VYAHPIPMGSLPQKDPPLTTP-----GTLISIDTPTTKSTGNTDNGLM	
B.napus_TC63534#1	(89)	VYAHPGFMP--QSQKGAALSTP-----GTPLNIDTPSKSTGNTENGLM	
A.lyrata_895903#1	(88)	VYAHGVQMGSQLQGVVSQATPG-----VTPLTMDA-PTNSAGNSDHGFM	
A.thaliana_AT4G01120.1#1	(89)	VYAHGVQMGSPQGPVVSQASG-----VTPLTIDA-PANSAGNSDHGFM	
B.napus_TC63500#1	(92)	VYAHPALQMGSPQGPASQATPV-----VATPLNLEAHPANSSGNTDQGF	
C.canephora_TC5243#1	(1)	-----	
C.roseus_AF084971#1	(86)	VYTHPGVPLGSHANAHAGATSPGATEAIAASPLSIDTPTKSSANGSQGLM	
N.tabacum_NP917548#1	(86)	VYAHPGVPIVSRPQAHVMTSSPAVSQTMDAASLSMDPSAKTSGDTNQG	
S.tuberosum_TC185019#1	(1)	-----	
N.tabacum_TC40642#1	(1)	-----	
S.lycopersicum_TC213303#1	(1)	-----	
C.maculosa_EH744483#1	(1)	-----	
C.maculosa_TA2192_215693#1	(84)	VYHPGVPLAGSP-----MSIDSPAKSSGNSDRGLL	
H.annuus_TC42219#1	(67)	VYAHPGVTVAGNPNTL-----VPIDSCAKSSGNPDRGLI	
H.tuberosus_TA2407_4233#1	(79)	VYAHPGVTVADNPNPAL-----IPIDSSAKSSGNPDRGLI	
C.sinensis_TC24527#1	(87)	VYAHPAVPL-----AVTPLNTEAPTCKSSGNADRGLA	
G.hirsutum_TC148476#1	(53)	VYAHPAVPLAS--HSLGVFSSP-----AAAGPVETPTKSPGNTDQGLM	
G.max_Glyma03g41590.4#1	(43)	VYTHPAVPIGPIGLTHSQGVSPSPA-----AGTPLSIEETPPKSSGNTDQGLM	
G.max_Glyma19g44190.1#1	(93)	VYTHPAVPIGPHTHSQGVSPSPA-----AGTPLSIEETPPKSSGNTDQGLM	
M.truncatula_AC137602_8.5#1	(90)	VYTHPAVPIGPHPHSQGISSPA-----TGTPLSIEETPPKSSGNTDQGLM	
G.max_Glyma07g06620.1#1	(90)	VYTHPAVAIGPHLHGQGVSSSPA-----VGTSSIESPTKLSGNTDQGLM	
G.max_Glyma16g03190.1#1	(90)	VYTHPAVAIGPHSHGQGVSPSPA-----AGTPSSVESPTKFSGNTNQG	
P.vulgaris_TC11785#1	(89)	VYTHPAVAIGPHSHGQGVSPSPA-----AGTPSSVDSPTKLSGNTDQGLM	
Similar_to_legume_bZIP	(26)	VYTHPAVAIGSNSNGQGISSPA-----AGTPTSIEETPTKSSGNTDQGLM	
P.trichocarpa_bZIP_like	(85)	VYAHPAVPIVSHPHGPGIVSSPA-----AGTLLSAETPTKSSGNTDQGLV	
Consensus	(101)	VY HPAV IG S P GT LSIDTP KSSGNTDQGLM	

FIGURE 2 B (continued)

14/43

	151			200
A.lyrata_322281#1	(133)	KKLKEFDGLAMSLGNNGPE--NGADEHKRNRNS	-----SET	
A.thaliana_AT2G46270.2#1	(134)	KKLKEFDGLAMSLGNNGPE--NGADEHKRNRNS	-----SET	
B.napus_TC63534#1	(131)	KKLKEFDGLAMSLGNNG-----NNGDEGKRNRNS	-----SET	
A.lyrata_895903#1	(133)	KKLKEFDGLAMSI SNNKVGSAEHSSEHRSSQRYIESNVVLI SIAPRMM	-----SS--EN	
A.thaliana_AT4G01120.1#1	(134)	KKLKEFDGLAMSI SNNKVGSAEHSSEHRSSQ	-----SS--EN	
B.napus_TC63500#1	(138)	KKLKEFDGLAMSI SNNKSGSGEHS-SEPKNSQ	-----SS--EN	
C.canephora_TC5243#1	(1)	-----	-----	
C.roseus_AF084971#1	(136)	NKLRGFDGLAMSI GNNGTDSADG-GTDHGISQS	-----GDT	
N.tabacum_NP917548#1	(136)	SKLKGSDGLGMSI GNCSVDNGDGDTHGSPQSDS	-----GQT	
S.tuberosum_TC185019#1	(1)	-----	-----	
N.tabacum_TC40642#1	(1)	-----	-----	
S.lycopersicum_TC213303#1	(1)	-----	-----	
C.maculosa_EH744483#1	(1)	-----	-----	
C.maculosa_TA2192_215693#1	(115)	KKLKGFDGLAMSI GNNGDGGG-GNENGISHS	-----GET	
H.annuus_TC42219#1	(102)	KKLKGIDELTMSI GNNN-----G-----ISNS	-----RDT	
H.tuberosum_TA2407_4233#1	(114)	KKLKGIDELTMSI GNNN-----G-----ISNS	-----RDT	
C.sinensis_TC24527#1	(118)	KKLKGIDGLAMSI GNASAE SAEGGAE-QRPSQ	-----SEA	
G.hirsutum_TC148476#1	(94)	KKLKGFDGLAISI GNGTAE NAEGRAK-PRPSHS	-----VET	
G.max_Glyma03g41590.4#1	(88)	KKLKEFDGLAMSI GNHGAESAERGGE-NRLSQS	-----VDT	
G.max_Glyma19g44190.1#1	(138)	KKLKEFDGLAMSI GNHGAESAEPGGE-NRLSES	-----VDT	
M.truncatula_AC137602_8.5#1	(135)	KKLKGFDGLAMSI GNHGAESAEPGAE-SRQSQS	-----VNT	
G.max_Glyma07g06620.1#1	(135)	KKSKGFDGLAMSI GNCAESAEGAE-NRQSQS	-----VDT	
G.max_Glyma16g03190.1#1	(135)	KKLKGFDGLAMSI GNCAESAERGAE-NRLSQS	-----VDT	
P.vulgaris_TC11785#1	(134)	KKLKGFDGLAMSI GNCAESAELGAE-NRLSQS	-----VDT	
Similar_to_legume_bZIP	(71)	KKLKGFDGLAMSI GNCAESAERGAE-NRLSRS	-----VDT	
P.trichocarpa_bZIP_like	(130)	NKLKGFDGLAMSI GNCAETVEGG-----RLSQS	-----VEI	
Consensus	(151)	KKLKGFDGLAMSI GN ES AE A R SQS	DT	

FIGURE 2 B (continued)

15/43

	201			
			250	
A.lyrata_322281#1	(167)	DGSTDGS	DGNTTGADEPKL	KRSREGTPTKDV
A.thaliana_AT2G46270.2#1	(168)	DGSTDGS	DGNTTGADEPKL	KRSREGTPTKDGK
B.napus_TC63534#1	(162)	DSSDGS	DGNTTGADEPKL	KRRREGTPTKDEE
A.lyrata_895903#1	(183)	LAMVVMY	SSFLPQGEQ	SRKIRRERSPTGE
A.thaliana_AT4G01120.1#1	(170)	DSSNGS	DGNTTGGEQ	SRKRRQRSPSTGE
B.napus_TC63500#1	(173)	DDSSNGS	DGNTTGGEQ	SRKRSREGSPNNDG
C.canephora_TC5243#1	(1)	-----	-----	-----
C.roseus_AF084971#1	(171)	EGSSDGS	NGTTSKAGQ	KNKKRSREGTPAND
N.tabacum_NP917548#1	(172)	EGSSDGS	NIHTAEVGE	KSKRSRETPNTSGD
S.tuberosum_TC185019#1	(1)	-----	-----	-----
N.tabacum_TC40642#1	(1)	-----	-----	-----
S.lycopersicum_TC213303#1	(1)	-----	-----	-----
C.maculosa_EH744483#1	(1)	-----	-----	-----
C.maculosa_TA2192_215693#1	(150)	EGSEGS	DGNTTEGGQ	NSGKRSREGSPK
H.annuus_TC42219#1	(127)	EGSEGS	DGNTS--CK	NGHKRIRGGSS
H.tuberosum_TA2407_4233#1	(139)	EGSEGS	DGNTS--CK	NEHKRIRGGSS
C.sinensis_TC24527#1	(152)	DGSTDGS	DGNTVRAGQ	SRKRSREGTPIAG
G.hirsutum_TC148476#1	(129)	AGSADGS	DGNTTGTDQ	SRKRSREGTPIAG
G.max_Glyma03g41590.4#1	(123)	EGSSDGS	DGNTSGANQ	SRKRSREGTPTDG
G.max_Glyma19g44190.1#1	(173)	EGSSDGS	DGNTSGANQ	TRRKRRSREGTPTDG
M.truncatula_AC137602_8.5#1	(170)	EGSSDGS	DGNTSGANQ	TRRKRRSREGTPTDG
G.max_Glyma07g06620.1#1	(170)	EGYSDGS	DGNTAGANQ	TKRRCREGTLLTDG
G.max_Glyma16g03190.1#1	(170)	EGSSDGS	DGNTAGANQ	TKRKRRSREGTPTDA
P.vulgaris_TC11785#1	(169)	EGSSDGS	DGNTAGANQ	TKMKRSREETS
Similar_to_legume_bZIP	(106)	EGSSDGS	DGNTTGTNG	TR-KRSRDG
P.trichocarpa_bZIP_like	(163)	EVSSDGI	DGNTTRG--	KK--RSREGT
Consensus	(201)	EGSSDGS	DGNTTGA Q	R KRSREGTPT

FIGURE 2 B (continued)

	251				300
A.lyrata_322281#1	(216)	TGVKLIQ	-----GSAILSPGVSAN	-----SNP	-----F
A.thaliana_AT2G46270.2#1	(217)	TGVKLIQ	-----GSGAILS	-----	-----
B.napus_TC63534#1	(212)	NVVKHSVQ	-----GGGAIVSAAGVS	-----ANSNP	-----T
A.lyrata_895903#1	(227)	NEKA	-----DVTMGTPVMP	-----TMMGFQN	-----
A.thaliana_AT4G01120.1#1	(214)	NEKP	-----DVTMGTPVMP	-----TAMSFQN	-----
B.napus_TC63500#1	(217)	SEKQ	-----AVTMGTPVMP	-----TVLDFP	-----
C.canephora_TC5243#1	(1)	-----	-----MGSVLSPNMTS	-----TLELRNPSGG	-----NMKTSP
C.roseus_AF084971#1	(221)	SEKAMRASKVPAAA	-----TEKVMGAVLS	-----PNMTT	-----ASELNRNPSAA
N.tabacum_NP917548#1	(221)	TKKETSI	-----AFNPGNIAEKVV	-----GTVFSP	-----TMTT
S.tuberosum_TC185019#1	(1)	-----	-----	-----	-----MTT
N.tabacum_TC40642#1	(1)	-----	-----	-----	-----TFLVKSSP
S.lycopersicum_TC213303#1	(1)	-----	-----MTGTGLSPCMTT	-----	-----LEMRNPASAH
C.maculosa_EH744483#1	(1)	-----	-----	-----	-----MKSSPTNG
C.maculosa_TA2192_215693#1	(200)	SKKVTGLTVTLP	-----KVSGKLGAAV	-----SANLTS	-----DLEIKNSPT
H.annuus_TC42219#1	(168)	SKKVTGVKILG	-----	-----KEVGAVISGDSGT	-----ELELKKSP
H.tuberosus_TA2407_4233#1	(180)	TKKVKGVKILG	-----	-----KEVGAVISGDSGT	-----ELELKKSP
C.sinensis_TC24527#1	(184)	-----	-----	-----VGPVLSPGMPT	-----KLELRNAPGM
G.hirsutum_TC148476#1	(169)	NQVAAGEVTATI	-----	-----SPKLI	-----GTVVSPGM
G.max_Glyma03g41590.4#1	(171)	SNKMLGVVPASV	-----	-----AGTIVGHVVS	-----SGMTT
G.max_Glyma19g44190.1#1	(221)	SNKMLAVVTAGV	-----	-----AGTIVGPVVS	-----SGMTT
M.truncatula_AC137602_8.5#1	(218)	SDKMMAVAPAGV	-----	-----TGQLVGPV	-----ASSAMTT
G.max_Glyma07g06620.1#1	(218)	SKKIVSATPASV	-----	-----AGTLVGPVVS	-----VVMAT
G.max_Glyma16g03190.1#1	(218)	SKKIVSATPASV	-----	-----AGTLVGPVVS	-----SGMAT
P.vulgaris_TC11785#1	(217)	SKMVM	-----SATPASV	-----	-----AGKLVGPV
Similar_to_legume_bZIP	(154)	SKKTVSVITSSA	-----	-----AENMVG	-----PVLSSG
P.trichocarpa_bZIP_like	(209)	S	-----	-----TDNVLRAA	-----VAVPG
Consensus	(251)	S	-----	-----K	-----
			-----	-----VG	-----VVS
			-----	-----M	-----T
			-----	-----L	-----LELRN

FIGURE 2 B (continued)

	350		
	301		
A.lyrata_322281#1	(239)	MSQSLAMVPPETWPNERELKRERRKQSNRESARRSRLRKQAEETEELARK	
A.thaliana_AT2G46270.2#1	(231)	-----PG-NERELKRERRKQSNRESARRSRLRKQAEETEELARK	
B.napus_TC63534#1	(240)	MSQSLAMVPPETWLNERELKRERRKQSNRESARRSRLRKQAEETEELARK	
A.lyrata_895903#1	(248)	----SAGMNGVQPWNEKEVKREKRKQSNRESARRSRLRKQAETEQLSVK	
A.thaliana_AT4G01120.1#1	(235)	----SAGMNGVQPWNEKEVKREKRKQSNRESARRSRLRKQAETEQLSVK	
B.napus_TC63500#1	(237)	----QFFPGAPHEVWNEKEVKREKRKQSNRESARRSRLRKQAEETEELSVK	
C.canephora_TC5243#1	(28)	-----VSEAWLNERELKRERRKQSNRESARRSRLRKQAEETEELAKK	
C.roseus_AF084971#1	(268)	VSQSCSSLPGETWLNERELKRERRKQSNRESARRSRLRKQAEETEELAKK	
N.tabacum_NP917548#1	(268)	VSRI SPAVPGEAWLNEREMKREKRKQSNRESARRSRLRKQGEAEELAIR	
S.tuberosum_TC185019#1	(12)	TSRI SPAVPGEVWLNERELKREKRKQSNRESARRSRLRKQAEAEELAVQ	
N.tabacum_TC40642#1	(1)	MFHTQPALPNEAWLNERELKREKRKQSNRESARRSRLRKQAEAEELAIR	
S.lycopersicum_TC213303#1	(31)	GSPLSPALPNETWLNERELKREKRKQSNRESARRSRLRKQAEAEELAIR	
C.maculosa_EH744483#1	(1)	-----MVPGESLLQNERELKRERRKQSNRESARRSRLRKQAEAEELAIK	
C.maculosa_TA2192_215693#1	(242)	ASATVAMVPGESLLQNERELKRERRKQSNRESARRSRLRKQAEAEELAIK	
H.annuus_TC42219#1	(200)	ASANMAMVPSLVLQNERELKREKRKQSNRESARRSRLRKQAEAEELGTR	
H.tuberosum_TA2407_4233#1	(212)	ASPNMSPVNSLVLQNERELKREKRKQSNRESARRSRLRKQAEAEELGTR	
C.sinensis_TC24527#1	(213)	VPQPCAVLPPETWIQNERELKRERRKQSNRESARRSRLRKQAEAEELSRK	
G.hirsutum_TC148476#1	(214)	MGVHCGVMPTEVWLQSERELKRERRKQSNRESARRSRLRKQAEETEELARK	
G.max_Glyma03g41590.4#1	(213)	APQPCPVLPAEAMVQNERELKRERRKQSNRESARRSRLRKQAEETEELARK	
G.max_Glyma19g44190.1#1	(263)	APQPCPVLPAETWLNQNERELKRERRKQSNRESARRSRLRKQAEETEELARK	
M.truncatula_AC137602_8.5#1	(263)	TPQPSAVLPPPEAWIQNERELKRERRKQSNRESARRSRLRKQAEAEELARK	
G.max_Glyma07g06620.1#1	(263)	APQPCAIVPNETCLQNERELKRERRKQSNRESARRSRLRKQAEETEELARK	
G.max_Glyma16g03190.1#1	(263)	AAQPCAVVRNETWLNQNERELKRERRKQSNRESARRSRLRKQAEETEELARK	
P.vulgaris_TC11785#1	(262)	APQPCAAVPPPEAWLNQNERELKRERRKQSNRESARRSRLRKQAEETEELARK	
Similar_to_legume_bZIP	(194)	APQPCGVLPPEAWMQNERELKRERRKQSNRESARRSRLRKQAEAEELARR	
P.trichocarpa_bZIP_like	(244)	IPQSGVVLPEAWLNQNELELQNERELKREKRKQSNRESARRSRLRKQAEAEELAHK	
Consensus	(301)	AVVP E WLQNERELKRERRKQSNRESARRSRLRKQAEETEELARK	

FIGURE 2 B (continued)

18/43

	351	400
A.lyrata_322281#1	(289)	VEALTAENMALRSELNQLNEKSDKLRGANATLLDKLKCSEPEKRV-----
A.thaliana_AT2G46270.2#1	(268)	VEALTAENMALRSELNQLNEKSDKLRGANATLLDKLKCSEPEKRV-----
B.napus_TC63534#1	(290)	VEALTAENMALRSELNQLNEKSNLIRGANATLLDKLKSSEPEK-----
A.lyrata_895903#1	(294)	VDALVAENMTLRSKLGLKNESEKLRLENEALLHQLKAQATG-----
A.thaliana_AT4G01120.1#1	(281)	VDALVAENMSLRSKLGLQNNSEKLRLENEAILDQLKAQATG-----
B.napus_TC63500#1	(283)	VDALVAENMTLRSKLGLQNDSEKLRLENEALLDQLKAQATG-----
C.canephora_TC5243#1	(70)	VQSLTAENLSLKSEIHKLTSESSERLKLLENATMMEKLNKLPQLGQTGNLSLS
C.roseus_AF084971#1	(318)	VQTLTAENMTLRSEINKLTENSE-HLRHESALLDKLKNARVMQAG--EMN
N.tabacum_NP917548#1	(318)	VQSLTSENGLKSEINNFTENSAKLKLENSALMERLQNKQRGQAEFEVTLG
S.tuberosum_TC185019#1	(62)	VQSLTSENALRLLEINKFTENSEKLVENAALMERLKNKQ-GQAKEVTLG
N.tabacum_TC40642#1	(51)	VQSLTGENMTLKSEINKLMENSEKLVENAALMEKLNNEQLSPTEEVSLG
S.lycopersicum_TC213303#1	(81)	VQALTGENLTLRSEINKLMDNSEKLVENATLMEKLNNEQLGQTEEVSLG
C.maculosa_EH744483#1	(45)	VESLTNENLTLKSEINRLTDNSEKLVENAKLQNAKLEKLNARQDTE-----
C.maculosa_TA2192_215693#1	(292)	VESLTNENLTLKSEINRLSDNSEKLVENAKLQNAKLEKLNARQDTEH-----
H.annuus_TC42219#1	(250)	VEALTSENLKLKSEINQLTVNATNLKLNQNAKLEKLNARQDTE-----
H.tuberosum_TA2407_4233#1	(262)	VEALTSENLKLKSEINQLTVNATNLKLNQNAKLEKLNARQDTE-----
C.sinensis_TC24527#1	(263)	VDSLIDENASLSEINQLTSENSEKLRQENAALLEKLSAQLGNKQEIIVLN
G.hirsutum_TC148476#1	(264)	VESLTSENAALRSEINQLTSEMSEKVRLENAILVEELKNAQLGHAQENILN
G.max_Glyma03g41590.4#1	(263)	VESLNAENATLKSEINRLTESSEKMRVENATLRGKLNQNAQLGQTEITLK
G.max_Glyma19g44190.1#1	(313)	VESLNAENATLKSEINRLTESSEKMRVENATLRGKLNQNAQLRQTEITLN
M.truncatula_AC137602_8.5#1	(313)	VESLNAESASLRSEINRLAENSERLRMENAAALKEFKIAKLGQPKIILT
G.max_Glyma07g06620.1#1	(313)	VDMLTAENVSLKSEIQLTGEGSEQMRMENSALREKLRNTQLGQREEIILS
G.max_Glyma16g03190.1#1	(313)	VEMLTAENVSLKSEITRLTEGSEQMRMENSALREKLRNTQLGPREEITLS
P.vulgaris_TC11785#1	(312)	VEMLTAENVSLKSEITQLTEGSEQMRMENSALREKLRNTQLGQREEIILD
Similar_to_legume_bZIP	(244)	VDALTAENLALKSEMNELAENSAKLVENATLKEKLENTQLGQTEEEIILN
P.trichocarpa_bZIP_like	(294)	VEVLTENMALQSEISQFTEKSEKLRLENAALTEKLNQNAQLGHAQEMILN
Consensus	(351)	VESLTAENLTLKSEINQLTSEKLRLENAALLEKLNQNAQLG EI L

FIGURE 2 B (continued)

	450		
	401		
A.lyrata_322281#1	(334)	-----SGKMLSRVKNSG-AGDKNKNQG-----DNDSKSTSKLYQL	
A.thaliana_AT2G46270.2#1	(313)	-----PANMLSRVKNSG-AGDKNKNQG-----DNDSNSTSKLHQL	
B.napus_TC63534#1	(333)	-----RVKSSGNGDDKKNKKG-----DNETNSTSKLHQL	
A.lyrata_895903#1	(336)	-----KTENLISRVDKNN-----VSGS-----KNVQHQL	
A.thaliana_AT4G01120.1#1	(323)	-----KTENLISRVDKNN-----VSGS-----KTVQHQL	
B.napus_TC63500#1	(325)	-----KTENLISGVDKNN-----VSGTSSSKNAEQQL	
C.canephora_TC5243#1	(120)	KFDEMRLOPVGTANLLARVNS-GSVDRNDEEG--EVFENTKSGAKLRQL	
C.roseus_AF084971#1	(365)	KYDELHRQPTGTADLLARVNS-GSTDKSNEEGGDFENRNSGTKLHQL	
N.tabacum_NP917548#1	(368)	KIGDKRLQPVSTADLLARVNS-GPLDRTNKDD--EIHENNTSGAKLHQL	
S.tuberosum_TC185019#1	(111)	MIDDKRLKPVSTADLLARVNNNGSFNRTNEDG--EVHDS-TSGAKLRQL	
N.tabacum_TC40642#1	(101)	KIDDKRVQPVGTANLLARVNS-GSLNRANEE--EYENNSSGAKLHQL	
S.lycopersicum_TC213303#1	(131)	KIDDKRLQPVGTVNLARVNS-GSSDTTNEGD--EYENNSSGAKLHQL	
C.maculosa_EH744483#1	(90)	PRLDPNGSSLSTANLLSRVNGSGART-DGDAEVYENNNNQNSGAKLRQL	
C.maculosa_TA2192_215693#1	(337)	PRLDPNGSSLSTANLLSRVDNGSGART-DGDVEVYENNNNQNSGAKLRQL	
H.annuus_TC42219#1	(292)	PRADKKGSSLSTANLLSRVDNRSG-----VVGEATSSGSGAPLYQL	
H.tuberosum_TA2407_4233#1	(304)	PRADKKGSSLSTANLLSRVDNRSG-----VVGDATSSGSGAPLHQL	
C.sinensis_TC24527#1	(313)	EDK--RVTPVSTENLLSRVTPVLLIETWRKXHPV-----	
G.hirsutum_TC148476#1	(314)	KKE-----DKE-----G--EMGEKRSDSGAKLHQL	
G.max_Glyma03g41590.4#1	(313)	IIDSQRATPVSTENLLSRVNN-SGSNDRTVEDENGFCENKPNKGAKLHQL	
G.max_Glyma19g44190.1#1	(363)	IIDSQRATPISSTENLLSRVNNSGSNDRTVEDENGFCENKPNKGAKLHQL	
M.truncatula_AC137602_8.5#1	(363)	NIDSQRTTPVSTENLLSRVNNSGSNDRTVEDENGYCDNKNPNSGAKLHQL	
G.max_Glyma07g06620.1#1	(363)	SIESKRAAPVSTENLLSRVNN-SSSNDRTTENENDFCENKPNKGAKLHQL	
G.max_Glyma16g03190.1#1	(363)	SIDSKRAAPVSTENLLSRVNN-SGANDRTAENENDICENKPNKGAKLHQL	
P.vulgaris_TC11785#1	(362)	SIDSKRSTPVSTENLLSRVNN-SSSNDRTSAENESDFCENKPNKGAKLHQL	
Similar_to_legume_bZIP	(294)	GMD-KRATPVSTENLLSRVND-SNSDDRAAEEENGFCENKPNKGAKLRQL	
P.trichocarpa_bZIP_like	(344)	IIDEH-RAPAVSTENLLSRVNN--S---AFEEEESDLYERNNSGAKLHQL	
Consensus	(401)	D R PVSTENLLSRVNN GS DR E NSGAKLHQL	

FIGURE 2 B (continued)

	451	466
A.lyrata_322281#1	(368)	LDTKPRANAVAAG---
A.thaliana_AT2G46270.2#1	(347)	LDTKPRAKAVAAG---
B.napus_TC63534#1	(362)	LDTKPRADGVAAR---
A.lyrata_895903#1	(361)	LNASPIITDPVAAS---
A.thaliana_AT4G01120.1#1	(348)	LNASPIITDPVAAS---
B.napus_TC63500#1	(355)	LNVSLRTDSVAAS---
C.canephora_TC5243#1	(167)	LDANPRTDAVAAG---
C.roseus_AF084971#1	(414)	LDASPRADAVAAG---
N.tabacum_NP917548#1	(415)	LDASHRTDAVAAR---
S.tuberosum_TC185019#1	(158)	LDASPRTDHAVAAR---
N.tabacum_TC40642#1	(148)	LDSSPRTDAVAAG---
S.lycopersicum_TC213303#1	(178)	LDTSVRTDAVAAG---
C.maculosa_EH744483#1	(139)	LDASPRTDAVAAG---
C.maculosa_TA2192_215693#1	(386)	LHASPRTDAVAAG---
H.annuus_TC42219#1	(334)	LDASPRADGVAAGAG-
H.tuberosum_TA2407_4233#1	(346)	LDASPRADAVAAGAG-
C.sinensis_TC24527#1	(347)	-----
G.hirsutum_TC148476#1	(337)	LDPSPRDDAVAAG---
G.max_Glyma03g41590.4#1	(362)	LDTSPRADAVAAG---
G.max_Glyma19g44190.1#1	(413)	LDTSPRADAVAAG---
M.truncatula_AC137602_8.5#1	(413)	LDASPRADAVAAG---
G.max_Glyma07g06620.1#1	(412)	LDTNPRADAVAAG---
G.max_Glyma16g03190.1#1	(412)	LDTNPRANAVAAG---
P.vulgaris_TC11785#1	(411)	LDTNPRADAVAAG---
Similar_to_legume_bZIP	(342)	LDTNPRANAVAAS---
P.trichocarpa_bZIP_like	(387)	LDASPRADAVAAG---
Consensus	(451)	LDASPRADAVAAG

FIGURE 2 B (continued)

21/43

Clade D	48	49	50	51	52	53	54	55	56	57	58
48. <i>P.taeda</i> TA23200_3352#1		23.9	17.9	15.2	21.7	12.0	23.1	13.8	14.8	22.0	24.3
49. <i>H.annuus</i> BU027457#1	40.8		19.7	33.7	36.5	22.0	47.2	14.0	14.1	34.0	36.3
50. <i>N.tabacum</i> BP133908#1	28.7	31.9		17.5	16.2	13.8	16.8	10.2	11.0	15.1	19.1
51. <i>H.petiolaris</i> TA3844_4234#1	31.5	42.7	24.2		41.4	19.4	38.3	18.5	35.7	38.9	33.7
52. <i>C.tinctorius</i> TA1437_4222#1	34.5	44.7	19.2	52.3		41.3	65.8	16.5	34.2	65.8	54.3
53. <i>C.endivia</i> EL359780#1	25.6	39.5	25.4	29.0	45.0		39.9	20.4	34.3	32.9	30.8
54. <i>H.tuberosus</i> TA2966_4233#1	37.2	53.0	21.1	48.8	74.2	48.8		16.9	28.3	59.5	55.1
55. <i>S.tuberosum</i> TC187892#1	17.5	25.4	22.8	23.4	20.1	29.3	20.7		42.2	20.4	17.9
56. <i>S.tuberosum</i> TC186959#1	32.3	26.6	20.6	45.2	44.7	49.5	41.8	42.7		40.3	31.9
57. <i>C.canephora</i> TC1061#1	33.7	42.4	20.7	50.5	76.9	40.2	69.7	22.0	49.2		58.4
58. <i>S.lycopersicum</i> likebZIP	40.5	46.5	22.9	48.6	68.2	41.5	71.2	21.8	45.8	72.1	
59. <i>S.lycopersicum</i> TC211600#1	32.9	37.7	20.3	53.1	71.1	33.7	56.3	22.3	51.7	74.6	80.0
60. <i>N.tabacum</i> TC76189#1	30.1	37.9	19.1	54.0	72.0	35.0	60.4	23.7	53.2	75.1	72.8
61. <i>C.roseus</i> AY027510#1	38.3	45.7	22.0	40.7	71.5	41.3	72.7	20.0	43.0	82.4	75.3
62. <i>C.sinensis</i> TC9189#1	22.9	30.8	19.3	42.7	31.8	38.7	24.6	38.6	58.7	33.1	27.8
63. <i>P.trifoliata</i> TA6299_37690#1	40.3	48.8	22.3	44.2	68.5	42.8	74.7	18.7	38.9	70.3	77.5
64. <i>E.esula</i> TC3985#1	22.0	25.4	18.6	37.9	32.4	26.5	22.8	37.7	57.3	32.5	26.4
65. <i>G.hirsutum</i> TC137570#1	26.0	26.7	19.5	46.4	40.8	44.3	38.2	29.4	67.4	43.0	38.0
66. <i>V.vinifera</i> GSVIVT00024984001#1	31.6	40.4	19.0	50.5	73.4	34.9	60.7	20.3	47.3	74.2	62.4
67. <i>F.vesca</i> TA11395_57918#1	37.1	48.4	22.2	47.6	65.2	42.9	73.0	21.5	44.4	67.5	75.0
68. <i>T.pratense</i> TA1890_57577#1	26.2	39.5	19.0	53.6	68.2	35.5	61.1	20.8	48.2	72.0	62.3
69. <i>M.truncatula</i> AC148484_8.5#1	27.9	39.4	18.8	52.4	70.6	33.5	60.3	20.9	49.1	71.5	60.9
70. <i>G.max</i> Glyma16g25600.2#1	27.3	40.1	18.1	54.9	70.3	35.0	60.2	20.5	50.7	75.1	62.9
71. <i>G.max</i> Glyma11g06960.1#1	27.6	40.7	18.0	51.2	67.4	32.0	58.1	18.9	46.2	67.7	58.4
72. <i>V.vinifera</i> GSVIVT00014657001#1	31.2	32.4	17.3	46.0	60.1	28.4	50.8	16.8	41.0	62.3	53.0
73. <i>C.clementina</i> TC35556#1	35.3	46.6	21.6	45.6	57.7	34.3	62.8	15.9	24.0	62.8	65.5
74. <i>P.trichocarpa</i> 719452#1	48.0	58.9	30.5	44.0	44.1	30.0	50.9	21.6	31.7	47.7	51.4
75. <i>B.napus</i> TC89867#1	33.5	38.4	18.7	54.8	63.7	32.6	56.5	19.0	48.1	65.6	58.4
76. <i>B.napus</i> TC63535#1	35.5	37.4	18.8	55.0	64.9	33.2	56.2	19.2	48.2	65.0	58.5
77. <i>B.napus</i> TC63581#1	34.2	37.7	19.2	55.0	62.2	32.6	56.5	18.5	46.6	66.9	57.8
78. <i>A.thaliana</i> AT4G36730.2#1	36.1	39.3	19.5	53.0	63.4	33.2	56.5	18.5	46.0	65.9	58.5
79. <i>A.lyrata</i> 944204#1	35.9	39.1	18.9	52.9	63.4	33.7	56.7	18.6	45.8	65.3	57.7
80. <i>Aquilegia</i> sp TC21139#1	29.2	32.6	17.3	46.0	58.1	26.6	50.9	15.5	40.6	62.5	51.2

FIGURE 3 A

22/43

Clade D	59	60	61	62	63	64	65	66	67	68	69
48. <i>P.taeda</i> _TA23200_3352#1	20.8	19.8	25.0	12.6	25.3	13.5	13.9	21.4	24.8	16.0	17.4
49. <i>H.annuus</i> _BU027457#1	29.3	31.2	35.9	17.6	38.9	16.4	13.7	33.4	39.1	29.6	30.4
50. <i>N.tabacum</i> _BP133908#1	17.0	16.1	17.9	11.0	17.3	11.8	11.7	14.5	16.9	13.7	13.4
51. <i>H.petiolaris</i> _TA3844_4234#1	45.0	46.8	32.2	31.1	34.4	30.3	34.1	42.3	35.3	42.6	43.7
52. <i>C.tinctorius</i> _TA1437_4222#1	54.8	56.6	61.7	25.2	58.8	25.3	31.5	61.4	55.8	52.9	53.8
53. <i>C.endivia</i> _EL359780#1	23.8	26.0	34.3	23.7	35.5	18.1	30.3	29.1	35.2	28.0	27.0
54. <i>H.tuberosus</i> _TA2966_4233#1	43.3	47.8	64.0	14.4	63.4	15.7	24.1	51.4	60.7	48.5	47.9
55. <i>S.tuberosum</i> _TC187892#1	20.0	21.4	17.2	32.2	15.3	33.5	24.2	17.3	16.5	16.8	15.0
56. <i>S.tuberosum</i> _TC186959#1	46.2	47.3	32.6	50.2	26.5	50.4	54.1	40.2	29.0	40.4	39.9
57. <i>C.canephora</i> _TC1061#1	60.0	63.1	73.7	27.0	62.1	26.5	33.3	63.7	58.3	58.0	57.2
58. <i>S.lycopersicum</i> likebZIP	79.8	66.7	62.9	19.0	65.3	18.3	25.1	52.5	61.7	47.2	46.1
59. <i>S.lycopersicum</i> _TC211600#1		82.6	53.7	33.0	52.1	33.8	38.5	65.3	49.7	56.8	58.1
60. <i>N.tabacum</i> _TC76189#1	88.3		55.3	32.8	50.1	32.5	38.4	64.8	49.0	57.5	59.0
61. <i>C.roseus</i> _AY027510#1	65.7	67.9		19.9	65.5	20.3	26.7	55.9	61.4	50.3	49.3
62. <i>C.sinensis</i> _TC9189#1	37.7	37.3	27.0		27.6	73.6	48.0	35.4	22.2	33.3	33.4
63. <i>P.trifoliata</i> _TA6299_37690#1	62.3	61.0	74.3	35.0		20.6	26.0	60.4	71.3	52.7	52.8
64. <i>E.esula</i> _TC3985#1	38.0	37.3	28.7	80.7	27.2		48.3	33.0	21.1	33.0	32.7
65. <i>G.hirsutum</i> _TC137570#1	44.9	45.4	39.3	57.5	38.2	56.1		39.8	26.2	38.6	38.9
66. <i>V.vinifera</i> _GSVIVT00024984001#1	77.2	75.8	65.1	39.0	67.3	37.9	46.4		53.4	63.6	63.5
67. <i>F.vesca</i> _TA11395_57918#1	60.6	59.8	72.0	29.8	79.9	29.5	40.4	61.5		52.4	52.4
68. <i>T.pratense</i> _TA1890_57577#1	71.1	72.0	63.0	39.8	62.3	39.2	47.3	73.6	63.0		89.7
69. <i>M.truncatula</i> _AC148484_8.5#1	73.7	74.6	62.6	40.3	62.1	37.9	47.4	75.8	62.6	92.1	
70. <i>G.max</i> _Glyma16g25600.2#1	76.0	76.3	65.3	41.2	64.7	39.2	49.9	77.7	62.6	85.8	89.7
71. <i>G.max</i> _Glyma11g06960.1#1	72.0	72.8	63.4	37.5	61.0	35.5	43.9	73.6	59.3	77.6	79.1
72. <i>V.vinifera</i> _GSVIVT00014657001#1	67.6	66.8	55.8	33.4	53.8	32.9	38.9	71.1	52.3	60.1	62.3
73. <i>C.clementina</i> _TC35556#1	52.0	50.9	64.7	23.7	79.5	20.8	32.9	56.0	65.7	52.1	52.4
74. <i>P.trichocarpa</i> _719452#1	41.1	41.0	49.7	28.9	55.8	25.8	27.6	44.8	52.7	39.5	39.1
75. <i>B.napus</i> _TC89867#1	63.7	66.5	61.9	38.4	61.3	36.8	46.8	66.5	60.3	66.3	67.4
76. <i>B.napus</i> _TC63535#1	65.1	66.5	61.7	39.0	62.0	36.4	47.0	67.0	60.7	67.5	67.4
77. <i>B.napus</i> _TC63581#1	64.0	67.6	59.4	37.4	60.1	34.5	47.3	64.8	59.4	66.3	66.2
78. <i>A.thaliana</i> _AT4G36730.2#1	65.1	67.1	62.6	38.3	61.3	36.4	46.3	67.3	60.4	66.0	66.5
79. <i>A.lyrata</i> _944204#1	64.0	66.2	62.8	37.8	62.5	36.5	47.4	67.3	61.5	67.5	67.9
80. <i>Aquilegia</i> _sp_TC21139#1	63.8	63.8	55.0	32.0	53.0	29.2	38.8	68.7	50.1	58.4	60.2

FIGURE 3 A (continued)

23/43

Clade D	70	71	72	73	74	75	76	77	78	79	80
48. P.taeda_TA23200_3352#1	15.4	17.2	20.8	21.5	33.2	19.9	20.3	20.6	20.5	20.4	20.3
49. H.annuus_BU027457#1	30.8	32.3	25.3	38.7	36.3	29.9	29.1	28.7	30.2	30.1	25.6
50. N.tabacum_BP133908#1	13.3	11.9	13.8	16.0	24.2	12.9	13.2	13.4	14.1	14.0	11.8
51. H.petiolaris_TA3844_4234#1	45.1	42.2	37.3	33.1	30.9	44.4	44.3	45.9	44.6	45.0	36.2
52. C.tinctorius_TA1437_4222#1	55.6	50.3	49.5	47.9	36.7	49.3	49.4	47.4	47.8	47.2	43.9
53. C.endivia_EL359780#1	27.9	24.8	22.4	28.2	15.4	23.5	23.5	23.2	24.7	24.2	21.4
54. H.tuberosus_TA2966_4233#1	49.3	46.3	41.5	52.9	40.9	41.3	41.8	41.2	42.2	41.4	39.6
55. S.tuberosum_TC187892#1	16.0	15.4	15.1	10.2	13.2	14.7	15.6	14.7	15.4	15.1	11.9
56. S.tuberosum_TC186959#1	42.9	39.3	35.7	15.0	15.8	39.3	39.0	38.6	38.8	38.0	32.7
57. C.canephora_TC1061#1	63.0	56.3	52.8	53.2	40.6	49.2	48.9	49.2	49.7	48.6	47.7
58. S.lycopersicumlikebZIP	49.7	46.2	41.6	54.0	40.8	41.1	40.0	39.0	41.8	40.8	37.5
59. S.lycopersicum_TC211600#1	62.8	58.7	54.4	42.8	32.2	50.9	51.1	49.7	52.6	51.3	47.6
60. N.tabacum_TC76189#1	63.7	57.9	54.4	41.1	33.3	54.0	53.1	54.4	53.7	53.3	47.8
61. C.roseus_AY027510#1	53.6	48.2	47.4	56.5	43.4	42.6	42.9	41.3	45.2	43.6	42.6
62. C.sinensis_TC9189#1	36.1	32.3	28.9	13.4	13.8	32.2	31.8	31.6	33.1	31.3	25.6
63. P.trifoliata_TA6299_37690#1	56.7	52.8	46.1	75.1	48.1	45.1	45.6	44.2	47.9	47.3	42.1
64. E.esula_TC3985#1	33.7	30.8	29.7	14.1	15.1	32.2	31.5	30.3	31.9	31.4	23.9
65. G.hirsutum_TC137570#1	41.7	37.2	33.5	20.1	14.3	36.4	36.6	37.3	36.5	37.4	31.6
66. V.vinifera_GSVIVT00024984001#1	68.6	61.2	59.1	48.5	37.5	53.6	54.1	52.7	56.3	56.0	52.9
67. F.vesca_TA11395_57918#1	54.9	50.7	43.8	56.0	43.2	44.9	45.0	43.6	45.0	45.0	39.7
68. T.pratense_TA1890_57577#1	78.5	66.2	50.9	42.3	33.2	51.9	51.4	51.7	51.6	52.7	46.4
69. M.truncatula_AC148484_8.5#1	80.4	68.3	52.0	44.2	33.9	52.1	52.5	51.6	52.8	54.5	47.7
70. G.max_Glyma16g25600.2#1		73.2	57.1	45.5	34.9	55.9	55.4	55.7	55.0	55.3	50.0
71. G.max_Glyma11g06960.1#1	81.4		52.2	41.7	31.6	49.6	49.2	49.0	50.4	50.1	47.7
72. V.vinifera_GSVIVT00014657001#1	65.6	62.1		38.4	29.8	46.4	46.9	45.9	49.9	48.3	54.5
73. C.clementina_TC35556#1	54.9	51.5	46.2		44.3	36.3	36.9	36.0	39.0	39.7	36.6
74. P.trichocarpa_719452#1	40.4	36.9	34.4	51.2		29.7	30.6	29.4	31.6	31.1	28.5
75. B.napus_TC89867#1	69.4	64.2	56.8	53.2	41.3		95.2	93.0	84.8	85.6	43.1
76. B.napus_TC63535#1	70.0	64.0	57.5	53.0	42.2	97.8		89.3	84.9	85.2	44.1
77. B.napus_TC63581#1	69.4	63.7	56.0	53.0	40.9	94.2	93.3		82.9	83.7	44.1
78. A.thaliana_AT4G36730.2#1	69.1	64.8	57.5	54.3	42.5	91.1	92.3	89.5		94.0	44.0
79. A.lyrata_944204#1	69.7	66.0	57.0	55.1	41.7	91.3	92.7	89.1	96.5		43.1
80. Aquilegia_sp_TC21139#1	62.0	60.2	69.3	48.3	35.1	57.4	59.2	58.7	56.8	57.9	

FIGURE 3 A (continued)

24/43

Clade B	11	12	13	14	15	16	17	18	19
11. G.max_Glyma19g44190.1#1		84.2	83.4	78.1	77.2	76.4	62.8	63.1	60.2
12. G.max_Glyma03g41590.4#1	85.4		73.6	70.8	69.1	69.7	71.5	58.9	54.1
13. M.truncatula_AC137602_8.5#1	89.6	78.6		73.5	73.7	74.4	60.4	63.1	59.7
14. G.max_Glyma16g03190.1#1	86.4	77.4	82.6		90.6	88.2	62.2	61.3	57.2
15. G.max_Glyma07g06620.1#1	85.6	75.7	83.5	93.4		88.0	61.7	61.0	56.6
16. P.vulgaris_TC11785#1	85.2	77.1	84.0	91.3	91.3		62.4	62.1	59.2
17. SimilartolegumebZIP	71.8	81.6	69.6	71.0	70.8	72.3		51.7	46.2
18. P.tricocarpabZIPlike	72.7	69.2	73.4	72.4	71.7	72.6	64.7		57.3
19. C.roseus_AF084971#1	73.2	65.0	73.2	73.0	70.9	72.1	59.2	70.7	
20. C.maculosa_TA2192_215693#1	68.7	65.1	68.5	67.5	64.2	64.8	58.3	70.2	70.0
21. C.maculosa_EH744483#1	28.0	31.0	27.3	27.4	25.9	26.7	31.9	29.3	28.2
22. H.tuberosus_TA2407_4233#1	59.1	59.4	60.9	58.5	58.3	58.9	58.1	61.9	59.9
23. H.annuus_TC42219#1	56.5	59.4	58.1	56.1	55.2	56.7	58.2	59.1	57.5
24. S.lycopersicum_TC213303#1	34.8	39.8	35.1	35.1	34.4	35.0	41.5	35.8	35.2
25. N.tabacum_TC40642#1	29.9	34.2	30.1	30.9	29.7	30.3	35.9	31.3	30.5
26. N.tabacum_NP917548#1	67.2	60.2	69.8	66.5	66.3	66.7	57.6	66.7	72.8
27. C.canephora_TC5243#1	33.6	38.2	32.2	32.8	32.3	32.9	39.0	34.6	35.9
28. S.tuberosum_TC185019#1	31.1	35.3	31.5	30.4	30.4	31.2	37.0	32.6	31.9
29. C.sinensis_TC24527#1	64.2	62.8	64.9	63.4	63.0	64.3	61.3	65.4	59.9
30. G.hirsutum_TC148476#1	65.9	73.3	63.8	62.0	61.3	62.4	69.5	67.4	59.9
31. A.thaliana_AT2G46270.2#1	62.6	61.8	62.8	59.9	59.4	60.3	59.6	64.2	62.9
32. A.lyrata_322281#1	65.4	62.4	66.4	63.7	63.2	63.6	59.7	64.4	66.9
33. B.napus_TC63534#1	65.2	63.1	64.9	61.1	61.1	61.0	59.4	63.2	64.6
34. A.thaliana_AT4G01120.1#1	60.2	58.6	58.4	55.0	55.9	56.3	54.2	58.6	56.3
35. A.lyrata_895903#1	56.5	53.2	52.9	52.4	53.1	52.2	48.3	54.9	51.9
36. B.napus_TC63500#1	60.0	58.8	57.6	57.1	58.5	58.4	56.7	59.4	56.1

FIGURE 3 B

25/43

Clade B	20	21	22	23	24	25	26	27	28
11. G.max_Glyma19g44190.1#1	56.0	23.9	48.7	46.8	28.2	25.2	54.2	28.7	25.1
12. G.max_Glyma03g41590.4#1	50.0	26.9	43.8	44.8	32.8	28.9	48.4	33.2	28.5
13. M.truncatula_AC137602_8.5#1	55.6	23.5	48.2	46.3	28.6	24.9	54.4	27.5	26.1
14. G.max_Glyma16g03190.1#1	52.9	22.1	47.1	45.3	26.6	25.0	53.1	27.6	23.3
15. G.max_Glyma07g06620.1#1	51.3	21.2	47.3	45.5	26.8	24.8	53.9	27.8	22.8
16. P.vulgaris_TC11785#1	53.3	22.2	48.5	47.7	27.6	25.8	54.4	28.1	23.8
17. SimilartolegumebZIP	43.9	26.0	41.1	42.6	33.5	30.4	44.7	33.0	29.2
18. P.triocarpabZIPlike	52.6	23.5	48.5	45.8	30.0	27.2	55.3	28.5	26.5
19. C.roseus_AF084971#1	55.3	22.8	47.1	45.7	28.4	25.2	59.9	30.1	25.3
20. C.maculosa_TA2192_215693#1		36.4	60.5	60.0	28.7	26.2	52.2	26.7	25.9
21. C.maculosa_EH744483#1	36.9		28.9	29.1	55.4	65.6	23.7	55.9	57.7
22. H.tuberosus_TA2407_4233#1	71.6	32.8		93.1	28.2	24.9	48.7	26.7	25.5
23. H.annuus_TC42219#1	70.1	33.0	94.7		28.5	25.4	45.5	27.6	26.7
24. S.lycopersicum_TC213303#1	35.2	64.7	34.4	34.8		72.1	32.2	68.1	67.2
25. N.tabacum_TC40642#1	31.9	78.8	30.6	32.2	77.4		28.6	65.4	67.6
26. N.tabacum_NP917548#1	67.0	28.3	60.2	57.4	36.8	32.3		28.1	32.4
27. C.canephora_TC5243#1	34.9	68.2	34.7	35.6	81.1	75.4	34.0		60.2
28. S.tuberosum_TC185019#1	33.7	73.1	32.5	33.3	78.4	80.1	36.3	77.1	
29. C.sinensis_TC24527#1	63.6	27.2	62.2	59.8	34.4	28.9	59.3	32.9	31.8
30. G.hirsutum_TC148476#1	64.3	30.7	63.6	64.8	36.7	30.9	54.6	36.4	33.0
31. A.thaliana_AT2G46270.2#1	64.3	29.8	63.1	61.3	33.7	31.2	57.1	34.5	32.3
32. A.lyrata_322281#1	66.3	27.4	60.5	58.4	34.7	29.7	59.7	33.2	31.3
33. B.napus_TC63534#1	64.1	25.7	61.8	59.9	32.4	27.3	56.9	31.0	29.1
34. A.thaliana_AT4G01120.1#1	56.5	26.7	58.3	53.6	32.8	28.1	55.3	31.1	29.2
35. A.lyrata_895903#1	53.8	25.7	54.2	52.8	31.4	25.7	53.2	30.6	28.4
36. B.napus_TC63500#1	61.3	26.7	56.7	52.3	32.4	27.2	54.3	30.0	31.1

FIGURE 3 B (continued)

26/43

Clade B	29	30	31	32	33	34	35	36
11. G.max_Glyma19g44190.1#1	55.8	56.2	52.4	54.4	51.7	46.4	43.2	46.0
12. G.max_Glyma03g41590.4#1	49.6	60.2	45.8	46.7	45.6	40.9	37.2	40.6
13. M.truncatula_AC137602_8.5#1	56.7	52.0	52.2	54.1	51.7	45.2	41.1	43.7
14. G.max_Glyma16g03190.1#1	53.9	51.2	48.9	50.9	49.1	41.5	39.2	41.6
15. G.max_Glyma07g06620.1#1	53.6	50.5	48.2	50.7	48.6	42.7	40.1	42.3
16. P.vulgaris_TC11785#1	54.5	51.3	49.3	51.8	50.2	43.2	40.5	43.3
17. SimilarartolegumebZIP	45.0	50.5	41.0	42.7	39.9	35.1	32.1	37.1
18. P.tricocarpabZIPlike	55.5	55.9	53.9	53.3	50.7	46.1	42.7	46.2
19. C.roseus_AF084971#1	48.4	49.2	51.2	52.7	50.8	43.8	42.5	42.4
20. C.maculosa_TA2192_215693#1	49.1	48.2	49.2	49.9	48.3	41.4	39.5	43.1
21. C.maculosa_EH744483#1	19.9	23.2	23.3	21.5	19.7	17.6	18.3	18.7
22. H.tuberosus_TA2407_4233#1	46.4	44.6	46.6	46.6	46.9	40.2	38.7	39.8
23. H.annuus_TC42219#1	45.1	45.5	44.2	44.8	45.4	38.5	38.0	37.3
24. S.lycopersicum_TC213303#1	24.9	28.6	24.9	26.1	23.6	22.2	22.0	24.1
25. N.tabacum_TC40642#1	22.6	24.9	23.1	21.1	20.1	18.2	17.4	19.0
26. N.tabacum_NP917548#1	48.3	44.5	43.4	43.8	42.9	38.7	37.7	37.9
27. C.canephora_TC5243#1	24.5	27.0	24.7	23.4	22.1	20.0	19.8	19.6
28. S.tuberosum_TC185019#1	23.3	27.0	22.5	22.4	22.1	19.4	20.0	21.1
29. C.sinensis_TC24527#1		52.5	49.2	49.3	48.9	45.5	42.4	50.0
30. G.hirsutum_TC148476#1	67.3		48.0	48.9	47.8	41.7	36.6	40.5
31. A.thaliana_AT2G46270.2#1	61.8	63.8		90.1	76.3	50.8	46.2	50.4
32. A.lyrata_322281#1	62.4	62.6	91.3		82.5	49.9	46.3	50.6
33. B.napus_TC63534#1	62.6	62.6	83.2	88.4		47.8	42.9	48.6
34. A.thaliana_AT4G01120.1#1	61.7	60.0	65.8	66.3	61.2		85.6	74.9
35. A.lyrata_895903#1	56.0	50.1	60.6	61.1	54.8	88.2		69.5
36. B.napus_TC63500#1	63.8	56.4	65.9	66.8	63.6	82.8	77.7	

FIGURE 3 B (continued)

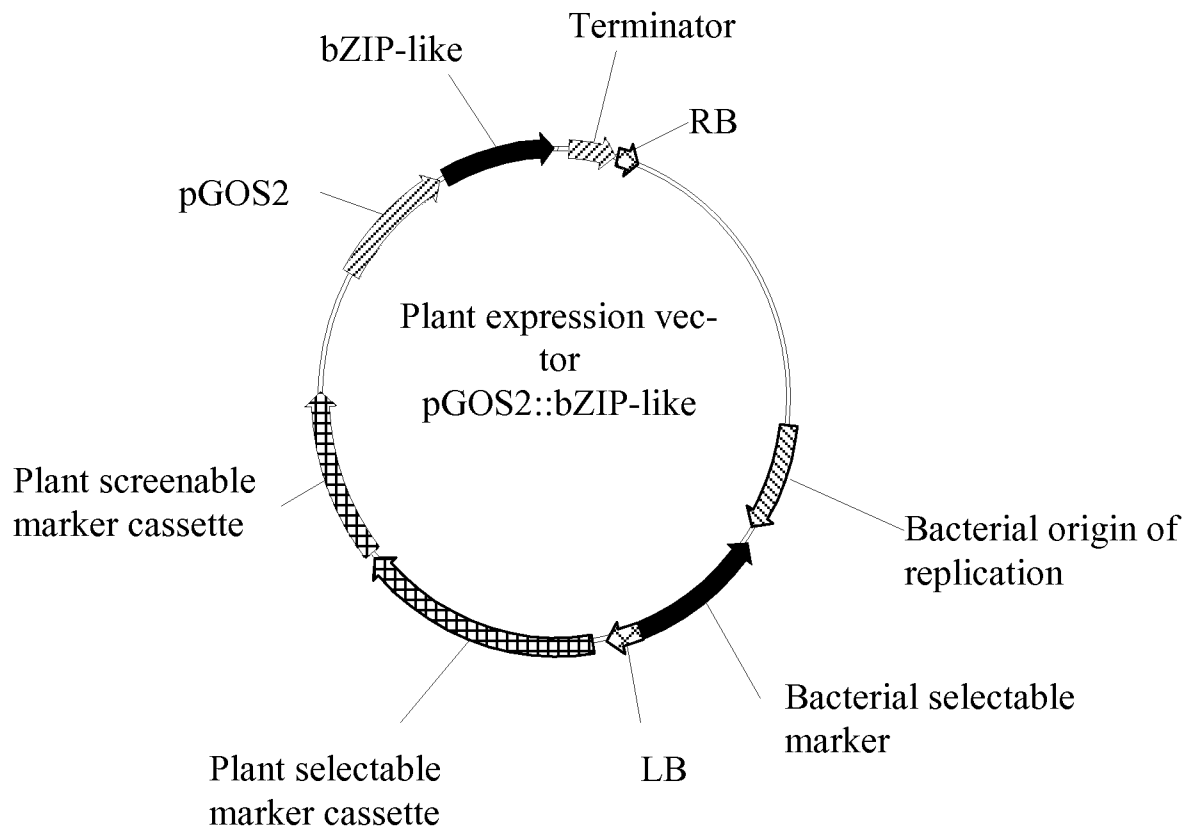


FIGURE 4

28/43

MERSAVFGGLQPNYLLYPSPNSSSLPFSHDRARLPNFSPPPSLSLKIHKQVSSCFK

AVSPFKRGAAFSDFHSDTFELADIDWDDLGFAYVPTDYMYSMKCTKGGNFSKGELQ

RYGNIELNPSAGV LNYGQGLFEGLKAYRKEDGNLLLF RPEENAMRMIMGAERMCMF
Motif 15 Motif 14SPTIDQFVDAVKATVLA ANKRWVPPGKGS LYIRPLLMGSGAVLGLAPAPEYTF LIY
Motif 13

VSPVGNFYFKEGVAPIHLIVEHELHRATPGGTGGVKTIGNYAAVLKAQSAAKARGFS

DVLYLDCVHKKYLEEVSSCNIFVVKGNSISTPAIKGTILPGITRKS IIDVARSQGF

QVEERLVTVDELDDADEVFCTGTAVVVSPVGSITYKGRVSYGVEGFGAVSQQLYS

VLTKLQMGLIEDKMNWTVELS

FIGURE 5

G.max_Glyma06g05280	-----MIQRTVSFPSL-----RKLLLRA
H.annuus_TC54245	-----MMIRQSPF-----LL--
A.thaliana_AT1G10070	-----MIKTITSLR-----KTLVLPL
P.trichocarpa_scaff_II.1054	-----MIQTNGLRSL-----VQSLRPI
A.thaliana_AT1G10060	-----MALR-----RCLPQY
A.thaliana_AT3G49680	-----MERAAILPSVNQN-----YLLCPSRAFS--TRLHSSTR
A.thaliana_AT5G65780	-----MERSAVASGFHRN-----YILCASRAATSTRLHSLSS
P.trichocarpa_BCAT4-like	-----MERSAVFGGLQPN-----YLLYSPNSSSLLPFSDDRAR
G.max_Glyma01g40420	-----MGKQKQKMESIRLI-----YPICPSRHSS--FLLSHQS
G.max_Glyma11g04870	-----MESSAVYGSIRPS-----YPICPSRRSS--SLLSHQS
M.truncatula_TC114768	-----MESSAALTSIRLT-----SSIRPSRFSS--PFLSPAF
S.lycopersicum_TC213629	-----MESAAVFAGLHPIPGHHNHLLGPSRTAI---KLLPPSI
H.vulgare_TC165564	-----MDCGTASHGALLA-----AAPLAGRRRRLLP
O.sativa_LOC_Os05g48450	-----MEYGAATRGAALLA-----AAPISGARRSWLP
Z.mays_TA12434_4577999	-----MEY-----GAVLA-----AAPLVAR-PNWLL
Hordeum_vulgare_PUT-169a-Horde	-----
Hordeum_vulgare_subsp_vulgare_	-----
O.sativa_Os03g0106400	-----MELHLTSRGALPL-----SPPLAQRRPHLS
O.sativa_LOC_Os04g47190	-----MELLPR--VGVAAP-----GPGRGGASPSPTR
T.aestivum_TC320973	-----MELRLRAPASPASA-----SPRGTSVSPSPRP
Zea_mays_GRMZM2G047347_T03	-----MELCTC--AARSS-----APPSSCRAAP--
P.patens_TC31354	MAVMCGIGLASSLLQESYMSVASSEAGRADVKRVSSSSSP-QLLQNGVG
P.patens_TC33668	MGMACGNRLASSLLQ-ESLMTVASSEARRKDARRVLSLSSPSQLRQSASG
H.vulgare_TC186077	-----MAVLSAKRVLPC-----ASAG--GVSGGLRA
T.aestivum_TC325793	-----MDVLSSAKRALPWG-----RTSAG--GVIGGLRA
O.sativa_LOC_Os03g12890	-----MAAAAAASSAKRALLP-----WARDAAHALARALQ
Zea_mays_GRMZM2G153536_T03	-----MAALTSAKGALLPS-----WARSSSSGHGGDLWR
A.thaliana_AT1G50090	-----MAPSVH-----PSSS
A.thaliana_AT1G50110	-----MAPSS-----PLRT
A.thaliana_AT3G19710	-----MAPSAQ-----PLPV
G.max_Glyma07g30510	-----MSPPS-----MLGN
G.max_Glyma08g06750	-----MVPPHGKALYIRP-----LLFGS
M.truncatula_AC159872_36	-----MAPPS-----ILRD
P.trichocarpa_804339	-----
P.trichocarpa_scaff_IX.827	-----MASSSTGPKKVLRSR-----FRQVV
M.truncatula_AC159872_55	-----MATSHQLPNN-----GKAS

FIGURE 6

G.max_Glyma06g05280
H.annuus_TC54245
A.thaliana_AT1G10070
P.trichocarpa_scaff_II.1054
A.thaliana_AT1G10060
A.thaliana_AT3G49680
A.thaliana_AT5G65780
P.trichocarpa_BCAT4-like
G.max_Glyma01g40420
G.max_Glyma11g04870
M.truncatula_TC114768
S.lycopersicum_TC213629
H.vulgare_TC165564
O.sativa_LOC_Os05g48450
Z.mays_TA12434_4577999
Hordeum_vulgare_PUT-169a-Horde
Hordeum_vulgare_subsp_vulgare_
O.sativa_Os03g0106400
O.sativa_LOC_Os04g47190
T.aestivum_TC320973
Zea_mays_GRMZM2G047347_T03
P.patens_TC31354
P.patens_TC33668
H.vulgare_TC186077
T.aestivum_TC325793
O.sativa_LOC_Os03g12890
Zea_mays_GRMZM2G153536_T03
A.thaliana_AT1G50090
A.thaliana_AT1G50110
A.thaliana_AT3G19710
G.max_Glyma07g30510
G.max_Glyma08g06750
M.truncatula_AC159872_36
P.trichocarpa_804339
P.trichocarpa_scaff_IX.827
M.truncatula_AC159872_55

GCSKSASSKIGT-----YNCFASQSSPL-----PSHNPSYRDDEY--
GLIQTISISKISA-----RCLTAQAASAL-----QENEPMIRHEEYA-
HLHIRTLOTFAK-----YNAQAASALRE-----ERKKPLYQNGD---
TSSLSLIVVATHHKQHLLFNKLADHIQTE-----AFGFLLFVGGDSSE
STSSYLSKIWG---FRMGTKAAASVVE-----EHVSGAEREDEE--
NLSPPSFASIKLQHSSSSVSSNGGISLT-----RCNAVSSNSSSTLV
LRNFPS--SSLRIRHCPSPISNFIVSEVRN-----RRCDAVS--SSTTDV
LPNFSPPSLSLKHKQVSSCFKAVSPFK-----RGAAFS--DTHSDT
PFLCEPSSLKLRKQFPLTSQNVLEAASPL-----RPSATLSSDDPYSET
PFLFEPSSLKLRKQFPLISQNFLEAASPL-----RPSATLSSDSYRET
PPK-PTSLSLKLQKQFPFTSQNVLQASNAL-----RPSASVSAS---EA
DKINFSPPLPKFKQKQSHFTSY-IGNSAINS-----GNSFRVASPASDVA
LSPPPSTPSIQIQNRLYSMSLLPLRKARGM-----GRCEASLASNYTQT
LSSPPSPPSIQIQNRLYSISLPL-KARGV-----RRCEASLASDYTKA
LSPPPLAPSIQIQNRLYSISFPL-KAGPV-----RACRA-LASNYTQT
-----MPTLHH-KAHTT-----VGCQASVASKYMET
-----MPTLHH-KAHTT-----VGCQASVASKYMET
LS----TPSLPIKNHTYVPPFFS-KAHCA-----IGCQASLATNYMET
RHRAPSHPILRSAAVCGAVAVCRGAVARR----SRWSTLVTAAYYTG
HPRLP SQIQKRLSG--SAVSVRRRGTAAARS----SLCSALMAASNYTGT
FPRVLSHRIWRSRGY--CTVCFTPSSPVAR----SRFSTLMTTAHNTGT
LRRTCMPAFFVTEERLRSLSLHNSHYHVRGSKVQQLHAVADALNQTSDL
GLRRRCRVPAFALIEDDSRSSVSQNATCHVRGSKVQPLNAVADALNQTSDL
LLG-----TDGGG-RSLLPSRWKSS-----LPQLDP-VDRSDE
LLG-----TDGGGRSLLPSRWKSS-----QPQLDP-VDRSDE
CGGGGLGLRGALPTAGGRWSLLQCRWRSS-----LPQLDS-ADRSDE
VLG-KALATAGGG-GGGGCSLLLPWWQSS-----LPQLDHVADRSNE
P-----LFTSKAD
-----TSETD
-----SVSD
R-----KDGSE-----IAVA
GSVMGIAPAPHCTFLIYTNPISNAYKCRVEF-----KTGADTVTQNTAG
T-----EDGSE-----SDMG
G-FVPLH-----
G-LLLPHSKSTPPPVSSTDDDKVTDDHQVS-----TDDHKVSTDEKVK
N-----RETE

FIGURE 6 (continued)

G.max_Glyma06g05280
H.annuus_TC54245
A.thaliana_AT1G10070
P.trichocarpa_scaff_II.1054
A.thaliana_AT1G10060
A.thaliana_AT3G49680
A.thaliana_AT5G65780
P.trichocarpa_BCAT4-like
G.max_Glyma01g40420
G.max_Glyma11g04870
M.truncatula_TC114768
S.lycopersicum_TC213629
H.vulgare_TC165564
O.sativa_LOC_Os05g48450
Z.mays_TA12434_4577999
Hordeum_vulgare_PUT-169a-Horde
Hordeum_vulgare_subsp_vulgare_
O.sativa_Os03g0106400
O.sativa_LOC_Os04g47190
T.aestivum_TC320973
Zea_mays_GRMZM2G047347_T03
P.patens_TC31354
P.patens_TC33668
H.vulgare_TC186077
T.aestivum_TC325793
O.sativa_LOC_Os03g12890
Zea_mays_GRMZM2G153536_T03
A.thaliana_AT1G50090
A.thaliana_AT1G50110
A.thaliana_AT3G19710
G.max_Glyma07g30510
G.max_Glyma08g06750
M.truncatula_AC159872_36
P.trichocarpa_804339
P.trichocarpa_scaff_IX.827
M.truncatula_AC159872_55

---ADVWDSLGLMPTDYMII TKCEGQ--NFGQQLSRYGNIELS
---ADIDWNNLGFGLKQTDYMYKSKCTKNN--TFEQQLVNYGNLELSPA
DVIADLDWNLGFLNPADYMYVMKCSKDG--EFTQGEELSPYGNLQLS
DEYAKVDWNLRFGITPADYMYTMKCSSDG--KFEQQLAPYGNVELS
--YADVWDLNGLFSLVRTDFMFA TKSCRDG--NFEQGLSRYGNIELNPA
TELAIDWDTVGFGLKPADYMYVMKCNIDG--EFSKGEIQRFGNIEISPS
TELAIDWDKIDFGLKPTDYMAMKCSRDG--EFSQQLQFPFNIDINPA
FELADIDWDLGFAYPTDYMYSMKCTKGG--NFSKGEIQRFGNIELNPS
IELADIEWDLGFLQPTDYMIMKCTRGG--TFSKGEIQRFGNIELNPS
IELADIEWDLGFLQPTDYMYSMKCTRGG--TFSKGEIQRFGNIELNPS
IELADIDWDLGFLQPTDYMIFMKCDQGG--TFSKGELKRFGNIELNPS
SELADIDWDLGFGFMPDYMYSMKCSQGE--NFSKGEIQRFGNIELS
SEFADLDWNLGFLVQTDYMYTAKCGPDG--NFDKGGMVPFPGPIEMNPA
SEVADLDWNLGFGIVQTDYMYITKCGQDG--NFSEGEIPIFGPIALNPS
SETVDLDWNLGFGIVQTDYMYIAKCGTDG--NFSEGEIPIFGPIALNPS
PEIVDLWNLGFLVNTDFMYMAKCGPDG--NFSKGEILPFGPIALS
PEIVDLWNLGFLVNTDFMYMAKCGPDG--NFSKGEILPFGPIALS
SAVADLDWNLGFLVQTDYMYIAKCGPDG--NFSKGEIPIFGPIALS
AELVDFNWE TLGFQVPTDFMYVMRCSEEG--VFTKGEIPIFGPIALS
PDLVDFWETLGFQLVPTDFMYIMKCSSDG--VFTKGEIPIFGPIALS
PDLVDFNWDLLGFQLIPTDFMYLMSCSSDG--VFMNGKLVPIGSIELNPA
DTLEGIDWDLNFGFLRPTDFMFMVKGDLG--NWQKGLRPFNLEVSPS
DTLQGDIDWDLNFGFLRPTDFMFMVKKGDIAG--NWQEGELVPIGNIELS
ESGGDIDWDLNFGFLTPTDYMVMRCSREE--GGFSRGEIPIFGPIALS
ESGGDIDWDLNFGFLTPTDYMVMRCSQEE--GGFSRGEIPIFGPIALS
ESGGEIDWDLNFGFLTPTDYMVMRCSLED--GVFSRGEIPIFGPIALS
ESGGEIDWDLNFGFLTPTDYMVMRCSLED--GVFSRGEIPIFGPIALS
EKYANVKWDELGFALVPTDYMVAKCKQGE--SFTGEIPIFGPIALS
EKYANVKWDELGFALVPTDYMVAKCKQGE--SFTGEIPIFGPIALS
EKYANVKWDELGFALVPTDYMVAKCKQGE--SFTGEIPIFGPIALS
EKYANVKWDELGFALVPTDYMVAKCKQGE--SFTGEIPIFGPIALS
ENYADINWDELGFLSVPDYMVMKCAKGD--KFSQGSIVPFGNIEISPS
ENYADINWDELGFLSVPDYMVMKCAKGD--KFSQGSIVPFGNIEISPS
ENYADINWDELGFLSVPDYMVMKCAKGD--KFSQGSIVPFGNIEISPS
ENYADINWDELGFLSVPDYMVMKCAKGD--KFSQGSIVPFGNIEISPS
-----INWDLNFGFLTPTDFMFLMKCPVGD--KYSEGLVPIGNIELS
SGGEDINWDLNFGFLTPTDFMFLMKCPVGD--KYSEGLVPIGNIELS
KIYANMDWDLKTCGVIPDYMIIKSNEDR--TYSNGTLVPIFGPIALS

... :
* : *
* : *
* : *
* : *

FIGURE 6 (continued)

G.max_Glyma06g05280	AGVLNYGQGLFEGTKAYRKENGG--LLLFRPEENAI	IRMKTGAQRMCMAS
H.annuus_TC54245	AGVLNYGQGLFEGTKAVRGEDGR--LLLFRPDQNA	IRMQIGAERMCMQSP
A.thaliana_AT1G10070	AGVLNYGQAIYEGTKAYRKENGK--LLLFRPDHNA	IRMKLGAERMLMPSP
P.trichocarpa_scaff_II.1054	A-----AGLYEGTKAYRTEDGR--LLLFRLDQNA	TRMKMGADRLCMACP
A.thaliana_AT1G10060	AGILNYGQGLIEGMKAYRGEDGR--VLLFRPELNAM	RMKIGAERMCMHSP
A.thaliana_AT3G49680	AGVLNYGQGLFEGLKAYRKKDGN--LLFRPEENAK	RMNRNGAERMCMPPAP
A.thaliana_AT5G65780	AGVLNYGQGLFEGLKAYRKQDGN--LLFRPEENAI	RMNRNGAERMCMPPSP
P.trichocarpa_BCAT4-like	AGVLNYGQGLFEGLKAYRKEDGN--LLFRPEENAM	RIMMGAERMCMPPSP
G.max_Glyma01g40420	AGVLNYGQGLFEGLKAYRKQDGS--LLFRPEENGL	RMQIGAERMCMPPSP
G.max_Glyma11g04870	AGVLNYGQGLFEGLKAYRKQDGS--LLFRPEENGL	RMQIGAERMCMPPSP
M.truncatula_TC114768	AGVLNYGQGLFEGLKAYRKDDGN--LLFRPEENAL	RMKTGAERMCMPPSP
S.lycopersicum_TC213629	AGILNYGQGLFEGLKAYRKHGDN--LLFRPEENAT	RLKMGAEERMCMPPSP
H.vulgare_TC165564	SGVLNYGQGLFEGLKAYRKTDDGS--LLFRPME	NAMRMQTGAERMCMPPAP
O.sativa_LOC_Os05g48450	SGVLNYGQGLFEGLKAYRTTDDGS--LLFRPEENAL	RMRTGAERMCMPPAP
Z.mays_TA12434_4577999	SGVLNYGQGLFEGLKAYRKTDDGS--LLFRPEENAE	RMRTGAERMCMPPAP
Hordeum_vulgare_PUT-169a-Horde	AGVLNYGQGLFEGLKAYRKTDDGS--VLLFRPEENAV	RMKNGSDRMCMPPAP
Hordeum_vulgare_subsp_vulgare_	AGVLNYGQGLFEGLKAYRKTDDGS--VLLFRPEENAV	RMKNGSDRMCMPPAP
O.sativa_Os03g0106400	AGVLNYGQGLFEGLKAYRKTDDGS--LLFRPEENAI	RMNRNGAERMCMPPAP
O.sativa_LOC_Os04g47190	AGVLNYGQGLLEGLRAHRKEDGS--VLLFRPDENA	LRMVGDRLCMPPAP
T.aestivum_TC320973	AAVLNYGQGLLEGLRAHRKEDGS--VIVFRPKENA	LRMIGADRLCMPPAP
Zea_mays_GRMZM2G047347_T03	AAVLNYGQGLLEGLRSHRKEGDS--LLFRPHENAR	MEIGADRLCMPPAP
P.patens_TC31354	AGVLNYGQGVFEGMKAYRTADDR--LILFRPEENAM	RMINGAERMSMPAP
P.patens_TC33668	AGVLNYGQGVFEGMKAYRTADDSDS--ILMFRPEENAL	RMVHGAERMSMPAP
H.vulgare_TC186077	SGVLNYGQGLFEGLKAYRRSDGAG-YMLFRPEENAR	RMQHGAGRMCMPSP
T.aestivum_TC325793	SGVLNYGQGLFEGLKAYRRADGPG-YMLFRPEENAR	RMQHGAGRMCMPAP
O.sativa_LOC_Os03g12890	SGVINYGQGLFEGLKAYRAANQDGSYMLFRPEENAR	RMQHGAGRMCMPSP
Zea_mays_GRMZM2G153536_T03	SGVLNYAQGLFEGMKAYRRPDRAG-YTLFRPEENAR	RMQHGAGRMCMPAP
A.thaliana_AT1G50090	AGILNYGQGLFEGLKAYRTEDGR--ITLFRPDQNA	IRMQTGADRLCMTTPP
A.thaliana_AT1G50110	SPILNYGQGLFEGLKAYRTEDDR--IRIFRPDQNA	LRMQTGAERLCMTTPP
A.thaliana_AT3G19710	AAVLQYGQGLYEGLKAYRTEDGR--LLFRPDQNG	LRLOAGADRLYMPYP
G.max_Glyma07g30510	AGILNYGQGLFEGLKAHRTEDGH--VLLFRPDENA	QRMKRGADRLCMPSP
G.max_Glyma08g06750	AGILNYGQGI FEGLKAYRTEDGC--LLFRPEENAO	RMKIGADRLCMPSP
M.truncatula_AC159872_36	AGILNYGQGI FEGLKAYRTEDGR--LLFRPEENAL	RMKMGADRLCMPSP
P.trichocarpa_804339	SSVLNYGQGLLEGLKAYRGDDNR--IRLFRPEQNAL	RMQMGAEERMCMSSP
P.trichocarpa_scaff_IX.827	SSVLNYGQGLFEGMKVYRREDDR--IMIIFRPEENAR	RMQMGAEERLLMQAP
M.truncatula_AC159872_55	SAVINYGQGLFEGMKAYRTKDDGN--VQLFRPEENAL	RMQMGAEERLLMPSP

FIGURE 6 (continued)

G.max_Glyma06g05280	FLIFASVVRNRYFKEGSAPLNLYVEENFDRASSR--GTGNVKTISNYAP--
H.annuus_TC54245	FLVYASPVGNRYFKEGTAPLNLYVNNEFHRTTRG--GAGGVKTI TNYAP--
A.thaliana_AT1G10070	FIVYASPVGNRYFKEGMAALNLYVEEYVRAAPG--GAGGVKSI TNYAP--
P.trichocarpa_scaff_II.1054	FLIYASPVGNRYFKEGLKPLNLYVEDEFHFRATRG--GAGGVKSI TNYAP--
A.thaliana_AT1G10060	FLVFGSPVQNYFKEGTAALNLYVEEVI PRAYLG--GTGGVKAISNYGP--
A.thaliana_AT3G49680	FIIYVSPVGNRYFKEGVAPINLI VENEFFHRATPG--GTGGVKTIGNYAA--
A.thaliana_AT5G65780	FLIFVSPVGNRYFKEGVAPINLI VETEFHRATPG--GTGGVKTIGNYAA--
P.trichocarpa_BCAT4-like	FLIYVSPVGNRYFKEGVAPIHLI VEHLELHRATPG--GTGGVKTIGNYAA--
G.max_Glyma01g40420	FLIYVSPVGNRYFKEGLAPINLI VENEFFHRATPG--GTGGVKTIGNYAA--
G.max_Glyma11g04870	FLIYVSPVGNRYFKEGLAPINLI VENEELHRATPG--GTGGVKTIGNYAA--
M.truncatula_TC114768	FLIYVSPVGNRYFKEGLAPINLI VESEELHRATPG--GTGGVKTIGNYAA--
S.lycopersicum_TC213629	FLIYVSPVGNRYFKEGLAPINLI VVETEMHRATPG--GTGGVKTIGNYAA--
H.vulgare_TC165564	FIIFASPVGNRYFKEGLAPINLI VEDKFHRATPG--GTGGVKTIGNYAS--
O.sativa_LOC_Os05g48450	FIIFVSPVGNRYFKEGLAPINLI VEDKFHRATPG--GTGSVKTIGNYAS--
Z.mays_TA12434_4577999	FIIFVSPVGNRYFKEGLAPINLI VVDKFHRATPG--GTGGVKTIGNYAS--
Hordeum_vulgare_PUT-169a-Horde	FLIYVSPVGNRYFKEGLAPINLI IEDNFHRAAPG--GTGGVKTIGNYAS--
Hordeum_vulgare_subsp_vulgare_	FLIYVSPVGNRYFKEGLAPINLI IEDNFHRAAPG--GTGGVKTIGNYAS--
O.sativa_Os03g0106400	FMI FVSPVGNRYFKEGLAPINLI I EENFHRAPG--GTGGVKTIGNYAS--
O.sativa_LOC_Os04g47190	FVVFACPVGHYFKDGLSPI SLLTEEEYQCAAPG--GTGDIKTI GNYAS--
T.aestivum_TC320973	FVVYVCPVGHYFKDGLSPI SLLTEEEYHRAAPG--GTGDIKTI GNYAS--
Zea_mays_GRMZM2G047347_T03	FIVFVCPVGHYFKGGLAPI SLLTEEEYHRAAPG--GTGDIKTI GNYAS--
P.patens_TC31354	FLIYVSPVGTIFYKGLSPI DLKVET YFHRAPG--GTGGVKTISNYAP--
P.patens_TC33668	FLVYVSPVGTIFYKGLSPI DLKVET YFHRAPG--GTGGVKTISNYAP--
H.vulgare_TC186077	FMIYAAPVGTIFYKEGMAA INLI VEEIHRAMPG--GTGGVKTISNYAP--
T.aestivum_TC325793	FMIYAAPVGTIFYKEGMAA INLI VEEIHRAMPG--GTGGVKSISNYAP--
O.sativa_LOC_Os03g12890	FLIYAAPVGTIFYKEGLAPINLI VEDSIHRAMPG--GTGGVKTITNYAP--
Zea_mays_GRMZM2G153536_T03	FLIYAAPVGNRYFKEGLAPINLI VVHDEFHRAAPG--GTGGVKTIANYAP--
A.thaliana_AT1G50090	FLIYTSFVGNRYHKAS-SGLNLKV DNHRRAHFG--GTGGVKSCTNYSF--
A.thaliana_AT1G50110	FLIYASPVGDYHKVS-SGLNLKV DHYHRAHSG--GTGGVKSCTNYSF--
A.thaliana_AT3G19710	FTAFACPVGRYHKDN-SGLNLKIEDQFRRAFPS--GTGGVKSITNYCP--
G.max_Glyma07g30510	FLIYCSFVGSYQKG---ALNLKVEDKLYRAISGCGGTGGIKSVTNYAP--
G.max_Glyma08g06750	LLIYCSFVTNRYHKG---SLNLKVESKFYRAISGTGTGGIKSVTNYAP--
M.truncatula_AC159872_36	FLIYCSFVGNRYHKG---RLNLKVEDKFHRSIAGSGGTGGIKSVTNYAP--
P.trichocarpa_804339	FLAYASPVGNRYFNG---PMHFVSEDKVYRAIPG--GTGGIKSI TNYSF--
P.trichocarpa_scaff_IX.827	FLIFASPIRNSYKSGIDAFNLSIETKLHRASPG--GTGGIKSI TNYAPVS
M.truncatula_AC159872_55	LLIIFTNPISNIYKQGQTSALNLI INENFPRAYPG--GTGGVKSISNYPL--

FIGURE 6 (continued)

G. max_Glyma06g05280	IFI AKGKCI STPAT - NGTILSGITRKS VIEIARDHGYQ - VEERAVAVDEL
H. annuus_TC54245	IFIVKGDVISTPST - VGTILEGIPRKS IIDIARALGYK - VEERLVAVDEL
A. thaliana_AT1G10070	VFVVKGR TISTPAT - NGTILEGITRKS VMEIASDQGYQ - VVEKAVHVDEV
P. trichocarpa_scaff_II.1054	IFLVKGNIISSPAT - SGTILPGVTRRS IIEIALDHGYQ - VEERAIPLDEL
A. thaliana_AT1G10060	IFLVKGN TIVTPAT - SGTILGGITR -----
A. thaliana_AT3G49680	IFIVKDNVISTPEI - KGTILPGITRKS MIDVARTQGFQ - VEERNVTVDEL
A. thaliana_AT5G65780	IFIVKDNVISTPEI - KGTILPGITRKS IIEVARSQGFK - VEERNVTVDEL
P. trichocarpa_BCAT4-like	IFVVKGN S ISTPAI - KGTILPGITRKS IIDVARSQGFQ - VEERLVTVDEL
G. max_Glyma01g40420	IFVVKGN IISTPAI - KGTILPGITRKS IIDVARSEGFQ - VEERLVSVDEL
G. max_Glyma11g04870	IFVVKGN VISTPAI - KGTILPGITRKS IIDVARSQGFQ - VEERLVSVDEL
M. truncatula_TC114768	IFVVKGN VISTPSI - KGTILPGITRKS IIDVARSQGFE - VEERLVAVDEL
S. lycopersicum_TC213629	VFIVKGN L I VTPAI - KGTILPGITRKS IIDVAISQGFE - VEERQVSVDEL
H. vulgare_TC165564	IFVVKGN VISTPAI - KGTILPGITRKS IIDVALSKGFQ - VEERPRVRG--
O. sativa_LOC_Os05g48450	IFVVKGN VISTPAV - KGTILPGITRKS IIDVALSKGFQ - VEERLVSVDEL
Z. mays_TA12434_4577999	VFVVKDN VISTPAI - KGTILPGITRKS IIDVALSKGFQ - VEERLVSVDEL
Hordeum_vulgare_PUT-169a-Horde	IFVVKGN AICTPAI - EGTILPGITRKS IIEVAESKGYK - VEERHVSVDEL
Hordeum_vulgare_subsp_vulgare_	IFVVKGN AICTPAI - EGTILPGITRKS IIEVAESKGYK - VEERHVSVDEL
O. sativa_Os03g0106400	IFIVKGN VISTPAI - KGTILPGITRKS IIEVAQRKGF - VEERLVSVDEL
O. sativa_LOC_Os04g47190	IFMVKDN IISTPLL - TGTVLPGITRRS IIEYARSLGFQ - VEECLITIDEL
T. aestivum_TC320973	IFMVKDN VISTPLL - TGTILPGITRRS IIEIASNLGIQ - VEERLAI AIDEL
Zea_mays_GRMZM2G047347_T03	IFMVKDN IISTPLL - TGTILPGITRRS VIEI SQNLGFQ - VEERLITIDEL
P. patens_TC31354	IFMVKGV IISTPEL - AGTILPGITRKS IIQLARSRGYE - VNERPVSVDEL
P. patens_TC33668	IFVVKDK VISTPEL - AGTILPGITRNS IIQLARSRGYE - VNERPVSVDEL
H. vulgare_TC186077	LFVVKGG AVATPAT - TAGTILPGVTRRS IIEELARDGQ - VEERLVSIDDL
T. aestivum_TC325793	LFVVKGG A IATPAT - EGTILPGVTRRS IIEELARDSGYQ - VEERLVSIDDL
O. sativa_LOC_Os03g12890	LFIVKDG VVATPAT - VGTILPGITRKS VIELARDRQ - VEERLVSIDDL
Zea_mays_GRMZM2G153536_T03	VFVVKGG VVATPDT - RGTILPGITRKS VIELARDRQ - VEERLVSIDDL
A. thaliana_AT1G50090	IFILKGN VISTPPT - SGTILPGITRKS ICELARDIGYE - VQERDLSVDEL
A. thaliana_AT1G50110	IFIVKGN VISTPPT - SGTILPGVTRKS ISELAHDIGYQ - VEERDVSVDEL
A. thaliana_AT3G19710	VFMLKGN VVSTPTI - AGTILPGVTRNCVMELCRDFGYQ - VEERTIPLVDF
G. max_Glyma07g30510	VFVLKDN AISTPAI - DGTILPGITRKS IIDIAIDLGYQ - VMERSVSVEEM
G. max_Glyma08g06750	VFVVKGN AICTPAT - NGAILPGITRKS IIEIALDMGYQ - VTERAISVEEM
M. truncatula_AC159872_36	IFVVKEN D IFTPAI - DGSILPGVTRKS IIDIAIDLGYK - VIERSISVEEM
P. trichocarpa_804339	IFVVKGN VISTPPI - AGTILPGITRKS IIEVASWLGQ - IEERAIPLLEL
P. trichocarpa_scaff_IX.827	VFMLKGN VISTPTI - LGTILPGITRKS IIEIAQDCGYE - VEEGRIPVEDV
M. truncatula_AC159872_55	AFIVKGV LSTAPT - LGTILPGVTRKS VIELARDLGYE - VMERKVSVEEL

FIGURE 6 (continued)

G.max_Glyma06g05280	IEADEVFCGTAVGVAPVGSITYQDKR-MEYITGSGT--ICQELNNTISG
H.annuus_TC54245	MEADEVFTTGTAVTVAIVGSIITYNGRR-VAYRTGDGL--VSQNLFKRLVG
A.thaliana_AT1G10070	MDADEVFCGTAVVAVPVGTITYQEKR-VEYKGTGDES--VCQKLRSLVVG
P.trichocarpa_scaff_II.1054	MDADEVFCGTAVGVAPVGTITYQDRR-VEYNVGEEES--VSQKLYSILEG
A.thaliana_AT1G10060	-----GRTKCSGRR-TEGS-----
A.thaliana_AT3G49680	LEADEVFCGTAVVSPVGSITYKGR-VSYGEGTFG-TVSKQLYTVLTS
A.thaliana_AT5G65780	VEADEVFCGTAVLSPVGSITYKSR-FSYGEDGFG-TVSKQLYTSLS
P.trichocarpa_BCAT4-like	LDAAEVFCGTAVVSPVGSITYKGR-VSYGVEGFG-AVSQQLYSLTK
G.max_Glyma01g40420	LDAAEVFCGTAVVSPVGSITYLGR-VTYG-DGIG-VVAQQLYTVLTR
G.max_Glyma11g04870	LDAAEVFCGTAVVSPVGSITYLDS-LFLL-ICCDFTLFYFLYTERLR
M.truncatula_TC114768	LEAAEVFCGTAVVSPVGSITYLGEK-KSYG-DGVG-AVSQQLYTGTR
S.lycopersicum_TC213629	LDAAEVFCGTAVVSPVGSITHQRR-VTYGNDGVG-LVVSQQLYSALTS
H.vulgare_TC165564	-----
O.sativa_LOC_Os05g48450	LEAAEVFCGTAVVSPVGSITYQGR-VEYAGNKGVGVSQQLYTSLS
Z.mays_TA12434_4577999	LDAAEVFCGTAVVSPVGSITYQGR-VEYG-HQGVGVVSQQLYTSLS
Hordeum_vulgare_PUT-169a-Horde	LDAAEVFCGTAVVSPVGSITYKGR-VKYDGNQGVGVVSQQLYTSLS
Hordeum_vulgare_subsp_vulgare_	LDAAEVFCGTAVVSPVGSITYKGR-VKYDGNQGVGVVSQQLYTSLS
O.sativa_Os03g0106400	LEAAEVFCGTAVVSPVGSITYLGR-VEY-GNQGVGVVSQQLYTSLS
O.sativa_LOC_Os04g47190	LDAAEVFCGTAVVSSVGCIVYQGR-VEYG-NQFRTVSQQLYSALTA
T.aestivum_TC320973	LDAAEVFCGTAVVSPVGSIVYHGR-VEYG-GKVGAVSQQLYSALTA
Zea_mays_GRMZM2G047347_T03	LGAAEVFCGTAVVSPVGSITYRGR-VEYGNQEAAGVVSQQLYAALTA
P.patens_TC31354	LAAEVFCGTAVVNPVGSITHGTNR-VQYN-NGAVGRVSQELYEALTT
P.patens_TC33668	LAAEVFCGTAVVNPVGSVTHGTR-VLYN-HGVVGGVSQELYEALTS
H.vulgare_TC186077	VGAAEVFCGTAVGVTPVSTITYQGR-HEFRTGEDT--LSRKLTYTTLTS
T.aestivum_TC325793	ISAAEVFCGTAVGITPVSTITYQGR-YEFRTGEDT--LSKKLYTALTS
O.sativa_LOC_Os03g12890	VGAAEVFCGTAVVAVPSSVTYHGQR-YEFRTGHD--LSQTLHTTTLTS
Zea_mays_GRMZM2G153536_T03	VAAAEVFCGTAVVAVPVSTVTYQGER-YEFRTGPD--VSQELYTTLTS
A.thaliana_AT1G50090	LEAAEVFCGTAVVIKAVETVTFHDKR-VKYRTGEEA--FSTKHLHLITN
A.thaliana_AT1G50110	LEAAEVFCGTAVVAVKAVETVTFHDKK-VKYRTGEEA--LSTKHLHMLTN
A.thaliana_AT3G19710	LDAAEAFCGTASIVTSIASVTFKDKK-TGFKTGEET--LAAKLYETLSD
G.max_Glyma07g30510	LGAAEMFCGTAVVNSVASVTYKETR-VDYKTGPAT--LSSKLRKTLVG
G.max_Glyma08g06750	LDAAEVFCGTAVVNSVSVTYKETR-TEYKTGPET--LSQKLRKTLVG
M.truncatula_AC159872_36	MSAAEVFCGTAVVNTSVASVTYKETR-AEYKTGAET--LSQKLRKTLVG
P.trichocarpa_804339	INVDEAFCSGTAIAIKPVGSVTYQGR-VEYKTGEGT--VSEKLCRRTLTG
P.trichocarpa_scaff_IX.827	LAAAEVFCGTAVVNTSVASITYQEQR-VEYKTGENT--VCHELRFTALTG
M.truncatula_AC159872_55	LEAAEVFCGTAVGISAVGSVTYKNKRCVTFKTGADT--VTKKLYDLITG

FIGURE 6 (continued)

G. max_Glyma06g05280	-IQGTIE-DKKGWIVEVD-----
H. annuus_TC54245	-IQVKVE-DKYTWIVDI-----
A. thaliana_AT1G10070	-IQGLIE-DNKGWTDIN-----
P. trichocarpa_scaff_II.1054	-IKGVIE-DKKGWIEIQ-----
A. thaliana_AT1G10060	-----
A. thaliana_AT3G49680	-LQMLIE-DNMKWTVNL-----
A. thaliana_AT5G65780	-LQMLSE-DNMNWTVQL-----
P. trichocarpa_BCAT4-like	-LQMLIE-DKMNWTVEL-----
G. max_Glyma01g40420	-LQMLTE-DEMNTVELR-----
G. max_Glyma11g04870	PYFPGPYL-SPFSPH-----
M. truncatula_TC114768	-LQMLAE-DNMNWTVELR-----
S. lycopersicum_TC213629	-LQMLSE-DKMGWIVELK-----
H. vulgare_TC165564	-----
O. sativa_LOC_Os05g48450	-LQMGQAE-DWLVTWVQL-----
Z. mays_TA12434_4577999	-LQMGQTE-DWMGWTVQLN-----
Hordeum_vulgare_PUT-169a-Horde	-LQMGHAE-DFMGWTVQLN-----
Hordeum_vulgare_subsp_vulgare_	-LQMGHAE-GPMGWTVQLN-----
O. sativa_Os03g0106400	-LQMGHVD-DCMGWTVELNQ-----
O. sativa_LOC_Os04g47190	-IQKGLVE-DSMGWTVQLN-----
T. aestivum_TC320973	-IQKGLVE-DSMGWSVQLN-----
Zea_mays_GRMZM2G047347_T03	-IQKGLTE-DSMGWTLQLT-----
P. patens_TC31354	-LQMGVSK-DEFDWMVVELV-----
P. patens_TC33668	-IQMGVSK-DEFDWMVVELA-----
H. vulgare_TC186077	-IQMGLAE-DKKGWTVAVD-----
T. aestivum_TC325793	-IQMGLAE-DKKGWTVAVD-----
O. sativa_LOC_Os03g12890	-IQMGLAE-DKKGWTVAVD-----
Zea_mays_GRMZM2G153536_T03	-IQMGMAAEDSKGTVAVE-----
A. thaliana_AT1G50090	-IQMGVVE-DKKGWMEIDHLVGTDSFPDET
A. thaliana_AT1G50110	-IQMGVVE-DKKGWMDIDPCQG-----
A. thaliana_AT3G19710	-IQTRVE-DTKGTVVEID-RQG-----
G. max_Glyma07g30510	-IQTGLE-DKKSWTVRVDSTI-----
G. max_Glyma08g06750	-IQTGCIE-DTKGTVRID-----
M. truncatula_AC159872_36	-IQTGCIE-DKKSWTVQVD-----
P. trichocarpa_804339	-IQTGLIE-DTMGWVVEIE-----
P. trichocarpa_scaff_IX.827	-IQTGLVE-DKKGWTVHLN-----
M. truncatula_AC159872_55	-IQTGLIE-DKKGWVVKID-----

FIGURE 6 (continued)

39/43

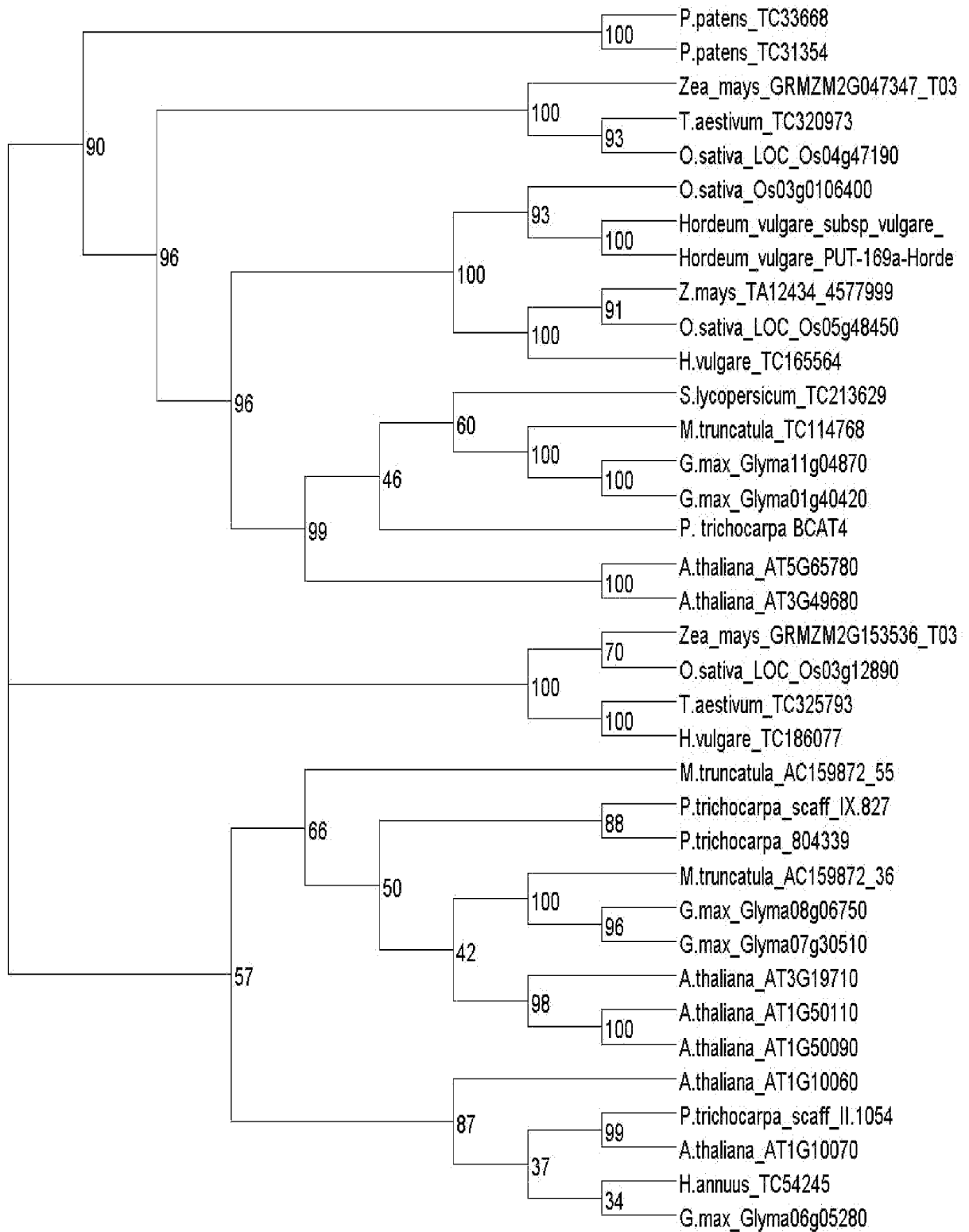


FIGURE 7

40/43

	1	2	3	4	5	6	7	8	9	10	11	12
1.		46.2	43.6	41.1	37.9	42.2	42.2	43.3	49.2	43.2	43.3	44.0
2.	58.5		50.1	51.5	46.5	56.7	54.2	57.1	64.7	55.7	58.2	54.1
3.	59.4	69.8		81.3	58.6	48.9	48.5	49.6	53.5	60.8	56.0	46.8
4.	59.6	70.4	89.9		62.1	48.2	48.1	50.0	52.3	61.0	54.4	47.1
5.	56.5	64.9	73.8	77.2		40.1	40.7	42.8	46.8	54.1	50.9	41.7
6.	53.8	72.6	66.1	66.3	59.3		79.3	72.6	57.2	51.7	51.9	69.2
7.	54.0	70.8	66.0	66.0	60.7	90.1		71.0	56.9	50.2	52.0	69.0
8.	53.3	72.6	66.6	67.3	61.5	83.5	83.1		58.8	52.4	53.7	85.7
9.	60.3	79.4	72.4	72.7	66.5	72.2	70.4	72.2		55.5	57.8	57.5
10.	59.1	72.2	76.3	77.7	74.1	66.8	67.2	66.1	71.1		71.0	48.7
11.	55.5	72.5	71.8	72.5	66.0	68.8	69.4	69.5	72.3	80.5		50.8
12.	54.5	70.3	64.5	63.5	59.9	79.4	80.2	89.1	70.6	64.7	65.9	
13.	58.9	77.6	71.3	71.1	65.5	69.5	69.6	74.6	76.8	70.3	69.0	70.3
14.	63.5	63.1	59.9	62.4	59.9	65.4	65.8	66.1	63.4	63.8	58.5	68.4
15.	56.2	75.1	69.3	69.0	63.7	73.8	74.2	74.8	75.3	71.8	72.5	71.0
16.	58.4	72.7	73.3	72.9	67.7	74.8	75.7	76.3	73.5	73.4	67.5	71.5
17.	58.6	72.4	73.0	72.6	67.4	74.6	75.4	76.0	73.2	73.2	67.3	71.5
18.	59.9	72.2	75.7	80.1	73.9	66.6	66.7	66.8	71.6	90.3	79.0	64.0
19.	55.9	72.9	74.7	76.3	71.8	64.9	65.3	65.6	71.4	74.9	67.8	63.0
20.	54.4	72.5	67.2	66.7	61.5	84.0	83.6	91.3	73.3	66.7	69.1	85.4
21.	53.6	71.6	65.9	65.6	59.4	75.5	75.7	73.8	71.6	68.0	68.3	71.2
22.	50.7	67.6	64.4	63.9	60.0	78.0	77.1	75.1	69.0	63.7	68.0	70.1
23.	53.5	70.9	67.2	67.2	62.1	79.9	81.2	82.1	72.4	68.0	68.7	78.1
24.	53.8	72.6	67.4	66.4	62.7	78.7	79.0	80.6	73.3	67.9	69.6	76.6
25.	50.0	65.8	60.7	61.4	55.3	72.1	73.5	70.3	65.8	60.5	63.7	67.8
26.	49.5	64.6	59.8	60.5	55.5	71.7	72.8	69.9	65.1	60.7	61.9	67.1
27.	60.5	71.1	72.2	77.5	73.2	65.4	64.6	65.1	69.8	76.9	68.3	61.8
28.	55.1	78.7	67.0	69.5	64.0	71.9	71.6	71.4	74.6	67.8	70.5	69.8
29.	52.4	69.9	64.4	67.0	61.2	69.4	69.4	66.7	68.4	67.9	69.4	65.6
30.	52.3	69.5	65.7	65.5	58.8	82.0	83.0	85.1	71.5	66.2	69.3	79.9
31.	55.7	72.6	66.8	67.1	60.5	84.5	85.1	87.7	71.9	67.8	70.0	84.0
32.	52.0	70.5	64.6	65.1	61.5	78.7	77.1	78.0	70.2	65.9	68.8	73.0
33.	57.1	75.4	68.7	68.7	62.9	73.8	74.0	74.1	72.9	71.4	73.0	71.3
34.	52.4	71.1	67.3	67.3	62.6	82.1	80.5	81.1	74.1	69.6	70.3	77.1
35.	50.9	67.9	64.9	63.5	60.5	76.5	77.1	76.8	68.9	66.9	68.9	71.3
36.	54.6	71.6	65.5	66.3	60.7	76.0	75.7	74.1	72.3	67.2	71.4	73.5

FIGURE 8

41/43

	13	14	15	16	17	18	19	20	21	22	23	24
1.	46.2	50.9	45.1	44.7	45.2	42.4	39.6	43.4	42.8	39.7	43.5	44.0
2.	62.2	47.8	58.3	58.5	58.2	54.4	51.7	57.3	58.2	51.7	55.5	57.6
3.	52.2	41.2	52.1	54.8	54.5	58.9	54.6	50.5	49.5	44.6	49.3	51.0
4.	52.7	41.9	53.9	54.2	53.9	62.0	54.4	49.5	50.7	45.4	49.9	49.8
5.	44.5	38.0	47.0	47.6	47.3	54.0	52.4	41.6	43.4	41.9	43.2	42.8
6.	54.8	55.5	57.8	63.8	63.5	51.1	48.3	74.0	59.1	56.4	68.8	68.2
7.	55.5	55.2	58.8	63.0	62.7	50.7	49.9	72.8	58.1	56.3	68.4	67.5
8.	58.7	56.9	59.7	66.7	66.5	51.3	49.5	82.9	60.4	56.8	70.6	69.4
9.	62.8	48.8	58.9	58.6	58.3	55.6	52.0	59.4	57.9	49.1	58.2	59.5
10.	54.4	45.0	53.0	57.8	57.6	76.4	54.5	51.8	51.5	48.0	51.7	51.0
11.	52.8	44.8	53.7	53.8	53.6	68.6	49.5	54.6	52.6	48.1	53.8	52.9
12.	55.4	58.3	58.0	61.9	61.9	48.3	47.3	77.5	58.0	53.1	66.9	64.3
13.		47.8	55.9	58.2	58.0	54.3	50.1	58.8	54.6	48.5	55.9	56.8
14.	64.2		48.7	57.8	57.8	43.7	41.1	57.2	49.2	48.7	72.4	65.2
15.	71.5	64.2		59.3	59.0	53.1	52.5	61.2	79.2	54.1	59.4	57.6
16.	73.2	73.2	75.1		99.7	56.4	54.5	65.5	57.9	58.6	72.4	77.1
17.	72.9	73.2	74.8	99.7		56.1	54.2	65.3	57.7	58.4	72.1	76.8
18.	69.7	63.9	70.5	72.6	72.3		53.9	51.8	50.2	45.6	51.8	50.1
19.	68.9	60.6	68.8	73.4	73.2	74.9		49.6	49.6	42.4	50.2	49.4
20.	72.5	67.6	75.2	75.7	75.5	66.7	65.9		60.6	56.6	71.5	69.6
21.	69.0	60.6	86.1	70.9	70.7	66.3	65.1	75.5		51.9	58.8	58.0
22.	65.4	64.6	71.5	75.1	74.9	62.9	63.2	75.9	70.7		59.3	59.7
23.	71.4	77.0	75.1	81.9	81.7	66.5	64.8	83.1	74.3	78.0		78.5
24.	71.1	73.1	75.6	84.9	84.7	66.7	65.2	80.6	73.3	77.1	87.3	
25.	64.6	59.4	68.9	67.6	67.4	60.5	61.6	71.5	71.0	66.2	70.8	71.2
26.	64.6	59.4	66.9	68.3	68.0	60.0	61.2	69.4	70.3	66.4	72.8	70.8
27.	70.5	63.7	67.0	72.3	72.1	76.2	72.3	65.0	64.9	62.4	65.0	66.2
28.	74.1	60.7	72.3	68.5	68.3	68.8	67.0	71.8	69.0	65.6	70.2	70.4
29.	65.6	57.2	70.1	65.1	64.8	67.2	65.6	69.1	70.8	66.3	67.0	68.2
30.	69.1	64.3	72.7	76.5	76.3	65.7	64.5	84.9	74.1	73.9	80.3	78.4
31.	70.5	68.3	74.3	75.8	75.5	67.6	64.6	86.2	73.3	76.3	81.8	80.1
32.	67.3	66.3	72.4	76.6	76.3	64.9	65.9	79.5	73.6	86.3	81.5	80.0
33.	71.2	62.4	96.0	75.2	74.9	70.7	67.9	76.0	85.6	71.5	74.6	75.6
34.	71.6	76.1	74.6	83.8	83.5	67.3	65.3	81.9	73.8	79.0	92.4	87.4
35.	66.7	63.7	72.6	76.5	76.3	65.4	65.7	76.2	72.6	82.9	77.3	77.3
36.	69.4	62.4	85.7	72.1	71.8	66.3	65.3	76.0	86.3	70.9	76.5	75.0

FIGURE 8 (continued)

42/43

	25	26	27	28	29	30	31	32	33	34	35	36
1.	39.3	38.1	40.0	43.3	38.7	43.8	45.3	40.5	44.4	41.1	40.2	43.0
2.	48.9	49.8	51.0	64.3	50.2	56.5	58.8	53.7	58.0	57.0	50.4	56.0
3.	44.7	45.1	53.4	51.0	46.7	50.1	49.4	46.4	51.1	49.6	45.6	48.6
4.	44.8	44.8	57.0	51.5	48.3	48.2	49.2	47.1	52.1	49.1	44.5	49.0
5.	37.9	38.8	52.3	46.7	44.2	40.2	42.5	42.1	44.5	44.3	40.2	42.3
6.	58.0	59.0	49.9	52.5	48.0	70.1	73.9	59.0	57.0	68.4	55.6	58.1
7.	57.4	57.1	48.7	52.8	48.3	69.3	73.0	57.6	57.0	66.4	57.5	56.1
8.	57.1	57.5	50.6	54.5	47.6	71.6	75.7	59.6	58.0	69.6	58.5	58.5
9.	51.6	50.9	51.5	60.0	49.2	58.1	60.4	51.1	57.6	57.4	50.5	57.8
10.	46.0	47.4	58.8	50.6	51.8	52.4	53.1	49.4	52.7	53.2	49.3	49.4
11.	47.6	48.0	54.3	54.9	52.0	53.4	56.3	49.2	53.5	53.5	49.4	52.4
12.	54.2	55.0	46.6	51.3	46.8	67.8	72.7	55.8	56.1	65.5	53.3	57.2
13.	50.2	50.7	52.0	59.1	46.4	56.6	58.1	52.6	53.9	56.9	49.4	54.6
14.	48.6	48.2	41.7	45.3	40.5	54.4	55.5	51.9	48.3	70.2	48.1	49.2
15.	53.1	53.3	50.4	55.6	51.8	58.2	59.4	55.0	92.5	58.7	52.6	77.0
16.	55.8	56.4	54.5	54.5	48.3	65.3	65.7	61.6	59.4	74.4	59.3	58.7
17.	55.6	56.1	54.2	54.2	48.1	65.1	65.5	61.3	59.1	74.2	59.0	58.5
18.	45.5	45.7	59.7	54.5	51.0	51.6	52.9	46.8	53.6	51.9	46.4	49.3
19.	44.4	44.0	51.1	47.6	50.4	47.3	48.4	46.7	51.3	50.6	44.7	48.9
20.	57.6	57.6	49.8	55.4	50.0	75.6	75.8	61.0	59.8	70.8	57.6	60.9
21.	53.8	55.9	49.3	53.4	49.5	58.9	58.6	53.2	77.7	59.5	52.1	77.8
22.	49.1	50.2	44.6	49.4	45.5	56.0	58.9	79.4	53.7	61.0	73.4	51.0
23.	57.4	57.7	48.4	53.4	48.5	68.6	68.8	63.3	58.6	88.3	58.5	59.4
24.	57.1	57.0	49.6	54.4	51.2	68.0	68.3	63.8	58.3	78.3	59.7	57.4
25.		83.8	45.7	49.1	46.3	57.1	57.4	52.2	52.0	55.6	50.1	54.5
26.	91.6		46.1	48.2	45.4	55.0	56.3	53.9	52.5	57.8	51.4	54.5
27.	59.6	59.4		50.3	53.1	48.2	50.1	46.8	49.9	49.4	45.9	48.1
28.	64.4	62.6	69.5		47.9	54.4	56.7	50.7	55.5	54.5	50.6	54.2
29.	64.6	63.7	67.5	68.9		49.4	49.4	46.3	50.9	49.0	47.3	50.2
30.	71.2	69.6	65.0	69.5	70.1		74.9	57.8	57.5	69.2	55.6	56.9
31.	69.9	68.7	65.4	72.2	67.7	86.1		59.9	58.8	67.6	58.1	56.9
32.	68.0	68.3	65.6	68.8	66.3	76.3	77.7		53.8	64.9	76.9	51.4
33.	67.6	66.2	66.9	72.7	69.6	73.1	73.8	72.4		58.0	52.3	77.2
34.	71.7	69.2	65.8	71.1	68.7	79.9	80.1	81.2	75.1		60.2	58.5
35.	67.4	67.1	64.4	68.4	67.0	75.8	76.0	84.1	71.9	78.8		51.1
36.	70.3	69.4	65.0	70.1	68.7	72.7	74.3	72.1	85.7	73.3	71.8	

FIGURE 8 (continued)

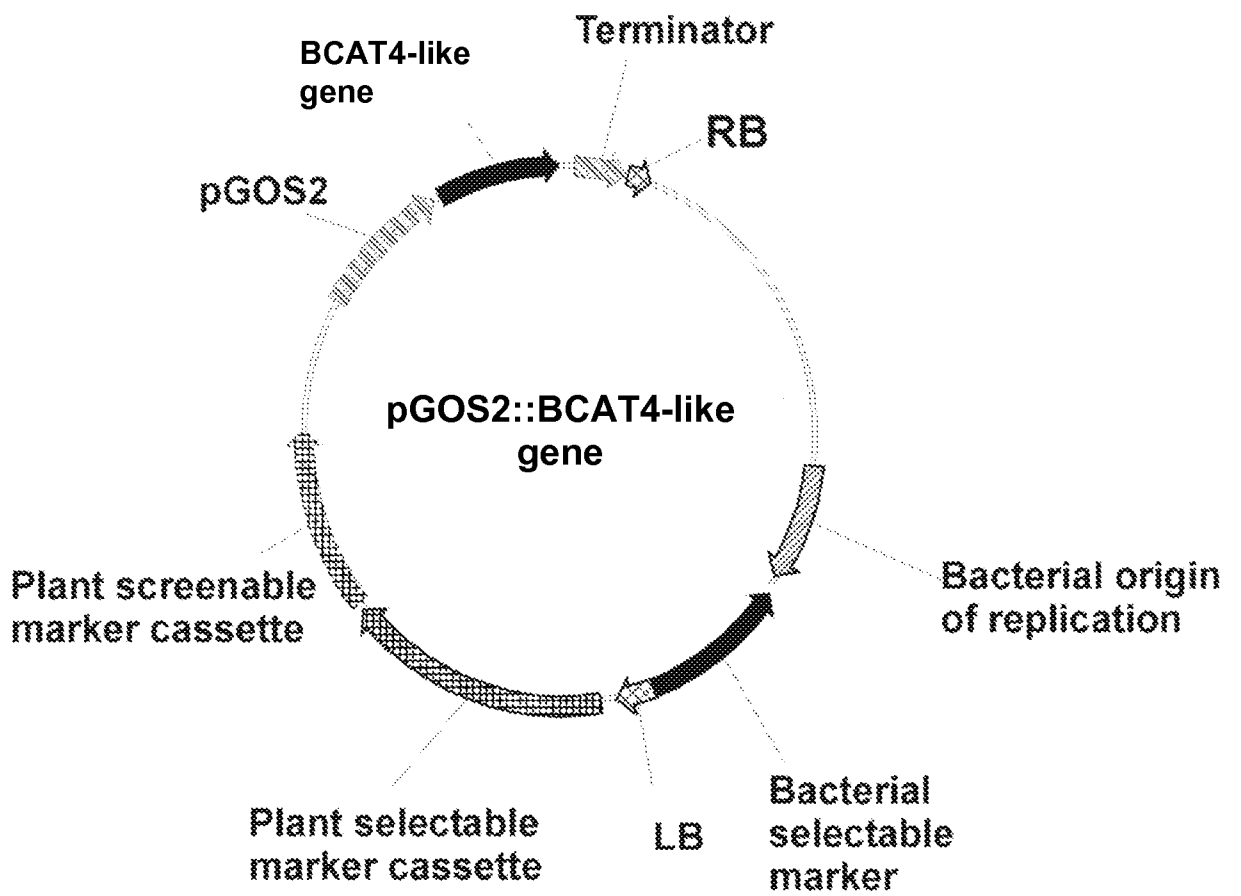


FIGURE 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2012/051938

A. CLASSIFICATION OF SUBJECT MATTER

See extra sheet

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: A01H; C12N; C07K

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRS; CNKI; WPI; EPODOC; ISI web of knowledge and keywords: plant, rice, yield, biomass, flowering time, enhance, bZIP, G-box, DNA binding protein, transgenic, etc.

GENBANK, EMBL: SEQ ID NOS:1, 2

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ABE, M. et al. FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. Science, Vol. 309, pages 1052-1056, 12 Aug. 2005 (12.08.2005) the Abstract; page 1052, right column, paragraph 2; page 1053, Table 1	1-9, 13-25
Y	the Abstract; page 1052, right column, paragraph 2; page 1053, Table 1	10-12

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&” document member of the same patent family</p>
--	---

Date of the actual completion of the international search 6 Sep. 2012 (06.09.2012)	Date of mailing of the international search report 20 Sep. 2012 (20.09.2012)
--	--

Name and mailing address of the ISA/CN
The State Intellectual Property Office, the P.R.China
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
100088
Facsimile No. 86-10-62019451

Authorized officer
SUN, Shangyu
Telephone No. (86-10)62412136

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2012/051938

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>GenBank Accession number CAA52896.1, 18 Apr. 2005 (18.04.2005), [retrieved on 06.09.2012], Retrieved from NCBI [online]: <URL: http://www.ncbi.nlm.nih.gov/protein/CAA52896.1></p> <p>the sequence</p>	10-12
A	<p>FUKAZAWA, J. et al. Repression of Shoot Growth, a bZIP Transcriptional Activator, Regulates Cell Elongation by Controlling the Level of Gibberellins. <i>The Plant Cell</i>, Vol. 12, pages 901-015, Jun. 2000 (06.2000)</p> <p>the whole text</p>	1-25
A	<p>ASSUNCAO, A. G. L. et al. Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. <i>Proceedings of the National Academy of Sciences</i>, Vol. 107, No. 22, pages 10296-10301, 1 Jun. 2010 (01.06.2010)</p> <p>the whole text</p>	1-25
A	<p>RODRIGUEZ-URIBE, L. et al. A root-specific bZIP transcription factor is responsive to water deficit stress in tepary bean (<i>Phaseolus acutifolius</i>) and common bean (<i>P. vulgaris</i>). <i>Journal of Experimental Botany</i>, Vol. 57, No. 6, pages 1391-1398, 10 Mar. 2006 (10.03.2006)</p> <p>the whole text</p>	1-25

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2012/051938

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See extra sheet.

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-25 (all partially), directed to the bZIP-like polypeptide represented by SEQ ID NO: 2 and subject-matter relating thereto.

- Remark on protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
 - The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
 - No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2012/051938

Continuation of: Box No. III

This Authority considers that there are 95 separate inventions covered by the claims indicated as follows:

Group I:

Invention 1: claims 1-25 (all partially)

a bZIP-like polypeptide according to SEQ ID NO: 2 and subject-matter relating thereto;

Inventions 2-59: claims 1-25 (all partially)

a bZIP-like polypeptide according to one of SEQ ID NOs: 4, 6, 8, ...118, and subject-matter relating thereto;

Group II:

Invention 60: claims 26-50 (all partially)

a BCAT4-like polypeptide according to SEQ ID NO:142 and subject-matter relating thereto;

Inventions 61-95: claims 26-50 (all partially)

a BCAT-like polypeptide according to one of SEQ ID NOs: 144, 146, 148, ...212, and subject-matter relating thereto.

The reason for which the inventions are not so linked as to form a single general inventive concept as required by Rule 13.1 PCT, are as follows:

The problem solved by the present inventions can be viewed as the provision of proteins for enhancing yield-related traits in plants. The solution as identified by the common concept linking the inventions together is to modulate expression of one of the bZIP-like peptides of Inventions 1-59 or the BCAT4-like peptide of Inventions 61-95 for enhancing yield-related traits in plants.

However, this concept cannot be regarded as the "single general inventive concept" required by Rules 13.1 PCT because it is not novel or inventive in view of the prior art.

D1 (ABE, M. et al. FD, a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex, Science, Vol.309, 1052-1056, 12 Aug. 2005) discloses FD, a bZIP protein, which mediates floral pathway and leads to early flowering time (see the whole text, especially Table 1 in Page 1053). Early flowering time is one of the yield-related traits. Hence D1 discloses the usage of bZIP protein in improving at least one of the yield-related traits.

SCHUSTER, J. et al. (Branched chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in Arabidopsis, The Plant Cell, Vol. 18, 2664-2679, Oct. 2006) discloses that BCAT4 is involved in glucosinolate (see the Abstract). TSAO, R. et. al. (Insecticidal Toxicities of Glucosinolate-containing Extracts from Crambe Seeds, J. Agric. Entomol. 13(2):109-120, Apr. 1996) discloses insecticidal toxicity of glucosinolates. In the light of SCHUSTER, J. and TSAO, R., it would be obvious for one in the art that modulating expression of BCAT4 in plants would improve the insect resistant and thereby increase the yield under biotic stresses.

Hence, there is no inventive concept linking the Inventions 1-95. The application hence does not meet the requirements of unity of invention as defined in Rules 13.1 and 13.2 PCT.

As no additional search fee has been paid, this international search report covers only the invention 1.

CLASSIFICATION OF SUBJECT MATTER

A01H 1/00 (2006.01) i

A01H 5/00 (2006.01) i

C12N 15/29 (2006.01) i

C12N 15/63 (2006.01) i

C07K 14/415 (2006.01) i