



- (51) **International Patent Classification:** Not classified
- (21) **International Application Number:** PCT/IB2014/061904
- (22) **International Filing Date:** 3 June 2014 (03.06.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13171889.2 13 June 2013 (13.06.2013) EP
13178738.4 31 July 2013 (31.07.2013) EP
- (71) **Applicant:** 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:** PLESCH, Gunnar; Plantagenhof 1, 14482 Potsdam (DE). BLAU, Astrid; Bahnhofstraße 110, 14532 Stahnsdorf (DE). HEROLD, Michael Manfred; Quitzowstr. 87, 10551 Berlin (DE). KAMLAGE, Beate; Varziner Str. 13/14, 12161 Berlin (DE). WENDEL, Birgit; Alt-Wittenau 67 B, 13437 Berlin (DE). PUZIO, Piotr; Wolvengracht 4, B-9030 Mariakerke (Gent) (BE). TAMAN-CHARDONNENS, Agnes; Brilliant Starstraat 16, NL-1611 DS Bovenkarspel (NL). HENDRIKS, Jan-neke; Fliedeweg 5, 14548 Schwielowsee (DE). MARTIN, Jerome; Weldadigheidstraat 16, B-9050 Gent (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).
- Published:**
- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
 - with sequence listing part of description (Rule 5.2(a))

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

(57) **Abstract:** The present invention relates generally to the field of molecular biology and concerns a method for enhancing various economically important yield-related traits in plants. More specifically, the present invention concerns a method for enhancing one or more yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a POI (Protein Of Interest) polypeptide. The present invention also concerns plants having modulated expression of a nucleic acid encoding a POI polypeptide, which plants have one or more enhanced yield-related traits compared with control plants. The invention also provides constructs, useful in performing the methods of the invention.



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

This application claims priority of applications with number EP 13171889.2 and EP 13178738.4, all of which are incorporated by reference in their entirety.

5

The present invention relates generally to the field of plant molecular biology and concerns a method for enhancing one or more yield-related traits in plants by modulating expression in a plant of a nucleic acid encoding a POI (Protein Of Interest) polypeptide. The present invention also concerns plants having modulated (i.e. increased or decreased) expression of a nucleic acid encoding a POI polypeptide, which plants have one or more one enhanced yield-related traits relative to corresponding wild type plants or other control plants. The invention also provides constructs useful in the methods, uses, plants, harvestable parts and products of the invention.

10

15

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.

20

25

35

A trait of 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.

30

A further trait of economic interest in a number of crops is the harvestable total sugar in the crop. An increase of the sugar concentration or sugar content in the harvestable parts of the crop is often desired.

40

Seed yield is an 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.

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, nutrient deficiency, extremes of temperature, chemical toxicity and oxidative stress. The ability to improve plant tolerance to abiotic stress would be of great economic advantage to farmers worldwide and would allow for the cultivation of crops during adverse conditions and in territories where cultivation of crops may not otherwise be possible.

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

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

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

Finding nucleic acids encoding a POI to enhance yield-related traits of plants is an ongoing challenge. The problem to be solved in this invention was to provide a method for enhancing one or more yield-related traits in plants by increasing the expression in a plant of a nucleic acid encoding a POI polypeptide.

The problem was solved by increasing expression of a nucleic acid encoding a POI being an "ATP:NAD(H) 2'-phosphotransferase" (also called ATP-NAD(H) kinase or ANK in the following) in plants, wherein said plants have one or more enhanced yield-related traits compared with control plants as a result of the increased expression of ANK as defined herein. In the following, ATP-NAD(H) kinase, also called ANK, means ATP-NAD⁺ kinase

and/or ATP-NADH kinase.

Members of the ANK family include ATP-NAD(H) kinases of enzyme class EC:2.7.1.23 and / or EC:2.7.1.86, which catalyses the phosphorylation of NAD(H) to NADP(H) utilising ATP and other nucleoside triphosphates as well as inorganic polyphosphate as a source of phosphorus. The reaction which is catalysed: $\text{ATP} + \text{NAD(H)} = \text{ADP} + \text{NADP(H)} + \text{H}^+$.

US 20100129886 discloses such an ANK in SEQ ID NO: 720 (called NADH kinase) which is encoded by the nucleic acid disclosed in SEQ ID NO: 719 (called POS5 gene). The sequences are disclosed in the context of a method of making recombinant yeast cells for the production of isobutanol in mitochondria with ATP-NAD(H) kinase activity.

The present invention concerns a method for enhancing one or more yield-related traits in plants by increasing the expression in a plant of a nucleic acid encoding an ANK polypeptide. The present invention also concerns plants having increased expression of a nucleic acid encoding an ANK polypeptide, which plants have one or more enhanced yield-related traits compared with control plants. The invention also provides hitherto unknown constructs comprising ANK-encoding nucleic acids, useful in performing the methods of the invention.

A preferred embodiment is a method for enhancing one or more yield-related traits in a plant relative to control plants, comprising the steps of increasing the expression, preferably by recombinant methods, in a plant of a nucleic acid encoding an ANK polypeptide preferably said nucleic acid is exogenous, wherein preferably the expression is under the control of a promoter sequence operably linked to the nucleic acid encoding the ANK polypeptide, and growing the plant.

These inventive methods comprise increasing the expression in a plant of a nucleic acid encoding an ANK polypeptide and thereby enhancing one or more yield-related traits of said plant compared to the control plant. "Thereby enhancing" is to be understood to include direct effects of increasing the expression of the ANK polypeptide as well as indirect effects as long as the increased expression of the ANK polypeptide encoding nucleic acid results in an enhancement of at least one of the yield-related traits. For example overexpression of a transcription factor A may increase transcription of another transcription factor B that in turn controls the expression of a number of genes of a given pathway leading to enhanced biomass or seed yield. Although transcription factor A does not directly enhance the expression of the genes of the pathway leading to enhanced yield-related traits, increased expression of A is the cause for the effect of enhanced yield-related-trait(s).

Enhanced yield-related traits comprise e.g. increased biomass (above-ground and / or below-ground) and / or increased seed yield, and / or increased sugar yield (either as harvestable sugar per plant, per fresh weight, per dry weight or per area) relative to control

plants.

Hence, it is an object of the invention to provide an expression cassette and a vector construct comprising a nucleic acid encoding an ANK polypeptide, operably linked to a beneficial promoter sequence. The use of such genetic constructs for making a transgenic plant having one or more enhanced yield-related traits relative to control plants is provided.

Also a preferred embodiment are transgenic plants transformed with one or more expression cassettes of the invention, and thus, expressing in a particular way the nucleic acids encoding an ANK protein, wherein the plants have one or more enhanced yield-related trait. Harvestable parts of the transgenic plants of the present invention and products derived from the transgenic plants and their harvestable parts are also part of the present invention.

Brief description of the several views of the drawings

Throughout the figures, for each sequence of Table 1 the short-name given in Table 1 below is used to represent the sequence.

Table 1: Examples of ANK nucleic acids and polypeptides

Source	Nucleic acid SEQ ID NO:	Protein SEQ ID NO:	Short-name in figures
SACCHAROMYCES CEREVISIAE	1	2	YPL188W
Pichia membranifaciens	3	4	H1
Saccharomyces cerevisiae S288c	5	6	H2
Ashbya gossypii ATCC 10895	7	8	H3
Scheffersomyces stipitis CBS 6054	9	10	H4
Meyerozyma guilliermondii ATCC 6260	11	12	H5
Lodderomyces elongisporus NRRL YB-4239	13	14	H6
Vanderwaltozyma polyspora DSM 70294	15	16	H7
Candida dubliniensis CD36	17	18	H8
Komagataella pastoris GS115	19	20	H9
Zygosaccharomyces rouxii CBS 732	21	22	H10
Lachancea thermotolerans CBS 6340	23	24	H11
Clavispora lusitaniae ATCC 42720	25	26	H12
Eremothecium cymbalariae DBVPG#7215	27	28	H13
Torulaspora delbrueckii	29	30	H14
Tetrapisispora phaffii CBS 4417	31	32	H15
Kazachstania africana CBS 2517	33	34	H16
Candida glabrata CBS 138	35	36	H17

Kluyveromyces lactis NRRL Y-1140	37	38	H18
Debaryomyces hansenii CBS767	39	40	H19
Yarrowia lipolytica CLIB122	41	42	H20
Saccharomyces cerevisiae (strain FostersB)	43	44	H21
Saccharomyces cerevisiae (strain Zymaflore VL3)	45	46	H22
Pichia angusta (strain ATCC 26012 / NRRL Y-7560 / DL-1)	47	48	H23
Naumovozya castellii (strain ATCC 76901 / CBS 4309 / NBRC 1992 / NRRL Y-12630)	49	50	H24
Naumovozya dairenensis (strain ATCC 10597 / BCRC 20456 / CBS 421 / NBRC 0211 / NRRL Y-12639)	51	52	H25

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

Fig. 1 shows the vector pMTX155 used for cloning the gene of interest for non-targeted expression according to Example 6.

Fig. 2 represents a multiple alignment of various ANK polypeptides using ClusalW according to Example 2. The single letter code for amino acids is used. White letters on black background indicate identical amino acids among the various protein sequences, white letters on grey background represent highly conserved amino acid substitutions. These alignments can be used for defining further motifs or signature sequences, when using conserved amino acids, i.e. those identical in the aligned sequences and/or those highly conserved. YPL188W equals the ANK polypeptide of SEQ ID NO: 2. The other sequences are identified by their respective short-name according to Table 1, which provides the details for each sequence such as organism and SEQ ID NOs.

Fig. 3 shows the MATGAT results for sequence identity analysis of Example 3. In the figure YPL188W equals the ANK polypeptide of SEQ ID NO: 2. The other sequences are identified by their respective short-name according to Table 1, which provides the details for each sequence such as organism and SEQ ID NO.

Fig. 4 shows the NEEDLE results for sequence identity analysis of Example 3. In the figure YPL188W equals the ANK polypeptide of SEQ ID NO: 2. The other sequences are identified by their respective short-name according to Table 1, which provides the details for each sequence such as organism and SEQ ID NO.

Fig. 5 provides tables showing the relations of the different SEQ ID NOs. to the lead sequence. YPL188W represents the ANK sequences of SEQ ID NO: 1 and 2.

Detailed description of the invention

- 5 The present invention shows that increasing expression in a plant of a nucleic acid encoding a POI polypeptide gives plants having one or more enhanced yield-related traits relative to control plants.

10 According to a first embodiment, the present invention provides a method for enhancing one or more yield-related traits in plants relative to control plants, comprising increasing expression in a plant of a nucleic acid encoding a POI polypeptide and optionally selecting for plants having one or more enhanced yield-related traits. According to another embodiment, the present invention provides a method for producing plants having one or more enhanced yield-related traits relative to control plants, wherein said method comprises

15 the steps of increasing expression in said plant of a nucleic acid encoding a POI polypeptide as described herein and optionally selecting for plants having one or more enhanced yield-related traits.

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

20

Any reference hereinafter to a “protein useful in the methods of the invention” is taken to mean a POI 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 POI polypeptide. In one embodiment any reference to a protein or nucleic acid “useful in the methods of the invention” is to be understood to mean proteins or nucleic acids “useful in the methods, constructs, plants, harvestable parts and products of the invention”. The nucleic acid to be introduced into a plant (and therefore useful in performing the methods of the invention) is any nucleic acid encoding the type of protein which will now be described, hereafter also named “*POI* nucleic acid” or “*POI* encoding nucleic acid” or “*POI* gene” or “gene of interest”.

25

30

Notably, the terms “POI encoding nucleic acid”, “POI nucleic acid”, “POI gene”, “POI nucleotide sequence” and “POI encoding nucleotide sequence” are used interchangeably herein.

35

A “POI polypeptide” as defined herein refers to any polypeptide preferably comprising at least one domain selected from the group of

- IPR002504 ATP-NAD/AcoX kinase;
- 40 • IPR016064 ATP-NAD kinase PpnK-type;
- IPR017437 ATP-NAD kinase, PpnK-type, all beta; and
- PFAM PF01513 “NAD_kinase”

when analysed with the InterProScan software tool, version 4.8 and InterPro release 42, and further conferring, more preferably under non-stress conditions, one or more enhanced yield-related traits relative to control plants when expression of said polypeptide is increased. In the following, "POI" and "ANK" (short for ATP-NAD(H) kinase) are used interchangeably. ANK may therefore replace POI wherever feasible.

Preferably the ANK polypeptide comprises one or more motifs and/or domains as defined herein. Motifs 1 to 7 were derived using the MEME algorithm version 3.5 (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. The motifs identified were converted into ProSite patterns using the software tool PRATT version 2.1 or manually. For details see Example 4.

Table 2: Pattern sequences according to Fuzzpro (see Example 4). The table shows the identified pattern sequences i.e. motifs of the ANK polypeptides in ProSite annotation and their location within SEQ ID NO: 2.

Motif	SEQ ID	Pattern-Sequence	Start match	End match	mis-matches
1	55	L-D-[IV]-[FY]-[IV]-[DGN]-x(3)-[FLM]-T-x-T-x-A-D-G-[ILV]-x-[FLV]-[AS]-[ST]-P-T-G-S-T-A-Y-S-L-S-[AS]-G-G-S-[IM]-[TV]-x-P-x-[IV]-x-[ACS]-[IV]-L-[LM]-T-P-[IV]-C-P- <u>R</u> -S-L-S-F-R-P-[LV]	241	300	0
2	56	V-x(0,3)-L-[AS]-F-[AS]-[LM]-G-T-L-G-F-L-L-P-F-[ADEN]-F-x(6)-F-x(2)-[MV]-x(2)-[GS]-[QR]-[AGS]-x(2)-[LV]-x-R-x-R-[IL]-[AEQ]-C	165	209	0
3	57	N-[IV]-x-[ILMV]-x-K-K-P-x(4)-[ATV]-x-[ADEQS]-A-x-[IV]-x-[FL]-I-x-H-[ILM]-H-x(2)-Y-P-x(2)-N-x(8,9)-E-[ILV]	69	111	0
4	58	T-G-x(3)-[DEQ]-I-x(2)-[KR]-x-[DE]-[LM]-[ILMV]-[IV]-[ST]-L-G-G-D-G-T-I-L-x-[AGS]-V-S-x-F	131	160	0
5	59	G-x(2)-C-[IV]-A-x(4)-[DQ]-W-x(2)-[DG]-[IL]-N-[DEGNQS]-L-L-G-F-N-S-[GS]-F	379	404	0
6	60	H-A-M-N-D-[IV]-x-[IL]-H-R-G-[GNS]-x-P-x-L	223	238	0
7	61	D-G-x-[APST]-x(3)-L-x(2)-[DG]-D-[EQS]-[ILV]-[HY]-[IV]	330	345	0

The ability of ANK to discriminate NADH from NAD⁺ is largely dependent on a single conserved Arg residue (position 293 is highlighted bold and underlined in Table 2 Motif 1).

Preferably, the position corresponding to this conserved Arg residue is not a site of mismatch, which is in correlation with the homologs H1 to H25 as given in their short-names in Table 1 (the respective pattern position is highlighted bold and underlined in Table 13; see Example 4).

5 In a further embodiment, the ANK polypeptide as used herein comprises at least one domain selected from the group of

- IPR002504 ATP-NAD/AcoX kinase;
- IPR016064 ATP-NAD kinase PpnK-type;
- IPR017437 ATP-NAD kinase, PpnK-type, all beta; and
- 10 • PFAM PF01513 "NAD_kinase"

when analysed with the InterProScan software tool, version 4.8 and InterPro release 42 (see Example 4 for details), and at least one of the motifs ANK pattern 1 (SEQ ID NO: 55), ANK pattern 2 (SEQ ID NO: 56), ANK pattern 3 (SEQ ID NO: 57), ANK pattern 4 (SEQ ID NO: 58), ANK pattern 5 (SEQ ID NO: 59), ANK pattern 6 (SEQ ID NO: 60), and ANK
15 pattern 7 (SEQ ID NO: 61), also called motif 1 to 7, respectively, as shown in Table 2, wherein in the letter-numbers combination

–x(a,b)–

of any motif the letter x stands for Xaa, i.e. any amino acid, and the integer numbers a and b give the minimum and the maximum number of Xaa that may be found after the amino
20 acid preceding the x. For example, S-x(0,3)-P indicates that following the amino acid Serine either one, two or three amino acids of any choice may be included before a Proline residue, or that no amino acid is to be found between the Serine and the Proline residue of this motif.

Consequently, the letters

25 –x(2)

indicate that exactly two amino acids of any type are found at this position of the motif. A single –x without a number in brackets- indicates that one amino acid residue of any type is present at this position of the motif.

30 Moreover any amino acid residue(s) replacing –x may be identical to or different from the amino acid residue preceding or succeeding it, or any other amino acid inserted instead of the –x at the same or any other position.

Residues within square brackets represent alternatives, e.g. the pattern Y-x(21,23)-[FW]

35 means that a conserved tyrosine is separated by minimum 21 and maximum 23 amino acid residues from either a phenylalanine or tryptophan.

In one preferred embodiment, the ANK polypeptide comprises one or more motifs selected from Motif 1, Motif 2, Motif 3, Motif 4, Motif 5, Motif 6, and Motif 7.

40

In still another embodiment, the ANK polypeptide comprises in increasing order of preference, at least 2, at least 3, at least 4, at least 5, at least 6 and/or at least 7 motifs.

In a preferred embodiment the ANK polypeptide comprises at least one or more of the following domains:

- IPR002504 ATP-NAD/AcoX kinase;
- IPR016064 ATP-NAD kinase PpnK-type;
- IPR017437 ATP-NAD kinase, PpnK-type, all beta; and
- PFAM PF01513 "NAD_kinase"

when analysed with the InterProScan software tool, version 4.8 and InterPro release 42 (see Example 4 for details), and

- 1) all of the motifs 1-7 according to Table 2 or
- 2) any 6 of the motifs listed under 1); or
- 3) any 5 of the motifs listed under 1); or
- 4) any 4 of the motifs listed under 1); or
- 5) any 3 of the motifs listed under 1); or
- 6) any two of the motifs listed under 1); or
- 7) one of the motifs as described herein above.

In another preferred embodiment the ANK polypeptide contains at least two of the Motifs 1 to 7 in the same order within the polypeptide sequence as found in the polypeptide of SEQ ID NO: 2, i.e. from the N-terminus to the C-terminus in the order of their starting position Motif 3, then Motif 4, then Motif 2, then Motif 6, then Motif 1, then Motif 7, and then Motif 5. In a further more preferred embodiment the ANK polypeptide contains all of the Motifs 1 to 7 in the same order within the polypeptide sequence as found in the polypeptide of SEQ ID NO: 2, i.e. from the N-terminus to the C-terminus in the order of their starting position Motif 3, then Motif 4, then Motif 2, then Motif 6, then Motif 1, then Motif 7, and then Motif 5.

In another preferred embodiment the ANK polypeptide contains when optimally aligned with the polypeptide sequence of SEQ ID NO: 2 an Arg at the position corresponding to position 293 of SEQ ID NO: 2.

Additionally or alternatively, the ANK 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: 2, provided that the homologous protein comprises any one or more of the motifs as outlined above. The overall sequence identity is determined using a global alignment algorithm, such as the Needleman Wunsch algorithm in the program GAP (GCG Wisconsin Package, Accelrys), preferably with default parameters and preferably with sequences of mature proteins (i.e. without taking into account secretion signals or transit peptides). In one embodiment the sequence identity

level is determined by comparison of the polypeptide sequences over the entire length of the sequence of SEQ ID NO: 2. Alternatively the sequence identity is determined by comparison of a nucleic acid sequence to the sequence encoding the mature protein in SEQ ID NO: 1. In another embodiment the sequence identity level of a nucleic acid sequence is determined by comparison of the nucleic acid sequence over the entire length of the coding sequence of the sequence of SEQ ID NO: 1.

In another embodiment, the sequence identity level is determined by comparison of one or more conserved domains or motifs in SEQ ID NO: 2 with corresponding conserved domains or motifs in other ANK polypeptides. Compared to overall sequence identity, the sequence identity will generally be higher when only conserved domains or motifs are considered. Preferably the motifs in a ANK 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: 55 to SEQ ID NO: 61 (Motifs 1 to 7). In other words, in another embodiment a method for enhancing one or more yield-related traits in plants is provided wherein said ANK 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 (or motif, respectively) starting with amino acid positions as given in Table 12 for the domains and for motifs in Table 2 ("start match" of Table 2 up to and including amino acid positions as given in column "end match" of Table 2 for motifs of SEQ ID NO: 2).

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

Preferably, the polypeptide sequence which when used in the construction of a phylogenetic tree, clusters with the group of ANK polypeptides comprising the amino acid sequence represented by SEQ ID NO: 2 rather than with any other group. In another embodiment the polypeptides of the invention when used in the construction of a phylogenetic tree cluster not more than 4, 3, or 2 hierarchical branch points away from the amino acid sequence of SEQ ID NO: 2.

Furthermore, ANK polypeptides (at least in their native form) typically have biological activity of an ATP-NAD(H) Kinase, hereinafter called ANK activity. Tools and techniques for measuring an ATP-NAD(H) kinase activity, also called ANK activity are well known in the art (e.g. Jacobsen and Jacobsen (1976), Arch Biochem Biophys 175: 627-34). Further details are provided in Example 12. In addition, nucleic acids encoding ANK polypeptides, when expressed in plants according to the methods of the present invention as outlined in Examples 6 to 11 and 13, give plants having increased yield-related traits, in particular biomass and/or sugars, preferably under non-stress conditions. Another function of the nucleic acid sequences encoding ANK polypeptides is to confer information for synthesis of

the ANK protein that increases yield or yield-related traits as described herein, when such a nucleic acid sequence of the invention is transcribed and translated in a living plant cell. In one embodiment, the ANK polypeptide, when expressed in plants according to the methods of the present invention as outlined in Examples 6 to 11 is mainly localized in mitochondria.

- 5 According one embodiment, there is provided a method for improving yield-related traits as provided herein in plants relative to control plants, comprising increasing expression in a plant of a nucleic acid encoding a POI polypeptide as defined herein. Preferably said one or more enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased seed yield relative to control
- 10 plants, and preferably comprise increased aboveground biomass, increased below-ground biomass, increased seed yield and/or increased sugar yield (either as harvestable sugar per plant, per fresh weight, per dry weight or per area) relative to control plants.

Increased sugar yield may be due to increased sugar content and / or increased sugar concentration per plant, per fresh weight, per dry weight and / or per area. In one preferred

15 embodiment the sugar yield of only the harvestable parts, more preferably the aboveground harvestable parts optionally excluding seed and / or the below-ground harvestable parts is increased. In another preferred embodiment the increased sugar yield is an increased yield of sucrose, glucose and / or fructose.

20 In an embodiment the nucleic acid sequences employed in the methods, uses, constructs, plants, harvestable parts and products of the invention are nucleic acid molecule selected from the group consisting of:

- (i) a nucleic acid represented by SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, or 51, preferably SEQ ID NO: 1;
- 25 (ii) the complement of a nucleic acid represented by SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, or 51, preferably SEQ ID NO: 1;
- (iii) a nucleic acid encoding an ANK polypeptide having in increasing order of preference at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%,
- 30 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 or 52, preferably SEQ ID NO: 2, comprising
- 35 one or more motifs having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one or more of the motifs given in SEQ ID NO: 55 to SEQ ID NO: 61, and further preferably conferring one or more enhanced yield-related traits relative to control plants; and
- 40 (iv) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (iii) under high stringency hybridization conditions and preferably confers one or more enhanced yield-related traits relative to control plants;

- (v) any of (i) to (iv) above wherein the nucleic acid encodes a polypeptide with ANK activity.

Or encode a polypeptide selected from the group consisting of:

- 5 (i) an amino acid sequence represented by SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 or 52, preferably SEQ ID NO: 2;
- 10 (ii) an amino acid sequence having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 or 52, preferably SEQ ID NO: 2, comprising one or more motifs having in
- 15 increasing order of preference, and additionally or alternatively comprising one or more motifs having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one or more of the motifs given in SEQ ID NO: 55 to SEQ ID NO: 61, and further preferably conferring one or more enhanced yield-related traits relative to control
- 20 plants; and
- (iii) an amino acid sequence having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,
- 25 97%, 98%, or 99% sequence identity to the amino acid sequence represented by SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 or 52, preferably SEQ ID NO: 2, comprising one or more motifs having in increasing order of preference, and additionally or alternatively comprising one or more motifs having in increasing order of preference at least 50%, 55%, 60%, 65%,
- 30 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one or more of the motifs given in SEQ ID NO: 55 to SEQ ID NO: 61, and wherein said ANK polypeptide contains at least two of the Motifs 1 to 7 in the same order within the polypeptide sequence as found in the polypeptide of SEQ ID NO: 2, i.e. from the N-terminus to the C-terminus in the order of their starting position Motif 3,
- 35 then Motif 4, then Motif 2, then Motif 6, then Motif 1, then Motif 7, and then Motif 5, and further preferably conferring one or more enhanced yield-related traits relative to control plants; and
- (iv) derivatives of any of the amino acid sequences given in (i) or (ii) above;
- (v) any of (i) to (iii) above wherein the amino acid sequence has ANK activity.
- 40

In one embodiment the nucleic acids useful in the methods of the invention are those listed in Tables IA as lead or homologue, or those encoding the protein sequences listed in tables

IIA as lead or homologues (Tables IA and IIA: see Figure 5), or those of the patterns according to SEQ ID NO: 55 to SEQ ID NO: 61 (see Table 2).

In one embodiment the nucleic acid sequences employed in the methods, constructs, plants, harvestable parts and products of the invention are sequences encoding but excluding those nucleic acids encoding the polypeptide sequences disclosed in US 20100129886, SEQ ID NO: 720.

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 ANK-encoding nucleic acid or ANK polypeptide as defined herein. The term "POI" or "POI polypeptide" as used herein also intends to include homologues as defined hereunder of SEQ ID NO: 2.

Examples of nucleic acids encoding ANK polypeptides are given in Table 1. Such nucleic acids are useful in performing the methods of the invention. The amino acid sequences given in Table 1 are example sequences of orthologues and paralogues of the ANK 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 *Saccharomyces cerevisiae* sequences.

With respect to the sequences of the invention or useful in the methods, constructs, plants, harvestable parts and products of the invention, in one embodiment a nucleic acid or a polypeptide sequence originating not from higher plants is used in the methods of the invention or the expression construct useful in the methods of the invention. In a preferred embodiment, the ANK polypeptide is of fungal, preferably of yeast origin. Non-plant polypeptides when expressed in plants are often not subject to the same endogenous regulatory and/or degradation mechanism as plant polypeptides. This can increase the biological activity and/or function of said ANK polypeptide in plants overexpressing the ANK polypeptide.

In one embodiment of the present invention, any reference to one or more enhanced yield-related trait(s) is meant to exclude the restoration of the expression and/or activity of the ANK polypeptide in a plant in which the expression and/or the activity of the ANK polypeptide has been reduced or disabled when compared to the original wildtype plant or original variety. For example, the overexpression of the ANK polypeptide in a knock-out mutant variety of a plant, wherein said ANK polypeptide or an orthologue or paralogue has been knocked-out is not considered enhancing one or more yield-related trait(s) within the meaning of the current invention when the expression and/or enzymatic activity level is

substantially the same as in the wildtype (i.e. non-mutant type).

The invention also provides ANK-encoding nucleic acids and ANK polypeptides useful in the methods, constructs, plants, harvestable parts and products of the invention and these are those sequences as described herein.

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

- (i) a nucleic acid represented by any one of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, or 51, preferably SEQ ID NO: 1;
- (ii) the complement of a nucleic acid represented by any one of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, or 51, preferably SEQ ID NO: 1;
- (iii) a nucleic acid having, in increasing order of preference at least 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with any of the nucleic acid sequences of Table 1, preferably SEQ ID NO: 1, and further preferably conferring one or more enhanced yield-related traits relative to control plants;
- (iv) a nucleic acid encoding the polypeptide as represented by any one of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, preferably as a result of the degeneracy of the genetic code, said isolated nucleic acid can be derived from a polypeptide sequence as represented by (any one of) SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, preferably SEQ ID NO: 2, and further preferably confers one or more enhanced yield-related traits relative to control plants;
- (v) a nucleic acid encoding an ANK 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: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, preferably SEQ ID NO: 2, and additionally or alternatively comprising one or more motifs having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one or more of the motifs given in SEQ ID NO: 55 to SEQ ID NO: 61, and further preferably conferring one or more enhanced yield-related traits relative to control plants; and
- (vi) a nucleic acid encoding an ANK 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: 2, 4, 6, 8, 10, 12, 14, 16, 18 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, preferably SEQ ID NO: 2, and additionally or alternatively comprising one or more motifs having in increasing order of preference at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one or more of the motifs given in SEQ ID NO: 55 to SEQ ID NO: 61, and wherein said ANK polypeptide contains at least two of the Motifs 1 to 7 in the same order within the polypeptide sequence as found in the polypeptide of SEQ ID NO: 2, i.e. from the N-terminus to the C-terminus in the order of their starting position Motif 3, then Motif 4, then Motif 2, then Motif 6, then Motif 1, then Motif 7, and then Motif 5, and further preferably conferring one or more enhanced yield-related traits relative to control plants; and
- (vii) a nucleic acid molecule which hybridizes with a nucleic acid molecule of (i) to (vi) under stringent hybridization conditions and preferably confers one or more enhanced yield-related traits relative to control plants; and
- (viii) any of (i) to (vii) above wherein the nucleic acid encodes a polypeptide with ANK activity.

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

- (i) an amino acid sequence represented by any one of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, preferably SEQ ID NO: 2;
- (ii) an amino acid sequence having, in increasing order of preference, at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence represented by any one of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, or 52, preferably SEQ ID NO: 2, and further preferably conferring one or more enhanced yield-related traits relative to control plants; and
- (iii) derivatives of any of the amino acid sequences given in (i) or (ii) above.
- (iv) any of (i) to (iii) above wherein the amino acid sequence has ANK activity.

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 1, the terms "homologue" and "derivative" being as defined herein. Also useful in the methods, constructs, plants, harvestable parts and products of the invention are nucleic acids encoding homologues and

derivatives of orthologues or paralogues of any one of the amino acid sequences given in Table 1. 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 ANK polypeptides, nucleic acids hybridising to nucleic acids encoding ANK polypeptides, splice variants of nucleic acids encoding ANK polypeptides, allelic variants of nucleic acids encoding ANK polypeptides and variants of nucleic acids encoding ANK polypeptides obtained by gene shuffling. The terms hybridising sequence, splice variant, allelic variant and gene shuffling are as described herein.

Nucleic acids encoding ANK 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 one or more yield-related traits in plants, comprising introducing, preferably by recombinant methods, and expressing in a plant a portion of any one of the nucleic acid sequences given in Table 1, or a portion of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table 1. Preferably, the portion of a nucleic acid encodes a polypeptide having ANK activity.

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

Portions useful in the methods, constructs, plants, harvestable parts and products of the invention, encode an ANK polypeptide as defined herein or at least part thereof, and have substantially the same biological activity as the amino acid sequences given in Table 1. Preferably, the portion is a portion of any one of the nucleic acids given in Table 1, or is a portion of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table 1. Preferably the portion is at least 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1608 consecutive nucleotides in length, the consecutive nucleotides being of any one of the nucleic acid sequences given in Table 1, or of a nucleic acid encoding an orthologue or paralogue of any one of the amino acid sequences given in Table 1. Most preferably the portion is a portion of the nucleic acid of SEQ ID NO: 1. Preferably, the portion encodes a fragment of an amino acid sequence which comprises any of the Motifs 1 to 7, and/or has biological activity of an ATP-NAD(H) Kinase (ANK activity), and/or has at least 37%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, or 99% sequence identity to SEQ ID NO: 2.

Another nucleic acid variant useful in the methods, constructs, plants, harvestable parts and products of the invention is a nucleic acid capable of hybridising, under reduced stringency conditions, preferably under stringent conditions, with a nucleic acid encoding an ANK polypeptide as defined herein, or with a portion as defined herein. According to the present invention, there is provided a method for enhancing one or more yield-related traits in plants, comprising introducing, preferably by recombinant methods, 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 1, or to the complement of a nucleic acid encoding an orthologue, paralogue or homologue of any one of the proteins given in Table 1.

Hybridising sequences useful in the methods, constructs, plants, harvestable parts and products of the invention encode an ANK polypeptide as defined herein, having substantially the same biological activity as the amino acid sequences given in Table 1. Preferably, the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding any one of the proteins given in Table 1, 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 1. Most preferably, the hybridising sequence is capable of hybridising to the complement of a nucleic acid encoding the polypeptide as represented by SEQ ID NO: 2 or to a portion thereof. In one embodiment, the hybridization conditions are of medium stringency, preferably of high stringency, as defined herein.

Preferably, the hybridising sequence encodes a polypeptide with an amino acid sequence which comprises any one or more of the Motifs 1 to 7, and/or has biological activity of an ATP-NAD(H) Kinase (ANK activity), and/or has at least 37%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 2.

In another embodiment, there is provided a method for enhancing one or more yield-related traits in plants, comprising introducing, preferably by recombinant methods, and expressing in a plant a splice variant of a nucleic acid encoding any one of the proteins given in Table 1, or a splice variant of a nucleic acid encoding an orthologue, paralogue or homologue of any of the amino acid sequences given in Table 1.

Preferred splice variants are splice variants of a nucleic acid represented by SEQ ID NO: 1, or a splice variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2. Preferably, the amino acid sequence encoded by the splice variant comprises any one or more of the Motifs 1 to 7, and/or has biological activity of an ATP-NAD(H) Kinase (ANK activity), and/or has at least 37%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 2.

In yet another embodiment, there is provided a method for enhancing one or more yield-related traits in plants, comprising introducing, preferably by recombinant methods, and expressing in a plant an allelic variant of a nucleic acid encoding any one of the proteins given in Table 1, or comprising introducing, preferably by recombinant methods, 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 1.

The polypeptides encoded by allelic variants useful in the methods of the present invention have substantially the same biological activity as the ANK polypeptide of SEQ ID NO: 2 and any of the amino acid sequences depicted in Table 1. Allelic variants exist in nature, and encompassed within the methods of the present invention is the use of these natural alleles. Preferably, the allelic variant is an allelic variant of SEQ ID NO: 1 or an allelic variant of a nucleic acid encoding an orthologue or paralogue of SEQ ID NO: 2. Preferably, the amino acid sequence encoded by the allelic variant comprises any one or more of the Motifs 1 to 7, and/or has biological activity of an ATP-NAD(H) Kinase (ANK activity), and/or has at least 37%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 2.

In another embodiment the polypeptide sequences useful in the methods, constructs, plants, harvestable parts and products of the invention have substitutions, deletions and/or insertions compared to the sequence of SEQ ID NO: 2, wherein the amino acid substitutions, insertions and/or deletions may range from 1 to 10 amino acids each.

In yet another embodiment, there is provided a method for enhancing one or more yield-related traits in plants, comprising introducing, preferably by recombinant methods, and expressing in a plant a variant of a nucleic acid encoding any one of the proteins given in Table 1, or comprising introducing, preferably by recombinant methods, 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 1, which variant nucleic acid is obtained by gene shuffling.

Preferably, the amino acid sequence encoded by the variant nucleic acid obtained by gene shuffling comprises any of the Motifs 1 to 7, and/or has biological activity of an ATP-NAD(H) Kinase (ANK activity), and/or has at least 37%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 2. 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.). ANK polypeptides differing from the sequence of SEQ ID NO: 2 by one or several amino acids (substitution(s), insertion(s) and/or deletion(s) as defined herein) may equally be useful to increase the yield of plants in the methods and constructs and plants of the invention. For

example, the nucleic acid encoding the ANK polypeptide of SEQ ID NO: 6 can be generated from the nucleic acid encoding the ANK polypeptide of SEQ ID NO: 2 by alteration of several nucleotides. To this end, SEQ ID NO: 5 can be derived from SEQ ID NO: 1 by altering the nucleic acid at position 539 from T to C by site-directed mutagenesis using standard PCR based methods as described herein.

Nucleic acids encoding ANK 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 ANK polypeptide-encoding nucleic acid is from fungal, preferably of yeast origin, more preferably originates from *Saccharomyces cerevisiae*.

The inventive methods for enhancing one or more yield-related traits in plants as described herein comprising introducing, preferably by recombinant methods, and expressing in a plant the nucleic acid(s) as defined herein, and preferably the further step of growing the plants and optionally the step of harvesting the plants or part(s) thereof.

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. DNA comprised within a cell, particularly a cell with cell walls like a plant cell, is better protected from degradation, damage and / or breakdown than a bare nucleic acid sequence. The same holds true for a DNA construct comprised in a host cell, for example a plant cell.

In a preferred embodiment the invention relates to compositions comprising the recombinant chromosomal DNA of the invention and/or the construct of the invention, and a host cell, preferably a plant cell, wherein the recombinant chromosomal DNA and/or the construct are comprised within the host cell, preferably within a plant cell or a host cell with a cell wall. In a further embodiment said composition comprises dead host cells, living host cells or a mixture of dead and living host cells, wherein the recombinant chromosomal DNA and/or the construct of the invention may be located in dead host cells and/or living host cell. Optionally the composition may comprise further host cells that do not comprise the recombinant chromosomal DNA of the invention or the construct of the invention. The compositions of the invention may be used in processes of multiplying or distributing the recombinant chromosomal DNA and/or the construct of the invention, and or alternatively to protect the recombinant chromosomal DNA and/or the construct of the invention from breakdown and/or degradation as explained above. The recombinant chromosomal DNA of the invention and/or the construct of the invention can be used as a quality marker of the compositions of the invention, as an indicator of origin and/or as an indication of producer.

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.

- 5 In another embodiment, the methods of the present invention may be performed under stress conditions, preferably under abiotic 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
- 10 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
- 15 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.
- 20 In a preferred embodiment the methods of the invention are performed using plants in need of increased abiotic stress-tolerance for example tolerance to drought, salinity and/or cold or hot temperatures and/or nutrient use due to one or more nutrient deficiency such as nitrogen deficiency.
- 25 Performance of the methods of the invention gives plants having one or more enhanced yield-related traits. In particular performance of the methods of the invention gives plants having increased sugars and/or increased yield, especially increased biomass and/or increased seed yield relative to control plants. The terms "yield" and "seed yield" are described in more detail in the "definitions" section herein.
- 30 Increase in biomass (weight) of one or more parts of a plant 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 roots such as taproots, stems, beets, leaves, flowers or seeds, and performance of the methods of the
- 35 invention results in plants having increased seed yield relative to the seed yield of control plants, and / or increased stem biomass relative to the stem biomass of control plants, and / or increased root biomass relative to the root biomass of control plants and / or increased beet biomass relative to the beet biomass of control plants. Moreover, it is particularly contemplated that the sugar content, in particular the glucose, fructose and/or sucrose
- 40 content in the stem (in particular of sugar cane plants) and/or in the root (in particular in sugar beets) is increased relative to the sugar content, in particular the glucose, fructose and/or sucrose content, in the stem and/or in the root of the control plant.

In a preferred embodiment throughout this application any reference to “root” as biomass or as harvestable parts or as organ e.g. of increased sugar content is to be understood as a reference to harvestable parts partly inserted in or in physical contact with the ground such as but not limited to beets and other hypocotyl areas of a plant, rhizomes, stolons or creeping rootstalks, but not including leaves, as well as harvestable parts belowground, such as but not limited to root, taproot, tubers or bulbs. In another embodiment aboveground parts or aboveground harvestable parts or aboveground biomass are to be understood as aboveground vegetative biomass not including seeds and/or fruits.

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

Performance of the methods of the invention results in plants having increased seed yield relative to the seed yield of control plants, and/or increased aboveground biomass, in particular stem biomass relative to the aboveground biomass, and in particular stem biomass of control plants, and/or increased root biomass relative to the root biomass of control plants and/or increased beet biomass relative to the beet biomass of control plants. Moreover, it is particularly contemplated that the sugar content (in particular the glucose, fructose and/or sucrose content) in the above ground parts, particularly stem (in particular of sugar cane plants) and/or in the belowground parts, in particular in roots including taproots and tubers, and/or in beets (in particular in sugar beets) is increased relative to the sugar content (in particular the glucose, fructose and/or sucrose content) in corresponding part(s) of the control plant

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

Performance of the methods of the invention gives plants grown under conditions of drought, increased yield-related traits relative to control plants grown under comparable

conditions. Therefore, according to the present invention, there is provided a method for increasing yield-related traits in plants grown under conditions of drought which method comprises increasing expression in a plant of a nucleic acid encoding an ANK polypeptide.

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

- In one embodiment of the invention, root biomass is increased, preferably beet and / or taproot biomass, more preferably in sugar beet plants, and optionally seed yield and / or
20 above ground biomass are not increased. In another embodiment of the invention, above ground biomass is increased, preferably stem, stalk and / or sett biomass, more preferably in Poaceae, even more preferably in a Saccharum species, most preferably in sugarcane, and optionally seed yield, below-ground biomass and / or root growth is not increased.

- 25 In a further embodiment the total harvestable sugar, preferably glucose, fructose and/or sucrose, is increased, preferably in addition to increased other yield-related traits as defined herein, for example biomass, and more preferably also in addition to an increase in sugar content, preferably glucose, fructose and/or sucrose content.

- 30 The invention also provides genetic constructs and vectors to facilitate introduction and/or expression in plants of nucleic acids encoding ANK 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
35 of the invention.

More specifically, the present invention provides a construct comprising:

- (a) an isolated nucleic acid encoding an ANK polypeptide as defined above;
(b) one or more control sequences capable of driving expression of the nucleic acid
40 sequence of (a); and optionally
(c) a transcription termination sequence.

Preferably, the nucleic acid encoding an ANK polypeptide is as defined above. The term “control sequence” and “termination sequence” are as defined herein.

In particular the genetic construct of the invention is a plant expression construct, i.e. a genetic construct that allows for the expression of the nucleic acid encoding an ANK polypeptide in a plant, plant cell or plant tissue after the construct has been introduced into this plant, plant cell or plant tissue, preferably by recombinant means. The plant expression construct may for example comprise said nucleic acid encoding an ANK polypeptide in functional linkage to a promoter and optionally other control sequences controlling the expression of said nucleic acid in one or more plant cells, wherein the promoter and optional the other control sequences are not natively found in functional linkage to said nucleic acid. In a preferred embodiment the control sequence(s) including the promoter result in overexpression of said nucleic acid when the construct of the invention has been introduced into a plant, plant cell or plant tissue.

The genetic construct of the invention may be comprised in a host cell - for example a 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 trait(s) to a plant when it has been introduced into said plant, which plant expresses the nucleic acid encoding the ANK polypeptide comprised in the genetic construct and preferably resulting in increased abundance of the ANK polypeptide. In another embodiment the genetic construct of the invention confers increased yield or yield-related trait(s) to a plant comprising plant cells in which the construct has been introduced, which plant cells express the ANK encoding nucleic acid comprised in the genetic construct.

The promoter in such a genetic construct may be a promoter not native to the nucleic acid described above, i.e. a promoter different from the promoter regulating the expression of the ANK encoding nucleic acid in its native surrounding. In a particular embodiment the nucleic acid encoding the ANK polypeptide useful in the methods, constructs, plants, harvestable parts and products of the invention is in functional linkage to a promoter resulting in the expression of the ANK encoding nucleic acid in

- aboveground biomass preferably the leaves and shoot, more preferably the stem, of monocot plants, preferably Poaceae plants, more preferably Saccharum species plants; and / or
- leaves, belowground biomass and / or root biomass, preferably tubers, taproots and / or beet organs, more preferably taproot and beet organs of dicot plants, more preferably

Solanaceae and / or Beta species plants.

The expression cassette or the genetic construct of the invention may be comprised in a host cell, plant cell, seed, agricultural product or plant.

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).

In one embodiment the ANK encoding nucleic acid is operably linked to a plant suitable promoter. Preferably, this promoter is a constitutive promoter. The constitutive promoter is preferably a ubiquitous constitutive promoter of high strength and/ or preferentially expressed in green tissue. More preferably the promoter is a Big35S promoter or a promoter of substantially the same strength and having substantially the same expression pattern (a functionally equivalent promoter). See the "Definitions" section herein for further examples of constitutive promoters.

It should be clear that the applicability of the present invention is not restricted to the ANK polypeptide-encoding nucleic acid represented by SEQ ID NO: 1, nor is the applicability of the invention restricted to the Big35S promoter when expression of an ANK polypeptide-encoding nucleic acid is driven by a constitutive promoter.

Yet another embodiment relates to genetic constructs useful in the methods, constructs, plants, harvestable parts and products of the invention wherein the genetic construct comprises the ANK encoding nucleic acid of the invention functionally linked to a promoter as disclosed herein above and further functionally linked to one or more of

1) nucleic acid expression enhancing nucleic acids (NEENAs):

- a) as disclosed in the international patent application published as WO 2011/023537 in table 1 on page 27 to page 28 and / or SEQ ID NO: 1 to 19 and / or as defined in items i) to vi) of claim 1 of said international application which NEENAs are herewith incorporated by reference; and / or
- b) as disclosed in the international patent application published as WO 2011/023539 in table 1 on page 27 and / or SEQ ID NO: 1 to 19 and / or as defined in items i) to vi) of claim 1 of said international application which NEENAs are herewith incorporated by reference; and / or
- c) as contained in or disclosed in:
 - i. the European priority application filed on 05 July 2011 as EP 11172672.5 in table 1 on page 27 and / or SEQ ID NO: 1 to 14937, preferably SEQ ID NO: 1 to 5, 14936 or 14937, and / or as defined in items i) to v) of claim 1 of said European priority application which NEENAs are herewith incorporated by reference; and / or

ii. the European priority application filed on 06 July 2011 as EP 11172825.9 in table 1 on page 27 and / or SEQ ID NO: 1 to 65560, preferably SEQ ID NO: 1 to 3, and / or as defined in items i) to v) of claim 1 of said European priority application which NEENAs are herewith incorporated by reference;
and / or

d) equivalents having substantially the same enhancing effect;
and / or

2) functionally linked to one or more Reliability Enhancing Nucleic Acid (RENA) molecule

a) as contained in or disclosed in the European priority application filed on 15 September 2011 as EP 11181420.8 in table 1 on page 26 and / or SEQ ID NO: 1 to 16 or 94 to 116666, preferably SEQ ID NO: 1 to 16, and / or as defined in point i) to v) of item a) of claim 1 of said European priority application which RENA molecule(s) are herewith incorporated by reference; or

b) equivalents having substantially the same enhancing effect.

A preferred embodiment of the invention relates to a nucleic acid molecule useful in the methods, constructs, plants, harvestable parts and products of the invention and encoding an ANK polypeptide of the invention under the control of a promoter as described herein above, wherein the NEENA, RENA and / or the promoter is heterologous to the ANK encoding nucleic acid molecule of the invention.

Optionally, one or more terminator sequences may be used in the construct introduced into a plant. Those skilled in the art will be aware of terminator sequences that may be suitable for use in performing the invention. Preferably, the construct comprises an expression cassette comprising a Big35S promoter (SEQ ID NO: 64; see also Kay R. et al, Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes, Science (1987) 236: 1299-1302; Omirulleh S. et al, Activity of a chimeric promoter with doubled CaMV 35S enhancer element in protoplast-derived cells and transgenic plants in maize, Plant Mol. Biol. (1993) 21: 415-428), operably linked to the nucleic acid encoding the ANK polypeptide. Furthermore, one or more sequences encoding selectable markers may be present on the construct introduced into a plant.

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

As mentioned above, a preferred method for increasing expression of a nucleic acid encoding an ANK polypeptide is by introducing, preferably by recombinant methods, and expressing in a plant a nucleic acid encoding an ANK polypeptide; however the effects of performing the method, i.e. enhancing one or more 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 one or more enhanced yield-related traits relative to control plants, comprising introduction and expression in a plant of any nucleic acid encoding an ANK polypeptide as defined herein.

More specifically, the present invention provides a method for the production of transgenic plants having one or more enhanced yield-related traits, particularly increased (seed) yield, which method comprises:

- (i) introducing and expressing in a plant or plant cell a recombinant ANK polypeptide-encoding nucleic acid or a genetic construct comprising an ANK polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development, preferably promoting plant growth and development of plants having one or more enhanced yield-related traits relative to control plants; and
- (iii) optionally selecting plants with increased yield-related trait(s) due to increased expression of the ANK polypeptide. Preferably, the introduction of the ANK polypeptide-encoding nucleic acid is by recombinant methods.

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

Cultivating the plant cell under conditions promoting plant growth and development, may or may not include regeneration and / or growth to maturity. Accordingly, in a particular embodiment of the invention, the plant cell transformed by the method according to the invention is regenerable into a transformed plant. In another particular embodiment, the plant cell transformed by the method according to the invention is not regenerable into a transformed plant, i.e. cells that are not capable to regenerate into a plant using cell culture techniques known in the art. While plants cells generally have the characteristic of totipotency, some plant cells cannot 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. One example concerns plant cells that do not sustain themselves through photosynthesis by synthesizing carbohydrate and protein from such inorganic substances as water, carbon dioxide and mineral salt.

In yet another embodiment the invention relates to transgenic plant cells and / or transgenic plant parts of the invention wherein said plant cells and / or plant parts are non-propagatable.

In a further embodiment the invention relates to dead plant cells comprising the construct, recombinant chromosomal DNA and / or polynucleotide and / or polypeptide of the invention. These dead cells cannot be used to regenerate a plant and are not photosynthetically active.

5

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

10

In one embodiment the methods of the invention are methods for the production of a transgenic Poaceae plant, preferably a Saccharum species plant, a transgenic part thereof, or a transgenic plant cell thereof, having one or more enhanced yield-related traits relative to control plants, comprising the steps of

15

- (i) introducing and expressing in a plant or plant cell a recombinant ANK polypeptide-encoding nucleic acid or a genetic construct comprising an ANK polypeptide-encoding nucleic acid; and
- (ii) cultivating the plant cell under conditions promoting plant growth and development, preferably promoting plant growth and development of plants having one or more enhanced yield-related traits relative to control plants; and
- (iii) optionally selecting plants with increased yield-related trait(s) due to increased expression of the ANK polypeptide, and further comprises the step of harvesting setts and/or gems from the transgenic plant and planting the setts and/or gems and growing the setts and/or gems to plants, wherein the setts and/or gems comprises the exogenous nucleic acid encoding the ANK polypeptide and the promoter sequence operably linked thereto.

20

25

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.

30

The present invention encompasses plants or parts thereof (including seeds) obtainable by the methods according to the present invention. The plants or plant parts or plant cells comprise a nucleic acid transgene encoding an ANK 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 substantially the same genotypic and / or phenotypic characteristic(s) as those produced by the parent in the methods according to the invention.

35

40

In a further embodiment the invention extends to seeds recombinantly comprising the

expression cassettes of the invention, the genetic constructs of the invention, or the nucleic acids encoding the ANK and / or the ANK polypeptides as described above. Typically a plant grown from the seed of the invention will also show enhanced yield-related traits.

- 5 The invention also includes host cells containing an isolated nucleic acid encoding an ANK 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
10 inventive method. In a particular embodiment the plant cells of the invention overexpress the nucleic acid molecule of the invention.

In a further embodiment the invention relates to a transgenic pollen grain comprising the construct of the invention and / or a haploid derivate of the plant cell of the invention.

- 15 Although in one particular embodiment the pollen grain of the invention cannot be used to regenerate an intact plant without adding further genetic material and / or is not capable of photosynthesis, said pollen grain of the invention may have uses in introducing the enhanced yield-related trait into another plant by fertilizing an egg cell of the other plant using a live pollen grain of the invention, producing a seed from the fertilized egg cell and
20 growing a plant from the resulting seed. Further pollen grains find use as marker of geographical and / or temporal origin.

- 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
25 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,
30 rapeseed, linseed, cotton, tomato, potato, Stevia species such as but not limited to Stevia rebaudiana 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,
35 emmer, spelt, einkorn, teff, milo and oats. In a particular embodiment the plants of the invention or used in the methods of the invention are selected from the group consisting of maize, wheat, rice, soybean, cotton, oilseed rape including canola, sugarcane, sugar beet and alfalfa. Advantageously the methods of the invention are more efficient than the known methods, because the plants of the invention have increased yield and / or tolerance to an
40 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, setts, sugarcane gems, roots, rhizomes, tubers and bulbs, which harvestable parts comprise a recombinant nucleic acid encoding an ANK polypeptide as defined herein. In particular, such harvestable parts are roots such as taproots, rhizomes, fruits, stems, beets, tubers, bulbs, leaves, flowers and / or seeds. In one
5 embodiment harvestable parts are stem cuttings (like setts or gems of sugar cane).

The invention furthermore relates to products derived or produced, preferably directly derived or directly produced, from one or more harvestable part(s) of such a plant, such as dry pellets, pulp pellets, pressed stems, setts, sugarcane gems, meal or powders, fibres, cloth, paper or cardboard containing fibres produced by the plants of the invention, oil, fat
10 and fatty acids, carbohydrates (including starches, paper or cardboard containing carbohydrates produced by the plants of the invention), sap, juice, molasses, syrup, chaff or proteins. Preferred carbohydrates are starches, cellulose, molasses, syrup and / or sugars, preferably sucrose. Also preferred products are residual dry fibers, e.g., of the stem (like bagasse from sugar cane after cane juice removal), molasses, syrups and / or filtercake,
15 preferably from sugarcane and / or sugar beet. Said products can be agricultural products.

In one embodiment the product comprises a recombinant nucleic acid encoding an ANK polypeptide and / or a recombinant ANK polypeptide for example as an indicator of the particular quality of the product. In another embodiment the invention relates to anti-
20 counterfeit milled seed, milled stem and / or milled root having as an indication of origin and / or as an indication of producer a plant cell of the invention and / or the construct of the invention, wherein milled root preferably is milled beet, more preferably milled sugar beet.

The invention also includes methods for manufacturing a product comprising a) growing the
25 plants of the invention and b) producing said product from or by the plants of the invention or parts thereof, including stem, sett, sugarcane gem, root, beet and/or 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
30 embodiment the method of the invention is a method for manufacturing cloth by a) growing the plants of the invention that are capable of producing fibres usable in cloth making, e.g. cotton, b) removing the harvestable parts as described herein from the plants, and c) producing fibres from said harvestable part and d) producing cloth from the fibres of c).

Another embodiment of the invention relates to a method for producing feedstuff for
35 bioreactors, fermentation processes or biogas plants, comprising a) growing the plants of the invention, b) removing the harvestable parts as described herein from the plants and c) producing feedstuff for bioreactors, fermentation processes or biogas plants. In a preferred embodiment the method of the invention is a method for producing alcohol(s) from plant material comprising a) growing the plants of the invention, b) removing the harvestable
40 parts as described herein from the plants and c) optionally producing feedstuff for fermentation process, and d) - following step b) or c) - producing one or more alcohol(s) from said feedstuff or harvestable parts, preferably by using microorganisms such as fungi,

algae, bacteria or yeasts, or cell cultures. A typical example would be the production of ethanol using carbohydrate containing harvestable parts, for example corn seed, sugarcane stem parts or beet parts of sugar beet. In one embodiment, the product is produced from the stem of the transgenic plant. In another embodiment the product is produced from the root, preferable taproot and / or beet of the plant.

In another embodiment the method of the invention is a method for the production of one or more polymers comprising a) growing the plants of the invention, b) removing the harvestable parts as described herein from the plants and c) producing one or more monomers from the harvestable parts, optionally involving intermediate products, d) producing one or more polymer(s) by reacting at least one of said monomers with other monomers or reacting said monomer(s) with each other. In another embodiment the method of the invention is a method for the production of a pharmaceutical compound comprising a) growing the plants of the invention, b) removing the harvestable parts as described herein from the plants and c) producing one or more monomers from the harvestable parts, optionally involving intermediate products, d) producing a pharmaceutical compound from the harvestable parts and / or intermediate products. In another embodiment the method of the invention is a method for the production of one or more chemicals comprising a) growing the plants of the invention, b) removing the harvestable parts as described herein from the plants and c) producing one or more chemical building blocks such as but not limited to Acetate, Pyruvate, lactate, fatty acids, sugars, amino acids, nucleotides, carotenoids, terpenoids or steroids from the harvestable parts, optionally involving intermediate products, d) producing one or more chemical(s) by reacting at least one of said building blocks with other building block or reacting said building block(s) with each other.

The present invention is also directed to a product obtained by a method for manufacturing a product, as described herein. In a further embodiment the products produced by the manufacturing methods of the invention are plant products such as, but not limited to, a foodstuff, feedstuff, a food supplement, feed supplement, fibre, cosmetic or pharmaceutical. In another embodiment the methods for production are used to make agricultural products such as, but not limited to, fibres, plant extracts, meal or presscake and other leftover material after one or more extraction processes, flour, proteins, amino acids, carbohydrates, fats, oils, polymers, vitamins, and the like. Preferred carbohydrates are sugars, preferably sucrose. In one embodiment the agricultural product is selected from the group consisting of 1) fibres, 2) timber, 3) plant extracts, 4) meal or presscake or other leftover material after one or more extraction processes, 5) flour, 6) proteins, 7) carbohydrates, 8) fats, 9) oils, 10) polymers e.g. cellulose, starch, lignin, lignocellulose, and 11) combinations and / or mixtures of any of 1) to 10). In a preferable embodiment the product or agricultural product does generally not comprise living plant cells, does comprise the expression cassette, genetic construct, protein and / or polynucleotide as described herein.

In yet another embodiment the polynucleotides and / or the polypeptides and / or the

constructs 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
5 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.

10 A further embodiment of the invention is a commercial package comprising

1. propagules of the plants of the invention, such as but not limited to setts or gems of sugarcane, and / or
2. comprising the plant cells of the invention, and / or
- 15 3. comprising the polynucleotides and /or the polypeptides and / or the constructs of the invention comprised in an agricultural product, and / or
4. comprising the recombinant chromosomal DNA of the invention.

A further embodiment of the invention is a protective covering comprising

- 20 1. propagules of the plants of the invention, such as but not limited to setts or gems of sugarcane, and / or
2. comprising the plant cells of the invention, and / or
3. comprising the polynucleotides and /or the polypeptides and / or the constructs of the invention comprised in an agricultural product, and / or
- 25 4. comprising the recombinant chromosomal DNA of the invention.

The protective covering is any kind of repository which allows safe-keeping of the material according to points 1 to 4 above. On the one hand the protective covering can be re-usable and / or re-sealable. On the other hand the protective covering can be of one-way nature
30 and /or biodegradable. Preferably, the protective covering is a commercial package. More preferably, the protective covering is testa.

The present invention also encompasses use of nucleic acids encoding ANK polypeptides as described herein and use of these ANK polypeptides in enhancing any of the
35 aforementioned yield-related traits in plants. For example, nucleic acids encoding ANK polypeptide described herein, or the ANK polypeptides themselves, may find use in breeding programmes in which a DNA marker is identified which may be genetically linked to an ANK polypeptide-encoding gene. The nucleic acids / genes, or the ANK polypeptides themselves may be used to define a molecular marker. This DNA or protein marker may
40 then be used in breeding programmes to select plants having one or more enhanced yield-related traits as defined herein in the methods of the invention. Furthermore, allelic variants of an ANK polypeptide-encoding nucleic acid / gene may find use in marker-assisted

breeding programmes. Nucleic acids encoding ANK 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.

5

In one embodiment, the total storage carbohydrate content of the plants of the invention, or parts thereof and in particular of the harvestable parts of the plant(s) is increased compared to control plant(s) and the corresponding plant parts of the control plants. Storage carbohydrates are preferably sugars such as but not limited to sucrose, fructose and glucose, and polysaccharides such as but not limited to starches, glucans and fructans.

10

The total storage carbohydrate content and the content of individual groups or species of carbohydrates may be measured in a number of ways known in the art. For example, the international application published as WO 2006066969 discloses in paragraphs [79] to [117] a method to determine the total storage carbohydrate content of sugarcane, including fructan content.

15

For sugarcane the following method can be used for sugar content analysis:

The transgenic sugarcane plants are grown for 10 to 15 months, either in the greenhouse or the field. Standard conditions for growth of the plants are used. Stalks of sugarcane plants which are 10 to 15 months old and have more than 10 internodes are harvested. After all of the leaves have been removed, the internodes of the stalk are numbered from top (= 1) to bottom (for example = 36). A stalk disc approximately 1-2 g in weight is excised from the middle of each internode. The stalk discs of 3 internodes are then combined to give one sample and frozen in liquid nitrogen. The fresh weight of the samples is determined. The extraction for the purposes of the sugar determination is done as described below.

20

25

For the sugar extraction, the stalk discs are first comminuted in a Waring blender (from Waring, New Hartford, Connecticut, USA). The sugars are extracted by shaking for one hour at 95°C in 10 mM sodium phosphate buffer pH 7.0. Thereafter, the solids are removed by filtration through a 30 µm sieve. The resulting solution is subsequently employed for the sugar determination (see herein below).

30

The glucose, fructose and sucrose contents in the extract obtained in accordance with the sugar extraction method described above is determined photometrically in an enzyme assay via the conversion of NAD⁺ (nicotinamide adenine dinucleotide) into NADH (reduced nicotinamide adenine dinucleotide). During the reduction, the aromatic character at the nicotinamide ring is lost, and the absorption spectrum thus changes. This change in the absorption spectrum can be detected photometrically. The glucose and fructose present in the extract is converted into glucose-6-phosphate and fructose-6-phosphate by means of the enzyme hexokinase and adenosin triphosphate (ATP). The glucose-6-phosphate is subsequently oxidized by the enzyme glucose-6-phosphate dehydrogenase to give 6-

35

40

phosphogluconate. In this reaction, NAD⁺ is reduced to give NADH, and the amount of NADH formed is determined photometrically. The ratio between the NADH formed and the glucose present in the extract is 1:1, so that the glucose content can be calculated from the NADH content using the molar absorption coefficient of NADH (at 340 nm 6.2 per mmol and per cm lightpath). Following the complete oxidation of glucose-6-phosphate, fructose-6-phosphate, which has likewise formed in the solution, is converted by the enzyme phosphoglucosomerase to give glucose-6-phosphate which, in turn, is oxidized to give 6-phosphogluconate. Again, the ratio between fructose and the amount of NADH formed is 1:1. Thereafter, the sucrose present in the extract is cleaved by the enzyme sucrase (Megazyme) to give glucose and fructose. The glucose and fructose molecules liberated are then converted with the abovementioned enzymes in the NAD⁺-dependent reaction to give 6-phosphogluconate. The conversion of one sucrose molecule into 6-phosphogluconate results in two NADH molecules. The amount of NADH formed is likewise determined photometrically and used for calculating the sucrose content, using the molar absorption coefficient of NADH.

Furthermore transgenic sugarcane plants may be analysed using any method known in the art for example but not limited to:

- The Sampling of Sugar Cane by the Full Width Hatch Sampler; ICUMSA (International Commission for Uniform Methods of Sugar Analysis, <http://www.icumsa.org/index.php?id=4>) Method GS 5-5 (1994) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Sampling of Sugar Cane by the Corer Method; ICUMSA Method GS 5-7 (1994) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of Sucrose by Gas Chromatography in Molasses and Factory Products - Official; and Cane Juice; ICUMSA Method GS 4/7/8/5-2 (2002) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of Sucrose, Glucose and Fructose by HPLC -in Cane Molasses-and Sucrose in Beet Molasses; ICUMSA Method GS 7/4/8-23 (2011) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of Glucose, Fructose and Sucrose in Cane Juices ,Syrups and Molasses, and of Sucrose in Beet Molasses by High Performance Ion Chromatography; ICUMSA Method GS 7/8/4-24 (2011) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>).

For crops other than sugarcane, similar methods are known in the art or can easily be adapted from a known method for another crop. For example, the storage carbohydrate content of sugar beet may be determined by any of methods described for sugarcane above with adaptations to sugar beet.

Further transgenic sugar beet plants may be analysed for biomass or their sugar content or other phenotypic parameters using any method known in the art for example but not limited to:

- The Determination of Glucose and Fructose in Beet Juices and Processing Products by an Enzymatic Method – ICUMSA (International Commission for Uniform Methods of Sugar Analysis, <http://www.icumsa.org/index.php?id=4>) Method GS 8/4/6-4 (2007) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of Mannitol, Glucose, Fructose, Sucrose and Raffinose in Beet Brei and Beet Juices by HPAEC-PAD; ICUMSA Method GS8-26 (2011) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of Sucrose, Glucose and Fructose by HPLC -in Cane Molasses-and Sucrose in Beet Molasses; ICUMSA Method GS 7/4/8-23 (2011) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of Glucose, Fructose and Sucrose in Cane Juices ,Syrups and Molasses, and of Sucrose in Beet Molasses by High Performance Ion Chromatography; ICUMSA Method GS 7/8/4-24 (2011) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of Glucose and Fructose in Beet Juices and Processing Products by an Enzymatic Method; ICUMSA Method GS 8/4/6-4 (2007) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>)
- The Determination of the Apparent Total Sugar Content of Beet Pulp by the Luff Schoorl Procedure; ICUMSA Method GS 8-5 (1994) available from Verlag Dr. Albert Bartens KG, Lückhoffstr. 16, 14129 Berlin (<http://www.bartens.com/>).

Further it is to be understood that “comprising” throughout this application may in one embodiment be replaced by “substantially consisting of”, preferably when “comprising” refers to the polynucleotides, constructs, recombinant chromosomal DNA and / or polypeptides of the invention. For example “comprising the POI encoding nucleic acid” may be replaced by “substantially consisting of the POI encoding nucleic acid”.

Moreover, the present invention relates to the following specific embodiments, wherein the expression “as defined in claim / item X” is meant to direct the artisan to apply the definition as disclosed in item / claim X. For example, “a nucleic acid as defined in item 1” has to be understood such that the definition of the nucleic acid as in item 1 is to be applied to the nucleic acid. In consequence the term “as defined in item” or “as defined in claim” may be replaced with the corresponding definition of that item or claim, respectively:

Additional or alternative embodiments:

1. A method for enhancing one or more yield-related traits in plants relative to control plants, comprising increasing expression in a plant of a nucleic acid encoding a ANK

polypeptide, wherein said ANK polypeptide comprises at least one domain selected from the group of

- IPR002504 ATP-NAD/AcoX kinase;
- IPR016064 ATP-NAD kinase PpnK-type;
- IPR017437 ATP-NAD kinase, PpnK-type, all beta; and
- PFAM PF01513 "NAD_kinase" domain represented by amino acid 69-356 in SEQ ID NO: 2

and enhancing one or more-yield-related traits of said plant compared to control plants.

2. Method according to embodiment 1, wherein said ANK polypeptide has ANK activity.
3. Method according to any of the preceding embodiments, wherein said increased expression is effected by introducing and expressing in a plant said nucleic acid encoding said ANK polypeptide.
4. Method according to any of the preceding embodiments, wherein said one or more enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased sugar yield and/or increased seed yield relative to control plants.
5. Method according to any of the preceding embodiments, wherein said one or more enhanced yield-related traits are obtained under non-stress conditions.
6. Method according to any one of embodiments 1 to 4, wherein said one or more enhanced yield-related traits are obtained under conditions of drought stress, salt stress or nitrogen deficiency.
7. Method according to any of the preceding embodiments, wherein said ANK polypeptide comprises:
 - (i) all of the motifs 1 - 7 according to SEQ ID NO: 55 to 61 or
 - (ii) any 6 of the motifs listed under (i); or
 - (iii) any 5 of the motifs listed under (i); or
 - (iv) any 4 of the motifs listed under (i); or
 - (v) any 3 of the motifs listed under (i); or
 - (vi) any two of the motifs listed under (i); or
 - (vii) one of the motifs as described under (i).
8. Method according to any one of the preceding embodiments, wherein said nucleic acid encoding an ANK is of fungal origin, preferably yeast origin, more preferably from *Saccharomyces cerevisiae*.
9. Method according to any one of the preceding embodiments, wherein said nucleic acid

encoding an ANK encodes any one of the polypeptides listed in Table 1 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with a complementary sequence of such a nucleic acid.

- 5 10. Method according to any one of the preceding embodiments, wherein said nucleic acid sequence encodes an orthologue or paralogue of any of the polypeptides given in Table 1.
- 10 11. Method according to any one of the preceding embodiments, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 2 or a polypeptide sequence which is at least 38%, 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% identical to SEQ ID NO: 2 over
15 the entire length of the sequence.
12. Method according to any one of the preceding embodiments, wherein the ANK polypeptide contains when optimally aligned with the polypeptide sequence of SEQ ID NO: 2 an Arg at the position corresponding to position 293 of SEQ ID NO: 2.
- 20 13. Method according to any one of the preceding embodiments, wherein said nucleic acid is operably linked to a plant suitable promoter, preferably a constitutive promoter, more preferably to a high strength constitutive promoter and / or preferentially expressed in green tissue, most preferably to a Big35S promoter.
- 25 14. Plant, or part thereof, or plant cell, obtainable by a method according to any one of the preceding embodiments, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an ANK polypeptide as defined in any of the embodiments 1, 2, and 7 to 12.
- 30 15. Construct comprising:
 - (i) nucleic acid encoding an ANK as defined in any of embodiments 1, 2, and 7 to 12;
 - (ii) one or more control sequences capable of driving expression of the nucleic acid
35 sequence of (i); and optionally
 - (iii) a transcription termination sequence.
16. Construct according to embodiment 15, wherein one of said control sequences is a ubiquitous constitutive promoter, preferably a high strength constitutive promoter and /
40 or a green tissue preferential promoter, more preferably to a Big35S promoter.
17. Use of a construct according to any of the embodiments 15 and 16 in a method for

making plants having one or more enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and / or increased biomass and / or increased sugar content relative to control plants.

- 5 18. A host cell, preferably a bacterial host cell, more preferably an Agrobacterium species host cell comprising the construct according to any of the embodiments 15 and 16.
19. Plant, plant part or plant cell transformed with a construct according to any of the
10 embodiments 15 and 16.
20. Method for the production of a transgenic plant having one or more enhanced yield-
related traits compared to control plants, preferably increased yield relative to control
plants, and more preferably increased seed yield and / or increased biomass and / or
increased sugar yield relative to control plants, comprising:
- 15 (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an
ANK polypeptide as defined in any of embodiments 1, 2, and 7 to 12; and
(ii) cultivating said plant cell or plant under conditions promoting plant growth and
development, particularly of plants having one or more enhanced yield-related
traits relative to control plants.
- 20 21. Transgenic plant having one or more enhanced yield-related traits relative to control
plants, preferably increased yield compared to control plants, and more preferably
increased seed yield and / or increased biomass and / or increased sugar yield,
resulting from increased expression of a nucleic acid encoding an ANK polypeptide as
25 defined in any of embodiments 1, 2, and 7 to 12 or a transgenic plant cell derived from
said transgenic plant.
22. Transgenic plant according to any of the embodiments 14, 19 and 21, or a transgenic
plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet
30 or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice,
maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or
oats.
23. Harvestable part of a plant according to any of the embodiments 14, 19, 21, and 22,
35 wherein said harvestable parts are preferably (shoot and / or root) biomass and / or
seeds, wherein the harvestable part comprises the ANK encoding nucleic acids and /
or ANK polypeptides as defined herein above.
24. A product derived from a plant according to any of the embodiments 14, 19, 21, and 22
40 and / or from harvestable parts of a plant according to embodiment 23, wherein the
product comprises the ANK encoding nucleic acids and / or ANK polypeptides as
defined herein above.

25. Use of a nucleic acid encoding an ANK polypeptide as defined in any of embodiments 1, 2 and 7 to 12 for enhancing one or more yield-related traits in plants compared to control plants, preferably for increasing yield, and more preferably for increasing seed yield and / or for increasing biomass and / or increasing sugar content in plants relative to control plants.
26. A method for manufacturing a product comprising the steps of growing the plants according to any of the embodiments 14, 19, 21 and 22 and producing said product from or by said plants; or parts thereof, including seeds.
27. Recombinant chromosomal DNA comprising the construct according to any of the embodiments 15 and 16.
28. A method for producing a transgenic seed, comprising the steps of (i) introducing into a plant the nucleic acid encoding an ANK as defined in any of embodiments 1, 2, and 7 to 12 or the construct as defined in any of the embodiments 15 and 16; (ii) selecting a transgenic plant having enhanced yield-related traits so produced by comparing said transgenic plant with a control plant; (iii) growing the transgenic plant to produce a transgenic seed, wherein the transgenic seed comprises the nucleic acid or the construct.
29. A method according to embodiment 28, wherein a progeny plant grown from the transgenic seed has increased expression of the polypeptide compared to the control plant.
30. Construct according to any of embodiments 15 and 16, preferably a plant expression construct, or recombinant chromosomal DNA according to embodiment 27 comprised in a host cell, preferably in a plant cell, more preferably in a crop plant cell.
31. A composition comprising the recombinant chromosomal DNA of embodiment 27 and / or the construct in any of embodiments 15 and 16, and a host cell, preferably a plant cell, wherein the recombinant chromosomal DNA and / or the construct are comprised within the host cell.
32. A transgenic pollen grain comprising the construct according to any of embodiments 15 and 16.
33. A protective covering comprising
- (i) propagules of the plants of any of embodiments 14, 19, 21, and 22, such as but not limited to setts of sugarcane and/or gems of sugarcane, and / or
 - (ii) the plant cells of any of the embodiments 14, 19, 21 and 22, and / or

(iii) the nucleic acid encoding the polypeptides as defined in any of embodiments 1, 2, and 7 to 12 and / or the polypeptides as defined in any of embodiments 1, 2, and 7 to 12 and / or the constructs of embodiments 15 and 16 comprised in an agricultural product, and / or

5 (iv) the recombinant chromosomal DNA of embodiment 27.

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. It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized. 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.

The term "nucleotide" refers to a nucleic acid building block consisting of a nucleobase, a pentose and at least one phosphate group. Thus, the term "nucleotide" includes a nukleosid monophosphate, nukleosid diphosphate, and nukleosid triphosphate.

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 substantially the same and functional activity as the unmodified protein from which they are derived.

“Homologues” of a gene encompass nucleic acid sequences with nucleotide substitutions, deletions and / or insertions relative to the unmodified gene in question and having substantially the same activity and / or functional properties as the unmodified gene from which they are derived, or encoding polypeptides having substantially the same biological and / or functional activity as the polypeptide encoded by the unmodified nucleic acid sequence.

Orthologues and paralogues are two different forms of homologues and encompass evolutionary concepts used to describe the ancestral relationships of genes or proteins.

Paralogues are genes or proteins within the same species that have originated through duplication of an ancestral gene; orthologues are genes or proteins 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 or a removal of one or more nucleotides from a nucleic acid.

An “insertion” refers to one or more amino acid residues being introduced into a predetermined site in a protein or to one or more nucleotides being introduced into a predetermined site in a nucleic acid sequence. Regarding 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. 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 3 below).

Table 3: 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).

“Derivatives” of nucleic acids include nucleic acids which may, compared to the nucleotide sequence of the naturally-occurring form of the nucleic acid comprise deletions, alterations,

or additions with non-naturally occurring nucleotides. These may be naturally occurring altered or non-naturally altered nucleotides as compared to the nucleotide sequence of a naturally-occurring form of the nucleic acid. A derivative of a protein or nucleic acid still provides substantially the same function, e.g., enhanced yield-related trait, when expressed or repressed in a plant respectively.

Functional fragments

The term "functional fragment" refers to any nucleic acid or protein which comprises merely a part of the fulllength nucleic acid or fulllength protein, respectively, but still provides substantially the same function e.g. enhanced yield-related trait(s) when overexpressed or repressed in a plant respectively. In cases where overexpression of nucleic acid is desired, the term "substantially the same functional activity" or "substantially the same function" means that any homologue and / or fragment provide increased / enhanced yield-related trait(s) when expressed in a plant. Preferably substantially the same functional activity or substantially the same function means at least 50%, at least 60%, at least 70%, at least 80 %, at least 90 %, at least 95%, at least 98 %, at least 99% or 100% or higher increased / enhanced yield-related trait(s) compared with the functional activity provided by the exogenous expression of the full-length ANK encoding nucleotide sequence or the ANK amino acid sequence.

Domain, Motif/Consensus sequence/Signature/Pattern

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

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

Reciprocal BLAST

Typically, this involves a first BLAST involving BLASTing (i.e. running the BLAST software with the sequence of interest as query sequence) a query sequence (for example using any of the sequences listed in Table 1) 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.

Transit peptide

A "transit peptide" (or transit signal, signal peptide, signal sequence) is a short (3-60 amino acids long) peptide chain that directs the transport of a protein, preferably to organelles within the cell or to certain subcellular locations or for the secretion of a protein. Transit peptides may also be called transit signal, signal peptide, signal sequence, targeting signals, or (subcellular) localization signals.

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:

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

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

^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; $\ln$, = 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.

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. Typical 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. In a preferred embodiment high stringency conditions mean hybridisation at 65°C in 0.1x SSC comprising 0.1 SDS and optionally 5x Denhardt's reagent, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.5% sodium pyrophosphate, followed by the washing at 65°C in 0.3x SSC.

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 substantially the same chromosomal position. Allelic variants encompass Single Nucleotide

- 5 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

- 10 Reference herein to an “endogenous” nucleic acid and / or protein refers to the nucleic acid and / or protein in question as found in a plant in its natural form (i.e., without there being any human intervention like recombinant DNA technology), 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
- 15 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.

Exogenous

- 20 The term “exogenous” (in contrast to “endogenous”) nucleic acid or gene refers to a nucleic acid that has been introduced in a plant by means of recombinant DNA technology. An “exogenous” nucleic acid can either not occur in this plant in its natural form, be different from the nucleic acid in question as found in the plant in its natural form, or can be identical to a nucleic acid found in the plant in its natural form, but not integrated within its natural
- 25 genetic environment.

The corresponding meaning of “exogenous” is applied in the context of protein expression. For example, a transgenic plant containing a transgene, i.e., an exogenous nucleic acid, may, when compared to the expression of the endogenous gene, encounter a substantial

30 increase of the expression of the respective gene or protein in total. A transgenic plant according to the present invention includes an exogenous ANK nucleic acid integrated at any genetic loci and optionally the plant may also include the endogenous gene within the natural genetic background.

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).

40

Expression cassette

“Expression cassette” as used herein is DNA capable of being expressed in a host cell or in

an in-vitro expression system. Preferably the DNA, part of the DNA or the arrangement of the genetic elements forming the expression cassette is artificial. The skilled artisan is well aware of the genetic elements that must be present in the expression cassette in order to be successfully expressed. The expression cassette comprises a sequence of interest to be expressed 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, one or more NEENA as described herein, and / or one or more RENA as described herein. 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 for increased expression/overexpression. 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 expression cassette may be integrated into the genome of a host cell and replicated together with the genome of said host cell.

Construct / genetic construct

This is DNA - artificial in part or total, or artificial in the arrangement of the genetic elements contained - capable of increasing or decreasing the expression of DNA and / or protein of interest typically by replication in a host cell and used for introduction of a DNA sequence of interest into a host cell or host organism. Replication may occur after integration into the host cell's genome or through the presence of the construct as part of a vector or an artificial chromosome inside the host cell.

Typically the construct / genetic construct is an expression construct and comprises one or more expression cassettes that may lead to overexpression (overexpression construct) or reduced expression of a gene of interest. A construct may consist of an expression cassette. The sequence(s) of interest is/are 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, one or more NEENA as described herein, and / or one or more RENA as described herein. 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 for increased expression/overexpression. 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.

Vector construct/ vector

This is DNA (such as, but not limited to plasmids or viral DNA) - artificial in part or total or artificial in the arrangement of the genetic elements contained - capable of replication in a host cell and used for introduction of a DNA sequence of interest into a host cell or host organism. A vector may be a construct or may comprise at least one construct. A vector may replicate without integrating into the genome of a host cell, e.g. a plasmid vector in a bacterial host cell, or it may integrate part or all of its DNA into the genome of the host cell and thus lead to replication and expression of its DNA. 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. Typically the vector comprises at least one expression cassette. The one or more sequence(s) 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, one or more NEENA as described herein and / or one or more RENA as described herein. 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.

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” or “promoter sequence” 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 herein, 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" or "functionally linked" is used interchangeably and, 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 direct transcription of the gene of interest.

The term "functional linkage" or "functionally linked" with respect to regulatory elements, is to be understood as meaning, for example, the sequential arrangement of a regulatory element (e.g. a promoter) with a nucleic acid sequence to be expressed and, if appropriate, further regulatory elements (such as e.g., a terminator, NEENA as described herein or a RENA as described herein) in such a way that each of the regulatory elements can fulfil its intended function to allow, modify, facilitate or otherwise influence expression of said nucleic acid sequence. As a synonym the wording "operable linkage" or "operably linked" may be used. The expression may result, depending on the arrangement of the nucleic acid sequences, in sense or antisense RNA. To this end, direct linkage in the chemical sense is not necessarily required. Genetic control sequences such as, for example, enhancer sequences, can also exert their function on the target sequence from positions which are further away, or indeed from other DNA molecules. Preferred arrangements are those in which the nucleic acid sequence to be expressed recombinantly is positioned behind the sequence acting as promoter, so that the two sequences are linked covalently to each other. The distance between the promoter sequence and the nucleic acid sequence to be expressed recombinantly is preferably less than 200 base pairs, especially preferably less than 100 base pairs, very especially preferably less than 50 base pairs. In a preferred embodiment, the nucleic acid sequence to be transcribed is located behind the promoter in such a way that the transcription start is identical with the desired beginning of the RNA of the invention. Functional linkage, and an expression construct, can be generated by means of customary recombination and cloning techniques as described (e.g., in Maniatis T, Fritsch EF and Sambrook J (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold

Spring Harbor Laboratory, Cold Spring Harbor (NY); Silhavy et al. (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY); Ausubel et al. (1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience; Gelvin et al. (Eds) (1990) Plant Molecular Biology Manual; Kluwer Academic
 5 Publisher, Dordrecht, The Netherlands). However, further sequences, which, for example, act as a linker with specific cleavage sites for restriction enzymes, or as a signal peptide, may also be positioned between the two sequences. The insertion of sequences may also lead to the expression of fusion proteins. Preferably, the expression construct, consisting of a linkage of a regulatory region for example a promoter and nucleic acid sequence to be
 10 expressed, can exist in a vector-integrated form and be inserted into a plant genome, for example by transformation.

Constitutive promoter

A “constitutive promoter” refers to a promoter that is transcriptionally active during most, but
 15 not necessarily all, phases of growth and development and under most environmental conditions, in at least one cell, tissue or organ. Table 4 below gives examples of constitutive promoters.

Table 4: Examples of constitutive promoters

Gene Source	Reference
Actin	McElroy et al, Plant Cell, 2: 163-171, 1990
HMGP	WO 2004/070039
CAMV 35S	Odell et al, Nature, 313: 810-812, 1985
CaMV 19S	Nilsson et al., Physiol. Plant. 100:456-462, 1997
GOS2	de Pater et al, Plant J Nov;2(6):837-44, 1992, WO 2004/065596
Ubiquitin	Christensen et al, Plant Mol. Biol. 18: 675-689, 1992
Rice cyclophilin	Buchholz et al, Plant Mol Biol. 25(5): 837-43, 1994
Maize H3 histone	Lepetit et al, Mol. Gen. Genet. 231:276-285, 1992
Alfalfa H3 histone	Wu et al. Plant Mol. Biol. 11:641-649, 1988
Actin 2	An et al, Plant J. 10(1); 107-121, 1996
34S FMV	Sanger et al., Plant. Mol. Biol., 14, 1990: 433-443
Rubisco small subunit	US 4,962,028
OCS	Leisner (1988) Proc Natl Acad Sci USA 85(5): 2553
SAD1	Jain et al., Crop Science, 39 (6), 1999: 1696
SAD2	Jain et al., Crop Science, 39 (6), 1999: 1696
nos	Shaw et al. (1984) Nucleic Acids Res. 12(20):7831-7846
V-ATPase	WO 01/14572
Super promoter	WO 95/14098
G-box proteins	WO 94/12015

Ubiquitous promoter

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

Developmentally-regulated promoter

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

Inducible promoter

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

15 Organ-specific/Tissue-specific promoter

- An “organ-specific” or “tissue-specific promoter” is one that is capable of preferentially initiating transcription in certain organs or tissues, such as the leaves, roots, seed tissue etc. For example, a “root-specific promoter” is a promoter that is transcriptionally active 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”.
- 20

Examples of root-specific promoters are listed in Table 5 below:

25 **Table 5: 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 6 to Table 9 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.

10 **Table 6: Examples of seed-specific promoters**

Gene source	Reference
seed-specific genes	Simon et al., Plant Mol. Biol. 5: 191, 1985;
	Scofield et al., J. Biol. Chem. 262: 12202, 1987.;
	Baszczynski et al., Plant Mol. Biol. 14: 633, 1990.
Brazil Nut albumin	Pearson et al., Plant Mol. Biol. 18: 235-245, 1992.
legumin	Ellis et al., Plant Mol. Biol. 10: 203-214, 1988.
glutelin (rice)	Takaiwa et al., Mol. Gen. Genet. 208: 15-22, 1986;
	Takaiwa et al., FEBS Letts. 221: 43-47, 1987.
zein	Matzke et al Plant Mol Biol, 14(3):323-32 1990
napA	Stalberg et al, Planta 199: 515-519, 1996.
wheat LMW and HMW glutenin-1	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
wheat SPA	Albani et al, Plant Cell, 9: 171-184, 1997
wheat α , β , γ -gliadins	EMBO J. 3:1409-15, 1984
barley ltr1 promoter	Diaz et al. (1995) Mol Gen Genet 248(5):592-8
barley B1, C, D, hordein	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250:750-60, 1996
barley DOF	Mena et al, The Plant Journal, 116(1): 53-62, 1998
blz2	EP99106056.7
synthetic promoter	Vicente-Carbajosa et al., Plant J. 13: 629-640, 1998.
rice prolamin NRP33	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice a-globulin G1b-1	Wu et al, Plant Cell Physiology 39(8) 885-889, 1998
rice OSH1	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117-8122,

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

Table 7: 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 8: 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 9: 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

5

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.

- 10 Examples of green tissue-specific promoters which may be used to perform the methods of the invention are shown in Table 10 below.

Table 10: 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 11 below.

Table 11: 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

10 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, genetic 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 man by non-natural, synthetic ("artificial") methods such as, for example, mutagenic treatment. Suitable methods are described, for example, in US 5,565,350, US 200405323 or WO 00/15815. Furthermore, 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 protein useful in the methods of the present invention, as defined above – becomes a recombinant expression cassette when this expression cassette is not integrated in the natural genetic environment but in a different genetic environment as a result of an isolation of said expression cassette from its natural genetic environment and re-insertion at a different genetic environment.

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 or cellular environment, respectively, and / or that has been modified by recombinant methods. An isolated nucleic acid sequence or isolated nucleic acid molecule is one that is not in its native surrounding or its native nucleic acid neighbourhood, yet it is physically and functionally connected to other nucleic acid sequences or nucleic acid molecules and is found as part of a nucleic acid construct, vector sequence or chromosome.

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.

As used herein, the term "transgenic" relating to an organisms e.g. transgenic plant refers to an organism, e.g., a plant, plant cell, callus, plant tissue, or plant part that exogenously contains the nucleic acid, construct, vector or expression cassette described herein or a part thereof which is preferably introduced by processes that are not essentially biological, preferably by Agrobacteria-mediated transformation or particle bombardment. A transgenic plant for the purposes of the invention is thus understood as meaning, as above, that the nucleic acids described herein are not present in, or not originating from the genome of said

plant, or are present in the genome of said plant but not at their natural genetic environment in the genome of said plant, it being possible for the nucleic acids to be expressed homologously or heterologously.

5 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
10 with subsequent translation. For the purposes of this invention, the original unmodulated expression may also be absence of any expression. The term “modulating the activity” or the term “modulating expression” with respect to the proteins or nucleic acids used in the methods, constructs, expression cassettes, vectors, plants, seeds, host cells and uses of the invention shall mean any change of the expression of the inventive nucleic acid
15 sequences or encoded proteins which leads to increased or decreased yield-related traits in 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.

20 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
25 protein. The process includes transcription of DNA and processing of the resulting mRNA product. The term “expression” or “gene expression” can also include the translation of the mRNA and therewith the synthesis of the encoded protein, i.e., protein expression.

Increased expression / overexpression

The term “increased expression”, “enhanced expression” or “overexpression” as used
30 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. Reference herein to “increased expression”, “enhanced expression” or “overexpression” is taken to mean an
35 increase in gene expression and / or, as far as referring to polypeptides, increased polypeptide levels and / or increased polypeptide activity, relative to control plants. The increase in expression, polypeptide levels or polypeptide activity is in increasing order of preference at least 10%, 20%, 30%, 40% or 50%, 60%, 70%, 80%, 85%, 90%, or 100% or even more compared to that of control plants. The increase in expression may be in
40 increasing order of preference at least 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, 2000%, 3000%, 4000% or 5000% or even more compared to that of control plants. In cases when the control plants have only very little expression, polypeptide

levels or polypeptide activity of the sequence in question and / or the recombinant gene is under the control of strong regulatory element(s) the increase in expression, polypeptide levels or polypeptide activity may be at least 100 times, 200 times, 300 times, 400 times, 500 times, 600 times, 700 times, 800 times, 900 times, 1,000 times, 2,000 times, 3,000 times, 5,000 times, 10,000 times, 20,000 times, 50,000 times, 100,000 times or even more compared to that of control plants.

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 increase 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., WO 9322443), 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).

To obtain increased expression or overexpression of a polypeptide most commonly the nucleic acid encoding this polypeptide is overexpressed in sense orientation with a polyadenylation signal. Introns or other enhancing elements may be used in addition to a promoter suitable for driving expression with the intended expression pattern. In contrast to this, overexpression of the same nucleic acid sequence as antisense construct will not result in increased expression of the protein, but decreased expression of the protein.

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

5 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 compared to that of control plants.

10 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
15 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
20 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.

25 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, preferably by recombinant methods, and expressing in a plant a genetic construct into which the nucleic acid (in this case a stretch of
30 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).

35 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
40 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).

5

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.

35

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

40

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
5 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 α -anomeric nucleic
10 acid sequence. An α -anomeric nucleic acid sequence forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -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
15 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
20 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.
25 US 4,987,071; and Cech et al. US 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
30 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
35 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
40 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
5 helical structures that prevent transcription of the gene in target cells. See Helene, C. (1991), *Anticancer Drug Res.* 6, 569-84; Helene et al. (1992), *Ann. N.Y. Acad. Sci.* 660, 27-36; and Maher (1992), *L.J. Bioassays* 14, 807-15.

Other methods, such as the use of antibodies directed to an endogenous polypeptide for
10 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.

15 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.

20 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 miRNAs have perfect or near-perfect complementarity with their target sequences. However, there are natural targets with up to five mismatches. They
25 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
30 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
35 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. (2005), *Dev. Cell* 8, 517-527). Convenient tools for design and generation of amiRNAs and their precursors are also available to the public (Schwab et al. (2006), *Plant Cell* 18, 1121-1133).

40 For optimal performance, the gene silencing techniques used for reducing expression in a plant of an endogenous gene requires the use of nucleic acid sequences from

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

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

Transformation

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

The transfer of foreign genes into the genome of a plant is called transformation.

Transformation of plant species is now a fairly routine technique. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable ancestor cell. The methods described for the transformation and regeneration of plants from plant tissues or plant cells may be utilized for transient or for stable transformation. Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts

(Krens, F.A. et al., (1982) Nature 296, 72-74; Negrutiu I et al. (1987) Plant. Mol. Biol. 8: 363-373); electroporation of protoplasts (Shillito R.D. et al. (1985) Bio/Technol 3, 1099-1102); microinjection into plant material (Crossway A et al., (1986) Mol. Gen Genet 202: 179-185); DNA or RNA-coated particle bombardment (Klein TM et al., (1987) Nature 327: 70) infection with (non-integrative) viruses and the like. Transgenic plants, including transgenic crop plants, are preferably produced via *Agrobacterium*-mediated transformation. An advantageous transformation method is the transformation *in planta*. To this end, it is possible, for example, to allow the agrobacteria to act on plant seeds or to inoculate the plant meristem with agrobacteria. It has proved particularly expedient in accordance with the invention to allow a suspension of transformed agrobacteria to act on the intact plant or at least on the flower primordia. The plant is subsequently grown on until the seeds of the treated plant are obtained (Clough and Bent, Plant J. (1998) 16, 735-743). Methods for *Agrobacterium*-mediated transformation of rice include well known methods for rice transformation, such as those described in any of the following: European patent application EP 1198985 A1, Aldemita and Hodges (Planta 199: 612-617, 1996); Chan et al. (Plant. Mol. Biol. 22 (3): 491-506, 1993), Hiei et al. (Plant. J. 6 (2): 271-282, 1994), which disclosures are incorporated by reference herein as if fully set forth. In the case of corn transformation, the preferred method is as described in either Ishida et al. (Nat. Biotechnol. 14(6): 745-50, 1996) or Frame et al. (Plant. Physiol. 129(1): 13-22, 2002), which disclosures are incorporated by reference herein as if fully set forth. Said methods are further described by way of example in B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press (1993) 128-143 and in Potrykus Annu. Rev. Plant Physiol. Plant Molec. Biol. 42 (1991) 205-225). The nucleic acids or the construct to be expressed is preferably cloned into a vector, which is suitable for transforming *Agrobacterium tumefaciens*, for example pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984) 8711). Agrobacteria transformed by such a vector can then be used in known manner for the transformation of plants, such as plants used as a model, like *Arabidopsis* (*Arabidopsis thaliana* is within the scope of the present invention not considered as a crop plant), or crop plants such as, by way of example, tobacco plants, for example by immersing bruised leaves or chopped leaves in an agrobacterial solution and then culturing them in suitable media. The transformation of plants by means of *Agrobacterium tumefaciens* is described, for example, by Höfgen and Willmitzer in Nucl. Acid Res. (1988) 16, 9877 or is known inter alia from F.F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press, 1993, pp. 15-38.

In addition to the transformation of somatic cells, which then have to be regenerated into intact plants, it is also possible to transform the cells of plant meristems and in particular those cells which develop into gametes. In this case, the transformed gametes follow the natural plant development, giving rise to transgenic plants. Thus, for example, seeds of *Arabidopsis* are treated with agrobacteria and seeds are obtained from the developing plants of which a certain proportion is transformed and thus transgenic (Feldman, KA and

Marks MD (1987). Mol Gen Genet 208:1-9; Feldmann K (1992). In: C Koncz, N-H Chua and J Shell, eds, Methods in Arabidopsis Research. Word Scientific, Singapore, pp. 274-289).

Alternative methods are based on the repeated removal of the inflorescences and incubation of the excision site in the center of the rosette with transformed agrobacteria, whereby transformed seeds can likewise be obtained at a later point in time (Chang (1994).

Plant J. 5: 551-558; Katavic (1994). Mol. Gen. Genet., 245: 363-370). However, an especially effective method is the vacuum infiltration method with its modifications such as the "floral dip" method. In the case of vacuum infiltration of *Arabidopsis*, intact plants under reduced pressure are treated with an agrobacterial suspension (Bechthold, N (1993). C R

Acad Sci Paris Life Sci, 316: 1194-1199), while in the case of the "floral dip" method the developing floral tissue is incubated briefly with a surfactant-treated agrobacterial suspension (Clough, SJ and Bent AF (1998) The Plant J. 16, 735-743). A certain proportion of transgenic seeds are harvested in both cases, and these seeds can be distinguished from non-transgenic seeds by growing under the above-described selective conditions. In

addition the stable transformation of plastids is of advantages because plastids are inherited maternally in most crops reducing or eliminating the risk of transgene flow through pollen. The transformation of the chloroplast genome is generally achieved by a process which has been schematically displayed in Klaus et al., 2004 (Nature Biotechnology 22 (2), 225-229). Briefly the sequences to be transformed are cloned together with a selectable

marker gene between flanking sequences homologous to the chloroplast genome. These homologous flanking sequences direct site specific integration into the plastome. Plastidal transformation has been described for many different plant species and an overview is given in Bock (2001) Transgenic plastids in basic research and plant biotechnology. J. Mol. Biol. 2001 Sep 21; 312 (3):425-38 or Maliga, P (2003) Progress towards commercialization of plastid transformation technology. Trends Biotechnol. 21, 20-28. Further biotechnological progress has recently been reported in form of marker free plastid transformants, which can be produced by a transient co-integrated marker gene (Klaus et al., 2004, Nature Biotechnology 22(2), 225-229).

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

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

plants. Alternatively, the transformed plants are screened for the presence of a selectable marker such as the ones described herein.

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).

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

T-DNA activation tagging

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

TILLING

The term “TILLING” is an abbreviation of “Targeted Induced Local Lesions In Genomes”

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

Homologous recombination

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

Yield-related Trait(s)

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 increased yield in rice), Water Use Efficiency (WUE), Nitrogen Use Efficiency (NUE), etc.

The term “one or more yield-related traits” is to be understood to refer to one yield-related trait, or two, or three, or four, or five, or six or seven or eight or nine or ten, or more than ten yield-related traits of one plant compared with a control plant.

Reference herein to “enhanced yield-related trait” is taken to mean an increase relative to

control plants in a yield-related trait, for instance in biomass, of a whole plant or of one or more parts of a plant, which may include (i) aboveground parts, preferably aboveground harvestable parts, and / or (ii) parts below ground, preferably harvestable parts below ground. In particular, such harvestable parts are roots such as taproots, stems, beets, tubers, leaves, flowers or seeds.

Throughout the present application the tolerance of and / or the resistance to one or more agrochemicals by a plant, e.g. herbicide tolerance, is not considered a yield-related trait within the meaning of this term of the present application. An altered tolerance of and / or the resistance to one or more agrochemicals by a plant, e.g. improved herbicide tolerance, is not an “enhanced yield-related trait” as used throughout this application.

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

- 5 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.

10

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.

20

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

40

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.

“Biotic stress” is understood as the negative impact done to plants by other living organisms, such as bacteria, viruses, fungi, nematodes, insects, other animals or other plants. “Biotic stresses” are typically those stresses caused by pathogens, such as bacteria, viruses, fungi, plants, nematodes and insects, or other animals, which may result in negative effects on plant growth and/ or yield.

“Abiotic stress” is understood as the negative impact of non-living factors on the living plant in a specific environment. Abiotic stresses or environmental stresses may be due to drought or excess water, anaerobic stress, salt stress, chemical toxicity, oxidative stress and hot, cold or freezing temperatures. The “abiotic stress” may be an osmotic stress caused by a water stress, 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.

Increase/Improve/Enhance

The terms “increase”, “improve” or “enhance” in the context of a yield-related trait 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% increase in the yield-related trait(s) (such as but not limited to 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

- 5 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
- 10 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

- 15 The term “biomass” as used herein is intended to refer to the total weight of a plant or plant part. Total weight can be measured as dry weight, fresh weight or wet weight. 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
 - 20 biomass, etc.;
 - aboveground harvestable parts such as but not limited to shoot biomass, seed biomass, leaf biomass, stem biomass, setts 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,
 - 25 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
 - 30 - propagules such as seed.

In a preferred embodiment throughout this application any reference to “root” as biomass or as harvestable parts or as organ e.g. of increased sugar content is to be understood as a reference to harvestable parts partly inserted in or in physical contact with the ground such

35 as but not limited to beets and other hypocotyl areas of a plant, rhizomes, stolons or creeping rootstalks, but not including leaves, as well as harvestable parts belowground, such as but not limited to root, taproot, tubers or bulbs. In another embodiment aboveground parts or aboveground harvestable parts or aboveground biomass are to be understood as aboveground vegetative biomass not including seeds and / or fruits.

40

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 an 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.

Host cell

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.

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.,
5 *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*,
10 *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.,
15 *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.,
20 *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.,
25 *Tripsacum dactyloides*, *Triticosecale rimpau*, *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.

30

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
35 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
40 plants, but also to plant parts, including seeds and seed parts.

Propagation material / Propagule

“Propagation material” or “propagule” is any kind of organ, tissue, or cell of a plant capable of developing into a complete plant. “Propagation material” can be based on vegetative reproduction (also known as vegetative propagation, vegetative multiplication, or vegetative cloning) or sexual reproduction. Propagation material can therefore be seeds or parts of the non-reproductive organs, like stem or leave. In particular, with respect to poaceae, suitable propagation material can also be sections of the stem, i.e., stem cuttings (like setts or sugarcane gems).

Non-propagative material

Non-propagative material is any kind of organ, tissue, or cell of a plant not capable of developing into a complete plant. E. g., dead cells cannot be used to regenerate a plant.

Stalk

A “stalk” is the stem of a plant belonging the Poaceae, and is also known as the “millable cane”. In the context of poaceae “stalk”, “stem”, “shoot”, or “tiller” are used interchangeably.

Sett

A “sett” is a section of the stem of a plant from the Poaceae, which is suitable to be used as propagation material. Synonymous expressions to “sett” are “seed-cane”, “stem cutting”, “section of the stalk”, and “seed piece”.

Gem

“Gem” or “sugarcane gem” is a part of the sugarcane stem that is cut, often in a round or oval shape with respect to the surface of the stem, and contains part of a node of the stem, preferably with a meristem, and is suitable for regeneration of a sugarcane plant.

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.

In particular, the plants used in the described experiments are used because Arabidopsis, tobacco, rice and corn plants are model plants for the testing of transgenes. They are widely used in the art for the relative ease of testing while having a good transferability of the results to other plants used in agriculture, such as but not limited to maize, wheat, rice, soybean, cotton, oilseed rape including canola, sugarcane, sugar beet and alfalfa, or other dicot or monocot crops.

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 SEQ ID NO: 1 and SEQ ID NO: 2

Sequences (full length cDNA, ESTs or genomic) related to SEQ ID NO: 1 and SEQ ID NO: 2 were identified amongst those maintained in the Entrez Nucleotides database at the National Center for Biotechnology Information (NCBI) using database sequence search tools, such as the Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215:403-410; and Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402). The program is used to find regions of local similarity between sequences by comparing nucleic acid or polypeptide sequences to sequence databases and by calculating the statistical significance of matches. For example, the polypeptide encoded by the nucleic acid of SEQ ID NO: 1 was used for the TBLASTN algorithm, with default settings and the filter to ignore low complexity sequences set off. The output of the analysis was viewed by pairwise comparison, and ranked according to the probability score (E-value), where the score reflect the probability that a particular alignment occurs by chance (the lower the E-value, the more significant the hit). In addition to E-values, comparisons were also scored by percentage identity. Percentage identity refers to the number of identical nucleotides (or amino acids) between the two compared nucleic acid (or polypeptide) sequences over a particular length. In some instances, the default parameters may be adjusted to modify the stringency of the search. For example the E-value may be increased to show less stringent matches. This way, short nearly exact matches may be identified.

Table 1 provides a list of nucleic acid sequences and polypeptides related to SEQ ID NO: 1 and SEQ ID NO: 2, respectively.

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 ANK polypeptide sequences

Alignment of the polypeptide sequences was performed using the ClustalW 1.83 algorithm of progressive alignment (Thompson et al. (1994) Nucleic Acids Res 22:4673). The source code for the stand-alone program is publicly available from the European Molecular Biology

Laboratory; Heidelberg, Germany. The analysis was performed using the default parameters of ClustalW v1.83 (gap open penalty: 10.0; gap extension penalty: 0.2; protein matrix: Gonnet; protein/DNA endgap: -1; protein/DNA gapdist: 4).

- 5 The ANK polypeptides are aligned in Figure 2. White letters on black background indicate identical amino acids among the various protein sequences, white letters on grey background represent highly conserved amino acid substitutions.

10 A phylogenetic tree of ANK polypeptides can be created using the program "SeaView" (Gouy et al, 2010: SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building, Molecular Biology and Evolution 27(2): 221-4). "SeaView" is publicly available ([Http://pbil.univ-lyon1.fr/software/seaview](http://pbil.univ-lyon1.fr/software/seaview)). The input sequences can be aligned using the ClustalW algorithm (integrated in SeaView) using default parameters as implemented in SeaView 4.4.1. Afterwards, the tree can be
15 created using NJ distance analysis (e.g. distance: Poisson, all gap sites ignored).

A consensus sequence can be derived from a multiple alignment of the sequences as listed in Table 1 and described herein. The letters represent the one letter amino acid code and indicate that the amino acids are conserved in at least 80% of the aligned proteins, whereas
20 the letter X stands for amino acids, which are not conserved in at least 80% of the aligned sequences. A consensus sequence can start with the first conserved amino acid in the alignment, and can end with the last conserved amino acid in the alignment of the investigated sequences.

25 Example 3: Calculation of global percentage identity between polypeptide sequences
Global percentages of identity between full length polypeptide sequences useful in performing the methods of the invention were determined using two methods:

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

Results of the MatGAT analysis are shown in Figure 3 with global identity percentages over the full length of the polypeptide sequences. Parameters used in the analysis were: Scoring matrix: Blosom50, First Gap: 12, Extending Gap: 2. The sequence
40 identity (in %) between the ANK polypeptide sequences useful in performing the methods of the invention can be as low as 38% (is generally higher than 50%) compared to SEQ ID NO: 2.

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

- (2) As an alternative and often preferable method global percentages of identity between full length polypeptide sequences useful in performing the methods of the invention were determined program "NEEDLE" from the EMBOSS software collection, version number 6.3.1.2 (The European Molecular Biology Open Software Suite; <http://www.ebi.ac.uk/Tools/psa/>; Trends in Genetics 16(6): 276 (2000); available from EMBL European Bioinformatics Institute, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, and <http://emboss.sourceforge.net/>).

Results of the analysis are shown in Figure 4 with global identity percentages over the full length of the polypeptide sequences. Parameters used in the analysis were: gapopen 10.0, gapextend 0.5, matrix: BLOSUM62 (abbreviated EBLOSUM62).

Example 4: Identification of patterns (also called motifs) 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 (the Wellcome Trust SANGER Institute, Hinxton, England, UK (<http://pfam.sanger.ac.uk/>)). Interpro is hosted at the European Bioinformatics Institute in the United Kingdom and can be searched via the public web servers (e.g. <http://www.ebi.ac.uk/Tools/pfa/iprscan/>).

For the identification of InterPro domains in the sequences of this application, the integrated software package "InterProScan" was used as implemented in the commercial software package "metalife" (metalife version 5.3, having attached InterPro release 42.0). The results of the InterProScan (see Zdobnov E.M. and Apweiler R.; "InterProScan - an integration platform for the signature-recognition methods in InterPro."; Bioinformatics, 2001, 17(9): 847-8; InterPro database, release 42.0) of the polypeptide sequence as represented by SEQ ID NO: 2 are presented in Table 12. Default parameters (DB genetic code = standard; transcript length = 20) were used.

HMMER is a collection profile hidden Markov methods for protein sequence analysis developed by Sean Eddy and co-workers (HMMER web server: interactive sequence similarity searching R.D. Finn, J. Clements, S.R. Eddy Nucleic Acids Research (2011) Web
5 Server Issue 39:W29-W37) and available from <http://hmmer.wustl.edu/> and <http://hmmer.janelia.org/>. For identification of domains in the sequences of this application, the hmmscan algorithm was used. Using program "hmmscan" from the HMMER 3.0 software collection to search the high quality section "PFAM-A" of PFAM release 27.0 of the Wellcome Trust SANGER Institute, Hinxton, England, UK (<http://pfam.sanger.ac.uk/>).
10 Domains were scored to be present in a given sequence when the reported E-value was 0.1 or lower and if at least 70% of the length of the PFAM domain model (HMM) was covered in the algorithm-produced alignment. Analysis of PFAM-domains for the ANK does result in one significant PFAM-A match hit: PFAM PF01513 ("ATP-NAD kinase"). The description of the PF01513 model (<http://pfam.sanger.ac.uk>) comprises matches with both
15 "ATP-NAD Kinases" (EC:2.7.1.23) and "ATP-NADH kinases" (EC:2.7.1.86).

In one embodiment an ANK 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%
20 sequence identity to a conserved domain from amino acid 69 to 356 in SEQ ID NO: 2.

Table 12: InterProScan results (major accession numbers) of the polypeptide sequence as represented by SEQ ID NO: 2

matches to: IPR002504: "ATP-NAD/AcoX kinase"			matches to IPR016064: ATP-NAD kinase, PpnK-type		matches to: IPR017437: ATP-NAD kinase, PpnK-type, all-beta	
query	HMMPhanter: PTHR20275 "ATP_NADK"	HMMPFam: PF01513 "NAD_kinase"	query	superfamily SSF111331 "ATP-NAD_kinase_PpnK-typ"	query	Gene3D G3DSA:2.60.200.30 "ATP-NAD_kinase_PpnK-typ"
(work_ID)	match	match	(work_ID)	match	(work_ID)	match
YPL188W	12..413	69..356	YPL188W	68..345	YPL188W	213..347
H1	42..206; 236..405	58..398	H1	57..401	H1	253..389
H2	12..413	69..356	H2	68..345	H2	213..347
H3	49..371	62..343	H3	61..376	H3	210..336
H4	22..317	69..328	H4	39..331	H4	189..319
H5	62..388	80..364	H5	63..393	H5	na
H6	1..436	152..451	H6	119..436	H6	na
H7	45..349	97..360	H7	67..348	H7	216..350
H8	99..149; 175..435	131..450	H8	101..435	H8	306..436
H9	45..349	58..361	H9	57..349	H9	219..351
H10	45..413	65..357	H10	64..345	H10	213..347
H11	43..414	95..358	H11	65..346	H11	214..348
H12	69..453	198..464	H12	168..453	H12	325..455

matches to: IPR002504: "ATP-NAD/AcoX kinase"			matches to IPR016064: ATP-NAD kinase, PpnK-type		matches to: IPR017437: ATP-NAD kinase, PpnK-type, all-beta	
query	HMMPhanter: PTHR20275 "ATP_NADK"	HMMPFam: PF01513 "NAD_kinase"	query	superfamily SSF111331 "ATP-NAD_kinase_PpnK-tyr"	query	Gene3D G3DSA:2.60.200.30 "ATP-NAD_kinase_PpnK-tyr"
H13	51..348	71..358	H13	70..348	H13	215..349
H14	46..418	70..363	H14	69..351	H14	219..353
H15	63..357	66..368	H15	75..356	H15	224..358
H16	54..354	108..365	H16	78..354	H16	222..356
H17	61..352	104..362	H17	74..351	H17	219..353
H18	60..357	109..369	H18	79..357	H18	225..359
H19	12..372; 66..457	69..356; 84..407	H19	68..345; 83..409	H19	213..347; 267..397
H20	1..355	63..362	H20	62..391	H20	226..356
H21	12..372	69..356	H21	68..345	H21	213..347
H22	12..372	69..356	H22	68..345	H22	213..347
H24	45..418	99..357	H24	69..346	H24	214..348
H23	11..357	61..360	H23	60..350	H23	219..352
H25	4..421	102..358	H25	72..349	H25	217..351

Identification of conserved motifs

Conserved patterns were identified with the software tool MEME version 3.5. MEME was developed by Timothy L. Bailey and Charles Elkan, Dept. of Computer Science and Engineering, University of California, San Diego, USA and is described by Timothy L. Bailey and Charles Elkan (Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994). The source code for the stand-alone program is public available from the San Diego Supercomputer center (http://meme.sdsc.edu). For identifying common motifs in all sequences with the software tool MEME, the following settings were used: -maxsize 500000, -nmotifs 15, -evt 0.001, -maxw 60, -distance $1e^{-3}$, -minsites number of sequences used for the analysis. Input sequences for MEME were non-aligned sequences in Fasta format. Other parameters were used in the default settings in this software version.

Prosite patterns for conserved domains were generated with the software tool PRATT version 2.1 or manually. PRATT was developed by Inge Jonassen, Dept. of Informatics, University of Bergen, Norway and is described by Jonassen et al. (I. Jonassen, J.F. Collins and D.G. Higgins, Finding flexible patterns in unaligned protein sequences, Protein Science 4 (1995), pp. 1587-1595; I. Jonassen, Efficient discovery of conserved patterns using a pattern graph, Submitted to CABIOS Febr. 1997]. The source code (ANSI C) for the stand-alone program is public available, e.g. at established Bioinformatic centers like EBI (European Bioinformatics Institute).

For generating patterns with the software tool PRATT, following settings were used: PL (max Pattern Length): 100, PN (max Nr of Pattern Symbols): 100, PX (max Nr of consecutive x's): 30, FN (max Nr of flexible spacers): 5, FL (max Flexibility): 30, FP (max Flex.Product): 10, ON (max number patterns): 50. Input sequences for PRATT were distinct regions of the protein sequences exhibiting high similarity as identified from software tool MEME. The minimum number of sequences, which have to match the generated patterns (CM, min Nr of Seqs to Match) was set to at least 80% of the provided sequences.

The presence of motifs, given in the PROSITE pattern format, within a given polypeptide sequence can be identified with program Fuzzpro, as implemented in the "The European Molecular Biology Open Software Suite" (EMBOSS), version 6.3.1.2 (Trends in Genetics 16 (6), 276 (2000)).

From a global alignment (as described in Example 2) of the ANK polypeptide (SEQ ID NO: 2) and the homologs, a 80% consensus sequence can be derived and seven protein patterns (SEQ ID NOs: 55 to 61) were derived using Fuzzpro (see Table 2). The letters represent the one letter amino acid code and indicate that the amino acids are conserved in at least 80% of the aligned proteins, whereas the letter X stands for amino acids, which are

not conserved in at least 80% of the aligned sequences. A consensus sequence will start with the first conserved amino acid in the alignment, and will end with the last conserved amino acid in the alignment of the investigated sequences. The number of given X indicates the distances between conserved amino acid residues, e.g. Y-x(21,23)-F means that conserved tyrosine and phenylalanine residues in the alignment are separated from each other by minimum 21 and maximum 23 amino acid residues in the alignment of all investigated sequences. Conserved domains were identified from all sequences and are described using a subset of the standard Prosite notation, e.g the pattern Y-x(21,23)-[FW] means that a conserved tyrosine is separated by minimum 21 and maximum 23 amino acid residues from either a phenylalanine or tryptophane. Patterns had to match at least 80% of the investigated proteins.

Comparing the protein pattern of SEQ ID NO: 55 with the ANK polypeptide according to SEQ ID NO: 1 and with the homologs H1 to H25, only 4 mismatches in maximum have been identified (see Table 13).

Table 13: Alignment of SEQ ID NO: 56 with ANK polypeptides

pattern SEQ ID	Start match	End match	mis- matches	Hit-Name	Hit-Sequenz
55	241	300	0	ANK	LDIFIDGEFLTRTTADGVALATPTGSTAYS LSAGGSIVSPLVPAILMTPICPR <u>SL</u> SFRPL
55	281	340	0	H1	LDVYINGNLLTRTTADGLIFATPTGSTAYS LSAGGSIVHPVVKICLLTPICPR <u>SL</u> SFRPL
55	241	300	0	H2	LDIFIDGEFLTRTTADGVALATPTGSTAYS LSAGGSIVSPLVPAILMTPICPR <u>SL</u> SFRPL
55	230	289	0	H3	LDIYVGKQFLTSTVADGVAVATPTGSTAYS LSAGGSIVSPQVPSILLTPICPR <u>SL</u> SFRPV
55	217	276	0	H4	LDIYIDNEFLTTTTADGIVFSTPTGSTAYS LSAGGSITHPLVPCILLTPICPR <u>SL</u> SFRPL
55	256	315	4	H5	LDVYLNDEFFTTTTGDGILCSSPTGSTAYS LSAGGSIVHPSVPCILLTPICPR <u>SL</u> LFRPV
55	335	394	2	H6	VDIYIDGEFFTTTYS DGLIFSTPTGSTAYS LSAGGSITHPAVPCILLTPICPR <u>SL</u> SFRPL
55	244	303	0	H7	LDVFIDGKYLTRTTADGVALSTPTGSTAYS LSAGGSIVSPLVPCILLTPICPR <u>SL</u> SFRPL
55	334	393	0	H8	LDIYIDNEFFTTTFADGVIFATPTGSTAYS LSSGGSITHPSVPCVLLTPICPR <u>SL</u> SFRPL
55	247	306	0	H9	LDIFIDGEFLTRTTADGISFSTPTGSTAYS LSAGGSIVHPLVKICLLTPICPR <u>SL</u> SFRPL
55	241	300	0	H10	LDIFIDGDFLTRTSADGVAFASTPTGSTAYS LSAGGSIVSPLVPSILLTPICPR <u>SL</u> SFRPL

55	242	301	0	H11	LDIFIDGEYLTRTTADGVTLATPTGSTAYS LSAGGSIVSPLVPSILLTPICPR <u>S</u> LSFRPL
55	353	412	0	H12	LDIYIDSEFLTTTTADGIILASPTGSTAYS LSAGGSITHPLVPCILMTPVCPR <u>S</u> LSFRPL
55	243	302	3	H13	MDITIDNEFLTSTTGDGVALSTPTGSTAYS LSAGGSIVSPLVPSILLTPICPR <u>S</u> LSFRPV
55	247	306	0	H14	LDIFIDGDFMTRTTADGIAFATPTGSTAYS LSAGGSIVSPLVPCILLTPICPR <u>S</u> LSFRPL
55	252	311	0	H15	LDVFIDGEYLTRTTADGIALATPTGSTAYS LSAGGSIVSPLVPSILLTPICPR <u>S</u> LSFRPL
55	250	309	0	H16	LDIFVDGEFLTRTIADGVVFSPTGSTAYS LSAGGSIVSPLIPSILLTPICPR <u>S</u> LSFRPL
55	247	306	0	H17	LDIYIDGEYLTRTTADGVTLSTPTGSTAYS LSAGGSIVSPLVPSILLTPICPR <u>S</u> LSFRPL
55	253	312	0	H18	LDIYIDGEFMTRTTADGVTLSTPTGSTAYS LSAGGSIVSPLVPSILLTPICPR <u>S</u> LSFRPL
55	295	354	2	H19	LDIYIDNEFLTTTTGDGLVFSTPTGSTAYS LSAGGSITHPLVPCIIITPICPR <u>S</u> LSFRPL
55	254	313	4	H20	LDIHVDGEFITRAIADGVTIATPTGSTAYS LSSGGSIVHPRVACILLTPICPR <u>S</u> LSFRPL
55	241	300	0	H21	LDIFIDGEFLTRTTADGVALATPTGSTAYS LSAGGSIVSPSVPAILMTPICPR <u>S</u> LSFRPL
55	241	300	0	H22	LDIFIDGEFLTRTTADGVALATPTGSTAYS LSAGGSIVSPLVPAILMTPICPR <u>S</u> LSFRPL
55	247	306	0	H23	LDVYVNGHFLTTTTADGLIFATPTGSTAYS LSAGGSMVHPVPCILLTPVCPR <u>S</u> LSFRPL
55	242	301	0	H24	LDIFIDGEYLTRTIADGVTLSTPTGSTAYS LSAGGSIVSPLVPAILLTPICPR <u>S</u> LSFRPL
55	245	304	0	H25	LDIFIDGEYLTRTIADGVTLSTPTGSTAYS LSAGGSIVSPLVPSILLTPICPR <u>S</u> LSFRPL

According to the scientific literature (Ando et al, 2011, Structural Determinants of Discrimination of NAD⁺ from NADH in Yeast Mitochondrial NADH Kinase POS5, J. Biol. Chem. 286(34): 29984-29992), the ability of ANK to discriminate NADH from NAD⁺ is

5 largely dependent on a single conserved Arg residue (position 293 in SEQ ID NO: 2). This conserved Arg residue is comprised in the match to custom pattern SEQ ID NO: 55 (the respective pattern position is highlighted bold and underlined in the above Table 13).

Notably, the position corresponding to this conserved Arg residue is not a site of mismatch in the homologs H1 to H25 (the respective pattern positions are highlighted bold and

10 underlined in the above Table 13).

In one embodiment a ANK polypeptide comprises a motif with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any of the seven protein patterns (SEQ ID NOs: 55 to 61).

5

Example 5: Topology prediction of the ANK polypeptide sequences

TargetP 1.1 (Olof Emanuelsson, Søren Brunak, and Gunnar von Heijne, J. Mol. Biol. (2000) 300: 1005-1016) predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal pre-sequences:

10 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

15 (RC) ranges from 1 to 5, where 1 indicates the strongest prediction. For the sequences predicted to contain an N-terminal pre-sequence a potential cleavage site can also be predicted. TargetP is maintained at the server of the Technical University of Denmark (see <http://www.cbs.dtu.dk/services/TargetP/> & "Locating proteins in the cell using TargetP, SignalP, and related tools", Olof Emanuelsson, Søren Brunak, Gunnar von Heijne, Henrik

20 Nielsen (2007), Nature Protocols 2, 953-971).

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). TargetP was used

25 as implemented in the commercial software "metalife" version 5.3.

The results of TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 2 are presented in Table 14 and Table 15. No cutoffs were defined. The abbreviations used in Table 14 and Table 15: Len, Length; cTP, Chloroplastic transit peptide; mTP,

30 Mitochondrial transit peptide, SP, Secretory pathway signal peptide, other, Other subcellular targeting, Loc, Predicted Location (M, mitochondrion; C, cytoplasm); RC, Reliability class; TPlen, Predicted transit peptide length.

Since ANK is a protein of non-plant origin and known as "Mitochondrial NADH kinase" in the scientific community, targetP prediction of N-terminal targeting peptides was done twice: once using the parameters for non-plants (for prediction of protein targeting in the donor organisms; see Table 14) and once using the parameters for plants (for prediction of the targeting of the heterogenously expressed proteins; see Table 15).

40 The subcellular localization of the polypeptide sequence as represented by SEQ ID NO: 2 is predicted to the mitochondrion using parameters for non-plants (Table 14). For plants, the

subcellular localization of the polypeptide sequence as represented by SEQ ID NO: 2 is predicted to the mitochondrion or the cytosol (Table 15).

In one embodiment, the polypeptide sequence as represented by SEQ ID NO: 2 is localized
5 in the mitochondrion.

Table 14: TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 2 “non-plant”

name	CL Length	cTP	mTP	SP	other	TP Length	Loc	RC
YPL188	414	0,689	0,047	0,254	0	17	M	3
H1	479	0,935	0,047	0,068	0	16	M	1
H2	414	0,689	0,047	0,254	0	17	M	3
H3	383	0,902	0,068	0,068	0	31	M	1
H4	382	0,214	0,045	0,85	0	0	—	2
H5	398	0,877	0,023	0,097	0	27	M	2
H6	535	0,772	0,038	0,254	0	43	M	3
H7	420	0,606	0,049	0,369	0	87	M	4
H8	530	0,713	0,044	0,34	0	116	M	4
H9	430	0,868	0,039	0,14	0	18	M	2
H10	421	0,736	0,052	0,25	0	13	M	3
H11	419	0,861	0,026	0,156	0	13	M	2
H12	530	0,88	0,015	0,221	0	12	M	2
H13	413	0,695	0,084	0,204	0	79	M	3
H14	422	0,701	0,07	0,213	0	18	M	3
H15	434	0,527	0,124	0,372	0	24	M	5
H16	423	0,75	0,046	0,339	0	34	M	3
H17	431	0,701	0,074	0,257	0	25	M	3
H18	420	0,919	0,025	0,086	0	33	M	1
H19	471	0,604	0,052	0,44	0	39	M	5
H20	399	0,861	0,054	0,118	0	27	M	2
H21	375	0,689	0,047	0,254	0	17	M	3
H22	375	0,689	0,047	0,254	0	17	M	3
H23	415	0,853	0,093	0,078	0	23	M	2
H24	418	0,917	0,026	0,092	0	18	M	1
H25	422	0,766	0,027	0,311	0	20	M	3

10 Table 15: TargetP 1.1 analysis of the polypeptide sequence as represented by SEQ ID NO: 2 “plant”

name	CL Length	cTP	mTP	SP	other	TP Length	Loc	RC
YPL188	414	0,101	0,777	0,004	0,064	17	M	2
H1	479	0,107	0,827	0,005	0,015	16	M	2

H2	414	0,101	0,777	0,004	0,064	17	M	2
H3	383	0,246	0,666	0,001	0,065	31	M	3
H4	382	0,091	0,181	0,063	0,74	0	—	3
H5	398	0,162	0,578	0,182	0,012	27	M	4
H6	535	0,772	0,317	0	0,024	72	C	3
H7	420	0,037	0,439	0,064	0,287	87	M	5
H8	530	0,194	0,332	0,015	0,065	116	M	5
H9	430	0,062	0,889	0,003	0,059	18	M	1
H10	421	0,022	0,909	0,014	0,088	13	M	1
H11	419	0,188	0,76	0,007	0,06	13	M	3
H12	530	0,036	0,936	0,014	0,089	12	M	1
H13	413	0,078	0,521	0,088	0,042	79	M	3
H14	422	0,031	0,847	0,015	0,099	18	M	2
H15	434	0,023	0,769	0,005	0,188	24	M	3
H16	423	0,059	0,759	0,021	0,109	34	M	2
H17	431	0,013	0,811	0,006	0,07	25	M	2
H18	420	0,018	0,909	0,027	0,03	33	M	1
H19	471	0,088	0,589	0,018	0,123	39	M	3
H20	399	0,05	0,941	0,006	0,065	27	M	1
H21	375	0,101	0,777	0,004	0,064	17	M	2
H22	375	0,101	0,777	0,004	0,064	17	M	2
H23	415	0,303	0,441	0,01	0,034	23	M	5
H24	418	0,624	0,783	0,005	0,011	18	M	5
H25	422	0,448	0,668	0,006	0,017	20	M	4

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

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

Example 6: Cloning of the ANK encoding nucleic acid sequence

The nucleotide sequence with SEQ ID NO: 1 as shown in Table 1, were amplified by PCR as described in the protocol of the *Pfu* Ultra, *Pfu* Turbo or Herculase DNA polymerase (Stratagene). The composition for the protocol of the *Pfu* Ultra, *Pfu* Turbo or Herculase DNA polymerase was as follows: 1 x PCR buffer (Stratagene), 0.2 mM of each dNTP, 100 ng genomic DNA of *Saccharomyces cerevisiae* (strain S288C; Research Genetics, Inc., now

Invitrogen), 50 pmol forward primer, 50 pmol reverse primer, with or without 1 M Betaine, 2.5 u *Pfu* Ultra, *Pfu* Turbo or Herculase DNA polymerase.

5 The amplification cycles were as follows:

1 cycle of 2-3 minutes at 94-95°C, then 25-36 cycles with 30-60 seconds at 94-95°C, 30-45 seconds at 50-60°C and 210-480 seconds at 72°C, followed by 1 cycle of 5-10 minutes at 72°C, then 4-16°C. ORF specific primer pairs for the genes to be expressed according to Figure 5 are SEQ ID NO: 53 and 54.

10

The following adapter sequences were added to *Saccharomyces cerevisiae* ORF specific primers for cloning purposes:

- i. forward primer: 5'-GGAATTCCAGCTGACCACC-3' (SEQ ID NO: 62)
- ii. reverse primer: 5'-GATCCCCGGGAATTGCCATG-3' (SEQ ID NO: 63)

15

These adaptor sequences allow cloning of the ORF into the various vectors containing the Resgen adaptors, including pMTX155.

Therefore for amplification and cloning of *Saccharomyces cerevisiae* SEQ ID NO: 1, a

20 primer consisting of the adaptor sequence i. and the ORF specific sequence SEQ ID NO: 53 and a second primer consisting of the adaptor sequence ii. and the ORF specific sequence SEQ ID NO: 54 was used.

For cloning of the *S. cerevisiae* ORF SEQ ID NO: 1 into the Resgen adaptor sequence of

25 MTX155, the vector DNA was treated with the restriction enzyme NcoI (MBI Fermentas). The reaction was stopped by inactivation at 70°C for 20 minutes and purified over QIAquick or NucleoSpin Extract II columns following the standard protocol (Qiagen or Macherey-Nagel).

30 Then the PCR-product representing the amplified ORF with the respective adapter sequences and the vector DNA were treated with T4 DNA polymerase according to the standard protocol (MBI Fermentas) to produce single stranded overhangs with the parameters 1 unit T4 DNA polymerase at 37°C for 2-10 minutes for the vector and 1-2 u T4 DNA polymerase at 15-17°C for 10-60 minutes for the PCR product representing SEQ ID

35 NO: 1. The reaction was stopped by addition of high-salt buffer and purified over QIAquick or NucleoSpin Extract II columns following the standard protocol (Qiagen or Macherey-Nagel).

Following these examples every sequence disclosed in Table 1 can be cloned by fusing the

40 adaptor sequences to the respective specific primers sequences as disclosed in SEQ ID NO: 53 and 54 using the respective vector shown in Table 16.

Table 16: Vector used for cloning the ORFs and shows their vector names (column A), the promoters they contain for expression of the ORFs (column B), the additional artificial targeting sequence column (C), the adapter sequence (column D), the expression type conferred by the promoter mentioned in column B (column E) and the figure number (column F).

A	B	C	D	E	F
Vector Name	Promoter Name	Target Sequence	Adapter Sequence	Expression Type	Figure
pMTX155	Big35S		Resgen	non targeted constitutive expression preferentially in green tissues	1

See Figure 1 for the vector pMTX155 which was used cloning the gene of interest for non-targeted expression.

The amplificate for YPL188W prepared as described above and empty vector were ligated according to standard practices.

The ligated constructs were transformed in the same reaction vessel by addition of competent *E. coli* cells (strain DH5alpha) and incubation for 20 minutes at 1°C followed by a heat shock for 90 seconds at 42°C and cooling to 1-4°C. Then, complete medium (SOC) was added and the mixture was incubated for 45 minutes at 37°C. The entire mixture was subsequently plated onto an agar plate with 0.05 mg/ml kanamycine and incubated overnight at 37°C.

The outcome of the cloning step was verified by amplification with the aid of primers which bind upstream and downstream of the integration site, thus allowing the amplification of the insertion. The amplifications were carried out as described in the protocol of Taq DNA polymerase (Gibco-BRL).

The amplification cycles were as follows:

1 cycle of 1-5 minutes at 94°C, followed by 35 cycles of in each case 15-60 seconds at 94°C, 15-60 seconds at 50-66°C and 5-15 minutes at 72°C, followed by 1 cycle of 10 minutes at 72°C, then 4-16°C.

Several colonies were checked, but only one colony for which a PCR product of the expected size was detected was used in the following steps.

A portion of this positive colony was transferred into a reaction vessel filled with complete medium (LB) supplemented with kanamycin and incubated overnight at 37°C.

The plasmid preparation was carried out as specified in the Qiaprep or NucleoSpin Multi-96 Plus standard protocol (Qiagen or Macherey-Nagel).

Example 7: Generation of transgenic Arabidopsis plants which express SEQ ID NO: 1 or any other sequence disclosed in Table 1.

1-5 ng of the plasmid DNA isolated was transformed by electroporation or transformation into competent cells of *Agrobacterium tumefaciens*, of strain GV 3101 pMP90 (Koncz and Schell, Mol. Gen. Gent. 204, 383 (1986)). Thereafter, complete medium (YEP) was added and the mixture was transferred into a fresh reaction vessel for 3 hours at 28°C. Thereafter, all of the reaction mixture was plated onto YEP agar plates supplemented with the respective antibiotics, e.g. rifampicine (0.1 mg/ml), gentamycine (0.025 mg/ml and kanamycine (0.05 mg/ml) and incubated for 48 hours at 28°C.

The agrobacteria that contains the plasmid construct were then used for the transformation of plants. A colony was picked from the agar plate with the aid of a pipette tip and taken up in 3 ml of liquid TB medium, which also contained suitable antibiotics as described above. The preculture was grown for 48 hours at 28°C and 120 rpm. 400 ml of LB medium containing the same antibiotics as above were used for the main culture. The preculture was transferred into the main culture. It was grown for 18 hours at 28°C and 120 rpm. After centrifugation at 4 000 rpm, the pellet was re-suspended in infiltration medium (MS medium, 10% sucrose).

In order to grow the plants for the transformation, dishes (Piki Saat 80, green, provided with a screen bottom, 30 x 20 x 4.5 cm, from Wiesauplast, Kunststofftechnik, Germany) were half-filled with a GS 90 substrate (standard soil, Werkverband E.V., Germany). The dishes were watered overnight with 0.05% Proplant solution (Chimac-Apriphar, Belgium). *Arabidopsis thaliana* C24 seeds (Nottingham Arabidopsis Stock Centre, UK; NASC Stock N906) were scattered over the dish, approximately 1 000 seeds per dish. The dishes were covered with a hood and placed in the stratification facility (8 h, 110 µmol/m²s⁻¹, 22°C; 16 h, dark, 6°C). After 5 days, the dishes were placed into the short-day controlled environment chamber (8 h, 130 µmol/m²s⁻¹, 22°C; 16 h, dark, 20°C), where they remained for approximately 10 days until the first true leaves had formed.

The seedlings were transferred into pots containing the same substrate (Teku pots, 7 cm, LC series, manufactured by Pöppelmann GmbH & Co, Germany).

Five plants were pricked out into each pot. The pots were then returned into the short-day

controlled environment chamber for the plant to continue growing.

After 10 days, the plants were transferred into the greenhouse cabinet (supplementary illumination, 16 h, 340 μ E/m²s, 22°C; 8 h, dark, 20°C), where they were allowed to grow for further 17 days.

For the transformation, 6-week-old *Arabidopsis* plants, which had just started flowering were immersed for 10 seconds into the above-described agrobacterial suspension which had previously been treated with 10 μ l Silwett L77 (Crompton S.A., Osi Specialties, Switzerland). The method in question is described by Clough J.C. and Bent A.F. (Plant J. 16, 735 (1998)).

The plants were subsequently placed for 18 hours into a humid chamber. Thereafter, the pots were returned to the greenhouse for the plants to continue growing. The plants remained in the greenhouse for another 10 weeks until the seeds were ready for harvesting.

Depending on the resistance marker used for the selection of the transformed plants the harvested seeds were planted in the greenhouse and subjected to a spray selection or else first sterilized and then grown on agar plates supplemented with the respective selection agent. Since the vector contained the bar gene as the resistance marker, plantlets were sprayed four times at an interval of 2 to 3 days with 0.02 % BASTA® and transformed plants were allowed to set seeds.

The seeds of the transgenic *Arabidopsis thaliana* plants were stored in the freezer (at -20°C).

Example 8: Plant transformation of crops

Rice transformation

The *Agrobacterium* containing the expression vector is used to transform *Oryza sativa* plants. Mature dry seeds of the rice japonica cultivar Nipponbare are dehusked. Sterilization is 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 are then germinated on a medium containing 2,4-D (2,4-dichlorophenoxy-acetic Acid 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 is used for co-cultivation. *Agrobacterium* is inoculated on AB medium with the appropriate antibiotics and cultured for 3 days at 28°C. The bacteria are then collected and suspended in liquid co-cultivation medium to a density (OD₆₀₀) of about 1. The calli are immersed in the suspension for 1 to

15 minutes. The callus tissues are 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 are grown on 2,4-D-containing medium for 10 to 14 days (growth time for indica: 3 weeks) under light at 28-32°C in the presence of a selection agent. During this period, rapidly growing resistant callus develop. After transfer of this material to regeneration media, the embryogenic potential is released and shoots develop in the next four to six weeks. Shoots are excised from the calli and incubated for 2 to 3 weeks on an auxin-containing medium from which they are transferred to soil. Hardened shoots are 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 are generated for one construct. The primary transformants are 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 are then harvested three to five months after transplanting. The method yields single locus transformants at a rate of over 50% (Aldemita and Hodges 1996, Chan et al. 1993, Hiei et al. 1994).

Corn transformation

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

Wheat transformation

Transformation of wheat is performed with the method described by Ishida et al. (1996)

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

Soybean transformation

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

Rapeseed / canola transformation

Cotyledonary petioles and hypocotyls of 5-6 day old young seedling are used as explants for tissue culture and transformed according to Babic et al. (1998, Plant Cell Rep 17: 183-188). The commercial cultivar Westar (Agriculture Canada) is the standard variety used for transformation, but other varieties can also be used. Canola seeds are surface-sterilized for *in vitro* sowing. The cotyledon petiole explants with the cotyledon attached are excised from the *in vitro* seedlings, and inoculated with *Agrobacterium* (containing the expression vector) by dipping the cut end of the petiole explant into the bacterial suspension. The explants are then cultured for 2 days on MSBAP-3 medium containing 3 mg/l BAP (6-Benzylamino-purine), 3% sucrose, 0.7% Phytagar at 23°C, 16 h 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.

5

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
10 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
15 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
20 plated on the same SH induction medium without acetosyringinone but with a suitable selection agent and suitable antibiotic to inhibit *Agrobacterium* growth. After several weeks, somatic embryos are transferred to BOi2Y development medium containing no growth regulators, no antibiotics, and 50 g/l sucrose. Somatic embryos are subsequently germinated on half-strength Murashige-Skoog medium. Rooted seedlings were
25 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
30 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 50 µg/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
35 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-
40 500 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 h 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 h, 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.

10 *Sugarbeet transformation*

Seeds of sugarbeet (*Beta vulgaris* L.) are sterilized in 70% ethanol for one minute followed by 20 min shaking in 20% Hypochlorite bleach e.g. Clorox® regular bleach (commercially available from Clorox, 1221 Broadway, Oakland, CA 94612, USA). Seeds are rinsed with sterile water and air dried followed by plating onto germinating medium (Murashige and Skoog (MS) based medium (Murashige, T., and Skoog, 1962. *Physiol. Plant*, vol. 15, 473-497) including B5 vitamins (Gamborg et al.; *Exp. Cell Res.*, vol. 50, 151-8.) supplemented with 10 g/l sucrose and 0,8% agar). Hypocotyl tissue is used essentially for the initiation of shoot cultures according to Hussey and Hepher (Hussey, G., and Hepher, A., 1978. *Annals of Botany*, 42, 477-9) and are maintained on MS based medium supplemented with 30 g/l sucrose plus 0,25 mg/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, 150 rpm) 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 30 s 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. 30 g/l sucrose followed by a non-selective period including MS based medium, 30 g/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 WO 9623891A.

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 0,5 cm 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 20 g/l sucrose, 500 mg/l casein hydrolysate, 0,8% agar and 5 mg/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, 20 g/l sucrose, 20 mg/l hygromycin and 500 mg/l cefotaxime). Tissue samples from regenerated shoots are used for DNA analysis. Other transformation methods for sugarcane are known in the art, for example from the in-ternational application published as WO 2010/151634A and the granted European patent EP 1831378.

For transformation by particle bombardment the induction of callus and the transformation of sugarcane can be carried out by the method of Snyman et al. (Snyman et al., 1996, S. Afr. J. Bot 62, 151-154). The construct can be cotransformed with the vector pEmuKN, which expressed the npt[pi] gene (Beck et al. Gene 19, 1982, 327-336; Gen-Bank Accession No. V00618) under the control of the pEmu promoter (Last et al. (1991) Theor.

Appl. Genet. 81, 581-588). Plants are regenerated by the method of Snyman et al. 2001 (Acta Horticulturae 560, (2001) 105-108).

Example 9: Bioanalytical Analysis

For the bioanalytical analyses of the transgenic *Arabidopsis thaliana* plants, the latter were grown uniformly in a specific culture facility. To this end the GS-90 substrate was introduced into the potting machine (Laible System GmbH, Singen, Germany) and filled into the pots. Thereafter, 35 pots were combined in one dish and treated with Previcur. For the treatment, 25 ml of Previcur were taken up in 10 l of tap. This amount was sufficient for the treatment of approximately 200 pots. The pots were placed into the Previcur solution and additionally irrigated overhead with tap water without Previcur. They were used within four days. For the sowing, the seeds, which had been stored in the refrigerator (at -20°C) were dispersed from the Eppendorf tubes into the pots. In total, approximately 5 to 12 seeds were distributed in the middle of the pot.

After the seeds had been sown, the dishes with the pots were covered with matching plastic hoods and placed into the stratification chamber for 4 days in the dark at 4°C. The humidity was approximately 90%. After the stratification, the test plants were grown for 22 to 23 days at a 16-h-light, 8-h-dark rhythm at 20°C, an atmospheric humidity of 60% and a CO₂ concentration of approximately 400 ppm. The light sources used were Powerstar HQI-T 250 W/D Daylight lamps from Osram, which generate a light resembling the solar color spectrum with a light intensity of approximately 220 µmolm⁻² s⁻¹.

Selection of transgenic plants was depending on the used resistance marker. In case of the bar gene as the resistance marker plantlets were sprayed three times at days 8-10 after sowing with 0.02% BASTA®, Bayer CropScience, Germany, Leverkusen. The resistance plants were thinned when they had reached the age of 14 days. The plants, which had grown best in the center of the pot were considered the target plants. All the remaining plants were removed carefully with the aid of metal tweezers and discarded.

During their growth, the plants received overhead irrigation with distilled water (onto the compost) and bottom irrigation into the placement grooves. Once the grown plants had reached the age of 23 or 24 days, they were harvested.

Metabolic analysis of transformed plants

The modifications identified in accordance with the invention, in the content of above-described metabolites, were identified by the following procedure.

a) Sampling and storage of the samples

Sampling was performed directly in the controlled-environment chamber. The plants were cut using small laboratory scissors, rapidly weighed on laboratory scales, transferred into a

pre-cooled extraction thimble and placed into an aluminium rack cooled by liquid nitrogen. If required, the extraction thimble can be stored in the freezer at -80°C. The time elapsing between cutting the plant to freezing it in liquid nitrogen amounted to not more than 10 to 20 seconds.

5 b) Lyophilization

During the experiment, care was taken that the plants either remained in the deep-frozen state (temperatures <-40°C) or were freed from water by lyophilization until the first contact with solvents.

10 The aluminum rack with the plant samples in the extraction thimbles was placed into the pre-cooled (-40°C) lyophilization facility. The initial temperature during the main drying phase was -35°C and the pressure was 0.120 mbar. During the drying phase, the parameters were altered following a pressure and temperature program. The final
15 temperature after 12 hours was +30°C and the final pressure was 0.001 to 0.004 mbar. After the vacuum pump and the refrigerating machine had been switched off, the system was flushed with air (dried via a drying tube) or argon.

 c) Extraction

20 Immediately after the lyophilization apparatus had been flushed, the extraction thimbles with the lyophilized plant material were transferred into the 5 ml extraction cartridges of the ASE device (Accelerated Solvent Extractor ASE 200 with Solvent Controller and AutoASE software (produced by Dionex, available from Thermo Fisher Scientific Inc.; 81 Wyman Street; Waltham, MA 02454; USA).

25 The 24 sample positions of an ASE device (Accelerated Solvent Extractor ASE 200 with Solvent Controller and AutoASE software (produced by Dionex, available from Thermo Fisher Scientific Inc.; 81 Wyman Street; Waltham, MA 02454; USA)) were filled with plant samples, including some samples for testing quality control.

30 The polar substances were extracted with approximately 10 ml of methanol/water (80/20, v/v) at T=70°C and p=140 bar, 5 minutes heating-up phase, 1 minute static extraction. The more lipophilic substances were extracted with approximately 10 ml of methanol / dichloromethane (40/60, v/v) at T=70°C and p=140 bar, 5 minute heating-up phase, 1
35 minute static extraction. The two solvent mixtures were extracted into the same glass tubes (centrifuge tubes, 50 ml, equipped with screw cap and pierceable septum for the ASE (produced by Dionex, available from Thermo Fisher Scientific Inc.; 81 Wyman Street; Waltham, MA 02454; USA).

40 The solution was treated with commercial available internal standards, such as ribitol, L-glycine-2,2-d2 (CAS No 4896-75-7), L-alanine-2,3,3,3-d4 (CAS No 18806-29-6),

methionine-d3 (CAS No 13010-53-2), Arginine-13C, Tryptophan-d5 (CAS No 62595-11-3), and α -methylglucopyranoside and methyl nonadecanoate, methyl undecanoate, methyl tridecanoate, methyl pentadecanoate, methyl nonacosanoate.

The total extract was treated with 8 ml of water. The solid residue of the plant sample and the extraction sleeve were discarded.

The extract was shaken and then centrifuged for 5 to 10 minutes at least 1400 g in order to accelerate phase separation. 1 ml of the supernatant methanol/water phase ("polar phase", colorless) was removed for the further GC analysis, and 1 ml was removed for the LC analysis. The remainder of the methanol / water phase was discarded. 0.5 ml of the organic phase ("lipid phase", dark green) was removed for the further GC analysis and 0.5 ml was removed for the LC analysis. All the portions removed were evaporated to dryness using the IR Dancer infrared vacuum evaporator (Hettich). The maximum temperature during the evaporation process did not exceed 40°C. Pressure in the apparatus was not less than 10 mbar.

d) Processing the lipid and polar phase for the LC/MS or LC/MS/MS analysis

The lipid extract, which had been evaporated to dryness was taken up in mobile phase. The polar extract, which had been evaporated to dryness was taken up in mobile phase.

e) LC-MS analysis

The LC part was carried out on a commercially available LCMS system from Agilent Technologies, USA. For polar extracts 10 μ l are injected into the system at a flow rate of 200 μ l/min. The separation column (Reversed Phase C₁₈) was maintained at 15°C during chromatography. For lipid extracts 5 μ l are injected into the system at a flow rate of 200 μ l/min. The separation column (Reversed Phase C₁₈) was maintained at 30°C. HPLC was performed with gradient elution.

The mass spectrometric analysis was performed on an Applied Biosystems API 4000 triple quadrupole instrument with turbo ion spray source. For polar extracts the instrument measures in negative ion mode in MRM-mode and fullscan mode from 100-1000 amu. For lipid extracts the instrument measures in positive ion mode in MRM-mode fullscan mode from 100-1000 amu. MS analysis is described in more detail in patent publication number WO 03/073464 (Walk and Dostler).

f) Derivatization of the lipid phase for the GC/MS analysis

For the trans-methanolysis, a mixture of 140 μ l of chloroform, 37 μ l of hydrochloric acid (37% by weight HCL in water), 320 μ l of methanol and 20 μ l of toluene was added to the evaporated extract. The vessel was sealed tightly and heated for 2 hours at 100°C, with shaking. The solution was subsequently evaporated to dryness. The residue was dried completely.

The methoximation of the carbonyl groups was carried out by reaction with methoxyamine hydrochloride (5 mg/ml in pyridine, 100 µl for 1.5 hours at 60°C) in a tightly sealed vessel. 20 µl of a solution of odd-numbered, straight-chain fatty acids (solution of each 0.3 mg/mL of fatty acids from 7 to 25 carbon atoms and each 0.6 mg/mL of fatty acids with 27, 29 and 31 carbon atoms in 3/7 (v/v) pyridine/toluene) were added as time standards. Finally, the derivatization with 100 µl of N-methyl-N-(trimethylsilyl)-2,2,2-trifluoroacetamide (MSTFA) was carried out for 30 minutes at 60°C, again in the tightly sealed vessel. The final volume before injection into the GC was 220 µl.

g) Derivatization of the polar phase for the GC/MS analysis

The methoximation of the carbonyl groups was carried out by reaction with methoxyamine hydrochloride (5 mg/ml in pyridine, 50 µl for 1.5 hours at 60°C) in a tightly sealed vessel. 10 µl of a solution of odd-numbered, straight-chain fatty acids (solution of each 0.3 mg/mL of fatty acids from 7 to 25 carbon atoms and each 0.6 mg/ml of fatty acids with 27, 29 and 31 carbon atoms in 3/7 (v/v) pyridine/toluene) were added as time standards. Finally, the derivatization with 50 µl of N-methyl-N-(trimethylsilyl)-2,2,2-trifluoroacetamide (MSTFA) was carried out for 30 minutes at 60°C, again in the tightly sealed vessel. The final volume before injection into the GC was 110 µl.

h) GC-MS Analysis

The GC-MS systems consist of an Agilent 6890 GC coupled to an Agilent 5973 MSD. The autosamplers are CompiPal or GCPal from CTC. For the analysis usual commercial capillary separation columns (30 m x 0.25 mm x 0.25 µm) with different poly-methyl-siloxane stationary phases containing 0% up to 35% of aromatic moieties, depending on the analysed sample materials and fractions from the phase separation step, are used (for example: DB-1ms, HP-5ms, DB-XLB, DB-35ms, Agilent Technologies). Up to 1 µl of the final volume is injected splitless and the oven temperature program is started at 70°C and ended at 340°C with different heating rates depending on the sample material and fraction from the phase separation step in order to achieve a sufficient chromatographic separation and number of scans within each analyte peak. Usual GC-MS standard conditions, for example constant flow with nominal 1 to 1.7 ml/min and helium as the mobile phase gas are used. Ionisation is done by electron impact with 70 eV, scanning within a m/z range from 15 to 600 with scan rates from 2.5 to 3 scans/sec and standard tune conditions.

Data analysis from Metabolic analysis of transformed plants

The samples were measured in individual series of 20 to 21 plant samples each (also referred to as sequences), each sequence containing at least 5 wild-type plants as controls. The peak area of each analyte was divided by the peak area of the respective internal standard. The data were standardized for the fresh weight established for the plant sample, respectively. The values calculated thus were related to the wild-type control group by being

divided by the mean of the corresponding data of the wild-type control group of the same sequence. The values obtained were referred to as ratio_by_WT, they are comparable between sequences and indicate how much the analyte concentration in the mutant differs in relation to the wild-type control. Appropriate controls were done before to proof that the vector and transformation procedure itself has no significant influence on the metabolic composition of the plants. Therefore the described changes in comparison with wildtypes were caused by the introduced gene constructs. 4 independent lines were analyzed for each construct, with at least nine samples (individual plants) per construct.

10 **Table 17: Increase in fructose, glucose and sucrose (median ratio_by_WT) in transgenic *A. thaliana***

Gene	SEQ ID NO (DNA)	SEQ ID NO (PRT)	Expression type	Targeting	Metabolite	ratio_by_WT (Median)	p- value
YPL188W	1	2	Constitutive expression preferentially in green tissues	No artificial targeting	Glucose	2.353	<0.001
YPL188W	1	2	Constitutive expression preferentially in green tissues	No artificial targeting	Fructose	1.462	0.002
YPL188W	1	2	Constitutive expression preferentially in green tissues	No artificial targeting	Sucrose	1.104	0.012

Example 10: Arabidopsis plant screening for yield increase under standardised growth conditions

15 In this experiment, a plant screening for yield increase measured as fresh-weight biomass increase under standardised growth conditions has been performed. Transgenic plants were grown as described in Example 9. Once the grown plants had reached the age of 23 or 24 days, they were harvested by measuring the fresh-weight of the aerial part (rosette) of the plant. Two successive experiments were conducted. In the first experiment, one
20 individual of each transformed event and two events per transformed construct was tested. In the second experiment at least four sibling events were put through a confirmation screen according to the same experimental procedures. Individuals of each event were grown, treated and measured as before. Biomass increase was calculated as ratio of the

median of the weights for transgenic plants compared to median of the weights of wild type control plants from the same experiment, grown in the same culture facility as the transformed plants and harvested on the same day.

5 **Table 18: Biomass production of transgenic *A. thaliana* developed under ambient growth conditions**

Gene	SEQ ID NO (DNA)	SEQ ID NO (PRT)	Expression type	Targeting	ratio_by_WT (Median)	p-value
YPL188W	1	2	Constitutive expression preferentially in green tissues	No artificial targeting	1.16	0.01

Example 11: Rice plant screening for yield increase under stress conditions

Drought screen

10 *Early drought screen*

T1 or T2 plants are germinated under normal conditions and transferred into potting soil as normally. After potting the plants in their pots are then transferred to a "dry" section where irrigation is withheld. Soil moisture probes are inserted in randomly chosen pots to monitor the soil water content (SWC). When SWC goes below certain thresholds, the plants are
15 automatically re-watered continuously until a normal level is reached again. The plants are then re-transferred again to normal conditions. The drought cycle is repeated two times during the vegetative stage with the second cycle starting shortly after re-watering after the first drought cycle is complete. The plants are imaged before and after each drought cycle.

20 The rest of the cultivation (plant maturation, seed harvest) is the same as for plants not grown under abiotic stress conditions. Growth and yield parameters are recorded as detailed for growth under normal conditions.

Reproductive drought screen

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

Nitrogen use efficiency screen

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

Salt stress screen

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

Example 12: Functional assay for the ANK polypeptide

There are several assays available to detect NAD(H) kinase activity, also called ANK activity in solution or cell extracts. Jacobson and Jacobson developed a standard protocol in 1976 which is available in Arch. Biochem. Biophys 175, 627-634. There are also many assay kits available commercially, e.g. the Elite™ NAD/NADH Assay Kit (red fluorescence) which is provided by eEnzyme®. The one skilled in the art is familiar with the technologies available and uses the one that best matches the facilities available to him.

Example 13: Arabidopsis plant stress screen

Transgenic *Arabidopsis thaliana* plants were generated as described in Example 7.

YPL188W was part of a screen as described in WO2004/092398 for Stress-Related Proteins (SRP) from Yeast which, upon expression of their coding genes in plants, confer increased tolerance and/or resistance to an environmental stress as compared to a corresponding non-transformed wild type plant cell. For this screen, transgenic *A. thaliana* plants were grown individually in pots containing a 4:1 (v/v) mixture of soil and quartz sand in a York growth chamber. Standard growth conditions were: photoperiod of 16 h light and 8 h dark, 20 °C, 60% relative humidity, and a photon flux density of 150 µE. To induce germination, sown seeds were kept at 4 °C, in the dark, for 3 days. Plants were watered daily until they were approximately 3 weeks old at which time drought was imposed by withholding water. Parallely, the relative humidity was reduced in 10% increments every second day to 20%. After approximately 12 days of withholding water, most plants showed visual symptoms of injury, such as wilting and leaf browning, whereas tolerant plants were identified as being visually turgid and healthy green in color. Plants were scored for symptoms of drought injury in comparison to neighbouring plants for 3 days in succession. Two successive experiments were conducted. In the first experiment, 10 independent T2 lines were sown for each gene being tested. The percentage of plants not showing visual symptoms of injury was determined. In the second experiment, at least 7 replicates of the

most tolerant line were grown and treated as before. The average and maximum number of days of drought survival after wild-type control had visually died were determined.

In the first experiment, after 12 days of drought, the control, non-transgenic *Arabidopsis thaliana* and most transgenic lines expressing other transgenes in the test showed extreme visual symptoms of stress including necrosis and cell death. Several plants expressing the genes retained viability as shown by their turgid appearance and maintenance of green color.

According to the results of the first experiment some major lines containing the specific SRP encoding yeast genes (including YPL188W) were identified, which showed the best results with regard to the average days of survival after wild type and/or the hit percentage.

In a second experiment these major lines were tested with multiple replicates (4-80 plants per line). The average number of days the plants of the major line survived longer than wild-type was measured. I.e., the number '1' means that, on average, the plants overexpressing this ORF, on average survived 1 day longer than wild-type. The value for WT in this column is '0'. The results for YPL188W are shown in table 19.

Table 19: Drought tolerance of transgenic *Arabidopsis thaliana* expressing the YPL188W-encoding gene from *Saccharomyces cerevisiae* after imposition of drought stress on 3 week old plants in a second experiment using several plants from one transgenic line. Drought tolerance is measured for the indicated number of transgenic plants (Plants tested) as the average number of days (Average days of survival after WT) that the transgenic plants survived after the control (untransformed wild type).

SEQ ID NO (DNA / PRT)	Gene	Plants tested	Average days of survival after WT
1 / 2	YPL188W	10	1.2

Claims

1. A method for enhancing one or more yield-related traits in plants relative to control plants, comprising the steps of
 - a. increasing expression in a plant of a nucleic acid encoding an ANK polypeptide, wherein said ANK polypeptide comprises at least one domain selected from the group of
 - IPR002504 ATP-NAD/AcoX kinase;
 - IPR016064 ATP-NAD kinase PpnK-type;
 - IPR017437 ATP-NAD kinase, PpnK-type, all beta; and
 - PFAM PF01513 "NAD_kinase" domain represented by amino acid 69-356 in SEQ ID NO: 2;and wherein the ANK polypeptide comprises at least 7, at least 6, at least 5, at least 4, at least 3 and/or at least 2 motifs according to SEQ ID NOs: 55 to 61;
 - b. and enhancing one or more-yield-related traits of said plant compared to control plants.
2. Method according to claim 1, wherein said ANK polypeptide comprises all of the domains from the group of IPR002504 ATP-NAD/AcoX kinase; IPR016064 ATP-NAD kinase PpnK-type; IPR017437 ATP-NAD kinase, PpnK-type, all beta; and PFAM PF01513 "NAD_kinase" domain represented by amino acid 69-356 in SEQ ID NO: 2 and wherein said ANK polypeptide comprises all of the motifs 1 to 7 according to SEQ ID NO: 55 to 61.
3. Method according to any of the preceding claims, wherein said ANK polypeptide has ANK activity.
4. Method according to any of the preceding claims, wherein said increased expression is effected by introducing and expressing in a plant said nucleic acid encoding said ANK polypeptide.
5. Method according to any of the preceding claims, wherein said one or more enhanced yield-related traits comprise increased yield relative to control plants, and preferably comprise increased biomass and/or increased sugar yield and/or seed yield relative to control plants.
6. Method according to any of the preceding claims, wherein said one or more enhanced yield-related traits are obtained under non-stress conditions.
7. Method according to any of the preceding claims, wherein said nucleic acid encoding an ANK polypeptide is of fungal origin, preferably yeast origin, more preferably from

Saccharomyces cerevisiae.

8. Method according to any one of the preceding claims, wherein said nucleic acid encoding an ANK polypeptide encodes any one of the polypeptides listed in Table 1 or is a portion of such a nucleic acid, or a nucleic acid capable of hybridising with a complementary sequence of such a nucleic acid.
9. Method according to any one of the preceding claims, wherein said nucleic acid encodes the polypeptide represented by SEQ ID NO: 2 or a polypeptide sequence which is at least 60% identical to SEQ ID NO: 2 over the entire length of the sequence.
10. Method according to any one of the preceding claims, wherein the ANK polypeptide contains when optimally aligned with the polypeptide sequence of SEQ ID NO: 2 an Arg at the position corresponding to position 293 of SEQ ID NO: 2.
11. Method according to any one of the preceding claims, wherein said nucleic acid is operably linked to a plant suitable promoter, preferably a constitutive promoter, more preferably to a high strength constitutive promoter and/or preferentially expressed in green tissue, most preferably to a Big35S promoter.
12. Plant, or part thereof, or plant cell, obtainable by a method according to any one of the preceding claims, wherein said plant, plant part or plant cell comprises a recombinant nucleic acid encoding an ANK polypeptide as defined in any of the claims 1, 2, 3, 7, 8, 9 and 10.
13. Construct comprising:
 - (i) nucleic acid encoding an ANK as defined in any of claims 1, 2, 3, 7, 8, 9 and 10;
 - (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 claim 13, wherein one of said control sequences is a ubiquitous constitutive promoter, preferably a high strength constitutive promoter and/or a green tissue preferential promoter, more preferably to a Big35S promoter.
15. Use of a construct according to any of the claims 13 and 14 in a method for making plants having one or more enhanced yield-related traits, preferably increased yield relative to control plants, and more preferably increased seed yield and/or increased biomass and/or increased sugar content relative to control plants.
16. A host cell, preferably a bacterial host cell, more preferably an Agrobacterium species

host cell comprising the construct according to any of the claims 13 and 14.

17. Plant, plant part or plant cell transformed with a construct according to any of the claims 13 and 14.

5

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

10

- (i) introducing and expressing in a plant cell or plant a nucleic acid encoding an ANK polypeptide as defined in any of claims 1, 2, 3, 7, 8, 9 and 10; and
- (ii) cultivating said plant cell or plant under conditions promoting plant growth and development, particularly of plants having one or more enhanced yield-related traits relative to control plants; and

15

- (iii) producing a transgenic plant having one or more enhanced yield-related traits compared to control plants.

19. Transgenic plant having one or more enhanced yield-related traits relative to control plants, preferably increased yield compared to control plants, and more preferably increased seed yield and/or increased biomass and/or increased sugar yield, resulting from increased expression of a nucleic acid encoding an ANK polypeptide as defined in any of claims 1, 2, 3, 7, 8, 9 and 10 or a transgenic plant cell derived from said transgenic plant.

20

25

20. Transgenic plant according to any of the claims 12, 17 and 19, or a transgenic plant cell derived therefrom, wherein said plant is a crop plant, such as beet, sugarbeet or alfalfa; or a monocotyledonous plant such as sugarcane; or a cereal, such as rice, maize, wheat, barley, millet, rye, triticale, sorghum, emmer, spelt, einkorn, teff, milo or oats.

30

21. Harvestable part of a plant according to any of the claims 12, 17, 19, and 20, wherein said harvestable parts are preferably (shoot and / or root) biomass and/or seeds, wherein the harvestable part comprises the ANK encoding nucleic acids and/or ANK polypeptides as defined herein above.

35

22. A product derived from a plant according to any of the claims 12, 17, 19, and 20 and/or from harvestable parts of a plant according to claim 21, wherein the product comprises the ANK encoding nucleic acids and/or ANK polypeptides as defined herein above.

40

23. Use of a nucleic acid encoding an ANK polypeptide as defined in any of claims 1, 2, 3,

7, 8, 9 and 10 for enhancing one or more yield-related traits in plants compared to control plants, preferably for increasing yield, and more preferably for increasing seed yield and/or for increasing biomass and/or increasing sugar yield in plants relative to control plants.

5

24. A method for manufacturing a product comprising the steps of growing the plants according to any of the claims 12, 17, 19 and 20 and producing said product from or by said plants; or parts thereof, including seeds.

10

25. Recombinant chromosomal DNA comprising the construct according to any of the claims 13 and 14.

15

26. A method for producing a transgenic seed, comprising the steps of (i) introducing into a plant the nucleic acid encoding an ANK as defined in any of claims 1, 2, 3, 7, 8, 9 and 10 or the construct as defined in any of the claims 13 and 14; (ii) selecting a transgenic plant having enhanced yield-related traits so produced by comparing said transgenic plant with a control plant; (iii) growing the transgenic plant to produce a transgenic seed, wherein the transgenic seed comprises the nucleic acid or the construct.

20

27. A method according to claim 26, wherein a progeny plant grown from the transgenic seed has increased expression of the polypeptide compared to control plant.

25

28. Construct according to any of the claims 13 and 14, preferably a plant expression construct, or recombinant chromosomal DNA according to claim 25 comprised in a host cell, preferably in a plant cell, more preferably in a crop plant cell.

30

29. A composition comprising the recombinant chromosomal DNA of claim 25 and/or the construct in any of claims 13 and 14, and a host cell, preferably a plant cell, wherein the recombinant chromosomal DNA and/or the construct are comprised within the host cell.

35

30. A transgenic pollen grain comprising the construct according to any of the claims 13 and 14.

40

31. A protective covering comprising

- (i) propagules of the plant of any of claims 12, 17, 19 and 20, such as but not limited to setts of sugarcane and/or gems of sugarcane, and/or
- (ii) the plant cells of any of the claims 12, 17, 19 and 20, and/or
- (iii) the nucleic acid encoding the polypeptides as defined in any of the claims 1, 2, 3, 7, 8, 9 and 10 and/or the polypeptides as defined in any of the claims 1, 2, 3,

- 7, 8, 9 and 10 and/or the constructs as defined in any of the claims 13 and 14
comprised in an agricultural product, and/or
(iv) the recombinant chromosomal DNA of claim 25.

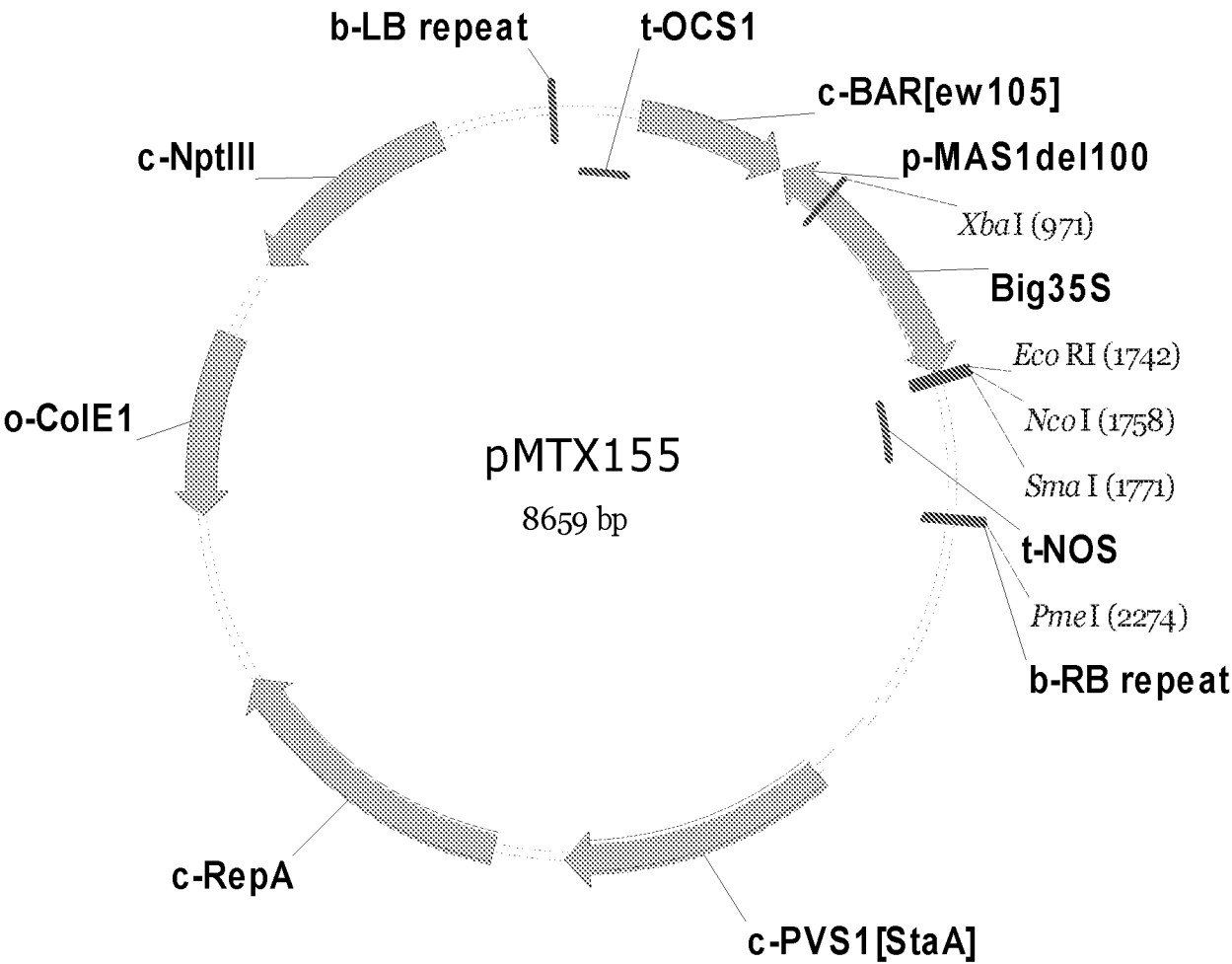


FIGURE 1

```

YPL188W : -----MFVVR---VKLNKPVKW : 13
H1       : -----MALVLRRF : 8
H2       : -----MFVVR---VKLNKPVKW : 13
H3       : -----MKK-----LQVWRSV : 10
H4       : ----- : -
H5       : -----MFCFRVH : 7
H6       : MFRTIRNNP SFFNKIFQRHSEPTSHYSAQSNKPMGDKSSSSRSMTTTLKSY : 50
H7       : -----MYLARTKFGI---FSCKHNF : 17
H8       : MFK-----IQFLFRPCLVNNHHQTNISRLIHFN-----SLHTITKST : 37
H9       : -----MLHVQR-----WSG : 9
H10      : -----MLLR-----VPLRL : 9
H11      : -----MIRVLR-----P--QCFFRR : 13
H12      : MLRIGANVGRRQAGTLRPRFWNKSWAHFGRSLHGGTEAQKKASKTCESTT : 50
H13      : -----MFSVSSK-----LWSLRCM : 14
H14      : -----MLLAR---IRLFSPLKS : 14
H15      : -----MYYPLRLITIRQHYFLSKPL : 20
H16      : -----MNILYCFEACQRGR---VISIVPRLA : 23
H17      : -----MLTSYRILPRR---VLLSGVNRA : 20
H18      : -----MVILLRGRLNVPSIWSLARQ : 20
H19      : -----MMLKTWKIP-----CYSYISNGIRM : 20
H20      : -----MRLLIARTG : 9
H21      : -----MFVVR---VKLNKPVKW : 13
H22      : -----MFVVR---VKLNKPVKW : 13
H23      : -----MISIRAPRV : 9
H24      : -----MRLAFSLRFP---ITTS----- : 14
H25      : -----MFLIRNR---VTTSRLIFS : 16

```

```

YPL188W : YRFYSTLDSH----- : 23
H1       : FSG--SVARA----- : 16
H2       : YRFYSTLDSH----- : 23
H3       : PRPWGARGYAG----- : 21
H4       : ----- : -
H5       : TFIFKPKFLP----- : 17
H6       : LQVCQSQSLSPNPLN----- : 64
H7       : VKFYSIATFG----- : 27
H8       : PLSSKLIIGKAQLS----- : 51
H9       : FIRLCSRNYS----- : 19
H10      : GRFYTSLEYF----- : 19
H11      : LHTVDPLN----- : 21
H12      : TNMGDTKINCNTSDRRMKIEKTAGQQSEKRDVCHERDQNEHNQASREWS : 100
H13      : VPTKVKKFFHNG----- : 25
H14      : VRYYSLEYF----- : 24
H15      : VRFLSNKTND----- : 30
H16      : RRSQCETAEYP----- : 33
H17      : IIRYYSFDIG----- : 30
H18      : QHVSGLFQCARFYS----- : 34
H19      : THIGCLRFTRNIS----- : 34
H20      : ITRPHSVQAR----- : 19
H21      : YRFYSTLDSH----- : 23
H22      : YRFYSTLDSH----- : 23
H23      : FWSRKLFSTS----- : 19
H24      : LRRFSSQESP----- : 24
H25      : IRKYSSLESFP----- : 27

```

FIGURE 2

```

YPL188W : -----SLKLCSGSKFVKIKPVNNLRSSSSADF : 50
H1 : -----TAPASLIKHLHPVSCIKQRCLPTY : 39
H2 : -----SLKLCSGSKFVKIKPVNNLRSSSSADF : 50
H3 : -----LPR-----AGAEVRAASATGVSRAPSF : 43
H4 : -----MAN-LIDIRSVTCIGNAKFPEY : 21
H5 : -----RFCISAQRYFSRPFNVRSCTELVSR--QDG : 45
H6 : -----LSSSSSSSSSSSGSGANHSILTIESCENLPKARLEPY : 100
H7 : -----ER-----TRYVKIKPVSDIKLGSNEDY : 49
H8 : -----LSTTPEKACP-----KMALTVQSCCTCLPTGKLPFY : 81
H9 : -----SQYVKILPVSC LAPRCHPEY : 39
H10 : -----AIRIRDYSKYVSLKSVQDLRLATSPDF : 46
H11 : -----VKLRDTSQFVRVKPVSKLRSISSPDF : 47
H12 : TKQNIQAQDCNKDETCQDSQCSQDKASVGEALCVRPCSSLRNRIVPEY : 150
H13 : -----VTRFCEMASGIHVKSVEALENSTLPDF : 52
H14 : -----SINIRDYSKFVSIKPVSDLRHSSSPDF : 51
H15 : -----EIVPTLLTQHLKVKPVMTLTASRSPGY : 57
H16 : -----STLFKETSKFVKIKPVRLKLSNAPDY : 60
H17 : -----GANVGKG-RFVKIKPVSLRASTADF : 56
H18 : -----EVRLRDSSKFVALKDVNDLRATSLPDF : 61
H19 : -----SLMRKSDILMSNDLLTIKSKHLRNSTFPEF : 65
H20 : -----RSTWIRLLSTEILHAEILPDRQSPHY : 45
H21 : -----SLKLCSGSKFVKIKPVNNLRSSSSADF : 50
H22 : -----SLKLCSGSKFVKIKPVNNLRSSSSADF : 50
H23 : -----VVSFSNVKIESVTCLKPRCFPOY : 42
H24 : -----LVNLRNTSKYVSIKPVTDLRSSAAEF : 51
H25 : -----LVELKNNSKFSVIKPVSNLRSSSTAEF : 54

```

```

YPL188W : VSPPMKQSLIMQN--LQNVYITKE TPSTREAVE T L ES BE : 98
H1 : VGQSNIVKSLVMTT--PSNVLIVKE HSKVLDAIT K L AN ES : 87
H2 : VSPPMKQSLIMQN--LQNVYITKE TPSTREAVE T L ES BE : 98
H3 : PCVPGRRALVMPQA--PANVFVVKEGSAETTA AIEL R M AC EG : 91
H4 : VQQSHKEHNVIWRT--LQNIYVVKE NYHVRDAVQ D I HH EA : 69
H5 : IGAIKPGVPVEWNP--LRNIYLVKE NPSVRDAIE H I GE BH : 93
H6 : VKSCCHRFNIIWRTS--PPKNVLMVKE DATVREAIQL N L VE EL : 149
H7 : ISSPSKQSLIMHN--LRNILITKE STTTTDAVK S L DS BE : 97
H8 : IKSSKRYNIVWSSSSPPTMIYIAKEGDASVRE IE N L CC PS : 131
H9 : ISSPHKEHNMIWTR--LRNVLVVKECQGHVLD VGL N I CE LS : 87
H10 : VTSAMKQSLIMQK--MCNVFITKE TETRASSVH SYL DSFE : 94
H11 : ISSPMKQSLIMHR--LQNVFVTKE ADNTRQAVE A L DS BE : 95
H12 : VGSPHKEHSIIWRT--LQNIYVVKE NEAAMA V T N I AE EA : 198
H13 : ISSPMKQSMVMQK--VRNVLVTRKE TKTKDAVEL SYM DS BE : 100
H14 : ITSPMKQSLIMHK--LQNVLVTKE TDTTREAVR T L DS BE : 99
H15 : VASPMKQSMIMSN--IRNVFITKE GTTTTREAVK T L DS BE : 105
H16 : ISSPMKHALIMPGS--LQNVLVTKE SCATKDAIT N L TT BE : 108
H17 : FTSPMKQSLVMRN--LQNVLVTKE TPTTREAVQ T L ES BE : 104
H18 : ISSPMKQSLIMHR--MCNVFLMKE TDTTREAVE T L DS BE : 109
H19 : VRSPMKQSNIMNS--LENIYVVKE DQDVRD TVR T I DN ES : 113
H20 : VQEST--SSSLVMDK--LENVLIVKE DHNVRESLIQMAS IQR ER : 92
H21 : VSPPMKQSLIMQN--LQNVYITKE TPSTREAVE T L ES BE : 98
H22 : VSPPMKQSLIMQN--LQNVYITKE TPSTREAVE T L ES BE : 98
H23 : AQNPHFFRNLMVQA--PSNVLIVKE RKEDVTF TAQ S I SS EN : 90
H24 : VASPMKQSLIMHT--LQNVLITKE TPTTREAVE T I DS BE : 99
H25 : VTSPMKQSLIMHK--LQNVLITKE TPTTREAVD T I DS BE : 102

```

FIGURE 2 (continued)

```

YPL188W : V V V QPD EEISQDFKS-----PLENDPNR : 125
H1       : V I V VPE EEELN---SIE-----RKSSDPDT : 112
H2       : V V V QPD EEISQDFKS-----PLENDPNR : 125
H3       : L V M AADT EELRAGLCA-----AAPG---- : 114
H4       : V V V NED DELVHEFSSGG-----CNDLDKST : 98
H5       : L V V GES VEELLQEVPAWNQIAS-----KMDISHKY : 126
H6       : C I V NED DELVNEVTT-----VNKIMDKSI : 177
H7       : I V V QPD VDEISKEFTM-----SPVQTPNE : 124
H8       : I V V NQE DELMHCLKTTTTIKQNSSSSISSSTSVSGKSIQELMDPLT : 181
H9       : C I L TEDIVKEIQDKLED-----DSKSGSSM : 114
H10      : I V V QPD DEIAQDFRS-----TPRCDPHQ : 121
H11      : I V V QPD EEIAQDFVK-----NPQSNPGQ : 122
H12      : L V V APD DELAQVYGDGA----- : 219
H13      : I V V QHD KELQEFPV-----KLPSS-QL : 126
H14      : I V V QPD EEIAQDFRS-----TPHCNPNE : 126
H15      : I V V QPP ISEISNDFKT-----FPECNPNH : 132
H16      : I V V LQPD EEVYQDFLT-----TANNDPNR : 135
H17      : I V V QPD EEISQDFKA-----TVHNDPNR : 131
H18      : V V V QPD EEISQDFRS-----MPKSNPNQ : 136
H19      : V V V SED DEIIHETNTGGSVTSIL-----KKQSPKFSSNS : 151
H20      : V I L EEH DEVQKCIGA-----AG : 113
H21      : V V V QPD EEISQDFKS-----PLENDPNR : 125
H22      : V V V QPD EEISQDFKS-----PLENDPNR : 125
H23      : CTV L TEE REFENENPEVF-----NKSGE-NV : 117
H24      : I I L LQPD EEIAQDFTS-----PLNCDFSR : 126
H25      : V I L QRD DEIAQDFKT-----TLNHDPNR : 129

```

```

YPL188W : PHIIY GPQCDEVNRPHELV LGGDGT I H G MFGNTQ PP L A A A G : 175
H1       : FIGIY GPLNEIISKPMI LSLGGDGT I R G LFSNTT PP L S S S G : 162
H2       : PHIIY GPQCDEVNRPHELV LGGDGT I H G MFGNTQ PP L A A A G : 175
H3       : -CVIY GTDSEAARA LLSLGGDGT I R A A G I FSEAR PP L A S S G : 163
H4       : KHIIY GKFGDEVDRHELV LGGDGT I R A G TFSNVT PP L S A A G : 148
H5       : NQTIY GSLSEIDKHELV LGGDGT I R A G SFSNEK PELL S A A G N : 176
H6       : QHVIY GETKDIDKHELV LGGDGT I R G LFSNVK PP L S A A G : 227
H7       : PHVIF GPCEADVSRHELV LGGDGT I H A G MFSNNRQ PP L A S S G : 174
H8       : DHVIY GKNEDEVDRHELV LGGDGT I H G LFSNVV PPIL S A A G : 231
H9       : THSIF GSMADTTKHELV LSLGGDGT I R A MFSNTI PPIL S Y S S G : 164
H10      : PHIIY GTDEEVSRHELV LGGDGT I R A G IFGNRQ PP L A S S G : 171
H11      : PHIIY GEMKDEVDRHELV LGGDGT I R A G LFSMQ PP L A S S G : 172
H12      : RHTIF GSVQDEVAKHELV LGGDGT I R A G AFSNGL PP L S A A G : 269
H13      : PCVIY GDTQDEVDRHELV LSLGGDGT I S A G MFSNTR PP L A S S G : 176
H14      : PHVIF GTEDEVAKHELV LGGDGT I R G MFSANKQ PP L A S S G : 176
H15      : PHVIF GENEDVTKHELV LGGDGT I H A G MFSHMQ PP L A S S G : 182
H16      : PHMIFSGTTDETSKAPHELV LGGDGT I H G MFKS-Q PP L A S S G : 184
H17      : PHLIF GTEEDVATRELV LGGDGT I H S G MFGDKIAPP L A S S G : 181
H18      : PHVIF GSDADVKKHELV LGGDGT I R S G LFSHTQ PP L A S S G : 186
H19      : KYVIY GELSDEVNKEHV LGGDGT I R A G TFSNAN PP L S A A G : 201
H20      : VTAIH GPGEVLRNKEHV LGGDGT I H A G MFSASGE PP L S S S G : 163
H21      : PHIIY GPQCDEVNRPHELV LGGDGT I H G MFGNTQ PP L A A A G : 175
H22      : PHIIY GPQCDEVNRPHELV LGGDGT I H G MFGNTQ PP L A A A G : 175
H23      : KHVIY GKSEEVSKHELV LSLGGDGT I R G LFSNTQ PPIL S S S G : 167
H24      : PHVIFSGTTEEVSRHELV LGGDGT I H G MFGNNO PP L A S S G : 176
H25      : PHVIFGTEEEVTTRELV LGGDGT I H G MFGNNO PP L A S S G : 179

```

FIGURE 2 (continued)

```

YPL188W : LGFL PF KEHKKV CE IS A C PTR EC LKK----- : 213
H1       : LGFL PF NNYAEA KOMFE SSI RERIEC IVKASPOSEALN--- : 209
H2       : LGFL PF KEHKKV CE IS A C PTR EC LKK----- : 213
H3       : LGFL PFE SEHAQALDD LQ AHC QSR VCRVLR----- : 201
H4       : LGFL PF KKASDT RM YE A A RNR EC VLDHYKHQG----- : 192
H5       : LGFL PF KTYKET RA HDNKTFA RKR QCVVKRKS GHVDES---- : 222
H6       : LGFL PF KNSMEC KS YEG A A RNR EC VIRKFVEQDVEDYKN : 277
H7       : LGFL PF CEHKKV DE IS A C PTR EC VVRKGE----- : 215
H8       : LGFL PF KNYKQT RE YEG S A RNR EC VIRKQIVKTLDDGER : 281
H9       : LGFL PF NNFKES NK YT A V PTR EC IVKKNNELPINT--- : 211
H10      : LGFL PF YCEHRRV EK IS A C PTR EC IIRNGC----- : 212
H11      : LGFL PF KEHKRV EE LM A C PTR EC VIRRGs----- : 213
H12      : LGFL PF ARFEEA RA FE S A PTR EC VVRSEALAKPP---- : 315
H13      : LGFL PFE KNYREV EM IS S C PTR EC LVR----- : 214
H14      : LGFL PF CEHKTI EQ IS A C PTR EC QCVVRGE----- : 217
H15      : LGFL PF SEHEVV NQ MN A C PTR EC FIVRKND----- : 223
H16      : LGFL PF KEHKEI NK LM A C PTR EC VIK----- : 222
H17      : LGFL PF KEHEKV SQ IS A C PTR EC VVR----- : 219
H18      : LGFL PFA KEHKKI EQ MT A C PTR EC LVR----- : 224
H19      : LGFL PF STFSES RM YE G A RHR EC VVRKSLNESTQT--- : 248
H20      : LGFL PF KDFKTA DM YS CASVVNPAR ACCKMSIRKEITHLP--- : 210
H21      : LGFL PF KEHKKV CE IS A C PTR EC LKK----- : 213
H22      : LGFL PF KEHKKV CE IS A C PTR EC LKK----- : 213
H23      : LGFL PF KDFKEA KQ FE ALM PREB EC IVKKS----- : 207
H24      : LGFL PF KEYKGV EK IR A C PTR EC FIK----- : 214
H25      : LGFL PF KEHKEI EK IR A C PTR EC LMK----- : 217

```

```

YPL188W : -----KDSNS----- : 218
H1       : -----CQRKDLETSYQNTRSLNAQEEVERLKRLSAAMDAPFDNLT : 249
H2       : -----KDSNS----- : 218
H3       : -----DGLPV----- : 206
H4       : -----C----- : 193
H5       : -----ENQAKDSNGH----- : 232
H6       : ETEAREEDVVDTRLKVL MVASGFGSKYDVQSVKG----- : 311
H7       : -----KAEDA----- : 220
H8       : ANKKLKTNGEKSISKLEEQS--SSSNGSRKIK----- : 312
H9       : -----EGGQPKGSNYSE----- : 223
H10      : -----SEDCS----- : 217
H11      : -----NGKEG----- : 218
H12      : -----QIAEYEIAHYKQHH----- : 329
H13      : -----NGKVT----- : 219
H14      : -----SVDTHN----- : 223
H15      : -----PKKIS----- : 228
H16      : -----KNDGE----- : 227
H17      : -----NGNST----- : 224
H18      : -----NGKTQ----- : 229
H19      : -----PITDATPIKNIRQSESGISPQEN----- : 271
H20      : -----SQSHIEHNSTHVYGNPDDYN----- : 230
H21      : -----KDSNS----- : 218
H22      : -----KDSNS----- : 218
H23      : -----TITDTNPKSMYKSGSDE----LS----- : 226
H24      : -----NGTTK----- : 219
H25      : -----NGSKD----- : 222

```

FIGURE 2 (continued)

```

YPL188W : ----- SIVT AMNDEFHHP NSPHITM D FNGGEFITRRTADGVA : 259
H1       : VSSELEALKKLLKI AMNDVHHP SLPGIVM DVYANGNLLITRTADGLI : 299
H2       : ----- SIVT AMNDEFHHP NSPHITM D FNGGEFITRRTADGVA : 259
H3       : ----- DGRWA AMNDVFIHP GAPHLAH D YVGKQFITSAVADGVA : 248
H4       : ----- VATMV AMNDESHHP SQPNITS D YNNEFITTTTADGIV : 235
H5       : ----- PHIMI AMNDESHHP SQPNITS D VYLNDEFETTATGDGIL : 274
H6       : ----- KRQMV AMNDETIHPASSENITAVD YNNGGEFETTYSDGLI : 353
H7       : ----- RASSI AMNDEFHHP SAPHLAY DVFNNGKYLTTRTADGVA : 262
H8       : ----- EMI AMNDVTIHP SSEPNITS D YNNEFEFTTFADGVI : 352
H9       : ----- FPTKV AMNDVHHP SIPTLTT D FNGGEFITRRTADGIS : 265
H10      : ----- LATSL AMNDLFHHP NSPHLAM D FNGGDFITRESADGVA : 259
H11      : ----- KSVAH AMNDEFHHP NSPHITM D FNGGEYLTTRTADGVT : 260
H12      : ----- NGSMV AMNDESHHP SQPNLIS D YNNESEFLTITTTDGII : 371
H13      : ----- RSCTL AMNDEFIHP NSPHLAHMD TYNNEFLTSTGTGDVA : 261
H14      : ----- LSTSL AMNDEFHHP DSEPHLAM D FNGGDFMTTRTADGIA : 265
H15      : ----- KLVSV AMNDEFHHP NABYLAY DVFNNGEYLTTRTADGIA : 270
H16      : ----- SSVI AMNDEFHHP SSEPHITM D FVNGEFLTRITDGUV : 268
H17      : ----- PIVA AMNDEFHHP NSPHITM D YNNGEYLTTRTADGVT : 265
H18      : ----- QTTTL AMNDEFHHP NSPHITM D YNNGEFMTTRTADGVT : 271
H19      : ----- VTTML AMNDESHHP SQPNLIS D YNNEFLTITTTGDGLV : 313
H20      : ----- LSPLTYAMNDENIHP AEPHITK D HVNGEFITRAITDGVT : 272
H21      : ----- SIVT AMNDEFHHP NSPHITM D FNGGEFLTRITDGVA : 259
H22      : ----- SIVT AMNDEFHHP NSPHITM D FNGGEFLTRITDGVA : 259
H23      : ----- QV AMNDVHHP SLPSLIN DVYVNGHFLTITTTDGLI : 265
H24      : ----- PIVT AMNDEFHHP NSPHLAM D FNGGEYLTTRITDGVT : 260
H25      : ----- PIVT AMNDEFHHP SSEPHIAM D FNGGEYLTTRITDGVT : 263

```

```

YPL188W : LAPTGSTAYSLS GGSNSEL EAMMTE CPRSL FREVPHS H R : 309
H1       : FAPTGSTAYSLS GGSNHEV KCMTE CPRSL FREVPLN H L : 349
H2       : LAPTGSTAYSLS GGSNSEL EAMMTE CPRSL FREVPHS H R : 309
H3       : VAPTGSTAYSLS GGSNSEC ESMTE CPRSL FREVPEST LLR : 298
H4       : FAPTGSTAYSLS GGSNTHEL ECMTE CPRSL FREVPESTCH M : 285
H5       : CSSPTGSTAYSLS GGSNHESE ECMTE CPRSL FREVPESTYKVM : 324
H6       : FAPTGSTAYSLS GGSNTHESA ECMTE CPRSL FREVPESTDM : 403
H7       : LAPTGSTAYSLS GGSNSEL ECMTE CPRSL FREVPESTH K : 312
H8       : FAPTGSTAYSLS GGSNTHESA ECVTE CPRSL FREVPESTDM : 402
H9       : FAPTGSTAYSLS GGSNHEL KCMTE CPRSL FREVPEAT H I : 315
H10      : FAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPHS H R : 309
H11      : LAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPHS H K : 310
H12      : LAPTGSTAYSLS GGSNTHEL ECMTEV CPRSL FREVPESTHVM : 421
H13      : LAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPESTY K : 311
H14      : FAPTGSTAYSLS GGSNSEL ECMTE CPRSL FREVPHS H K : 315
H15      : LAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPHS HVK : 320
H16      : FAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPHS H K : 318
H17      : LAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPHS Y K : 315
H18      : LAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPESTH K : 321
H19      : FAPTGSTAYSLS GGSNTHEL ECMTE CPRSL FREVPESTH V : 363
H20      : LAPTGSTAYSLS GGSNHEP ACMETE CPRSL FREVPESTH K : 322
H21      : LAPTGSTAYSLS GGSNHESE EAMMTE CPRSL FREVPHS H R : 309
H22      : LAPTGSTAYSLS GGSNSEL EAMMTE CPRSL FREVPHS H R : 309
H23      : FAPTGSTAYSLS GGSNHEV ECMTEV CPRSL FREVPESTH K : 315
H24      : LAPTGSTAYSLS GGSNSEL EAMMTE CPRSL FREVPHS H K : 310
H25      : LAPTGSTAYSLS GGSNSEL ESMTE CPRSL FREVPHS H K : 313

```

FIGURE 2 (continued)

```

YPL188W : IKIGSKLN--CKPVNSVVK SVDGIP CD DV DRY INEVGTIYI--- : 354
H1       : IKVIGKENVKIDYTKCNAK SIDGIP LKMVP DRY IISVSRLNSVN : 399
H2       : IKIGSKLN--CKPVNSVVK SVDGIP CD DV DRY INEVGTIYI--- : 354
H3       : LVIGAKSA--CDPAAIKLCMSVDGVSKPE SV DLT TDEV----- : 338
H4       : IRLS-----ELNRNSSIE TIDGIP RD LP DSR VSEKGTIYV--- : 326
H5       : MRLT-----DGSRNSGIM NIDGIP PE RH DRY ISIE----- : 359
H6       : IKLS-----ENNRNSFIE TIDGIS ED HP DLT ITSETIVSNGKIM : 447
H7       : IQVGAQSQ--FDPNDHEIN SVDGVPKET KV DRY INEVGTIYV--- : 357
H8       : IRLS-----ESNRNCRIE TIDGIT PD HP DV ITSEVAITSG--- : 443
H9       : VRVVAQDV--SR--ECSVK SIDGVP VG SQD DRY VSETGTIYL--- : 358
H10      : VRVGSKAT--CGPANAMVK SVDGIP ED RI DRY VNEVGTIYL--- : 354
H11      : IKVGCKAS--CGPDNKLVR SIDGVPLED NV DRY VNEVGTIYV--- : 355
H12      : IKLS-----DSNRNGSIKM NIDGIP CD KP DRY VNEGTIFI--- : 462
H13      : VVISSRVA--HDFDECVLK SIDGIP ED RL DRY RSEAGTIYV--- : 356
H14      : VRIGSKKT--CGPANTVVK SVDGIP PD RI DRY VNEVGTIYL--- : 360
H15      : IRISTRKT--YNLENNKIN SIDGRPYDS DV DRY INETGTIYL--- : 365
H16      : IRVCSKMN--EKTGNNSVNISVDGVA ND EV DRY IVNEVGTIYI--- : 363
H17      : IKVESKMN--MNVANHIVK SIDGIP ED VA DRY INEVGTIYI--- : 360
H18      : IKIGSKHT--GGPDGRVVK SIDGIP EDVYVND DRY VNEVGTIYV--- : 366
H19      : IKLS-----DKNRNASIK NIDGIS CD QP DRY AGENDSALA--- : 404
H20      : ITAS-----SESRRGAE SVDGIAGLVRPSDK L ESETG----- : 359
H21      : IKIGSKLN--CKPVNSVVK SVDGIP CD DV DRY INEVGTIYI--- : 354
H22      : IKIGSKLN--CKPVNSVVK SVDGIP CD DV DRY INEVGTIYI--- : 354
H23      : VIVRSKG---LSGHDCSAK SIDGIP LK SA DRY ISESGSLDPS- : 361
H24      : IKVSSKMN--CKNLSNVVM SIDGVP HD QI DRY INEVGTIYV--- : 355
H25      : IRVGSKMN--CNHRNNVVM SIDGIP CD QI DRY INEIGTIYV--- : 358

```

```

YPL188W : -----DGTQLPTTRKTENDFNNSKK----- : 374
H1       : DDEDDIASGTTADAPDCVNASTTVSKESKTKSLGRRRGVCK-----R : 441
H2       : -----DGTQLPTTRKTENDFNNSKK----- : 374
H3       : -----STAHP----- : 344
H4       : -----PGQHEPPTTLRHTN--TANTNS----- : 347
H5       : -----PSQRS----- : 364
H6       : LGGVGVGGGVGGKGVGGVCSSSSVSGGICKIKEEEPEEVYVNGQDVV--- : 494
H7       : -----NGSELPASCSS- GKFKKKPG----- : 376
H8       : -GQSLLSSTGGGNKTTSTSTSACTSSGSSGSTSSSPSGSGNNTSGS--- : 489
H9       : -----PDTKLPPVLESHRVGQHI SDHHDSKIAKVTNKGID : 393
H10      : -----NGSQLPSTSNKRQVSEERRKS----- : 374
H11      : -----NGSQLPSTTDAATIKRRGCD----- : 375
H12      : -----PGKHQPPGSLRNREEFDHYVEPGHFYFQTGKG--- : 494
H13      : -----EGSIVPNRKDTSGLSNMKDP----- : 376
H14      : -----NGSQVPS-ADKKEVRERRKT----- : 379
H15      : -----NGSELPPEYNSPDLFKRKSK----- : 385
H16      : -----DGTQVPPSKKLADETLPNSTS----- : 384
H17      : -----DGTQFPPIAAKKQTADKATITGKT-----PDKD : 387
H18      : -----NGSKLP--IPQEAIKSQKTS----- : 384
H19      : -----TSSRSEKVGSGST SENEQTDSTDSTPSSKQDL---- : 435
H20      : ----- : -
H21      : -----DGTQLPTTRKTENDFNNSKK----- : 374
H22      : -----DGTQLPTTRKTENDFNNSKK----- : 374
H23      : -----WNLPTSISLKKKTT-----T : 376
H24      : -----DGTQLPTLAKNSNGEENQTLKKR-----E--- : 379
H25      : -----DGTQLPTATINNSNSQND SKISK-----K--- : 382

```

FIGURE 2 (continued)

```

YPL188W : -PKRS-IY-V-KTEM-IRGI-ELGGFNS-RLTKRQTDND----- : 414
H1       : TAERS-VW-VVQSKG-VNGI-GMGGFNLG-KSSKSNK----- : 479
H2       : -PKRS-IY-V-KTEM-IRGI-ELGGFNS-RLTKRQTDND----- : 414
H3       : --RDAGVY-V-RSEMT-TRGI-ELGGFNG-RRGGRRTHPPS----- : 383
H4       : ----K-IW-I-RSEMT-TKDI-ELGGFNS-OGKKQYHM----- : 382
H5       : -----LW-I-RSRH-TKDI-ELGGFNL-RRHGGVRRS----- : 398
H6       : YNDKN-IW-V-TNEM-AKDL-SGGFNS-RRDCKGKKLHL----- : 535
H7       : -VKNS-IY-V-KTEM-AQGI-ELGGFNS-RFSHIRNKIDDTNT----- : 420
H8       : YGDKN-IW-V-TNQNC-SKDL-SGGFNS-RRDCKGKRLHL----- : 530
H9       : KTHKS-IW-V-KSEM-TTGI-MLGGFNS-RGFQKP----- : 430
H10      : -IKNS-IY-I-KTEM-TRGL-ELGGFNS-RLTKPQKEEBLERDLTD-- : 421
H11      : -IKNA-VY-V-KTEM-TRGI-ELGGFNS-RFMSQREKEGENSK----- : 419
H12      : TEVKK-LF-V-NTES-IKGI-QLGGFNS-RGCKM----- : 530
H13      : --KST-IY-V-RTEM-IRGI-ELGGFNG-RAISHKKP----- : 413
H14      : -IKNS-IY-V-RTEM-THGI-ELGGFNS-RFTRNSKNENKDK----- : 422
H15      : -IKYS-IY-V-KTEM-SLGI-ELGGFNS-RFKPAKTKNDKPSNCSFEE : 434
H16      : -KQRA-IY-I-KSEM-IKGI-DGGFNG-RFTRRKQDN----- : 423
H17      : VQKRS-IY-V-KTEM-IKGI-ELGGFNS-RFIHRGKDDDEKK----- : 431
H18      : -LKNS-IY-V-KTEM-IKGI-ELGGFNS-RFGSKS----- : 420
H19      : --CCK-IW-V-RSEMT-TKDI-ELGGFNS-KASKSKL----- : 471
H20      : --HNS-IW-V-KTDR-VSGL-GGGFNS-GKGGEASGDVA----- : 399
H21      : -T----- : 375
H22      : -A----- : 375
H23      : DEAKT-LW-V-RTKG-VTGI-MLGGFNG-KSINDAHP----- : 415
H24      : -SKNS-IY-V-KSEM-TRGI-ELGGFNG-RFIRRSKDN----- : 418
H25      : -IKKS-IY-V-KTEM-TRGI-ELGGFNS-RFVRNKKDSE----- : 422

```

FIGURE 2 (continued)

MatGat (matrix = BLOSUM50; first_gap = 12; extending_gap = 2)	2	3	4	5	6	7	8	9	10	11	12	13
1. YPL188W	41,7	99,8	46	52	42,8	38,1	64	39,5	48,2	66,8	67,4	41,7
2. H1		41,5	36,6	42	37,2	38,5	42	38,6	43,7	40	40,9	34,7
3. H2			45,8	51,8	42,6	37,9	63,7	39,3	48	66,6	67,1	41,5
4. H3				42,1	41,9	32,2	45,4	34	39,5	46,1	48,2	33,9
5. H4					51,4	44,8	49,1	45,5	51,2	49,6	49,9	46,7
6. H5						37,5	41,5	40,2	42,7	39,9	41,3	39
7. H6							37,9	52,9	40,7	35,9	39,2	41
8. H7								38,6	49,7	62,4	63,7	39,1
9. H8									39,5	36,9	39	38,5
10. H9										47,8	49,7	42
11. H10											68,2	38,9
12. H11												41,3
13. H12												
14. H13												
15. H14												
16. H15												
17. H16												
18. H17												
19. H18												
20. H19												
21. H20												
22. H21												
23. H22												
24. H23												
25. H24												
26. H25												

FIGURE 3

MatGat (matrix = BLOSUM50; first_gap = 12; extending_gap = 2)	14	15	16	17	18	19	20	21	22	23	24	25	26
1. YPL188W	59,1	69,3	61,2	64	70,6	67,3	45,3	41,9	90,1	90,3	47	72,1	73,7
2. H1	40,3	41,9	39,1	40,5	41,3	41,6	43,2	38,8	38,7	38,7	50,4	42,2	40,9
3. H2	58,9	69	61	63,8	70,3	67,1	45,1	41,7	89,9	90,1	46,7	71,8	73,5
4. H3	49	47,5	43,1	46,3	45,4	47,9	38,5	40	43,8	43,6	41,5	45,2	45,2
5. H4	49,4	49,1	47,4	48,7	49,3	50,9	54,4	42,6	48,7	48,9	50,2	50,2	51,6
6. H5	42,3	42,6	39,6	41,1	40,5	42,2	44,8	39,2	42,4	42,2	44,1	41,4	42,4
7. H6	37,7	36,7	37	37,9	37,6	38,6	45,3	34,5	35,7	35,7	38,1	37,8	39,8
8. H7	57,6	65,6	65,1	58	62,2	61,7	44,1	41	58,5	58,5	48,4	64	62,8
9. H8	38,4	37,2	39	39	38,5	38,4	46,2	34,9	36,4	36,4	40	38,5	39,5
10. H9	48,8	50,6	48,3	49,5	48,8	50,4	47,1	42,3	44,4	44,6	52,2	49,6	50,6
11. H10	57,5	76,3	59,7	57,8	62,7	64,9	43,7	41,8	60,9	61,2	45,4	66,6	67,4
12. H11	57,9	69,8	62,2	60,7	64,1	70,2	44	42,2	62,6	62,9	46,3	66,3	64,4
13. H12	38,7	40,1	39,6	39,7	42,1	41,6	44,7	32,8	38,9	39,1	39,9	40	40,3
14. H13		57,8	53,3	51,9	57,1	60,9	45,5	39,3	53,8	54,1	46,9	58,1	56,6
15. H14			62,9	62,1	66,6	69,4	46,2	43,2	64	64,2	48,8	68,9	69,1
16. H15				54,2	60,4	59,8	44,4	40,2	56,2	56,4	46,7	59,4	61,4
17. H16					62,5	59,4	43,1	38,6	58,3	58,5	46,5	66,1	64,3
18. H17						65,1	44,8	40,6	64,1	64,4	45,2	68,9	70,3
19. H18							47,5	41,9	63,3	63,5	47,8	65,3	64,7
20. H19								40,9	42	42,4	46,7	46,3	47,5
21. H20									40,6	40,8	42,2	42,4	42,6
22. H21										99,5	44,8	66	67,1
23. H22											44,8	66,3	67,3
24. H23												47,4	47,6
25. H24													79,3
26. H25													

FIGURE 3 (continued)

needle (EBLOSUM62, gap_open 10.0, gap_extend 0.5)	YPL188W	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13
YPL188W	100	40,7	99,8	45,2	51,9	42,3	37,3	61,9	38,4	47,8	66,6	66,8	40,8	58,2
H1	40,7	100	40,5	36	41,3	35,7	37,6	41,8	37,3	42,3	40	40	33,6	39,5
H2	99,8	40,5	100	45	51,7	42,1	37,2	61,7	38,2	47,6	66,4	66,6	40,7	57,9
H3	45,2	36	45	100	41,9	40,3	31,1	43	32,5	39	45,2	47	31,6	47,9
H4	51,9	41,3	51,7	41,9	100	50,6	44,2	48,7	44,7	50,7	49,4	50	46	49
H5	41,7	35,7	41,5	40,3	50,6	100	36,6	41,3	40,2	41,3	39	40,3	34,9	42,5
H6	37,3	37,6	37,2	31,1	44,2	36,6	100	35	52,5	39,4	35,8	36,6	40,4	36,3
H7	61,9	41,8	61,7	43	48,7	41,3	35	100	36,2	49	61,1	62,6	38,3	57,7
H8	38,4	37,3	38,2	32,5	44,7	40,2	52,5	36,2	100	37,1	36,3	37,9	36,3	36,3
H9	47,8	42,3	47,6	39	50,7	41,3	39,4	49	37,1	100	47,3	49,8	41,3	46,6
H10	66,6	40	66,4	45,2	49,4	39	35,8	61,1	36,3	47,3	100	67,5	36,3	56,8
H11	66,8	40	66,6	47,1	50	40,3	36,6	62,6	37,9	49,8	67,5	100	39,2	57,1
H12	40,8	33,6	40,7	31,6	46	34,9	40,4	38,3	36,3	41,3	36,3	39,2	100	37,5
H13	58,2	39,5	57,9	47,9	49	42,5	36,3	57,7	36,3	46,6	56,8	57,1	37,5	100
H14	68,9	41,5	68,6	46,8	48,6	42,8	35,2	65,3	36,9	49,2	75,9	69,3	38,7	57,3
H15	60,6	38,3	60,3	42,3	46,6	39,2	36,8	64,9	38,3	47,1	58,9	61,4	37	52
H16	62,6	40,1	62,3	45	48,8	41,2	36,8	56,9	38,5	48,2	57,5	59,5	37,9	51
H17	70,1	40,1	69,9	44,2	48,6	39,6	37,1	62	37,4	48,8	61,4	62,6	38,7	57,2
H18	67,1	41,5	66,8	46,1	50,3	39,7	37	61,5	37,8	48,9	65,1	70,2	38,5	59,4
H19	44,6	41,3	44,4	35,8	54,4	44,4	42,9	40,9	45,6	45,4	43,4	43,4	42,4	44,9
H20	41,7	38,9	41,5	39,5	41,6	37,9	32,2	40,5	32,7	41,5	41,7	41,8	31	37,9
H21	90,1	37,1	89,9	43	48	40,6	33	56,6	35,4	42,6	60,9	62,1	36,8	51,8
H22	90,3	37,3	90,1	42,5	48,2	40,4	33	56,8	35,3	43	61,2	62,3	37	52
H23	46,9	50	46,7	40,8	50,5	43,8	36,3	47,1	39	51,6	44,2	46,3	37,3	46,8
H24	71,8	41,6	71,6	43,9	49,3	40,8	37,2	63,7	37,8	48,7	66,4	65,6	37,7	57,9
H25	73,7	40	73,5	44,1	49,5	40,5	38,3	62,5	37,9	49,3	66,1	64,7	38,3	56,3

FIGURE 4

needle (EBLOSUM62, gap_open 10.0, gap_extend 0.5)	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24	H25
YPL188W	68,9	60,6	62,6	70,1	67,1	44,6	41,7	90,1	90,3	46,9	71,8	73,7
H1	41,5	38,3	40,1	40,1	41,5	41,3	38,9	37,1	37,3	50	41,9	40
H2	68,6	60,3	62,3	69,9	66,8	44,4	41,5	89,9	90,1	46,7	71,6	73,5
H3	46,8	42,3	45	44,2	46,1	35,8	39,5	43	42,5	40,8	43,9	44,1
H4	48,6	46,6	48,8	48,6	50,3	54,4	41,6	48	48,2	50,5	49,3	49,5
H5	42,8	39,2	41,2	39,6	39,7	44,4	37,9	40,1	39,9	43,8	40,8	40,5
H6	35,2	36,8	36,8	37,1	37	42,9	32,2	33	33	36,3	37,2	38,3
H7	65,3	64,9	56,9	62	61,5	40,9	40,5	56,6	56,8	47,1	63,7	62,5
H8	36,9	38,3	38,5	37,4	37,8	45,6	32,7	35,4	35,3	39	37,8	38,2
H9	49,2	47,1	48,2	48,8	48,9	45,4	41,5	42,6	43	51,6	48,7	49,3
H10	75,9	58,9	57,5	61,4	65,1	43,4	41,7	60,9	61,2	44,3	66,4	66,1
H11	69,3	61,4	59,5	62,6	69	43,4	41,6	62,1	62,3	46,3	65,6	64,7
H12	38,7	37	37,9	38,7	38,5	42,4	31	36,8	37	37,3	37,7	38,3
H13	57,3	52	51	57,2	59,4	44,9	37,9	51,8	52	46,8	57,9	56,3
H14	100	61,1	61,7	66,1	68,2	45,8	41,3	62	63,5	47,9	68,6	68,5
H15	61,1	100	51,3	59,1	56,7	43	39,3	54,8	55	46,5	59,6	60,2
H16	61,7	51,3	100	62,3	59,4	40,5	37,6	56,2	56,4	45,6	64,2	63,7
H17	66,1	59,1	62,3	100	63,8	44,2	40	61,4	62,2	44,9	67,8	69,5
H18	68,2	56,7	59,4	63,8	100	46,4	41,9	62,8	63,1	46,9	65,1	64,3
H19	45,8	43,5	40,5	44,1	46,4	100	39,7	39,8	42,3	46,4	45,6	44,8
H20	40,9	39,3	37,6	40	41,9	39,7	100	40,4	39,7	42,2	39,8	41,2
H21	62	54,8	56,2	61,4	62,8	39,8	40,4	100	99,5	43,9	65,3	66,4
H22	63,5	55	56,4	62,2	63,1	42,3	39,7	99,5	100	43,7	65,6	66,6
H23	47,9	46,5	45,6	44,9	44,5	45,8	42,2	43,9	43,7	100	47,2	47
H24	68,6	59,6	64,2	67,8	65,1	45,6	39,8	65,3	65,6	47,2	100	79,1
H25	68,5	60,2	63,7	69,5	64,3	44,8	41,2	66,4	66,6	47	79,1	100

FIGURE 4 (continued)

Table IA: Nucleic acid sequence ID numbers

	1.	2.	3.	4.	5.
Appli- cation	Hit	Locus	Organism	Lead SEQ ID	SEQ IDs of Nucleic Acid Homologs
1	1	YPL188W	S.cerevisiae	1	3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51

Table IIA: Amino acid sequence ID numbers

	1.	2.	3.	4.	5.
Appli- cation	Hit	Locus	Organism	Lead SEQ ID	SEQ IDs of Polypeptide Homologs
1	1	YPL188W	S. cerevisiae	2	4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52

Table III: Primer nucleic acid sequence ID numbers

	1.	2.	3.	4.	5.
Appli- cation	Hit	Locus	Organism	Lead SEQ ID	SEQ IDs of Primers
1	1	YPL188W	S. cerevisiae	1	53, 54

Table IV: Consensus amino acid sequence ID numbers

	1.	2.	3.	4.	5.
Appli- cation	Hit	Locus	Organism	Lead SEQ ID	SEQ IDs of Pattern Sequences
1	1	YPL188W	S. cerevisiae	2	55, 56, 57, 58, 59, 60, 61

FIGURE 5