

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

03 December 2020 (03.12.2020)



(10) International Publication Number

WO 2020/239680 A2

(51) International Patent Classification:

A01H 1/02 (2006.01)

C07K 14/415 (2006.01)

A01H 1/08 (2006.01)

C12N 15/82 (2006.01)

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))

(21) International Application Number:

PCT/EP2020/064420

(22) International Filing Date:

25 May 2020 (25.05.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

19176646.8

25 May 2019 (25.05.2019)

EP

(71) Applicant: **KWS SAAT SE & CO. KGAA** [DE/DE];

Grimsehlstraße 31, 37574 Einbeck (DE).

(72) Inventors: **KLOIBER-MAITZ, Monika**; Im Pohlschen

Bock 2, 37574 Einbeck (DE). **OUZUNOVA, Milena**;

Max-Born-Ring 2c, 37077 Göttingen (DE). **BOLDUAN,**

Christof; Hulderson 22a, 37574 Einbeck (DE). **WIECK-**

HORST, Silke; Langer Wall 16a, 37574 Einbeck (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, DJ, DK, DM, DO,

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

KR, KW, KZ, LA, LC, LK, LR, LS, 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, ST, SV, SY, TH, TJ, TM, TN, TR,

TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, ST, 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).

Declarations under Rule 4.17:

— as to the identity of the inventor (Rule 4.17(i))

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

(54) Title: HAPLOID INDUCTION ENHANCER

(57) Abstract: The present invention relates to methods for haploid induction, based on the HIR8 QTL on maize chromosome 9, as well as associated genes and markers. The invention further relates to identification of haploid inducers, as well as methods for transgenically generating haplotype inducers as well as the combination of haplotype induction and gene editing.



WO 2020/239680 A2

HAPLOID INDUCTION ENHANCER

FIELD OF THE INVENTION

The invention relates to the field of plant biotechnology, more particularly genetic engineering
5 and gene editing, as well as plant breeding. In particular, the invention provides nucleic acids
which, after transcription or after expression in a plant, are suitable for enhancing the haploid
induction capability of a haplotype inducer plant, such as caused by a mutated patatin-like
phospholipase (PLA1).

10 BACKGROUND OF THE INVENTION

Commonly, during haploid induction, both parent lines used in the induction cross are both
diploids, so their gametes (egg cells and sperm cells) are haploids. Haploid induction is
frequently a medium to low penetrance trait of the inducer line, so the resulting progeny,
depending on the species or situation, may be either diploid (if no genome loss takes place) or
15 haploids (if genome loss does indeed take place). If the parent line that is crossed to the
haploid inducer is not diploid, but rather a tetraploid, hexaploid, or other plant of higher ploidy,
the term haploid induction is something of a misnomer, because the haploid progeny
produced will have a gametic chromosome number, and thus would not really be haploids, but
rather diploids (if the parent is tetraploid) or triploids (if the parent is hexaploid) and so on.
20 Therefore, as used herein, haploids possess half the number of chromosomes of either parent;
thus haploids of diploid organisms (e.g., maize) exhibit monoploidy; haploids of tetraploid
organisms (e.g., ryegrasses) exhibit diploidy; haploids of hexaploid organisms (e.g., wheat)
exhibit triploidy.

- Haploid induction can occur during self-pollination or intercrossing of two lines within the same species, or it can occur during wide crosses, where it can be viewed as a hybridization barrier, preventing the formation of interspecific hybrids. In maize, the most commonly employed method of inducing haploids is through the use of an intraspecific haploid inducer male line, which is primarily triggered by rearrangements of, mutations in, and/or recombinations, insertion, or deletions within a region of chromosome 1, specifically the PHOSPHOLIPASE A1 (PLA1) gene, also known as NOT LIKE DAD1 (NLDI) and MATRILINEAL (MATL) (WO 2016/075255 A1; Kelliher et al. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature*, 542(7639), 105.).
- 10 Haploid inducer (HI) maize lines contain a quantitative trait locus (QTL) on Chromosome 1 responsible for at least 66% of the variation in haploid induction. The QTL causes haploid induction at different rates when it is introgressed into various backgrounds. All maize haploid inducer lines used in the seed industry are derivatives of the founding HI line, known as Stock6, and all have the haploid inducer chromosome 1 QTL mutation.
- 15 In maize, haploid seed or embryos are specifically produced by making crosses between a haploid inducer male (i.e., haploid inducer pollen) and virtually any ear that one chooses - the ear could be of any inbred, hybrid, or other germplasm. Haploids are produced when the haploid inducer pollen DNA is not fully transmitted and/or maintained through the first cell divisions of the embryos. The resulting phenotype is not fully penetrant, with some ovules containing haploid embryos, and others containing diploid embryos, aneuploid embryos, chimeric embryos, or aborted embryos. The haploid kernels have embryos that contain only the maternal DNA plus normal triploid endosperm. After haploid induction, haploid embryos or seed are typically segregated from diploid and aneuploid siblings using a phenotypic or genetic marker screen and grown or cultured into haploid plants. These plants are then
- 20 converted either naturally or via chemical manipulation (e.g., using an anti-microtubule agent such as colchicine) into doubled haploid (DH) plants which then produce inbred seed.
- 25 Plant breeding is facilitated by the use of doubled haploid (DH) plants. The production of DH plants enables plant breeders to obtain inbred lines without multi- generational inbreeding, thus decreasing the time required to produce homozygous plants. DH plants provide an invaluable tool to plant breeders, particularly for generating inbred lines, QTL mapping, cytoplasmic conversions, trait introgression, and F2 screening for high throughput trait improvement. A great deal of time is spared as homozygous lines are essentially generated in one generation, negating the need for multigenerational single-seed decent (conventional inbreeding). In particular, because DH plants are entirely homozygous, they are very amenable
- 30

to quantitative genetics studies. The production of haploid seed is critical for the doubled haploid breeding process. Haploid seed are produced on maternal germplasm when fertilized with pollen from a gynogenetic inducer, such as Stock 6 and Stock 6-derivative lines.

5 Recently, it has been tried to transfer the HI system from maize to sorghum (WO 2018/158301 A1) and other crops. However, the achieved induction rates were rather low and reached in maximum approximately 2%. From the commercial perspective this is usually not sufficient for setting-up a new HI system.

10 Additionally, in principle it is possible to develop new haploid inducer line which are not directly derived from Stock 6, however for that it is crucial to fully understand the HI system and all its genetic components. A second QTL on chromosome 9 (HIR8), increasing the induction rate to approx. 10% was recently published (Liu et al. (2015) Fine mapping of qhir8 affecting in vivo haploid induction in maize. Theor Appl Genet. 2015 Dec;128(12):2507-15). HIR8 locus had been narrowed down to a 789 kb region flanked by markers 4292232 and umc1867. Within the 789 kb intervals, there were 35 genes according to the B73 reference
15 sequence annotated. However, the region is non-conserved and marker development in this region was not possible and candidate genes had not been selected based on expression and regulatory sequences or any other analyses.

In order to improve in vivo haploid induction based on the mutated PLA1 gene in maize and to transfer the system to other crops, it is important to know the relevant genes and their
20 functional mutations or haplotypes.

It is an object of the present invention to provide means and methods which may be used to produce an *in vivo* haploid inducer, or to increase the induction capability of an inducer genotype and/or to produce a haploid plant.

25 SUMMARY OF THE INVENTION

The present invention is based on the identification of genes for the haploid inducer locus *qhir8*. Four genes could be identified that, individually or in any combination, are responsible for and/or involved in haploid induction capability, in particular in enhancing the haploid induction capability of an inducer plant caused by a mutated patatin-like phospholipase (PLA1).
30 Accordingly, (increased) haploid induction capability may be achieved via manipulation of the identified genes, in particular in an inducer plant comprising a mutated patatin-like phospholipase (PLA1).

- Identification of the HIR8 genes helps to improve the maize inducer and to develop an effective inducer in Sorghum or other species, in particular where the phospholipase system is functional for maternal induction. In particular Sorghum is of interest. As already shown in WO 2018/158301 A1 different mutations in PLA1 gene of Sorghum are able to confer the capability of a haploid inducer to a sorghum plant. Due to the knowledge of the genetic nature of the HIR8 locus, the haploid inducer rates in Sorghum based on mutated PLA1 can be further increased. Another application of the knowledge of the joint effect of HIR1 and HIR8 is the use haploid induction simultaneously with gene editing. The disclosure of WO 2018/102816 relates to this concept and is herein incorporated by reference in its entirety.
- 10 The present invention is of particular relevance, given the lack of markers in the non-conserved regions. The original markers used for haplotyping and annotation were based on B73. The most interesting regulatory regions were missed working only with B73. Further, the previous failure to develop a TILLING population of RWS, necessitated identification of functional mutants instead of knockout mutants in a non-inducing line.
- 15 The invention in an aspect relates to means and methods for enhancing the induction capability of a haploid inducer plant, in particular caused by a mutated patatin-like phospholipase (PLA1). More particularly, nucleic acids are provided, which, after transcription or after expression in a plant, are suitable for enhancing the induction capability of an inducer plant caused by a mutated patatin-like phospholipase (PLA1) through manipulation of the identified gene(s). The invention also relates to methods for producing a plant which is a carrier of an enhancer for the induction capability of an inducer plant, in particular caused by a mutated patatin-like phospholipase (PLA1) by mutagenizing an endogenous DNA sequence that is identical to a nucleic acid described herein or by transforming a plant with the nucleic acid(s) described herein, as well as methods and molecular markers for identifying such carriers. The invention further relates to methods for producing plants that are suitable for use as haploid inducer, as well as the haploid inducer plants and parts thereof. The invention further relates to methods for producing haploid plants or double haploid plants, in particular by crossing a plant with a haploid inducer plant according to the invention. The invention also relates to haploid plants and double haploid plants. The invention also relates to methods of editing plant genomic DNA using a haploid inducer plant according to the invention and transforming it so that it contains DNA coding for cellular machinery capable of editing genes.
- 20
- 25
- 30
- The present invention is in particular captured by any one or any combination of one or more of the below numbered statements 1 to 64, with any other statement and/or embodiments.

1. A nucleic acid molecule, comprising a nucleotide sequence, which:

- (i) is a sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 10, 12, 15, 18, 21, 30, and 35, or a functional fragment thereof, preferably SEQ ID NOs: 1, 4, 10, 15, 30, and 35;
- (ii) has a coding sequence selected from the group consisting of SEQ ID NOs: 2, 6, 13, 16, 19, 22 and 36, preferably SEQ ID NOs: 2, 16, and 36; or
- (iii) is complementary to the sequence from (i) or (ii); or
- (iv) is at least 80% identical to the sequence from (i) or (ii); or
- (v) encodes a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 3, 7, 9, 11, 14, 17, 20, 23 and 37, or a functional part of the protein, preferably SEQ ID NOs: 3, 11, 17, and 37, or a functional part of the protein; or
- (vi) encodes a protein comprising an amino acid sequence which is at least 80% identical to a sequence selected from the group consisting of SEQ ID NOs: 3, 7, 9, 11, 14, 17, 20, 23 and 37, or a functional part of the protein, preferably SEQ ID NOs: 3, 11, 17, and 37, or a functional part of the protein; or
- (vii) hybridizes with a sequence from (iii) under stringent conditions.

2. A nucleic acid molecule, comprising a nucleotide sequence, which:

- (i) is a sequence selected from the group consisting of SEQ ID NOs: 24, 27, and 32, or a functional fragment thereof, preferably SEQ ID NO: 24, or a functional fragment thereof; or
- (ii) has a coding sequence selected from the group consisting of SEQ ID NOs: 25, 28 and 33; preferably SEQ ID NO: 25; or
- (iii) is complementary to the sequence from (i) or (ii); or
- (iv) is at least 80% identical to the sequence from (i) or (ii); or
- (v) encodes a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 26, 29 and 34, or a functional part of the protein, preferably SEQ ID NO: 26, 29 and 34 or a functional part of the protein; or
- (vi) encodes a protein comprising an amino acid sequence which is at least 80% identical to a sequence selected from the group consisting of SEQ ID NOs: 26, 29 and 34, or a functional part of the protein, preferably SEQ ID NO: 26, or a functional part of the protein; or
- hybridizes with a sequence from (iii) under stringent conditions.

3. A nucleic acid molecule encoding an RNA that has a double-stranded portion, wherein at least one strand of the double-stranded portion has a nucleotide sequence which is homologous or identical to at least 14 successive nucleotides of a coding sequence of a nucleic acid that:

- (i) has a sequence selected from the group consisting of SEQ ID Nos: 25, 28, and 33, or a fragment thereof, in a sense or anti-sense orientation; or
- (ii) is complementary to the sequence from (i); or
- (iii) is at least 80% identical to the sequence from (i); or
- 5 (vi) hybridizes with the sequence from (ii) under stringent conditions.
- 4. The nucleic acid according to statement 3, which is an RNAi molecule.
- 5. The nucleic acid molecule according to any of statements 1 or 3-4, wherein said nucleic acid, after transcription or expression in a plant, or the nucleic acid molecule according to statement 2, wherein said nucleotide, after silencing in a plant, is suitable for enhancing the haploid
- 10 induction capability of a haploid inducer plant.
- 6. The nucleic acid molecule according to statement 5, wherein said haploid inducer plant comprises the HIR1 QTL.
- 7. The nucleic acid molecule according to statement 5 or 6, wherein said haploid inducer plant is selected from or derived from Stock 6, RWK, RWS, UH400, AX5707RS, NP2222-mai/.
- 15 8. The nucleic acid molecule according to any of statements 5 to 7, wherein said haploid inducer plant comprises a mutated patatin-like phospholipase (PLA1).
- 9. The nucleic acid molecule according to statement 8, wherein said PLA1 has a sequence identical to the PLA1 sequence of Stock 6, RWK, RWS, UH400, AX5707RS, or NP2222-mai/.
- 10. The nucleic acid molecule according to statement 8, wherein said PLA1 sequence
- 20 (i) having the nucleotide sequence of SEQ ID NO: 41 or 47 or a sequence which is at least 80% identical to SEQ ID NO: 41 or 47,
- (ii) having the coding sequence of SEQ ID NO: 42 or 48 or a sequence which is at least 80% identical to SEQ ID NO: 42 or 48, or
- (iii) encoding the amino acid sequence of SEQ ID NO: 43, 49 or 50 or a sequence which is at
- 25 least 80% identical to SEQ ID NO: 43, 49 or 50,
- comprises one or more mutation, preferably producing a change in the transcription or expression rate of the PLA1 sequence in a plant, in comparison to a non-mutagenized wild-type plant, or a change in the activity or stability of the encoded amino acid sequence in a plant, in comparison to a non-mutagenized wild-type plant, preferably wherein the mutation causes the
- 30 property of a haploid inducer plant or increases the induction capability of a haploid inducer plant,
- preferably wherein the mutation is
- (A) a deletion or mutation in SEQ ID NO: 41, 42, 47 or 48, preferably which leads to an early stop codon causing a shortening of the encoded amino acid sequence, wherein

- preferably the amino acid sequence at the C-terminal end is shortened by 5 or more amino acids, and/or
- (B) a deletion or mutation in SEQ ID NO: 47 or 48 preferably resulting in an amino acid sequence having SEQ ID NO: 46, or a deletion in SEQ ID NO: 41 or 42 preferably resulting
- 5 in an amino acid sequence having SEQ ID NO: 40;
- (C) an alteration in the coding sequence of SEQ ID NO: 42 or 48 which causes an amino acid exchange between amino acid positions 57 and 289 of SEQ ID NO: 43, 49, or 50;
- (D) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 58, wherein arginine is replaced by glutamine (R58Q);
- 10 (E) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 74, wherein aspartate is replaced by asparagine (D74N);
- (F) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 78, wherein glycine is replaced by arginine (G78R);
- (G) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or
- 15 50 at position 160, wherein valine is replaced by isoleucine (V160I);
- (H) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 288, wherein serine is replaced by leucine (S288L); and/or
- (I) is a knock-out mutation.
11. The nucleic acid according to statement 8, wherein said PLA1 has a sequence as set forth in
- 20 any of SEQ ID NO: 38-40 or SEQ ID NO: 44-46.
12. A vector comprising the nucleic acid according to any of statements 1 to 11.
13. The vector according to statement 12, which is an expression vector.
14. The vector according to statement 13, which is a conditional expression vector.
15. The vector according to statement 13 or 14, which is a tissue-specific expression vector.
- 25 16. The vector according to any of statements 13 to 15, which is a pollen-specific expression vector.
17. The vector according to any of statements 13 to 16, which is an inducible expression vector.
18. A plant cell comprising the nucleic acid or vector according to any of statements 1 to 17.
19. A plant or a part thereof comprising the plant cell according to statement 18.
- 30 20. The plant cell, plant, or plant part according to statement 18 or 19, which is transgenic or gene-edited.
21. A method for producing a (transgenic) plant, preferably a (transgenic) plant which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of statements 5 to 11 comprising:

a1) transforming at least one plant cell with a nucleic acid molecule according to statement 1 or 3 or a vector according to any of statements 11 to 17 for overexpression of myosin heavy chain,

5 a2) transforming at least one plant cell with a nucleic acid molecule according to statement 2 or a vector according to any of statements 11 to 17 for repression of aquaporin, or

a3) transforming at least one plant cell with a nucleic acid molecule according to statement 1 or 3 or a vector according to any of statements 11 to 17 for expression of PPR corresponding to RWS PPR, or

10 a4) transforming at least one plant cell with a nucleic acid molecule according to statement 1 or 3 or a vector according to any of statements 11 to 17 for expression of the gene set forth in SEQ ID NO: 35; and

b) optionally regenerating a plant from the at least one transgenic plant cell of a) or a transgenic progeny thereof.

22. A method for producing a (mutated) plant, preferably a (mutated) plant which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of statements 5 to 11

comprising the following steps:

a) mutagenizing plant cells and subsequent regenerating of plants from the mutagenized plant cells or mutagenizing plants, and

20 b) identifying a regenerated or mutagenized plant which:

has at least one mutation in an endogenous DNA sequence that is identical to a nucleic acid according to statement 1 or 2, or

has at least one mutation in a regulatory sequence of said endogenous DNA sequence, wherein said mutation produces an increase in the transcription or expression rate of the endogenous DNA sequence that is identical to a nucleic acid according to statement 1 in the identified plant or a decrease in the transcription or expression rate of the endogenous DNA sequence that is identical to a nucleic acid according to statement 2 in the identified plant, in comparison to a non-mutagenized wild-type plant, or wherein said mutation produces an increased activity or stability of a protein or polypeptide encoded by the endogenous DNA sequence that is identical to a nucleic acid according to statement 1 in the identified plant or an decreased activity or stability of a protein or polypeptide encoded by the endogenous DNA sequence that is identical to a nucleic acid according to statement 2 in the identified plant, in comparison to a non-mutagenized wild-type plant,

preferably wherein said at least one mutation causes the property of an enhancer of the haploid induction capability of a haploid inducer plant, preferably a haploid inducer plant as defined in any of statements 5 to 11.

23. A method for producing a plant, preferably a plant which comprises or which is carrier of
5 an enhancer for the haploid induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of statements 5 to 11 comprising overexpressing the myosin heavy chain gene or protein, or increasing the myosin heavy chain mRNA or protein stability or functionality.

24. A method for producing a plant, preferably a plant which comprises or which is carrier of
10 an enhancer for the haploid induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of statements 5 to 11 comprising repressing the aquaporin gene or protein, or decreasing the aquaporin mRNA or protein stability or functionality.

25. A method for producing a plant, preferably a plant which comprises or which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably a haploid
15 inducer plant as defined in any of statements 5 to 11 comprising mutating the PPR gene, preferably resulting in a PPR protein having the sequence as set forth in SEQ ID NO: 17.

26. The method according to any of statements 21 to 25, wherein the plant cell of step a) is a cell comprising a mutated patatin-like phospholipase (PLA1), preferably as defined in any one of the statements 9-11 .

20 27. The method according to any of statements 21 to 26, wherein the method further comprises a step of crossing the plant with a plant comprising a mutated patatin-like phospholipase (PLA1) and selecting a plant from the progeny which comprises A) the mutated patatin-like phospholipase (PLA1) and B) the nucleic acid or vector introduced in step a) of the method of statement 21, the at least one mutation created in the method of statement 22, or
25 the manipulation of any of statements 23 to 25.

28. A plant produced by the method according to any of statements 21 to 26, or a part thereof.

29. A plant, preferably a Zea mays or Sorghum bicolor plant, comprising

i) the nucleotide sequence according to SEQ ID NO: 10, or

ii) a nucleotide sequence encoding the amino acid according to SEQ ID NO: 11, or

30 iii) a nucleotide sequence encoding a myosin heavy chain having at position 325 referenced to SEQ ID NO: 11 or 14 an amino acid different to alanine, preferably valine.

30. A plant, preferably a Zea mays or Sorghum bicolor plant, comprising

i) the nucleotide sequence according to SEQ ID NO: 38 or 39, or

ii) a nucleotide sequence encoding the amino acid according to SEQ ID NO: 40.

35 31. The plant according to statement 30 or 31, which is transgenic or gene edited.

32. The plant according to statement 30 or 31, which is obtained by introgression.

33. The plant according to statement 30 or 31, which is mutagenized.

34. Method for identification/selection of a an enhancer for the haploid induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of statements 5 to 11, comprising steps of

a) optionally isolating (DNA from) at least one cell of a plant, preferably a maize plant or sorghum plant,

b) determining (on (DNA isolated from) the at least one plant cell)

(i) the presence of the nucleic acid molecule according to SEQ ID NOs: 1, 10, 15, or 35

10 (ii) the presence of the nucleic acid molecule of SEQ ID NO: 4 or 30, or

(iii) the increased stability or activity or overexpression of the nucleic acid molecule according to statement 1 or the myosin heavy chain protein, or

(iv) the decreased stability or activity or suppression of the nucleic acid molecule according to statement 2 or the aquaporin protein,

15 c) identify/select the plant from which the at least one cell had been isolated.

35. The method according to statement 34, wherein step b)(iii) or step b)(iv) comprises comparing the expression level and/or activity of said one or more gene with a predetermined threshold.

20 36. A method for the production of a haploid plant or doubled haploid plant, comprising the following steps:

a) crossing a plant according to any one of statements 19, 20, or 28 to 31 with a plant of the same genus,

b) selecting a fertilized haploid seed or embryo, and

c) generating a haploid plant from the seed or embryo selected in step b),

25 d) optionally, doubling the set of chromosomes via a chromosome doubling agent or via spontaneous doubling, optionally wherein the doubling agent is colchicine, pronamide, dithiopyr, or trifluralin.

37. A haploid or doubled haploid plant produced by the method according to statement 36.

30 38. An isolated polynucleic acid specifically hybridising with a nucleotide sequence molecule of SEQ ID NO: 1, 2, 4, 10, 15, 16, 17, 24, 25, 30, 35, 36, 38, or 39; or the complement or the reverse complement thereof.

39. The isolated polynucleic acid according to statement 38 which is a primer or probe.

40. A method of editing or modifying plant genomic DNA or RNA, comprising:

35 a) obtaining or providing a first plant, wherein the first plant is a haploid inducer plant as obtained by a method according to any of statements 21 to 27 or 36, or a plant according to

any of statements 19, 28 to 33, or 37, and wherein said first plant is capable of expressing a DNA or RNA modification enzyme and optionally at least one guide nucleic acid;

b) obtaining a second plant, wherein the second plant comprises the plant genomic DNA or RNA which is to be edited or modified;

5 c) pollinating the second plant with pollen from the first plant; and

d) selecting at least one haploid progeny produced by the pollination of step (c) wherein the haploid progeny comprises the genome of the second plant but not the first plant, and the genome of the haploid progeny has been edited or modified by the DNA or RNA editing or modification enzyme and optional at least one guide nucleic acid delivered by the first plant.

10 41. The method of statement 40, wherein the DNA or RNA editing or modification enzyme is a site-directed nuclease selected from the group consisting of meganucleases (MNs), zinc-finger nucleases (ZFNs), transcription-activator like effector nucleases (TALENs), Cas nucleases, such as Cas9 nuclease, Cpf1 nuclease, Cas13 nuclease, dCas9-FokI, dCpf1-FokI, chimeric Cas9-cytidine deaminase, chimeric Cas9-adenine deaminase, chimeric FENI-FokI, and Mega-TALs, a
15 nickase Cas9 (nCas9), chimeric dCas9 non-FokI nuclease and dCpf1 non-FokI nuclease;

optionally wherein said DNA or RNA modification enzyme is as defined in any of statements 50 to 62.

42. The method of statement 41, wherein the at least one guide nucleic acid is a guide RNA.

43. The method of any of statements 40 to 42, wherein the edited or modified haploid
20 progeny is treated with a chromosome doubling agent, thereby creating an edited doubled haploid progeny.

44. The method of statement 43, wherein the chromosome doubling agent is colchicine, pronamide, dithiopyr, trifluralin, or another known anti-microtubule agent.

45. A gene-edited or modified plant produced by the method of any one of statements 40 to
25 44.

46. A molecular marker which is able to hybridize to nucleic acid molecule as defined in statement 34 b) (i) or (ii), or a molecular marker which is able to detect a polymorphism associated with or closely linked to nucleic acid molecule as defined in statement 34 b) (i) or (ii). In certain embodiment, such polymorphism is a SNP, deletion or insertion, exemplary
30 polymorphisms are given in table 2.

47. Use of the molecular according to statement 46 or the isolated polynucleic acid according to statement 38 or 39 for the identification of the plant of any of statements 19, 20, or 28 to 31, 37, or 45.

48. Use of a nucleic acid molecule according to any of statements 1 or 3-11 for enhancing the haploid induction capability of an inducer plant, preferably a haplotype inducer plant as defined in any of statements 5 to 11.
49. A plant or plant part obtainable by the method according to any of statements 21 to 27, 36, 5 or 40, or a plant according to any of statements 19, 28 to 33, 37, or 45 comprising or expressing a DNA or RNA modification enzyme, or a nucleic acid encoding a DNA or RNA modification enzyme, and optionally a guide nucleic acid.
50. The plant or plant part according to statement 49, wherein said DNA or RNA modification enzyme is a ZFN, TALEN, meganuclease, or a Cas effector protein.
- 10 51. The plant or plant part according to statement 49 or 50, wherein said DNA or RNA modification enzyme is a Cas effector protein.
52. The plant or plant part according to statement 51, wherein said Cas effector protein is a class 2 effector protein.
53. The plant or plant part according to statement 52, wherein said Cas effector protein is a 15 type II, type V (such as type Va or type Vb), or type VI (such as type Via or type Vlb) effector protein.
54. The plant or plant part according to any of statements 50 to 53, wherein said Cas effector protein is Cas9, Cas12 (such as Cas12a or Cas12b), or Cas13 (such as Cas13a or Cas13b).
55. The plant or plant part according to any of statements 49 to 54, wherein said guide nucleic 20 acid is a guide RNA.
56. The plant or plant part according to any of statements 49 to 55, further comprising or expressing a tracr sequence.
57. The plant or plant part according to any of statements 49 to 56, comprising a CRISPR/Cas system.
- 25 58. The plant or plant part according to any of statements 50 to 57, wherein said Cas effector protein is mutated.
59. The plant or plant part according to statement 58, wherein said Cas effector protein is catalytically inactive or is a nickase.
60. The plant or plant part according to statement 58 or 59, wherein said Cas effector protein 30 has increased or decreased target recognition specificity, increased or decreased nuclease activity, increased or decreased stability, and/or altered PAM recognition.
61. The plant or plant part according to any of statements 50 to 60, wherein said Cas effector protein is fused to a (functional) heterologous domain.
62. The plant or plant part according to statement 61, wherein said heterologous domain 35 comprises base editing activity, nucleotide deaminase activity, methylase activity, demethylase

activity, translation activation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, chromatin modifying or remodeling activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, and nucleic acid binding activity.

63. The method according to any of statements 21 to 27, further comprising the step of introducing in the plant or plant cell DNA or RNA modification enzyme as defined in any of statements 49 to 57.

64. The method according to any of statements 21 to 27, wherein said plant or plant cell comprises or expresses a DNA or RNA modification enzyme as defined in any of statements 49 to 57.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Protein/sequence map of a fragment of the myosin (*Zea mays*) protein indicating the mutation V325A caused by a mutation in the coding sequence of the myosin gene. Protein sequences of wildtype (SEQ ID NO: 9) and mutated (SEQ ID NO: 11) myosin fragments are shown.

DETAILED DESCRIPTION OF THE INVENTION

Before the present system and method of the invention are described, it is to be understood that this invention is not limited to particular systems and methods or combinations described, since such systems and methods and combinations may, of course, vary. It is also to be understood that the terminology used herein is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

The terms “comprising”, “comprises” and “comprised of” as used herein are synonymous with “including”, “includes” or “containing”, “contains”, and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps. It will be appreciated that the terms “comprising”, “comprises” and “comprised of” as used herein comprise the terms “consisting of”, “consists” and “consists of”, as well as the terms “consisting essentially of”, “consists essentially” and “consists essentially of”.

The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

The term "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, is meant to encompass variations of +/-20% or less, preferably +/-10% or less, more preferably +/-5% or less, and still more preferably +/-1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

Whereas the terms "one or more" or "at least one", such as one or more or at least one member(s) of a group of members, is clear per se, by means of further exemplification, the term encompasses inter alia a reference to any one of said members, or to any two or more of said members, such as, e.g., any ≥ 3 , ≥ 4 , ≥ 5 , ≥ 6 or ≥ 7 etc. of said members, and up to all said members.

All references cited in the present specification are hereby incorporated by reference in their entirety. In particular, the teachings of all references herein specifically referred to are incorporated by reference.

Unless otherwise defined, all terms used in disclosing the invention, including technical and scientific terms, have the meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. By means of further guidance, term definitions are included to better appreciate the teaching of the present invention.

Standard reference works setting forth the general principles of recombinant DNA technology include Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Current Protocols in Molecular Biology, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1992 (with periodic updates) ("Ausubel et al. 1992"); the series Methods in Enzymology (Academic Press, Inc.); Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990; PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995); Harlow and Lane, eds. (1988) Antibodies, a Laboratory Manual; and Animal Cell Culture (R.I. Freshney, ed. (1987). General principles of microbiology are set forth, for example, in Davis, B. D. et al., Microbiology, 3rd edition, Harper & Row, publishers, Philadelphia, Pa. (1980).

In the following passages, different aspects of the invention are defined in more detail. Each aspect so defined may be combined with any other aspect or aspects unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or advantageous may be combined with any other feature or features indicated as being preferred or advantageous.

Reference throughout this specification to “one embodiment” or “an embodiment” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment” or “in an embodiment” in various places
5 throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of
10 different embodiments are meant to be within the scope of the invention, and form different embodiments, as would be understood by those in the art. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

In the following detailed description of the invention, reference is made to the accompanying drawings that form a part hereof, and in which are shown by way of illustration only of specific
15 embodiments in which the invention may be practiced. It is to be understood that other embodiments may be utilised and structural or logical changes may be made without departing from the scope of the present invention. The following detailed description, therefore, is not to be taken in a limiting sense, and the scope of the present invention is defined by the appended claims.

20 Preferred statements (features) and embodiments of this invention are set herein below. Each statements and embodiments of the invention so defined may be combined with any other statement and/or embodiments unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or advantageous may be combined with any other feature or features or statements indicated as being preferred or advantageous.

25 The terms “haploid inducer” and “haploid inductor” are used as synonyms herein and refer to a plant that is capable of producing fertilized seeds or embryos which have a single (haploid) chromosome set from a crossing with a plant of the same genus, preferably a plant of the same species, which is not a haploid inducer. Haploid induction is frequently a medium to low penetrance trait of the inducer line, so the resulting progeny, depending on the species or
30 situation, may be either diploid (if no genome loss takes place) or haploids (if genome loss does indeed take place). If the parent line that is crossed to the haploid inducer is not diploid, but rather a tetraploid, hexaploid, or other plant of higher ploidy, the term haploid induction is something of a misnomer, because the haploid progeny produced will have a gametic chromosome number, and thus would not really be haploids, but rather diploids (if the parent

is tetraploid) or triploids (if the parent is hexaploid) and so on. Therefore, as used herein, haploids possess half the number of chromosomes of either parent; thus haploids of diploid organisms (e.g., maize) exhibit monoploidy; haploids of tetraploid organisms (e.g., ryegrasses) exhibit diploidy; haploids of hexaploid organisms (e.g., wheat) exhibit triploidy.

5 According to the invention, any haploid inducer can be used. In certain embodiments, a haploid inducer as referred to herein can be a plant selected and/or derived from the lines Stock 6, RWK, RWS, UH400, AX5707RS, NP2222-mai/, or any other haploid inducer line, in particular a line comprising the HIR1 QTL. In certain embodiments, a haploid inducer as referred to herein can be a mutant plant selected from the plants as disclosed in US
10 2017/327832 A1 (or a plant of a different species having the corresponding mutation(s)), incorporated herein by reference in its entirety, in particular the tilling mutants having a D74N (exchange of aspartate at Position 74 for asparagin) or G78R (exchange of glycine at Position 78 for arginine) in the PLA1 gene of maize. It will be understood that corresponding amino acids can be mutated in other plants. In certain embodiments, a haploid inducer as referred to
15 herein can be a mutant plant selected from the plants as described in WO 2018/158301 (or a plant of a different species having the corresponding mutation(s)), incorporated herein by reference in its entirety. In certain embodiments, a haploid inducer as referred to herein can be a plant with mutated PLA1 according to SEQ ID NOs: 38-40 or 44-46, or as defined herein elsewhere. In certain embodiments, a haploid inducer as referred to herein can be a mutant
20 plant selected from the plants as described in Kalinowska et al. (2018) "State-of-the-art and novel developments of in vivo haploid technologies"; Theor. Appl. Genetics; 132(3):593-605; incorporated herein by reference in its entirety.

As used herein, PLA1 refers to patatin-like phospholipase, also known as NOT LIKE DAD1 (NLDI) and MATRILINEAL (MATL), as for instance defined in Kelliher et al. (2017). MATRILINEAL, a
25 sperm-specific phospholipase, triggers maize haploid induction. Nature, 542(7639), 105. The skilled person will understand that PLA1 may originate from maize but equally the PLA1 orthologues may be used herein, such as Sorghum PLA1.

In certain embodiments, wild type PLA1 or PLA1 originating from a line which is not a haploid inducer (or a haploid inducer which is not caused by PLA1 mutations) is PLA1 from PH207 or
30 B73, i.e. PLA1 having a genomic, coding, or protein sequence (substantially) identical to the PLA1 sequence from PH207 or B73, or an orthologue thereof. In certain embodiments, wild type PLA1 or PLA1 originating from a line which is not a haploid inducer (or a haploid inducer which is not caused by PLA1 mutations) is PLA1 having a genomic, coding, or protein sequence which is at least 80% identical (preferably over the entire length) to the PLA1 sequence from

PH207 or B73 (or an orthologue thereof), preferably at least 85%, such as at least 86%, 87%, 88%, 89%, more preferably at least 90%, such as at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. In certain embodiments, wild type PLA1 or PLA1 originating from a line which is not a haploid inducer (or a haploid inducer which is not caused by PLA1 mutations) is PLA1 having a genomic sequence which is at least 80% identical (preferably over the entire length) to the PLA1 sequence as set forth in SEQ ID NO: 41 or 47 (or an orthologue thereof), preferably at least 85%, such as at least 86%, 87%, 88%, 89%, more preferably at least 90%, such as at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. In certain embodiments, wild type PLA1 or PLA1 originating from a line which is not a haploid inducer (or a haploid inducer which is not caused by PLA1 mutations) is PLA1 having a coding sequence which is at least 80% identical (preferably over the entire length) to the PLA1 sequence as set forth in SEQ ID NO: 42 or 48 (or an orthologue thereof), preferably at least 85%, such as at least 86%, 87%, 88%, 89%, more preferably at least 90%, such as at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. In certain embodiments, wild type PLA1 or PLA1 originating from a line which is not a haploid inducer (or a haploid inducer which is not caused by PLA1 mutations) is PLA1 having a protein sequence which is at least 80% identical (preferably over the entire length) to the PLA1 sequence as set forth in SEQ ID NO: 43, 49, or 50 (or an orthologue thereof), preferably at least 85%, such as at least 86%, 87%, 88%, 89%, more preferably at least 90%, such as at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. It will be understood that such "wild type" PLA1 variants (such as orthologues) are not haploid inducing variants. In particular, "wild type" PLA1 preferably excludes any of the PLA1 mutations described below. Preferably, such variants have comparable expression levels, stability, and/or activity as PLA1 from PH207 or B73, i.e. expression, stability, and/or activity does not differ by more than 100%, preferably does not differ by more than 50%.

In certain embodiments, mutated PLA1 as referred to herein causes the property of a haploid inducer plant or increases the induction capability of a haploid inducer plant. In certain embodiments PLA1 having the wild type sequence as set forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 is mutated.

In certain embodiments, mutated PLA1 as referred to herein has an altered (mRNA or protein) expression level compared to (unmutated) wild type PLA1 (such as PH207 or B73 PLA1). In certain embodiments, mutated PLA1 as referred to herein has an altered activity compared to (unmutated) wild type PLA1 (such as PH207 or B73 PLA1). In certain embodiments, mutated PLA1 as referred to herein has an altered stability compared to (unmutated) wild type PLA1 (such as PH207 or B73 PLA1). In certain embodiments, mutated PLA1 as referred to herein has a

reduced (mRNA or protein) expression level compared to (unmutated) wild type PLA1 (such as PH207 of B73 PLA1). In certain embodiments, mutated PLA1 as referred to herein has a reduced activity compared to (unmutated) wild type PLA1 (such as PH207 of B73 PLA1). In certain embodiments, mutated PLA1 as referred to herein has a reduced stability compared to
5 (unmutated) wild type PLA1 (such as PH207 of B73 PLA1). Altered or decreased expression, activity, and stability are to be interpreted as defined herein elsewhere. In certain embodiments PLA1 having the wild type sequence as set forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 is mutated to have an altered (decreased) expression, activity, and/or stability.

In certain embodiments, mutated PLA1 as referred to herein has an altered or reduced
10 expression level, activity, and/or stability as described above, and causes the property of a haploid inducer plant or increases the induction capability of a haploid inducer plant. In certain embodiments PLA1 having the wild type sequence as set forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 is mutated to have an altered (decreased) expression, activity, and/or stability, and causes the property of a haploid inducer plant or increases the induction capability of a haploid
15 inducer plant.

In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein is a truncated PLA1. In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises one or more deletion such as an internal deletion of one or more nucleotides or amino acids,
20 or a 3' and/or 5' or N-terminal and/or C-terminal deletion of one or more nucleotides or amino acids. In certain embodiments, mutated PLA1 (genomic sequence or coding sequence) as referred to herein comprises a premature stop codon. In certain embodiments, mutated PLA1 (genomic sequence or coding sequence) as referred to herein comprises a frame shift. In certain embodiments, mutated PLA1 (genomic sequence or coding sequence) as referred to
25 herein comprises a frame shift resulting in a premature stop codon. In certain embodiments, mutated PLA1 (genomic sequence or coding sequence) as referred to herein comprises a (point) mutation resulting in a premature stop codon. In certain embodiments, mutated PLA1 (genomic sequence or coding sequence) as referred to herein comprises a deletion resulting in a premature stop codon. In certain embodiments PLA1 having the wild type sequence as set
30 forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 is truncated, has a deletion, frame shift, point mutation, premature stop codon, as defined above.

In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation (such as deletion, truncation, or frame shift) wherein the amino acid sequence at the C-terminal end is shortened by 5 or more amino

acids, preferably 10 or more amino acids, such as 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 50 or more, 60 or more, 70 or more, 80 or more, 90 or more, or 100 or more amino acids. In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation (such as deletion, truncation, or frame shift) wherein the amino acid sequence at the C-terminal end is shortened by 5 to 100 amino acids, preferably 10 to 100 amino acids, such as 15 to 100 amino acids, 10 to 90 amino acids, 15 to 90 amino acids, 10 to 80 amino acids, 15 to 80 amino acids, 10 to 70 amino acids, 15 to 70 amino acids, 10 to 60 amino acids, 15 to 60 amino acids. In certain embodiments PLA1 having the wild type sequence as set forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 is C-terminally shortened as defined above.

In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a destabilizing mutation, i.e. a mutation affecting protein stability. In certain embodiments PLA1 having the wild type sequence as set forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 comprises a destabilizing mutation.

In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a point mutation. In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation in which one or more amino acids are exchanged. In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation which causes an amino acid exchange between amino acid positions 57 and 289, referenced to SEQ ID NO: 43, 49, or 50, i.e. between positions 57 and 289 of PLA1 having a protein sequence as set forth in SEQ ID NO: 43, 49, or 50 or the corresponding positions in a PLA1 variant or orthologue as described herein elsewhere. In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation which causes an amino acid exchange at position 58, referenced to SEQ ID NO: 43, 49, or 50, preferably wherein the amino acid is replaced by glutamine (58Q), preferably wherein arginine is replaced by glutamine (R58Q). In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation which causes an amino acid exchange at position 74, referenced to SEQ ID NO: 43, 49, or 50, preferably wherein the amino acid is replaced by asparagine (74N), preferably wherein aspartate is replaced by asparagine (D74N). In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation which causes an amino acid exchange at position 78, referenced to SEQ ID NO: 43, 49, or 50, preferably wherein the amino acid is

replaced by arginine (78R), preferably wherein glycine is replaced by arginine (G78R). In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation which causes an amino acid exchange at position 160, referenced to SEQ ID NO: 43, 49, or 50, preferably wherein the amino acid is replaced by
5 isoleucine (160I), preferably wherein valine is replaced by isoleucine (V160I). In certain embodiments, mutated PLA1 (genomic sequence, coding sequence, or protein sequence) as referred to herein comprises a mutation which causes an amino acid exchange at position 288, referenced to SEQ ID NO: 43, 49, or 50, preferably wherein the amino acid is replaced by leucine (288L), preferably wherein serine is replaced by leucine (S288L). In certain
10 embodiments PLA1 having the wild type sequence as set forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 comprises a point mutation/amino acid exchange as described above.

In certain embodiments, mutated PLA1 as referred to herein has a protein sequence as set forth in SEQ ID NO: 40 or 46, or a sequence at least 80%, preferably at least 85%, such as at least 90% or 95% identical (preferably over the entire length) to SEQ ID NO: 40 or 46 (with the
15 proviso that such variant is a haploid inducer), or an orthologue comprising the corresponding mutation.

In certain embodiments, mutated PLA1 (genomic sequence) as referred to herein comprises a mutation in one or more of its regulatory sequences (in particular the promoter). In certain embodiments, mutated PLA1 (genomic sequence) as referred to herein, comprises a mutation
20 in one or more of its regulatory sequences (in particular the promoter) which affect transcription. In certain embodiments, mutated PLA1 (genomic sequence) as referred to herein, comprises a mutation in one or more of its regulatory sequences (in particular the promoter) which reduce or (substantially) eliminate transcription. In certain embodiments PLA1 having the wild type sequence as set forth in SEQ ID NOs: 41-43 or SEQ ID NOs: 47-50 comprises
25 a mutation in one or more of its regulatory sequences as defined above.

In certain embodiments, mutated PLA1 as referred to herein is knocked out or knocked down.

As used herein, the "haploid induction rate" or "induction rate" of a haploid inducer refers to the number or percentage of fertilized seeds or embryos that have a haploid chromosome set and which have arisen from a crossing of the haploid inducer with a plant of the same genus
30 (preferably, a plant of the same species) which is not a haploid inducer.

"Enhancing or increasing the haploid induction capability of a haploid inducer" or "mediating the property of an enhancer of the haploid induction capability of an a haploid inducer" means that via the use of a nucleic acid as described herein, in particular an inducing-promoting

nucleic acid as described herein or a nucleic acid encoding a double-stranded RNA as described herein, the haploid induction rate of a haploid inducer, in particular a haploid inducer comprising a mutated patatin-like phospholipase (PLA1), can be increased with at least 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8% or 0.9%, preferably with at least 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5% or 5%, more preferably with at least 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30% or 50%. The number of fertilized seeds or embryos which have a haploid chromosome set and which have arisen from a crossing of the haploid inducer with a plant of the same genus (preferably, a plant of the same species) which is not a haploid inducer may thus be higher by at least 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8% or 0.9%, preferably at least 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5% or 5%, more preferably, at least 6%, 7%, 8%, 9%, 10%, 15%, 20%, 30% or 50%, than the number of haploid fertilized seeds or embryos which is achieved without the use of the nucleic acid as described herein.

The term "plant" includes whole plants, including descendants or progeny thereof. The term "plant part" includes any part or derivative of the plant, including particular plant tissues or structures, plant cells, plant protoplast, plant cell or tissue culture from which plants can be regenerated, plant calli, plant clumps and plant cells that are intact in plants or parts of plants. Non-limiting examples of plant parts include fertilized or unfertilized seeds, embryos, pollen, ovules, leaves, roots, root tips, anthers, flowers, fruits, stems, shoots, and the like. Plant parts may include processed plant parts or derivatives, including flower, oils, extracts etc.

In certain embodiments, the plant part or derivative comprises, consists of, or consists essentially of one or more, preferably all of stalks, leaves, and cobs. In certain embodiments, the plant part or derivative is leaves. In certain embodiments, the plant part or derivative is stalks. In certain embodiments, the plant part or derivative is cobs. In certain embodiments, the plant part or derivative comprises, consists of, or consists essentially of one or more, preferably all of stalks and leaves. In certain embodiments, the plant part or derivative comprises, consists of, or consists essentially of one or more, preferably all of stalks, and cobs. In certain embodiments, the plant part or derivative comprises, consists of, or consists essentially of one or more, preferably all of leaves and cobs. In certain embodiments, the plant part or derivative is not (functional) propagation material, such as germplasm, a seed, or plant embryo or other material from which a plant can be regenerated. In certain embodiments, the plant part or derivative does not comprise (functional) male and female reproductive organs. In certain embodiments, the plant part or derivative is or comprises propagation material, but propagation material which does not or cannot be used (anymore) to produce or generate new plants, such as propagation material which have been chemically, mechanically or

otherwise rendered non-functional, for instance by heat treatment, acid treatment, compaction, crushing, chopping, etc. in certain preferred embodiments, the plant part is corn cobs or stover.

[illegible]

20 As used herein, “maize” refers to a plant of the species *Zea mays*, preferably *Zea mays* ssp
mays.

“B73” is a maize breeding line that is used as a model genotype in maize genetics and was used to create the first maize reference sequence. In certain embodiments, the B73 reference genome as defined herein is the B73 AGPv2 genome and refers to the assembly B73 RefGen_v2 (also known as AGPv2, B73 RefGen_v2) as provided on the Maize Genetics and Genomics Database (https://www.maizegdb.org/genome/genome_assembly/B73%20RefGen_v2).

“RWS” is a haploid inducer line based on Stock 6, and has the haploid inducer chromosome 1
30 QTL mutation (HI1).

“W22” is a non-haploid inducer line but carrier of the enhancer allele on HIR8 locus.

As used herein, “sorghum” refers to a plant of the genus *Sorghum*, and includes without limitation *Sorghum bicolor*, *Sorghum sudanense*, *Sorghum bicolor* × *Sorghum sudanense*,

Sorghum × *alrum* (*Sorghum bicolor* × *Sorghum halepense*), *Sorghum arundinaceum*, *Sorghum* × *drummondii*, *Sorghum halepense* and/or *Sorghum propinquum*.

A "functional fragment" of a nucleotide sequence as used herein means a segment of a nucleotide sequence which has the functionality identical or comparable to the complete nucleotide sequence from which the functional fragment originates. As such, the functional fragment may possess a nucleotide sequence which is identical or homologous to the complete nucleotide sequence over a length of at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 94%, 96%, 97%, 98% or 99%. Furthermore, a "functional fragment" of a nucleotide sequence may also mean a segment of a nucleotide sequence which alters the functionality of the total nucleotide sequence, e.g., in the course of post-transcriptional gene silencing. As such, the functional fragment of a nucleotide sequence may include at least 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25, preferably at least 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120 or 140, more preferably at least 160, 180, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or 1000 successive nucleotides of the complete nucleotide sequence.

A "functional part" of a protein means a segment of a protein, or a section of the amino acid sequence, that encodes for the protein, wherein the segment may exert functionality identical or comparable to the entire protein in a plant cell. A functional part of a protein has, over a length of at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 94%, 96%, 97%, 98%, or 99%, an identical or—under conservative and semi-conservative amino acid exchanges—similar amino acid sequence to that of the protein from which the functional part originates.

As used herein, the term "sequence identity" refers to the degree of identity between any given nucleic acid sequence and a target nucleic acid sequence. Percent sequence identity is calculated by determining the number of matched positions in aligned nucleic acid sequences, dividing the number of matched positions by the total number of aligned nucleotides, and multiplying by 100. A matched position refers to a position in which identical nucleotides occur at the same position in aligned nucleic acid sequences. Percent sequence identity can also be determined for any amino acid sequence. To determine percent sequence identity, a target nucleic acid or amino acid sequence is compared to the identified nucleic acid or amino acid sequence using the BLAST 2 Sequences (BL2seq) program from the stand-alone version of BLASTZ containing BLASTN and BLASTP. This stand-alone version of BLASTZ can be obtained from Fish & Richardson's web site (World Wide Web at fr.com/blast) or the U.S. government's National Center for Biotechnology Information web site (World Wide Web at ncbi.nlm.nih.gov). Instructions explaining how to use the BL2seq program can be found in the readme file accompanying BLASTZ. BL2seq performs a comparison between two sequences using either the

BLASTN or BLASTP algorithm.

BLASTN is used to compare nucleic acid sequences, while BLASTP is used to compare amino acid sequences. To compare two nucleic acid sequences, the options are set as follows: -i is set to a file containing the first nucleic acid sequence to be compared (e.g. , C:\seq1.txt); -j is set to a file containing the second nucleic acid sequence to be compared (e.g. , C:\seq2.txt); -p is set to blastn; -o is set to any desired file name (e.g. , C:\output.txt); -q is set to -1 ; -r is set to 2; and all other options are left at their default setting. The following command will generate an output file containing a comparison between two sequences: C:\B12seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt -q -1 -r 2. If the target sequence shares homology with any portion of the identified sequence, then the designated output file will present those regions of homology as aligned sequences. If the target sequence does not share homology with any portion of the identified sequence, then the designated output file will not present aligned sequences. Once aligned, a length is determined by counting the number of consecutive nucleotides from the target sequence presented in alignment with the sequence from the identified sequence starting with any matched position and ending with any other matched position. A matched position is any position where an identical nucleotide is presented in both the target and identified sequences. Gaps presented in the target sequence are not counted since gaps are not nucleotides. Likewise, gaps presented in the identified sequence are not counted since target sequence nucleotides are counted, not nucleotides from the identified sequence. The percent identity over a particular length is determined by counting the number of matched positions over that length and dividing that number by the length followed by multiplying the resulting value by 100. For example, if (i) a 500-base nucleic acid target sequence is compared to a subject nucleic acid sequence, (ii) the B12seq program presents 200 bases from the target sequence aligned with a region of the subject sequence where the first and last bases of that 200-base region are matches, and (iii) the number of matches over those 200 aligned bases is 180, then the 500-base nucleic acid target sequence contains a length of 200 and a sequence identity over that length of 90% (i.e. , $180 / 200 \times 100 = 90$). It will be appreciated that different regions within a single nucleic acid target sequence that aligns with an identified sequence can each have their own percent identity. It is noted that the percent identity value is rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 are rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 are rounded up to 78.2. It also is noted that the length value will always be an integer.

An "isolated nucleic acid sequence" or "isolated DNA" refers to a nucleic acid sequence which is no longer in the natural environment from which it was isolated, e.g. the nucleic acid

sequence in a bacterial host cell or in the plant nuclear or plastid genome. When referring to a "sequence" herein, it is understood that the molecule having such a sequence is referred to, e.g. the nucleic acid molecule. A "host cell" or a "recombinant host cell" or "transformed cell" are terms referring to a new individual cell (or organism) arising as a result of at least one nucleic acid molecule, having been introduced into said cell. The host cell is preferably a plant cell or a bacterial cell. The host cell may contain the nucleic acid as an extra-chromosomally (episomal) replicating molecule, or comprises the nucleic acid integrated in the nuclear or plastid genome of the host cell, or as introduced chromosome, e.g. minichromosome.

In certain embodiments, the nucleic acid molecule as described herein comprises less than 50000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises less than 40000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises less than 30000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises less than 25000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises less than 20000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises less than 15000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises less than 10000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises less than 5000 nucleotides. In certain embodiments, the nucleotide molecule as described herein comprises at least 100 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 50000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 40000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 30000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 25000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 20000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 15000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 10000 nucleotides. In certain embodiments, the nucleic acid molecule as described herein comprises at least 100 nucleotides and less than 5000 nucleotides.

When reference is made to a nucleic acid sequence (e.g. DNA or genomic DNA) having "substantial sequence identity to" a reference sequence or having a sequence identity of at

least 80%>, e.g. at least 85%, 90%, 95%, 98%> or 99%> nucleic acid sequence identity to a reference sequence, in one embodiment said nucleotide sequence is considered substantially identical to the given nucleotide sequence and can be identified using stringent hybridisation conditions. In another embodiment, the nucleic acid sequence comprises one or more
5 mutations compared to the given nucleotide sequence but still can be identified using stringent hybridisation conditions.

The terms “hybridizing” and “hybridization” refer to a process in which a single-stranded nucleic acid molecule is added to a nucleic acid strand that is complementary to the greatest possible extent, i.e., enters into base pairing. Standard methods for hybridization are described
10 in Sambrook et al. 2001, for example. Preferably at least 60%, more preferably at least 65%, 70%, 75%, 80%, or 85%, even more preferably at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% of the bases of the nucleic acid molecule enter into a base pairing with the nucleic acid strand that is complementary to the greatest possible extent. The possibility of such an addition depends upon the stringency of the hybridization conditions. The term
15 “stringency” relates to the hybridization conditions. High stringency is present when a base pairing is made more difficult; low stringency is present if a base pairing is made easier. For example, the stringency of the hybridization conditions depends upon the salt concentration, or ion strength, and the temperature. In general, the stringency may be increased by increasing the temperature and/or decreasing the salt content. With “stringent hybridization
20 conditions” is meant herein those conditions under which a hybridization predominantly occurs only between homologous nucleic acid molecules. The term “hybridization conditions” thereby relates not only to the conditions prevailing in the actual addition of the nucleic acids, but also to the conditions prevailing in the following washing steps. Stringent hybridization conditions are, for example, conditions under which, predominantly, only those nucleic acid
25 molecules are hybridized that have at least 70%, preferably at least 75%, at least 80%, at least 85%, at least 90% or at least 95% sequence identity. Stringent hybridization conditions are, for example, hybridization in 4×SSC at 65° C., and subsequent repeated washing in 0.1×SSC at 65° C. for approximately 1 hour in total. The term “stringent hybridization conditions” that is used here may also mean hybridization at 68° C. in 0.25 M sodium phosphate, pH 7.2, 7% SDS, 1
30 mM EDTA and 1% BSA for 16 hours, and subsequent washing twice with 2×SSC and 0.1% SDS at 68° C. A hybridization preferably occurs under stringent conditions.

As used herein, “increase in the expression rate” or “increased expression rate” or “activation of the expression” or “overexpression” or a comparable phrase in certain embodiments means an increase in the expression rate of a nucleotide sequence by more than 10%, 15%, 20%, 25%

or 30%, preferably by more than 40%, 50%, 60%, 70%, 80%, 90% or 100%, more preferably by more than 150%, 200%, 250%, 300%, 500%, or 1000%, in comparison to the specified reference, such as a plant not comprising the genetic or otherwise modifications according to the invention as described herein elsewhere (e.g. overexpression of the myosin heavy chain),
5 or a reference plant, such as BL73. The increase in the expression rate preferably leads to a change of the phenotype of a plant in which the expression rate is increased. In the context of the present invention, an altered phenotype may be the enhanced induction capability of a haploid inductor caused by a mutated palatin-like phospholipase (PLA1).

10 “Increase in the transcription rate” or “increased transcription rate” or a comparable phrase in certain embodiments means an increase in the transcription rate of a nucleotide sequence by more than 10%, 15%, 20%, 25% or 30%, preferably by more than 40%, 50%, 60%, 70%, 80%, 90% or 100%, more preferably by more than 150%, 200%, 250%, 300%, 500%, or 1000%, in comparison to the specified reference, such as a plant not comprising the genetic or otherwise modifications according to the invention as described herein elsewhere (e.g. overexpression of
15 the myosin heavy chain), or a reference plant, such as BL73. The increase in the transcription rate preferably leads to a change of the phenotype of a plant in which the transcription rate is increased. In the context of the present invention, an altered phenotype may be the enhanced induction capability of a haploid inductor caused by a mutated palatin-like phospholipase (PLA1).

20 As used herein, “reducing the expression rate” or “reduction in the expression rate” or “suppression of the expression,” “reduced expression rate,” or “repression” or a comparable phrase in certain embodiments means a reduction in the expression rate of a nucleotide sequence by more than 10%, 15%, 20%, 25% or 30%, preferably by more than 40%, 45%, 50%, 55%, 60% or 65%, more preferably by more than 70%, 75%, 80%, 85%, 90%, 92%, 94%, 96% or
25 98% in comparison to the specified reference, such as a plant not comprising the genetic or otherwise modifications according to the invention as described herein elsewhere (e.g. repression of the aquaporin protein or gene), or a reference plant, such as BL73. However, it may also mean that the expression rate of a nucleotide sequence is reduced by 100%. The reduction in the expression rate preferably leads to a change of the phenotype of a plant in
30 which the expression rate is reduced. In the context of the present invention, an altered phenotype may be the enhanced induction capability of a haploid inductor caused by a mutated palatin-like phospholipase (PLA1).

“Reduction in the transcription rate” or “reduced transcription rate” or a comparable phrase in certain embodiments means a reduction in the transcription rate of a nucleotide sequence by

more than 10%, 15%, 20%, 25% or 30%, preferably by more than 40%, 45%, 50%, 55%, 60% or 65%, more preferably by more than 70%, 75%, 80%, 85%, 90%, 92%, 94%, 96% or 98% in comparison to the specified reference, such as a plant not comprising the genetic or otherwise modifications according to the invention as described herein elsewhere (e.g. repression of the aquaporin protein or gene), or a reference plant, such as BL73. However, it may also mean that the transcription rate of a nucleotide sequence is reduced by 100%. The reduction in the transcription rate preferably leads to a change of the phenotype of a plant in which the transcription rate is reduced. In the context of the present invention, an altered phenotype may be the enhanced induction capability of a haploid inducer caused by a mutated palatin-like phospholipase (PLA1).

As used herein, "increased (protein) activity" refers to increased activity of about at least 10%, preferably at least 30%, more preferably at least 50%, such as at least 20%, 40%, 60%, 80% or more, such as at least 85%, at least 90%, at least 95%, or more. As used herein, reduced (protein) activity refers to decreased activity of about at least 10%, preferably at least 30%, more preferably at least 50%, such as at least 20%, 40%, 60%, 80% or more, such as at least 85%, at least 90%, at least 95%, or more. Activity is (substantially) absent or eliminated if activity is reduced at least 80%, preferably at least 90%, more preferably at least 95%. In certain embodiments, activity is (substantially) absent, if no activity, in particular the wild type or native protein activity, can be detected. (Protein) activity levels can be determined by any means known in the art, depending on the type of protein, such as by standard detection methods, including for instance enzymatic assays (for enzymes), transcription assays (for transcription factors), assays to analyse a phenotypic output, etc. Activity may be compared to a reference as defined above.

As used herein, "increased stability" may refer to increased protein stability or increased RNA, such as mRNA stability. Stability of proteins or RNA can be determined by means known in the art, such as determination of protein/RNA half-life. Increased protein or RNA stability in certain embodiments means an increase of stability of about at least 10%, preferably at least 30%, more preferably at least 50%, such as at least 20%, 40%, 60%, 80% or more, such as at least 85%, at least 90%, at least 95%, or more, such as 100%, 200%, 300%, 400%, 500% or more. Increased protein or RNA stability in certain embodiments means an increase in half-life of about at least 10%, preferably at least 30%, more preferably at least 50%, such as at least 20%, 40%, 60%, 80% or more, such as at least 85%, at least 90%, at least 95%, or more, such as a two-fold, three-fold, four-fold, 5-fold or more increase in half-life. Stability may be compared to a reference as defined above.

“RNA interference” or “RNAi” is a biological process in which RNA molecules inhibit gene expression or translation, by neutralizing targeted mRNA molecules. Two types of small ribonucleic acid (RNA) molecules – microRNA (miRNA) and small interfering RNA (siRNA) – are central to RNA interference. RNAs are the direct products of genes, and these small RNAs can bind to other specific messenger RNA (mRNA) molecules and either increase or decrease their activity, for example by preventing an mRNA from being translated into a protein. The RNAi pathway is found in many eukaryotes, including animals, and is initiated by the enzyme Dicer, which cleaves long double-stranded RNA (dsRNA) molecules into short double-stranded fragments of about 21 nucleotide siRNAs (small interfering RNAs). Each siRNA is unwound into two single-stranded RNAs (ssRNAs), the passenger strand and the guide strand. The passenger strand is degraded and the guide strand is incorporated into the RNA-induced silencing complex (RISC). Mature miRNAs are structurally similar to siRNAs produced from exogenous dsRNA, but before reaching maturity, miRNAs must first undergo extensive post-transcriptional modification. A miRNA is expressed from a much longer RNA-coding gene as a primary transcript known as a pri-miRNA which is processed, in the cell nucleus, to a 70-nucleotide stem-loop structure called a pre-miRNA by the microprocessor complex. This complex consists of an RNase III enzyme called Drosha and a dsRNA-binding protein DGCR8. The dsRNA portion of this pre-miRNA is bound and cleaved by Dicer to produce the mature miRNA molecule that can be integrated into the RISC complex; thus, miRNA and siRNA share the same downstream cellular machinery. A short hairpin RNA or small hairpin RNA (shRNA/Hairpin Vector) is an artificial RNA molecule with a tight hairpin turn that can be used to silence target gene expression via RNA interference. The most well-studied outcome is post-transcriptional gene silencing, which occurs when the guide strand pairs with a complementary sequence in a messenger RNA molecule and induces cleavage by Argonaute 2 (Ago2), the catalytic component of the RISC. It will be understood that the RNAi molecules can be applied as such to/in the plant, or can be encoded by appropriate vectors, from which the RNAi molecule is expressed. Delivery and expression systems of RNAi molecules, such as siRNAs, shRNAs or miRNAs are well known in the art.

The term “regulatory sequence” as used herein relates to a nucleotide sequence which affects the specificity and/or the expression strength, e.g., in that the regulatory sequence mediates a defined tissue specificity. Such a regulatory sequence may be located upstream of the transcription initiation point of a minimal promoter, but also downstream of it, e.g., as in a transcribed, but untranslated, leader sequence or within an intron.

As used herein, a “vector” has its ordinary meaning in the art, and may for instance be a

plasmid, a cosmid, a phage or an expression vector, a transformation vector, shuttle vector, or cloning vector; it may be double- or single-stranded, linear or circular; or it may transform a prokaryotic or eukaryotic host, either via integration into its genome or extrachromosomally. The nucleic acid according to the invention is preferably operatively linked in a vector with one
5 or more regulatory sequences which allow the transcription, and, optionally, the expression, in a prokaryotic or eukaryotic host cell. A regulatory sequence—preferably, DNA—may be homologous or heterologous to the nucleic acid according to the invention. For example, the nucleic acid is under the control of a suitable promoter or terminator. Suitable promoters may be promoters which are constitutively induced (example: 35S promoter from the “Cauliflower
10 mosaic virus” (Odell et al., 1985); those promoters which are tissue-specific are especially suitable (example: Pollen-specific promoters, Chen et al. (2010), Zhao et al. (2006), or Twell et al. (1991)), or are development-specific (example: blossom-specific promoters). Suitable promoters may also be synthetic or chimeric promoters which do not occur in nature, are composed of multiple elements, and contain a minimal promoter, as well as—upstream of the
15 minimum promoter—at least one cis-regulatory element which serves as a binding location for special transcription factors. Chimeric promoters may be designed according to the desired specifics and are induced or repressed via different factors. Examples of such promoters are found in Gurr & Rushton (2005) or Venter (2007). For example, a suitable terminator is the nos-terminator (Depicker et al., 1982). The vector may be introduced via conjugation,
20 mobilization, biolistic transformation, agrobacteria-mediated transformation, transfection, transduction, vacuum infiltration, or electroporation.

As used herein, the term “operatively linked” or “operably linked” means connected in a common nucleic acid molecule in such a manner that the connected elements are positioned and oriented relative to one another such that a transcription of the nucleic acid molecule may
25 occur. A DNA which is operatively linked with a promoter is under the transcriptional control of this promoter.

As used herein the term "transformation" refers to the transfer of isolated and cloned genes into the DNA, usually the chromosomal DNA or genome, of another organism.

"Transgene" refers to a genetic locus comprising a DNA sequence, such as a recombinant gene,
30 which has been introduced into the genome of a plant cell or plant by transformation, such as Agrobacterium mediated transformation. A plant cell or plant comprising a transgene stably integrated into its genome is referred to as “transgenic plant cell” or “transgenic plant”.

The term "locus" (loci plural) means a specific place or places or a site on a chromosome where

for example a QTL, a gene or genetic marker is found. As used herein, the term "quantitative trait locus" or "QTL" has its ordinary meaning known in the art. By means of further guidance, and without limitation, a QTL may refer to a region of DNA that is associated with the differential expression of a quantitative phenotypic trait in at least one genetic background, e.g., in at least one breeding population. The region of the QTL encompasses or is closely linked to the gene or genes that affect the trait in question. An "allele of a QTL" can comprise multiple genes or other genetic factors within a contiguous genomic region or linkage group, such as a haplotype. An allele of a QTL can denote a haplotype within a specified window wherein said window is a contiguous genomic region that can be defined, and tracked, with a set of one or more polymorphic markers. A haplotype can be defined by the unique fingerprint of alleles at each marker within the specified window. A QTL may encode for one or more alleles that affect the expressivity of a continuously distributed (quantitative) phenotype. In certain embodiments, the QTL, marker, allele, or gene as described herein may be homozygous. In certain embodiments, the QTL, marker, allele, or gene as described herein may be heterozygous.

As used herein, the term "homozygote" refers to an individual cell or plant having the same alleles at one or more or all loci. When the term is used with reference to a specific locus or gene, it means at least that locus or gene has the same alleles. As used herein, the term "homozygous" means a genetic condition existing when identical alleles reside at corresponding loci on homologous chromosomes. As used herein, the term "heterozygote" refers to an individual cell or plant having different alleles at one or more or all loci. When the term is used with reference to a specific locus or gene, it means at least that locus or gene has different alleles. As used herein, the term "heterozygous" means a genetic condition existing when different alleles reside at corresponding loci on homologous chromosomes. In certain embodiments, the QTL, marker(s), or gene(s) as described herein is/are homozygous. In certain embodiments, the QTL, marker(s), or gene(s) as described herein are heterozygous. In certain embodiments, the QTL, marker(s), or gene(s) allele(s) as described herein is/are homozygous. In certain embodiments, the QTL, marker(s), or gene(s) allele(s) as described herein are heterozygous.

As used herein, the term "allele" or "alleles" refers to one or more alternative forms, i.e. different nucleotide sequences, of a locus.

As used herein, the term "mutant alleles" or "mutation" of alleles include alleles having one or more mutations, such as insertions, deletions, stop codons, base changes (e.g. , transitions or transversions), or alterations in splice junctions, which may or may not give rise to altered

gene products. Modifications in alleles may arise in coding or non-coding regions (e.g. promoter regions, exons, introns or splice junctions).

A "marker" is a (means of finding a position on a) genetic or physical map, or else linkages among markers and trait loci (loci affecting traits). The position that the marker detects may be
5 known via detection of polymorphic alleles and their genetic mapping, or else by hybridization, sequence match or amplification of a sequence that has been physically mapped. A marker can be a DNA marker (detects DNA polymorphisms), a protein (detects variation at an encoded polypeptide), or a simply inherited phenotype (such as the 'waxy' phenotype). A DNA marker can be developed from genomic nucleotide sequence or from expressed nucleotide sequences
10 (e.g., from a spliced RNA or a cDNA). Depending on the DNA marker technology, the marker may consist of complementary primers flanking the locus and/or complementary probes that hybridize to polymorphic alleles at the locus. The term marker locus is the locus (gene, sequence or nucleotide) that the marker detects. "Marker" or "molecular marker" or "marker locus" may also be used to denote a nucleic acid or amino acid sequence that is sufficiently
15 unique to characterize a specific locus on the genome. Any detectable polymorphic trait can be used as a marker so long as it is inherited differentially and exhibits linkage disequilibrium with a phenotypic trait of interest.

Markers that detect genetic polymorphisms between members of a population are well-established in the art. Markers can be defined by the type of polymorphism that they detect
20 and also the marker technology used to detect the polymorphism. Marker types include but are not limited to, e.g., detection of restriction fragment length polymorphisms (RFLP), detection of isozyme markers, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), detection of simple sequence repeats (SSRs), detection of amplified variable sequences of the plant genome, detection of self-sustained
25 sequence replication, or detection of single nucleotide polymorphisms (SNPs). SNPs can be detected e.g. via DNA sequencing, PCR-based sequence specific amplification methods, detection of polynucleotide polymorphisms by allele specific hybridization (ASH), dynamic allele-specific hybridization (DASH), molecular beacons, microarray hybridization, oligonucleotide ligase assays, Flap endonucleases, 5' endonucleases, primer extension, single
30 strand conformation polymorphism (SSCP) or temperature gradient gel electrophoresis (TGGE). DNA sequencing, such as the pyrosequencing technology has the advantage of being able to detect a series of linked SNP alleles that constitute a haplotype. Haplotypes tend to be more informative (detect a higher level of polymorphism) than SNPs.

A "marker allele", alternatively an "allele of a marker locus", can refer to one of a plurality of

polymorphic nucleotide sequences found at a marker locus in a population. With regard to a SNP marker, allele refers to the specific nucleotide base present at that SNP locus in that individual plant.

"Fine-mapping" refers to methods by which the position of a QTL can be determined more accurately (narrowed down) and by which the size of the introgression fragment comprising the QTL is reduced. For example Near Isogenic Lines for the QTL (QTL-NILs) can be made, which contain different, overlapping fragments of the introgression fragment within an otherwise uniform genetic background of the recurrent parent. Such lines can then be used to map on which fragment the QTL is located and to identify a line having a shorter introgression fragment comprising the QTL.

"Marker assisted selection" (of MAS) is a process by which individual plants are selected based on marker genotypes. "Marker assisted counter-selection" is a process by which marker genotypes are used to identify plants that will not be selected, allowing them to be removed from a breeding program or planting. Marker assisted selection uses the presence of molecular markers, which are genetically linked to a particular locus or to a particular chromosome region (e.g. introgression fragment, transgene, polymorphism, mutation, etc), to select plants for the presence of the specific locus or region (introgression fragment, transgene, polymorphism, mutation, etc). For example, a molecular marker genetically (closely) linked to a digestibility QTL as defined herein, can be used to detect and/or select plants comprising the QTL on chromosome 7. The closer the genetic linkage of the molecular marker to the locus (e.g. about 7cM, 6 cM, 5 cM, 4 cM, 3 cM, 2 cM, 1 cM, 0.5 cM or less), the less likely it is that the marker is dissociated from the locus through meiotic recombination. Likewise, the closer two markers are linked to each other (e.g. within 7 or 5 cM, 4 cM, 3 cM, 2 cM, 1 cM or less) the less likely it is that the two markers will be separated from one another (and the more likely they will co-segregate as a unit). A marker "within 7 cM or within 5 cM, 3 cM, 2 cM, or 1 cM" of another marker refers to a marker which genetically maps to within the 7 cM or 5 cM, 3 cM, 2 cM, or 1 cM region flanking the marker (i.e. either side of the marker). Similarly, a marker within 5 Mb, 3 Mb, 2.5 Mb, 2 Mb, 1 Mb, 0.5 Mb, 0.4 Mb, 0.3 Mb, 0.2 Mb, 0.1 Mb, 50 kb, 20 kb, 10 kb, 5 kb, 2 kb, 1 kb or less of another marker refers to a marker which is physically located within the 5 Mb, 3 Mb, 2.5 Mb, 2 Mb, 1 Mb, 0.5 Mb, 0.4 Mb, 0.3 Mb, 0.2 Mb, 0.1 Mb, 50 kb, 20 kb, 10 kb, 5 kb, 2 kb, 1 kb or less, of the genomic DNA region flanking the marker (i.e. either side of the marker). "LOD-score" (logarithm (base 10) of odds) refers to a statistical test often used for linkage analysis in animal and plant populations. The LOD score compares the likelihood of obtaining the test data if the two loci (molecular marker loci and/or a phenotypic

trait locus) are indeed linked, to the likelihood of observing the same data purely by chance. Positive LOD scores favor the presence of linkage and a LOD score greater than 3.0 is considered evidence for linkage. A LOD score of +3 indicates 1000 to 1 odds that the linkage being observed did not occur by chance.

5 A "marker haplotype" refers to a combination of alleles at a marker locus.

A "marker locus" is a specific chromosome location in the genome of a species where a specific marker can be found. A marker locus can be used to track the presence of a second linked locus, e.g., one that affects the expression of a phenotypic trait. For example, a marker locus can be used to monitor segregation of alleles at a genetically or physically linked locus.

10 A "marker probe" is a nucleic acid sequence or molecule that can be used to identify the presence of a marker locus, e.g., a nucleic acid probe that is complementary to a marker locus sequence, through nucleic acid hybridization. Marker probes comprising 30 or more contiguous nucleotides of the marker locus ("all or a portion" of the marker locus sequence) may be used for nucleic acid hybridization. Alternatively, in some aspects, a marker probe
15 refers to a probe of any type that is able to distinguish (i.e., genotype) the particular allele that is present at a marker locus.

The term "molecular marker" may be used to refer to a genetic marker or an encoded product thereof (e.g., a protein) used as a point of reference when identifying a linked locus. A marker can be derived from genomic nucleotide sequences or from expressed nucleotide sequences
20 (e.g., from a spliced RNA, a cDNA, etc.), or from an encoded polypeptide. The term also refers to nucleic acid sequences complementary to or flanking the marker sequences, such as nucleic acids used as probes or primer pairs capable of amplifying the marker sequence. A "molecular marker probe" is a nucleic acid sequence or molecule that can be used to identify the presence of a marker locus, e.g., a nucleic acid probe that is complementary to a marker locus sequence.
25 Alternatively, in some aspects, a marker probe refers to a probe of any type that is able to distinguish (i.e., genotype) the particular allele that is present at a marker locus. Nucleic acids are "complementary" when they specifically hybridize in solution, e.g., according to Watson-Crick base pairing rules. Some of the markers described herein are also referred to as hybridization markers when located on an indel region, such as the non- collinear region
30 described herein. This is because the insertion region is, by definition, a polymorphism vis a vis a plant without the insertion. Thus, the marker need only indicate whether the indel region is present or absent. Any suitable marker detection technology may be used to identify such a hybridization marker, e.g. SNP technology is used in the examples provided herein.

"Genetic markers" are nucleic acids that are polymorphic in a population and where the alleles of which can be detected and distinguished by one or more analytic methods, e.g., RFLP, AFLP, isozyme, SNP, SSR, and the like. The terms "molecular marker" and "genetic marker" are used interchangeably herein. The term also refers to nucleic acid sequences complementary to the genomic sequences, such as nucleic acids used as probes. Markers corresponding to genetic polymorphisms between members of a population can be detected by methods well-established in the art. These include, e.g., PCR-based sequence specific amplification methods, detection of restriction fragment length polymorphisms (RFLP), detection of isozyme markers, detection of polynucleotide polymorphisms by allele specific hybridization (ASH), detection of amplified variable sequences of the plant genome, detection of self-sustained sequence replication, detection of simple sequence repeats (SSRs), detection of single nucleotide polymorphisms (SNPs), or detection of amplified fragment length polymorphisms (AFLPs). Well established methods are also known for the detection of expressed sequence tags (ESTs) and SSR markers derived from EST sequences and randomly amplified polymorphic DNA (RAPD).

A "polymorphism" is a variation in the DNA between two or more individuals within a population. A polymorphism preferably has a frequency of at least 1 % in a population. A useful polymorphism can include a single nucleotide polymorphism (SNP), a simple sequence repeat (SSR), or an insertion/deletion polymorphism, also referred to herein as an "indel". The term "indel" refers to an insertion or deletion, wherein one line may be referred to as having an inserted nucleotide or piece of DNA relative to a second line, or the second line may be referred to as having a deleted nucleotide or piece of DNA relative to the first line.

"Physical distance" between loci (e.g. between molecular markers and/or between phenotypic markers) on the same chromosome is the actually physical distance expressed in bases or base pairs (bp), kilo bases or kilo base pairs (kb) or megabases or mega base pairs (Mb).

"Genetic distance" between loci (e.g. between molecular markers and/or between phenotypic markers) on the same chromosome is measured by frequency of crossing-over, or recombination frequency (RF) and is indicated in centimorgans (cM). One cM corresponds to a recombination frequency of 1%. If no recombinants can be found, the RF is zero and the loci are either extremely close together physically or they are identical. The further apart two loci are, the higher the RF.

A "physical map" of the genome is a map showing the linear order of identifiable landmarks (including genes, markers, etc.) on chromosome DNA. However, in contrast to genetic maps, the distances between landmarks are absolute (for example, measured in base pairs or

isolated and overlapping contiguous genetic fragments) and not based on genetic recombination (that can vary in different populations).

An allele "negatively" correlates with a trait when it is linked to it and when presence of the allele is an indicator that a desired trait or trait form will not occur in a plant comprising the allele. An allele "positively" correlates with a trait when it is linked to it and when presence of the allele is an indicator that the desired trait or trait form will occur in a plant comprising the allele.

A centimorgan ("cM") is a unit of measure of recombination frequency. One cM is equal to a 1 % chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing over in a single generation.

As used herein, the term "chromosomal interval" designates a contiguous linear span of genomic DNA that resides in planta on a single chromosome. The genetic elements or genes located on a single chromosomal interval are physically linked. The size of a chromosomal interval is not particularly limited. In some aspects, the genetic elements located within a single chromosomal interval are genetically linked, typically with a genetic recombination distance of, for example, less than or equal to 20 cM, or alternatively, less than or equal to 10 cM. That is, two genetic elements within a single chromosomal interval undergo recombination at a frequency of less than or equal to 20% or 10%.

The term "closely linked", in the present application, means that recombination between two linked loci occurs with a frequency of equal to or less than about 10% (i.e., are separated on a genetic map by not more than 10 cM). Put another way, the closely linked loci co-segregate at least 90% of the time. Marker loci are especially useful with respect to the subject matter of the current disclosure when they demonstrate a significant probability of co-segregation (linkage) with a desired trait (e.g., resistance to gray leaf spot). Closely linked loci such as a marker locus and a second locus can display an inter-locus recombination frequency of 10% or less, preferably about 9% or less, still more preferably about 8% or less, yet more preferably about 7% or less, still more preferably about 6% or less, yet more preferably about 5% or less, still more preferably about 4% or less, yet more preferably about 3% or less, and still more preferably about 2% or less. In highly preferred embodiments, the relevant loci display a recombination a frequency of about 1 % or less, e.g., about 0.75% or less, more preferably about 0.5% or less, or yet more preferably about 0.25% or less. Two loci that are localized to the same chromosome, and at such a distance that recombination between the two loci occurs at a frequency of less than 10% (e.g., about 9 %, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1 %, 0.75%, 0.5%,

0.25%, or less) are also said to be "proximal to" each other. In some cases, two different markers can have the same genetic map coordinates. In that case, the two markers are in such close proximity to each other that recombination occurs between them with such low frequency that it is undetectable.

- 5 "Linkage" refers to the tendency for alleles to segregate together more often than expected by chance if their transmission was independent. Typically, linkage refers to alleles on the same chromosome. Genetic recombination occurs with an assumed random frequency over the entire genome. Genetic maps are constructed by measuring the frequency of recombination between pairs of traits or markers. The closer the traits or markers are to each other on the
- 10 chromosome, the lower the frequency of recombination, and the greater the degree of linkage. Traits or markers are considered herein to be linked if they generally co-segregate. A 1/100 probability of recombination per generation is defined as a genetic map distance of 1.0 centiMorgan (1.0 cM). The term "linkage disequilibrium" refers to a non-random segregation of genetic loci or traits (or both). In either case, linkage disequilibrium implies that the relevant
- 15 loci are within sufficient physical proximity along a length of a chromosome so that they segregate together with greater than random (i.e., non-random) frequency. Markers that show linkage disequilibrium are considered linked. Linked loci co-segregate more than 50% of the time, e.g., from about 51 % to about 100% of the time. In other words, two markers that co-segregate have a recombination frequency of less than 50% (and by definition, are separated
- 20 by less than 50 cM on the same linkage group.) As used herein, linkage can be between two markers, or alternatively between a marker and a locus affecting a phenotype. A marker locus can be "associated with" (linked to) a trait. The degree of linkage of a marker locus and a locus affecting a phenotypic trait is measured, e.g., as a statistical probability of co-segregation of that molecular marker with the phenotype (e.g., an F statistic or LOD score).
- 25 The genetic elements or genes located on a single chromosome segment are physically linked. In some embodiments, the two loci are located in close proximity such that recombination between homologous chromosome pairs does not occur between the two loci during meiosis with high frequency, e.g., such that linked loci co-segregate at least about 90% of the time, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.75%, or more of the time. The
- 30 genetic elements located within a chromosomal segment are also "genetically linked", typically within a genetic recombination distance of less than or equal to 50cM, e.g., about 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0.75, 0.5, 0.25 cM or less. That is, two genetic elements within a single chromosomal segment undergo recombination during

meiosis with each other at a frequency of less than or equal to about 50%, e.g., about 49%, 48%, 47%, 46%, 45%, 44%, 43%, 42%, 41%, 40%, 39%, 38%, 37%, 36%, 35%, 34%, 33%, 32%, 31%, 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25% or less.

5 "Closely linked" markers display a cross over frequency with a given marker of about 10% or less, e.g., 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25% or less (the given marker locus is within about 10 cM of a closely linked marker locus, e.g., 9, 8, 7, 6, 5, 4, 3, 2, 1, 0.75, 0.5, 0.25 cM or less of a closely linked marker locus). Put another way, closely linked marker loci co-segregate at least about 90% the time, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

10 99%, 99.5%, 99.75%, or more of the time. As used herein, the terms "introgression", "introgressed" and "introgressing" refer to both a natural and artificial process whereby chromosomal fragments or genes of one species, variety or cultivar are moved into the genome of another species, variety or cultivar, by crossing those species. The process may optionally be completed by backcrossing to the recurrent parent. For example, introgression of

15 a desired allele at a specified locus can be transmitted to at least one progeny via a sexual cross between two parents of the same species, where at least one of the parents has the desired allele in its genome. Alternatively, for example, transmission of an allele can occur by recombination between two donor genomes, e.g., in a fused protoplast, where at least one of the donor protoplasts has the desired allele in its genome. The desired allele can be, e.g.,

20 detected by a marker that is associated with a phenotype, at a QTL, a transgene, or the like. In any case, offspring comprising the desired allele can be repeatedly backcrossed to a line having a desired genetic background and selected for the desired allele, to result in the allele becoming fixed in a selected genetic background. The process of "introgressing" is often referred to as "backcrossing" when the process is repeated two or more times. "Introgression

25 fragment" or "introgression segment" or "introgression region" refers to a chromosome fragment (or chromosome part or region) which has been introduced into another plant of the same or related species either artificially or naturally such as by crossing or traditional breeding techniques, such as backcrossing, i.e. the introgressed fragment is the result of breeding methods referred to by the verb "to introgress" (such as backcrossing). It is

30 understood that the term "introgression fragment" never includes a whole chromosome, but only a part of a chromosome. The introgression fragment can be large, e.g. even three quarter or half of a chromosome, but is preferably smaller, such as about 15 Mb or less, such as about 10 Mb or less, about 9 Mb or less, about 8 Mb or less, about 7 Mb or less, about 6 Mb or less, about 5 Mb or less, about 4 Mb or less, about 3 Mb or less, about 2.5 Mb or 2 Mb or less,

35 about 1 Mb (equals 1,000,000 base pairs) or less, or about 0.5 Mb (equals 500,000 base pairs)

or less, such as about 200,000 bp (equals 200 kilo base pairs) or less, about 100,000 bp (100 kb) or less, about 50,000 bp (50 kb) or less, about 25,000 bp (25 kb) or less. In certain embodiments, the introgression fragment comprises, consists of, or consists essentially of the QTL according to the invention as described herein.

5 A genetic element, an introgression fragment, or a gene or allele conferring a trait (such as improved digestibility) is said to be "obtainable from" or can be "obtained from" or "derivable from" or can be "derived from" or "as present in" or "as found in" a plant or plant part as described herein elsewhere if it can be transferred from the plant in which it is present into
10 techniques without resulting in a phenotypic change of the recipient plant apart from the addition of the trait conferred by the genetic element, locus, introgression fragment, gene or allele. The terms are used interchangeably and the genetic element, locus, introgression fragment, gene or allele can thus be transferred into any other genetic background lacking the trait. Not only plants comprising the genetic element, locus, introgression fragment, gene or
15 allele can be used, but also progeny/descendants from such plants which have been selected to retain the genetic element, locus, introgression fragment, gene or allele, can be used and are encompassed herein. Whether a plant (or genomic DNA, cell or tissue of a plant) comprises the same genetic element, locus, introgression fragment, gene or allele as obtainable from such plant can be determined by the skilled person using one or more techniques known in the
20 art, such as phenotypic assays, whole genome sequencing, molecular marker analysis, trait mapping, chromosome painting, allelism tests and the like, or combinations of techniques. It will be understood that transgenic plants may also be encompassed.

As used herein the terms "genetic engineering", "transformation" and "genetic modification" are all used herein as synonyms for the transfer of isolated and cloned genes into the DNA,
25 usually the chromosomal DNA or genome, of another organism.

The present invention in an aspect relates to means and methods suitable for enhancing the haploid induction capability of an inducer plant. The haploid inducer plant in certain embodiments may be a plant selected and/or derived from the lines Stock 6, RWK, RWS, UH400, AX5707RS, NP2222-mai/, or any other haploid inducer line. In certain embodiments, a
30 haploid inducer as referred to herein can be a mutant plant selected from the plants as disclosed in US 2017/327832 A1 (or a plant of a different species having the corresponding mutation(s)), incorporated herein by reference in its entirety, in particular the tilling mutants having a D74N (exchange of aspartate at Position 74 for asparagin) or G78R (exchange of glycine at Position 78 for arginine) in the PLA1 gene. In certain embodiments, a haploid inducer

as referred to herein can be a mutant plant selected from the plants as described in WO 2018/158301 (or a plant of a different species having the corresponding mutation(s)), incorporated herein by reference in its entirety. In certain embodiments, a haploid inducer as referred to herein can be a plant with mutated PLA1 according to SEQ ID NOs: 38-40. In certain

5 embodiments, a haploid inducer as referred to herein can be a mutant plant selected from the plants as described in Kalinowska et al. (2018) "State-of-the-art and novel developments of in vivo haploid technologies"; Theor. Appl. Genetics; 132(3):593-605; incorporated herein by reference in its entirety; preferably wherein haploid induction is caused by a mutated patatin-like phospholipase (PLA1).

10 In an aspect, the invention provides a nucleic acid which, after transcription or expression in a plant or after silencing in a plant, is suitable for enhancing the haploid induction capability of an inducer plant, such as caused by a mutated patatin-like phospholipase (PLA1). On the other hand, an endogenous DNA sequence in the genome of a plant, or in the genome of a plant haploid inductor, which is identical to one of the nucleic acids according to the invention, may

15 also be modified such that the property of a haploid inductor is mediated, or the induction capability of the haploid inductor is increased, after transcription or expression of the endogenous DNA sequence. The nucleic acid of the present invention is preferably an isolated nucleic acid which is extracted from its natural or original environment (genetic context). A nucleic acid may be double-stranded or single-stranded, and linear or circular. It may thereby

20 be genomic DNA, synthetic DNA, cDNA, or an RNA type (for example, lncRNA, siRNA, or miRNA), wherein the nucleobase uracil occurs in RNA instead of the nucleobase thymine.

The nucleic acid according to the invention may be used as a transgene. On the other hand, an endogenous DNA sequence in the genome of a plant, or in the genome of a haploid inductor, preferably a haploid inductor comprising a mutation in a patatin-like phospholipase (PLA1),

25 which is identical to one of the nucleic acids according to the invention, may also be modified such that the induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) can be enhanced, after transcription or expression of the endogenous DNA sequence or after silencing of the endogenous DNA sequence, and optionally after crossing with a plant comprising a mutated patatin-like phospholipase (PLA1).

30 The nucleic acid of the present invention is preferably an isolated nucleic acid which is extracted or isolated from its natural or original environment (genetic context). A nucleic acid may be double-stranded or single-stranded, and linear or circular. It may thereby be genomic DNA, synthetic DNA, cDNA, or an RNA type (for example, lncRNA, siRNA, or miRNA), wherein the nucleobase uracil occurs in RNA instead of the nucleobase thymine.

In particular embodiments, the nucleic acid which, after transcription or expression in a plant, is preferably suitable for enhancing the induction capability of an inducer plant, such as caused by a mutated patatin-like phospholipase (PLA1), may be a nucleic acid comprising a nucleotide sequence which:

- 5 (i) is a sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 10, 12, 15, 18, 21, 30 and 35, or a functional fragment thereof, preferably SEQ ID NOs: 1, 4, 10, 15, 30 and 35, or a functional fragment thereof; or
- (ii) has a coding sequence selected from the group consisting of SEQ ID NOs: 2, 6, 13, 16, 19, 22 and 36, preferably SEQ ID NOs: 2, 16, and 36; or
- 10 (iii) is complementary to the sequence from (i) or (ii); or
- (iv) is at least 80%, 82%, 84%, 86%, 88%, preferably at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, or more preferably at least 97%, 97.5%, 98%, 98.5%, 99%, or 99.5% identical to the sequence from (i) or (ii); or
- (v) encodes a protein comprising an amino acid sequence selected from the group
- 15 consisting of SEQ ID NOs: 3, 7, 9, 11, 14, 17, 20, 23 and 37, or a functional part of the protein, preferably SEQ ID NOs: 3, 11, 17, and 37, or a functional part of the protein; or
- (vi) encodes a protein comprising an amino acid sequence which is at least 80%, 82%, 84%, 86%, 88%, preferably at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, or more
- 20 preferably at least 97%, 97.5%, 98%, 98.5%, 99%, or 99.5% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 3, 7, 9, 11, 14, 17, 20, 23 and 37, or a functional part of the protein, preferably SEQ ID NOs: 3, 11, 17, and 37, or a functional part of the protein; or
- (vii) hybridizes with a sequence from (iii) under stringent conditions.
- 25 This nucleic acid, or an RNA encoded by the nucleic acid or a protein or polypeptide encoded by the nucleic acid, has a positive effect on the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1), and therefore, in the following, a nucleic acid which is defined here is designated as an induction-promoting nucleic acid. Non-limiting examples of uses of the nucleic acid in order to enhance the induction
- 30 capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1), may include transgenically or endogenously increasing the expression rate of the nucleic acid or the activity or stability of the encoded protein or of the encoded part of the protein. Additional methods and uses of the induction-promoting nucleic acid, as well as substances which comprise the induction-promoting nucleic acid, are disclosed further below.

In particular embodiments, the nucleic acid which, after silencing in a plant, is preferably suitable for enhancing the induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) may be a nucleic acid comprising a nucleotide sequence which:

- 5 (i) is a sequence selected from the group consisting of SEQ ID NOs: 24, 27 and 32, or a functional fragment thereof, preferably SEQ ID NO: 24, or a functional fragment thereof ; or
- (ii) has a coding sequence selected from the group consisting of SEQ ID NO: 25, 28 and 33, preferably SEQ ID NO: 25; or
- 10 (iii) is complementary to the sequence from (i) or (ii); or
- (iv) is at least 80%, 82%, 84%, 86%, 88%, preferably at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, or more preferably at least 97%, 97.5%, 98%, 98.5%, 99%, or 99.5% identical to the sequence from (i) or (ii); or
- (v) encodes a protein comprising an amino acid sequence selected from the group
- 15 consisting of SEQ ID NOs: 26, 29 and 34, or a functional part of the protein, preferably SEQ ID NO: 26, or a functional part of the protein; or
- (vi) encodes a protein comprising an amino acid sequence which is at least 80%, 82%, 84%, 86%, 88%, preferably at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, or more preferably at least 97%, 97.5%, 98%, 98.5%, 99%, or 99.5% identical to an amino
- 20 acid sequence selected from the group consisting of SEQ ID NOs: 26, 29 and 34, or a functional part of the protein, preferably SEQ ID NO: 26, or a functional part of the protein; or
- (vii) hybridizes with a sequence from (iii) under stringent conditions.

This nucleic acid, or an RNA encoded by the nucleic acid or a protein or polypeptide encoded

25 by the nucleic acid, has a negative effect on the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1), and therefore, in the following, a nucleic acid which is defined here is designated as an induction-inhibiting nucleic acid. Non-limiting examples of uses of the nucleic acid in order to enhance the induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase

30 (PLA1), may include transgenically or endogenously reducing the expression rate of the nucleic acid or the activity or stability of the encoded protein or of the encoded part of the protein. Additional methods and uses of the induction-inhibiting nucleic acid, as well as substances which comprise the induction-inhibiting nucleic acid, are disclosed further below.

In other embodiments, the nucleic acid which, after transcription or expression in a plant, is

- preferably suitable for enhancing the induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) may be a nucleic acid that encodes for an RNA that has a double-stranded portion, wherein at least one strand of the double-stranded portion has a nucleotide sequence which is homologous or identical to at least 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25, preferably at least 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, or 140, more preferably at least 160, 180, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 successive nucleotides of a coding sequence of a nucleic acid that:
- (i) has a sequence selected from the group consisting of SEQ ID Nos: 25, 28 and 33, or a fragment thereof, in a sense or anti-sense orientation; or
 - (ii) is complementary to the sequence from (i); or
 - (iii) is at least 80%, 82%, 84%, 86%, or 88%, preferably at least 90%, 91%, 92%, 93%, 94%, 95%, or 96%, more preferably at least 97%, 97.5%, 98%, 98.5%, 99%, or 99.5% identical to the sequence from (i); or
 - (iv) hybridizes with the sequence from (ii) under stringent conditions.
- In post-transcriptional gene silencing, as, for example, in the RNAi approach or miRNA approach (Fire et al., 1998), such a nucleic acid may be used to suppress the expression of the induction-inhibiting nucleic acid described above.

Table 1: Sequence index and sequence association of the nucleotide and amino acid sequences.

SEQ ID NO:	description	Sequence type	Source
1	GRMZM2G435294 Myosin (<i>Zea mays</i>)	genomic DNA	RWS and W22
2	GRMZM2G435294 Myosin (<i>Zea mays</i>)	cDNA	RWS and W22
3	GRMZM2G435294 Myosin (<i>Zea mays</i>)	polypeptide	RWS and W22
4	upstream region of myosin gene (<i>Zea mays</i>)	genomic DNA	RWS and W22
5	GRMZM2G435294 Myosin (<i>Zea mays</i>)	genomic DNA	B73
6	GRMZM2G435294 Myosin (<i>Zea mays</i>)	cDNA	B73
7	GRMZM2G435294 Myosin (<i>Zea mays</i>)	polypeptide	B73
8	Myosin fragment (<i>Zea mays</i>)	genomic DNA24	PH207
9	Myosin fragment (<i>Zea mays</i>)	polypeptide	PH207
10	Myosin A325V fragment (<i>Zea mays</i>)	genomic DNA	TILLING
11	Myosin A325V fragment (<i>Zea mays</i>)	polypeptide	TILLING
12	Myosin (<i>Sorghum bicolor</i>)	genomic DNA	
13	Myosin (<i>Sorghum bicolor</i>)	cDNA	
14	Myosin (<i>Sorghum bicolor</i>)	polypeptide	

SEQ ID NO:	description	Sequence type	Source
15	GRMZM2G094586 PPR (<i>Zea mays</i>)	genomic DNA	RWS and W22
16	GRMZM2G094586 PPR (<i>Zea mays</i>)	cDNA	RWS and W22
17	GRMZM2G094586 PPR (<i>Zea mays</i>)	polypeptide	RWS and W22
18	GRMZM2G094586 PPR (<i>Zea mays</i>)	genomic DNA	B73
19	GRMZM2G094586 PPR (<i>Zea mays</i>)	cDNA	B73
20	GRMZM2G094586 PPR (<i>Zea mays</i>)	polypeptide	B73
21	PPR (<i>Sorghum bicolor</i>)	genomic DNA	
22	PPR (<i>Sorghum bicolor</i>)	cDNA	
23	PPR (<i>Sorghum bicolor</i>)	polypeptide	
24	GRMZM2G081239 Aquaporin (<i>Zea mays</i>)	genomic DNA	RWS and W22
25	GRMZM2G081239 Aquaporin (<i>Zea mays</i>)	cDNA	RWS and W22
26	GRMZM2G081239 Aquaporin (<i>Zea mays</i>)	polypeptide	RWS and W22
27	GRMZM2G081239 Aquaporin (<i>Zea mays</i>)	genomic DNA	B73
28	GRMZM2G081239 Aquaporin (<i>Zea mays</i>)	cDNA	B73
29	GRMZM2G081239 Aquaporin (<i>Zea mays</i>)	polypeptide	B73
30	upstream region of aquaporin gene (<i>Zea mays</i>)	genomic DNA	RWS and W22
31	upstream region of aquaporin gene (<i>Zea mays</i>)	genomic DNA	B73
32	aquaporin (<i>Sorghum bicolor</i>)	genomic DNA	
33	aquaporin (<i>Sorghum bicolor</i>)	cDNA	
34	aquaporin (<i>Sorghum bicolor</i>)	polypeptide	
35	gene in upstream region of aquaporin gene (<i>Zea mays</i>)	genomic DNA	RWS
36	gene in upstream region of aquaporin gene (<i>Zea mays</i>)	cDNA	RWS
37	gene in upstream region of aquaporin gene (<i>Zea mays</i>)	polypeptide	RWS
38	mutated PLA1 (<i>Zea mays</i>)	genomic DNA	PH207 (mutated HIR1)
39	mutated PLA1 (<i>Zea mays</i>)	cDNA	PH207 (mutated HIR1)
40	mutated PLA1 (<i>Zea mays</i>)	polypeptide	PH207 (mutated HIR1)
41	native PLA1 (<i>Zea mays</i>)	genomic DNA	PH207
42	native PLA1 (<i>Zea mays</i>)	cDNA	PH207
43	native PLA1 (<i>Zea mays</i>)	polypeptide	PH207
44	PLA1 (<i>Zea mays</i>)	genomic DNA	RWS (mutated

SEQ ID NO:	description	Sequence type	Source
			HIR1)
45	PLA1 (<i>Zea mays</i>)	cDNA	RWS (mutated HIR1)
46	PLA1 (<i>Zea mays</i>)	polypeptide	RWS (mutated HIR1)
47	PLA1 (<i>Zea mays</i>)	genomic DNA	B73
48	PLA1 (<i>Zea mays</i>)	cDNA	B73
49	PLA1 (<i>Zea mays</i>)	polypeptide 1	B73
50	PLA1 (<i>Zea mays</i>)	polypeptide 2	B73
51	aquaporin marker allele 1	DNA	RWS with HIR8 locus
52	aquaporin marker allele 1	DNA	genotype without enhancer at HIR8 locus
53	aquaporin marker allele 2	DNA	RWS with HIR8 locus
54	aquaporin marker allele 2	DNA	genotype without enhancer at HIR8 locus
55	aquaporin marker allele 3	DNA	RWS with HIR8 locus
56	aquaporin marker allele 3	DNA	genotype without enhancer at HIR8 locus
57	aquaporin marker allele 4	DNA	RWS with HIR8 locus
58	aquaporin marker allele 4	DNA	genotype without enhancer at HIR8 locus
59	aquaporin marker allele 5	DNA	RWS with HIR8 locus
60	aquaporin marker allele 5	DNA	genotype without enhancer at HIR8 locus
61	aquaporin marker allele 6	DNA	RWS with HIR8 locus
62	aquaporin marker allele 6	DNA	genotype without enhancer at HIR8 locus
63	aquaporin marker allele 7	DNA	RWS with HIR8 locus
64	aquaporin marker allele 7	DNA	genotype without

SEQ ID NO:	description	Sequence type	Source
			enhancer at HIR8 locus
65	aquaporin marker allele 7 polymorphism	DNA	RWS with HIR8 locus
66	aquaporin marker allele 7 polymorphism	DNA	genotype without enhancer at HIR8 locus
67	myosin marker allele 1	DNA	RWS with HIR8 locus
68	myosin marker allele 1	DNA	genotype without enhancer at HIR8 locus
69	myosin marker allele 2	DNA	RWS with HIR8 locus
70	myosin marker allele 2	DNA	genotype without enhancer at HIR8 locus
71	myosin marker allele 2 polymorphism	DNA	RWS with HIR8 locus
72	myosin marker allele 3	DNA	RWS with HIR8 locus
73	myosin marker allele 3	DNA	genotype without enhancer at HIR8 locus
74	myosin marker allele 3 polymorphism	DNA	RWS with HIR8 locus

In certain embodiments, the haploid induced plant comprises the HIR1 QTL (*qhir1*), as for instance described in Nair et al. (2017) "Dissection of a major QTL *qhir1* conferring maternal haploid induction ability in maize"; Theor Appl Genet; 130(6):1113-1122, incorporated herein
5 by reference in its entirety. In certain preferred embodiments, the haploid induced plant comprises the HIR1 QTL sub-region *qhir11*. The skilled person will understand that a comparable or corresponding QTL in plants other than maize can also be used.

In certain embodiments, the haploid induced plant is selected from, or is derived from the following lines: Stock 6, RWK, RWS, UH400, AX5707RS, NP2222-mai/. It will be understood that
10 "derived from" in this context means that the haploid induction phenotype is exhibited in the plant.

In certain embodiments, the haploid inducer plant comprises a mutated PLA1 gene, as defined

herein elsewhere. The mutated PLA1 gene may result in a gain-of-function, such as a dominant negative phenotype, (partial) loss-of-function, and may result in a knock-out or a truncated PLA1.

In certain embodiments, the PLA1 gene has a sequence, such as a genomic sequence, mRNA
5 sequence, or preferably protein sequence identical to the corresponding sequence of the PLA1 gene in lines Stock 6, RWK, RWS, UH400, AX5707RS, NP2222-mai/.

In certain embodiments, the PLA1 gene has one or more of the following mutations: D74N, G78R, R58Q, V162I, S288L, or Q369stop.

In certain embodiments, the PLA1 gene has a genomic sequence as set forth in SEQ ID NO: 38.

10 In certain embodiments, the PLA1 gene has a cDNA sequence as set forth in SEQ ID NO: 39. In certain embodiments, the PLA1 gene has a protein sequence as set forth in SEQ ID NO: 40.

In certain embodiments, the PLA1 gene has a genomic sequence as set forth in SEQ ID NO: 44.

In certain embodiments, the PLA1 gene has a cDNA sequence as set forth in SEQ ID NO: 45. In certain embodiments, the PLA1 gene has a protein sequence as set forth in SEQ ID NO: 46.

15 In a further aspect, the present invention relates to a vector which comprises a nucleic acid as described herein, preferably an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein.

The vector may be a plasmid, a cosmid, a phage or an expression vector, a transformation vector, shuttle vector, or cloning vector; it may be double- or single-stranded, linear or circular.

20 The vector may transform a prokaryotic or eukaryotic host, either via integration into its genome or extrachromosomally.

In certain embodiments, the vector is an expression vector. The nucleic acid is preferably operatively linked in a vector with one or more regulatory sequences which allow the transcription, and optionally the expression, in a prokaryotic or eukaryotic host cell. A
25 regulatory sequence may be homologous or heterologous to the nucleic acid. For example, the nucleic acid is under the control of a suitable promoter or terminator. Suitable promoters may be promoters which are constitutively induced, for example, the 35S promoter from the "Cauliflower mosaic virus" (Odell et al., 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter.) Tissue-specific promoters, e.g. pollen-

30 specific promoters as described in Chen et al. (2010. Molecular Biology Reports 37(2):737-744), Zhao et al. (2006. Planta 224(2): 405-412), or Twell et al. (1991. Genes & Development 5(3): 496-507), are particularly suitable, as are development-specific promoters, e.g. blossom-

specific promoters. Suitable promoters may also be synthetic or chimeric promoters which do not occur in nature, and which are composed of multiple elements. Such synthetic or chimeric promoter may contain a minimal promoter, as well as at least one cis-regulatory element which serves as a binding location for special transcription factors. Chimeric promoters may be designed according to the desired specifics and can be induced or repressed via different factors. Examples of such promoters are found in Gurr & Rushton (2005. Trends in Biotechnology 23(6): 275-282) or Venter (2007. Trends in Plant Science: 12(3): 118-124). For example, a suitable terminator is the nos-terminator (Depicker et al., 1982. Journal of Molecular and Applied Genetics 1(6): 561-573).

- 10 In certain embodiments, the vector is a conditional expression vector. In certain embodiments, the vector is a constitutive expression vector. In certain embodiments, the vector is a tissue-specific expression vector, such as a pollen-specific expression vector. In certain embodiments, the vector is an inducible expression vector. All such vectors are well-known in the art.

Methods for preparation of the described vectors are commonplace to the person skilled in the art (Sambrook et al., 2001).

- Also envisaged herein is a host cell, such as a plant cell, which comprises a nucleic acid as described herein, preferably an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein, or a vector as described herein. The host cell may contain the nucleic acid as an extra-chromosomally (episomal) replicating molecule, or comprises the nucleic acid integrated in the nuclear or plastid genome of the host cell, or as introduced chromosome, e.g. minichromosome.

The host cell may be a prokaryotic (for example, bacterial) or eukaryotic cell (for example, a plant cell or a yeast cell). For example, the host cell may be an agrobacterium, such as *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. Preferably, the host cell is a plant cell.

- A nucleic acid described herein or a vector described herein may be introduced in a host cell via well-known methods, which may depend on the selected host cell, including, for example, conjugation, mobilization, biolistic transformation, agrobacteria-mediated transformation, transfection, transduction, vacuum infiltration, or electroporation. In particular, methods for introducing a nucleic acid or a vector in an agrobacterium cell are well-known to the skilled person and may include conjugation or electroporation methods. Also methods for introducing a nucleic acid or a vector into a plant cell are known (Sambrook et al., 2001) and may include diverse transformation methods such as biolistic transformation and agrobacterium-mediated

transformation.

In particular embodiments, the present invention relates to a transgenic plant cell which comprises a nucleic acid as described herein, in particular an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein, as a transgene or a vector as described herein. In further embodiments, the present invention relates to a transgenic plant or a part thereof which comprises the transgenic plant cell.

For example, such a transgenic plant cell or transgenic plant is a plant cell or plant which is, preferably stably, transformed with a nucleic acid as described herein, in particular an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein, or a vector as described herein.

Preferably, the nucleic acid in the transgenic plant cell is operatively linked with one or more regulatory sequences which allow the transcription, and optionally the expression, in the plant cell. A regulatory sequence may be homologous or heterologous to the nucleic acid. The total structure made up of the nucleic acid according to the invention and the regulatory sequence(s) may then represent the transgene.

A part of a transgenic plant may be, for example, a fertilized or unfertilized seed, an embryo, a pollen, a tissue, an organ, or a plant cell, wherein the fertilized or unfertilized seed, the embryo, or the pollen are generated in the transgenic plant, and the nucleic acid as described herein, in particular an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein, is integrated into its genome as a transgene or the vector. The term transgenic plant as used herein also includes a descendant of the transgenic plant described herein in whose genome the nucleic acid as described herein, in particular an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein, is integrated as a transgene or the vector.

Also envisaged herein are plant cells that endogenously comprise a nucleic acid as described herein, in particular an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein, as well as a plant or a part thereof comprising such plants cells.

Further disclosed herein is a protein or a polypeptide encoded by a nucleic acid as described herein. The protein or polypeptide may be suitable for enhancing the induction capability of a plant inducer, preferably caused by a mutated patatin-like phospholipase (PLA1). The protein or polypeptide encoded by an induction-promoting nucleic acid described herein is preferred. The protein or polypeptide may comprise an amino acid sequence selected from

SEQ ID NOs: 3, 7, 9, 11, 14, 17, 20, 23, or 37, preferably SEQ ID NOs: 3, 11, 17, or 37, or from SEQ ID NOs: 26, 29 or 34.

In a further aspect, the present invention relates to methods for the production of a plant comprising or which is carrier of an enhancer for the haploid induction capability of an inducer
5 plant, such as caused by a mutated patatin-like phospholipase (PLA1).

In particular, a method is provided herein for the production of a (transgenic) plant, such as a plant which comprises or which is a carrier of an enhancer for the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1), said method comprising the following steps:

- 10 a) provision of a nucleic acid as described herein, such as an induction-promoting nucleic acid or a nucleic acid encoding a double-stranded RNA as described herein; or provision of a vector as described herein comprising said nucleic acid,
- b) transformation, in particular stable transformation, of at least one plant cell with the nucleic acid or the vector provided in step a),
- 15 c) regeneration of a transgenic plant from the at least one transgenic plant cell obtained in step b) or a transgenic progeny thereof, and
- d) optionally growing the transgenic plant under conditions that are suitable for (over)expression of the transformed nucleic acid.

The introduced or transformed nucleic acid may be (over)expressed in a non-tissue-specific
20 way, or in tissue-specific way, preferably in pollen. Preferably, the (over)expression of the introduced or transformed nucleic acid appears at least in pollen of the transgenic plant or in a tissue of pollen of the transgenic plant.

In certain embodiments, the method may further comprise the step e) of identification of a modified expression pattern in the transgenic plant obtained in step c) or in a plant part
25 thereof, preferably in pollen of the transgenic plant or in a tissue of pollen of the transgenic plant.

The expression pattern is preferably altered to the effect that:

- (I) the transcription or expression rate of the introduced induction-promoting nucleic acid is increased in the transgenic plant in comparison to a wild-type plant which, for example, was
30 regenerated from an isogenic, untransformed plant cell, and/or
- (II) due to post-transcriptional gene silencing, the expression rate of an endogenous DNA sequence having a nucleotide sequence that is identical to the induction-inhibiting nucleic acid as described herein is reduced, e.g. via a double-stranded RNA which is encoded by the introduced nucleic acid encoding a double-stranded RNA, in the transgenic plant in comparison

to a wild-type plant which, for example, was regenerated from an isogenic, untransformed plant cell.

A verification of the transcription rate may take place via qRT-PCR, for example.

In embodiments of the method, two or more of the induction-promoting nucleic acids and
5 nucleic acids encoding a double-stranded RNA may be provided and the at least one plant cell may be transformed with said two or more nucleic acids. These two or more nucleic acids may be comprised in one or more vectors.

In embodiments, one or more additional nucleic acids which are known to be suitable for the generation of a haploid inducer (for example, manipulated *cenH3* gene (Ravi & Chan, 2010.
10 Nature 464(7288): 615-618) or for enhancing the induction capability of a haploid inducer may be provided and transformed into the at least one plant cell, in addition to the induction-promoting nucleic acid and/or the nucleic acid encoding a double-stranded RNA as described herein.

Also envisaged herein is a transgenic plant which can be or is produced with this method,
15 including a descendant of the transgenic plant which comprises the introduced nucleic acid as a transgene, or a part of this plant. A part of the plant may be a fertilized or unfertilized seed, an embryo, a pollen, a tissue, an organ, or a plant cell, wherein the fertilized or unfertilized seed, the embryo, or the pollen are generated at the transgenic plant, and the introduced nucleic acid is integrated into its genome as a transgene or the vector.

Also provided herein is a method for the production of a (mutated) plant, such as a (mutated)
20 plant which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1), said method comprising the following steps:

a) mutagenization of plant cells and subsequent regeneration of plants from the mutagenized
25 plant cells or mutagenization of plants, and

b) identification of a plant obtained in step a) which:

has at least one mutation in an endogenous DNA sequence which is identical to an induction-promoting or an induction-inhibiting nucleic acid as described herein, or

has at least one mutation in a regulatory sequence (for example, a promoter, enhancer,
30 terminator, or intron) of said endogenous DNA sequence,

wherein said mutation produces a change in the transcription or expression rate of the endogenous DNA sequence in the identified plant, in comparison to a non-mutagenized wild-

type plant, or a change in the activity or stability of a protein or polypeptide encoded by the endogenous DNA sequence in the identified plant, in comparison to a non-mutagenized wild-type plant,

- preferably wherein the at least one mutation causes the property of an enhancer of the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) in the identified plant.

Preferably, the change in the transcription rate or expression rate, or the change in the activity or stability, appears at least in a pollen of the identified plant or in a tissue of a pollen of the identified plant.

- 10 In preferred embodiments, step b) encompasses the identification of a plant obtained in step a) which (i):

has at least one mutation in an endogenous DNA sequence that is identical to an induction-inducing nucleic acid as described herein, or

- has at least one mutation in a regulatory sequence (for example, a promoter, enhancer, terminator, or intron) of said endogenous DNA sequence,

wherein said mutation effects an increase in the transcription or expression rate of said endogenous DNA sequence or an increase in the activity or stability of a protein or polypeptide encoded by said endogenous DNA sequence; and/or

which (ii):

- 20 has at least one mutation in an endogenous DNA sequence that is identical to an induction-inhibiting nucleic acid as described herein, or

has at least one mutation in a regulatory sequence (for example, a promoter, enhancer, terminator, or intron) of said endogenous DNA sequence,

- wherein said mutation effects a reduction in the transcription or expression rate of said endogenous DNA sequence or a reduction in the activity or stability of a protein or polypeptide encoded by said endogenous DNA sequence,

wherein the at least one mutation from i) and/or ii) causes the property of an enhancer of the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) in the identified plant.

- 30 As used herein, a "mutation" refers to a modification at the DNA level, and includes changes in the genetics and/or epigenetics. An alteration in the genetics may include an insertion, a

deletion, an introduction of a stop codon, a base change (e.g. , transition or transversion), or an alteration in splice junctions. These alterations may arise in coding or non-coding regions (e.g. promoter regions, exons, introns or splice junctions) of the endogenous DNA sequence. For example, an alteration in the genetics may be the exchange of at least one nucleobase in
5 the endogenous DNA sequence or in a regulatory sequence of the endogenous DNA sequence. If such a nucleobase exchange takes place in a promoter, for example, this may lead to an altered activity of the promoter, since, for example, cis-regulator elements are modified such that the affinity of a transcription factor to the mutated cis-regulatory elements is altered in comparison to the wild-type promoter, so that the activity of the promoter with the mutated
10 cis-regulatory elements is increased or reduced, depending upon whether the transcription factor is a repressor or inductor, or whether the affinity of the transcription factor to the mutated cis-regulatory elements is intensified or weakened. If such a nucleobase exchange occurs, e.g., in an encoding region of the endogenous DNA sequence, this may lead to an amino acid exchange in the encoded protein, which may produce an alteration in the activity
15 or stability of the protein, in comparison to the wild-type protein. An alteration in the epigenetics may take place via an altered methylation pattern of the DNA.

Mutagenesis may be performed in accordance with any of the techniques known in the art. As used herein, “mutagenization” or “mutagenesis” includes both conventional mutagenesis and location-specific mutagenesis or “genome editing” or “gene editing”. In conventional
20 mutagenesis, modification at the DNA level is not produced in a targeted manner. The plant cell or the plant is exposed to mutagenic conditions, such as TILLING, via UV light exposure or the use of chemical substances (Till et al., 2004). An additional method of random mutagenesis is mutagenesis with the aid of a transposon. Location-specific mutagenesis enables the introduction of modification at the DNA level in a target-oriented manner at predefined
25 locations in the DNA. For example, TALENS, meganucleases, homing endonucleases, zinc finger nucleases, or a CRISPR/Cas System as further described herein may be used for this.

In embodiments, the identification of a plant in step b) takes place with the aid of molecular markers or probes. Such markers may be based upon an SNP and may be specific to the mutation and include, for example, KASP and TaqMan markers. DNA probes are primers or
30 primer pairs which may be used in a PCR reaction. For example, TILLING mutants may be verified or identified by sequencing the target gene in a TILLING population, or via additional methods that verify the mispairings in the DNA, e.g., melting point analyses or use of mispairing-specific nucleases. Mutants generated by means of transposons may also be verified by use of transposon-specific primers and target gene-specific primers in PCR, across

the entire population and subsequent sequencing of PCR products. For example, a change in the expression rate in pollen may be determined with RT-PCR; the change in the stability may be determined by examining ubiquitin binding locations and prediction of changes to the tertiary structure, for example. Furthermore, recombinant expression of the wild-type proteins, and the corresponding mutant proteins and subsequent biochemical activity tests, are also suitable.

In certain embodiments the invention relates to a molecular marker which is able to hybridize to nucleic acid molecule as described above, or a molecular marker which is able to detect a polymorphism genetically associated with or closely linked to nucleic acid molecule as defined above. In particular embodiment, such polymorphism is a SNP, deletion or insertion, exemplary polymorphisms are given in table 2.

Table 2: Polymorphic traits/polymorphisms for detection of gene enhancing haploid induction capability.

Marker	HIR8 locus carrying enhancer for haploid induction	HIR8 locus not carrying enhancer for haploid induction
Aquaporin01	A	G
Aquaporin02	T	T
Aquaporin03	G	G
Aquaporin04	G	G
Aquaporin05	C	C
Aquaporin06	deletion	GGCCCGGCC
Aquaporin07	SEQ ID NO: 65 GTGCATGTTGTATATTGCTAATGTAATTAAATCCAAGGAGTTCATAG TTTAAATTAGCTCTTGAGCATGATATCAATAACTATGGTATTAGAACT ATTAACAACATAAGCATTAGATGATGCAATATGTCTATTATTATTATG ACTAGCGACACCAACATGAGAATAAATAGCACTTGCAAAATTAATTAT TAGAGTCATAAGATGTACTTGCAAGATATACCTTGAGAAAAAATTGTA GTAGTCTTGTGGTTCAACATCTTGGTACTCCATGATCTATAAACTTAC CATCTTGTGAATTTCTCGATGACAAGAGCAACGCTTTGCAAAAACCG CAAGCAAAAGCACAATGGGTCAAACTAGACGCAATCGGCCTTTTCGG CCAAATGTAATAGTCATAGAAAGCATAAAGATAGATCAACCAAAAATA ACTATAGGGTAGATAGATTTAGCTACAAAGTAGAGCCAAAGCAATCGT GTACTCCCTCCGTTTCGTTTTAGTTATTGCTGGATAGTGCAAATTTGA ACTATCCAGCGACAACATAAAAGAAACGGAGGGAGTACTAGTGTAGTC ACCAATAAATGTGTTGACGCCAAATTTGGATCACACACCAATCAGGGCT AGATCACTTGGGCTACCATAGATCAAAAGAACAACAAGGTCGGTTA TCAGACTTATATAGTGAAGCACAAAGAAGTTTATGAGCAAGCCACCAA AGAATAATATAGCGCTTACATGGCAAAGAACAACACTAGTTTTTAGGCAA AGTAGAAAAGAAAATTGTGAAAGCTCTCACCGTTGACGCACCCAATGA GGGCAACTACGTCATCGAATTATCCTAGGTTAATCAGCAAGCATAGGA TGGCATCCTTCATCATTTAGATCAAGTGATGAACAACACAAGGTTGAAA GAGGGCAAAAAGTAGTAGATGTAAACAGACTTATTGGATGTTGGAT GTATCAATCGACTGTGATCCGCTCCTTTACATAGCTGACATCGGTCTT ATCCTAGTAGGAAATAAGCCTACATCATAATCTCCAAGTTTATCATGT	SEQ ID NO: 66 CCTAACCCCAATA TGATAACTCATG CGCCGTTCCCTGA GCGGCATAATTT TTTACTTTAGTT CTTCTGATTTGG TCATGATTTATA TTCGGAATTTAA ACTAAAAATTGT CTAATTATTCAA CAATCCAAAAAC CTTATTGTAGAC AATTTCCTAGTT TTTATATGGCTG CTTTTGCCGACA CGCTGAAGCCAG AAAAGTTCAATG GTATGCACTTTA AGAGGTGGCAAG TCAAGGCCACGC TCTGGCTTACTG CTATGAATGTGT

CAAAAATATATCTAAACACATTTAGAAACTAATCGATCCGGAATAGAT ATAAAACTAGTCTAGCTGGGTCCCAAACCGACTTTCCCGGTTTTTCACG GGAGCTGGTCAGCCACTTAAGAAGTAAAACCAGCCTAATCGGTTATAT TGCGAGTCCAACATAGAAATTCTTTTCGAAAACATATTTATTTCTAAGC TCCACATAAGGCAATCCATATATGTATTTTGATCATCTCGATGAGAAA AACGCAACATTAAGGCTCATTTCTTTTCATTCAAAGAAGTTATAAAATCC AACTAGAGCGGTTATAAGGGTCTGTTTGGAATGACTCTACTCCACCTT GGAGCAGCTATGCTCCAAACTCTAGGGTGCAGCAACTCCACTGTGGA GTTTTAGTGGTGTGTTGATAGAAAATGGTTCATAGCTCCATGAACAAGG TTTTACAAGCGAAAGTACATCCTTACCCACGGTTGCCGGTGGAGGGAC GAGGAGGCACAACAGTGAGCGAAAAAAGGATGAGGCATGATGGCCCCGA GGGAGGAGGAGACACGTGAGAGGAACCGTGCGGCCATAGACCGAGAAG GTGCAGAAATGGATAAGAGAAAAGGTTCTAGAACAGAGAAAAGAACAAACC TTTTTGTGTAATGAGTGGCAGTGATGGGTAAATACTTGTTGTGCGGCAT TTCGAGACCGGGGGGTCCCTAAGCCGACGAGTGAATGTGCGCCGCGTG TCCAGCCCAGATGGGTGACGCGAGGCCGAGCGCGAAGGGG- AAGTGAGGTGGCCGGAGACAGGCGTGAGAGAGGTGGAATCCCGCGGC CTTCTGTGTTCTGTCGCCGCGCCAGGTGGGTGCGCTTGCAGTAGGGGGT TACAAGCGTCCACGCGGGAGAGGGAGCGAGCGGCCTCACGCGAGCGCT GTCTCCTCCTCGTCCCCACGCGGCCAACCTTCTCTAAGAGGGCCCTG TCCCTTCCCTTTTATAGGCGTAAGGAGAGGATCCAGGTGTACAATGGGG GGTATAGCAGAGTGCTACGTGTCTAGCGGAGGAGAGCTAGCGCCCTAA GTACATGCCGTGCTGGCAGCCGGAGAGATTTTGACACCCAGCTGGTGT GATGTGCTGGCCGTGCGAGGAGTGTTGGAGCCTGGCGGAGGGACAAC GTCGGAGCTGTTGAGTCTTGCTGACGTCTTGTCTCCGTAAGGGG GCTGAGAGCCGCCGTGCTCACAGAGTATGCGGGGCGCCATCATTGCCT TTCTGGCGGAGCGAGCCAGATGGGACGCCGGTCTTGTTCCCATGGCC CGAGTCAGCTTGGGGTAGGGTGATGATGGCGTCTCCTGTTGACGTGGC TGGTCTGCGCCCTAGGTTGGGTGATGTGGAAGCTCCTCCGAAGCCGAG GTCGAGTCTGTCTTCCGTGGCCGAGGTCGAGTCCGAGCCCCTGGTTCG GGCTAGGCGGAGACCGTGGGTGAGGCCAGGGCGGAGTTGAGGCCCTG GGGTCGGGCGGAGCGGAGTTTGTGCTCTTCCGGGGCTGAGCCCAAGTC CGAGCCCTGGGTGCGGCGAAGCGGAGTTCGCCGTCTTCCGGGGCTTAG CCCGAGTCCGAGCCCTGGGTGCGGCGGAGCGGAGTTCGCCGTCTTCC GGGCTTAGCCCGAGTCCGAGCCCTGGGTGCGGCGGAGCGGAGTTCGCC GTCTTCCGGGGCTTAGCCCGAGTCCGAGCCCTGGGTGCGGCGGAGCGG AGTTCGCCGTCTTCCGGGGCTTAGCCCGAGTCCGAGCCCTGGGTGCGG CGGAGCGGAGTTCGCCGTCTTCCGGGGCTTAGCCCGAGTCCGAGCCCT GGGTCGGGCGGAGCAGAGTTCGCCGTCTTCCGGGGCTTAGCCCGAGTC CGAGCCCTGGGTGCGGCGAAGCGGAGCTTCTATGGTGCCTGCGGCCG GGCTGACTGCCTGTCAGCCTCACTCTGTCAAGTGGCACCGCAGTCGG AGCGGCGCAGGCGGCGCTATCTTCTGTGAGCCGGTCAGTGGAGCGG CGGAGTGACGGCGGTCACTTCGGCTCTGTGCGCTGGGGGCGCGCGTC AGGATAAAGGTGTCAGGCCACCTTTGCATTAAATGCTCCTGCAATTTG GTCGGTCGGTGCGGCGATTTGGTCAGGGTTGCTTCTTGGCGAAGACAG GGCTTCGGGCGAGCCGGAATATGTTGCCCGCTGGAGGGGG- CCTCGGGCGAGACGGAGATCCTCCGGGGTGGGTGCGCTGCCCTTGTCCGAGG CTAGGCTCGGGCGAGGCGTGATCGAATCCCTCGAATGGACTGATTCCCT GACTTAATCGCACCCATCAGGCCCTTGCAGCTTTATGCTGATGGGGGT TACTAGCTGAGAAATTAGGAGCCTTGAGGGTACCCCTAATTATGGTCCC CGACAGTAGCCCCCGAGCCTCGAAGGGAGTGTTAGCACTCGCTTGGAG GCTTTCCGCCCACTTTTTTGAAGGGGACAGCCTTTCTCGGTTGCGT TTTGTTCCGGTGGGTGCGCGGAGCGCACCCGCCGGGTGTAGCCCCCG AGGCTTCGGAGGAGTGGTTTCACTCCTCCGAGGCTTAATGCCTCGCG TAATGCTTCGGCTGGTCTGGTCATCCCTCATGCGAGCTGGCCGTAGCC CGGTTGTACGGTGGGTCCCAAGTTCTCGGGCTGGTATGTTGACGCTG TCAACGGTTCCGGCCGAGCCGGGTTTGCGAGAGCAGCCCCGAGCCTC TGCACAGGGCGAGAGGGCGATCAGGGACAGACTTGACTTTTTTACATA CGCCCCGTGCGTGCCTTTCCGCAAGGAGGAGGGGGGAAAGCGCCATGT TGCCCTCGATGGGCGCCGAACATGGTGTCTCCGGTGAGCTGCAAGCGG GTAATCCGAGTGGACGTCCGTGCCCCGTTCTGTTAGGGGTGCGCTAGGG	TCCATGTTAGTA AGGCAAACCTGA GGGCGCACTGAC TTCTGAACAGGA GAAAGAGTACGA TCATGCCAATAC TATCTTTACGGG AGCAGTTCTTAG CGCCCTTGTTGA CCGTCTGGTTGA TGCGAATATTCA GCACACAGACGG GAAAGAGTTGTG GGATGCACTTAC TACCAAGTATGG TACATCAGATGC TGGCAGTGACCT GTATATCATGGA GAGCTTTTATGG TTATAAGATGGC TGATAATCGCTC TATTGTAGAGCA AGCTCATGAAAT ACAGTGTATAGC CAAGGAGCTCGA CCACCTTAAGAT AGTCCTTCCTGA CCAATTTGTGGC TGGATGCATTAT TGCAAAGTTGCC TTCTACATGGAG GAACTTCGCCAC ATCTCTGAAACA TAAGAGATAGGA GATATCAGTTGA AATTCTGATAGC GTCTCTGGATGT TGAGTAGAAAGC TCGAGCTAAGGA TACGGGTTCTAA AGGAGGCGAGGG CCACTCCAGCGC CAACATGGTTCA AAGGAACCACAA CAAGGGCAAAGG AAAGCCAAAATC TAACAAGCCAAA CAAAACATCAA CTTCAAGAAGAA GAAGAACAAGGC TGAACGACATG TTCCGCATGTGG CGAGACGGGTCA TTTTGCTAAGAA TTGCTCTGATCG AGCGGATTGTCA TGGCAAACAGGG AAATGTCAACAC AGTGATCGCTAG CAATGAGGAAGA CAAAGGGTATGA
--	---

GCCCAGAGGCGCGCCCAAAAGTACCTGCGGGTGATCTGCCGGACCCGG TCCCCTGGCGACGGGGTCCGAGGGCTCGATGCCTCCCTCTGATGGGAT TCCGTTACAAGATCGTTCCCGCTGGTCTCGGAAATGTCTAGGGTACC TCGGGAGCGTAGCCCGAGCCTTGGTTATGTATCGAACGTACCCATGGT CATCCCTCGCTCTGTGTCTGAGGCGGGCTGTGAACCCCTTCGGGGGCCAG CCTTCGAACCCCTGATCAGTAGTGGGCGTGGAGCCCGAGTAGCCTGAG GCGGCCGTGGAACCCCTTCGAGGGGCCGGCCTTCGAACCTCTGACCAG TAGTGGGTGTAGAGCCACGCGATCTGAGGCGGCTGTGAACCCCTTCC GAGGGGCCAGCCTTCGAACCTCTGATCAGTAGGGAGGCTCGGAGCCTG GTTCTTACAGAGGAAGGATCCTTTTCGGGGTATCCCTTTCCCGGTC CCTGTTGCAAGAGAGAGAAAGAGGAAAAAAGGAAAAGGATACGAAATC GAACGACGCGGCGTACCTTTTTTACGCGGCCATTATGGCGAAGGCGA AGCGTCGCCCCTTCTCCTGCCAGAGGCGCCGCTGTCCCGCCGCGGA GTTAATGCGACGGGGCGAGTGGTTCGCGAGGCGGCTGTTGCGCGTGCG CGAGCCGTTTCGAGGAACGGAACACGGGCGCGTTGTCTTCACGCCGTGA GAGAGGGTTCTCTCGTCGCCCCCTAGATGGGACGTAGGCTTGGCTGACG ACGTGACCGCTGCTCCTGCCCGCCTGCCACCGCCATTACTGCCGGCCC ATTTTTGGCCGTATTGACCGTCGCGCCAGGCTGGCGCTGCTGGGTGCT GCGCTGGGTGCGCTCGAGTCGCGGTATTGGTTCGCGAGTCGAGGAGCG GCGTGGTGGCGCAAGTGGCGGTGCGGCTGCATGCACGAAGCATTGCG TCGCGCGGTTGCATGACGCGTGGGCGCTGGGCGCTCCATGCCGGGCGTGT TGGGAGTCGGAGAAGTGCGCCCACTTGGCGCGGTTGCATGCCGCTGC ATGGCTGCCCGCCCTTTTCGCCCGCTGGTCTGGGCAAAAGTGGAGAGTC GCTTGTAACCGCTGGGTGGTTGTACGCACCGCGCACCGCGGTTTGGCT TCTTCCGCTCTGAGCCGGCTTGCATGACGTGTGGGACCCAGCCCCGCG GCCGTAGGGGAAGACCTTGGAGCGTGTGGAGAAGACTCAGCCCGTGA CGGCTGGGGACGCAAGTAGGGAGAGTCGCGCTTTAAAAGGAGGGTGACC CCCTTGGAAGGCGACCATGTCTTCGCGCTCCCTTATGCATCGTGTCTT TCTACCTTCCAAGCCCCCGGATGGGGGATTCCCGCCGTCTTTCGCGTT CGTCGTTGGAGGAACGCAACTCCGCGGGAGTTGGTACCTTTCAACCAT CGTTTGGCTTCAAGGATTTTCATCATACAGCCCGGTTGCACCCCCCCC CCCGCCGGTGGTCACCCAAAACGGTGACCTCCAGCCCTGGATGGGGA GAGGCGAACCGGGCTGTGATCTTGGTCCCGCCCTCAGCCTCGAGGATG TTCGTATCCTTGTCTGGGGCGGGGGCGAGGCGAGCCGGGGCTCTACC TCCCGCGCGGGCCGGCTGGCCACCTCTTCTTCCAGCTTCTGGTGGTGG CAACCATCCTTCAGCTCTGTAGAGGAGAGCTCCTCCAGCCATGCCGGG GAAGGCGAACTGTTGCTGCCCACTAGGATGCAACATTCCGCCCTCTT CCTCGCTTTCGTGGCGAGGACGGGAGCGAGGACCTGCCGGTGCATTTT GGAGCGGCCCGCTCTTTGGCTTTAATAGCTGGTTCGCGTCCCTTGAGC GGGAACGCGAACGAGAGCCCTCCGGCGGGCGTGTCCGTCTTGGGACCA TGGCTGCTGCTGCAGAGGTGCTGGCGGGTCCCCCGGTGCTTGTGCGC CCGCAGGCTCCCCGGTGGCGAGGTTGTTCATACCTGTGGGGACGGAAC CAGAGTTCGCTGTGATGGCACTTCGAATGCCAGTGTCTTTTGTTCAT TGTGGCTGTGCGGGCCCTGAACATGTATGTAATTTCCGACGGAGCCGT GTTTTTTCTCATTTTCGAGCACTAAGACTCGCCTGTTGGTTGTCTGA ACCGCTTACCAAGCGTGAGTCGCCCCGTGTAAAGGTGACGAGTGAGG TATCCGTATCCCGGAGGCGTAGGAGTCCCTCGGCTCGGTGAGCCTTGT TGTCCGAGGCTTCTCTTGCTCAGTTAAAGGAACTCCTCGGCCGCTCTT CGATGAGCCGAGGCCAGGGGTAGCGGTGTGAGCGCGAACAGAGGCAGA GTTGGCTCGAGAGGAAGACCTGGTCGGCCGAGCCTGGCCGGGTGCTC CGTTAGCGGGACTGACGCTGGAATTGACCAGCCGAGGCCTCGGGTCGG GCTGACGTCTTCGGAGGACAACTGGCCGAGGCCCGGGGTGACCGGTC GAGCCGCTGCTCGGGCCGGATTCCCGGAGAAGACCTGGCGGCGACG GCCCAGGCTTGGTGATGACGCCGTCTTCCAGCGGAGGTCTTCGGAC CGCGTGGCGTCCGAGGCTGGGTGCGACTTCGCCGAAGGTGTTGTGCA CGCCGAGGGTGTGCTGCTCCCTTCAAGCGTCAAGATCCGAGCCTGCA GGATCGGATCGTCTTGTAGCGTGTGTTTCTACGGCCGCGGAGGCCAA AACACACCCCTCGCCGTGTTGTGAAGCTGCGTCTCTTTTCTCTTGCTT CGAGTACCCGGACTTTTTTGTGCGTAACAGAATTGTTTGTGCGAGCGA GAGCTGCTTCTCACGGAAGGTGATGAGTGAGGTATCCGTATCCCGGAG GCGTAGGAGTCCCTCGGCTCGGTGAGCCTTGCCGCTTACGCGCACTCT	TAATTTACCTTT CATCTTCTCAGT ATTTCAATCACC TAGTTGGTGGCT TGATACTGGTGC TAATGTTTATGT GTGTTCTGACAT TAATTTGTCTC TTCTTATCAGGG TGCTCGGGATTG TTCCGTCCTAAT GGGGAATGGGTG ACATGCTTCTGT TCATGGCACTGG CACGGTGGATCT GAAGTTTACTTC GGAAAAGATCGT GCAACATGTCCC TTCTATACAAA GAATCTCGTTAG CGGAACCCCTTCT ATGTAGAGATGG GTTCAAGGTAGT TTTGGAGTCCAA TAAATTAGTTGT GTCTAAGTATGG ACAATTTATTGG TAAAGTCTATGA TTGCGGAGGCTT GTTCCGCTTTTC TCTGATAGATTT CAATAATAAGTC TGTGAACCATAT TTGTGCTAATGT TGATGATCTTGC GAGTATTTGGCA TTCTCGTTTTGTG TCATATTAATTT TGGCTCTTTGTC TCGGCTTTCAAC CATGAGTTTAAT TCCGAATATCAC CATAGTCAAAGG TTCTAAGTGCCA TAGTTGTGTGCA GTCGAAGCAACC TCGAAAGTCTCA CAAGGCTGCTGA GGAGAGACACCT GGCACCTCTAGA ACTCATACATTC TGATCTTTGTGA GATGAATGGTGT GTTGAAAAAAGG TGGCAAGAGATA CTTCTATGACATT GATTGATGATGC GTCTAGATTTTTT GCTATGTATACT TTCTAAAAACTA AAGATGAGGCTT
---	---

TACCCGTCCATGGGGTTCTGTACCCGACACAGTCAAGAAGGCTCGAAG AAATCGCTCCGGCAGAAGAGCTTCCGAACGTGAAGACTTGTTCCGGTCCG CGGAATCACTTATCCGAACGTGAGTTACTTATCGCAGAAGGTGATGAG TGAGGTATCCGTATCCCGGAGGCGTAGGAGTCCCTCGGCTCGGTGAGC CTTGCGTGTCTTACGTGTACTCCGCCGTTTTTCAGGATCCACTTTTCAAG TAGTCAAAAAGCACGAAAGACGTTCTGGCAGAAGGGATCTTTTTTTCGA GGAAAAAATTCGACGCAGAGGGGTTCCCCC- TTTTAGCCCCCGAGGGAGGGTCAAGGCTTGTCCGAGGCGAGGCCGACCC TTCCCTTGATGACTAAACTTTGCGTGGGAGTGAGGTATATGAACAACTT GAAAACATCTTAAGGGTAGAAGCGACGTAGCTGTTGGATGTTCCAAGC GTTGCTGTAGACCTCGCCTTGACTGTTGGCCAGCTTGACGTTCCGGG CTTCAGAACCTTTGGCGATGACGAATGGCCCCCTCCAGGGGGGCGTGAG CTTGTCGCCGCCCTCGGGCGTCTTGTCGCAGCCGAAGCACCAGGTGCGC CACCTGGAGGTCTCGGGACCGGACCCCTCGGGCGTGGTAGCGTCGCAG GGACTGCTGGTACCGCGCCGAGTGTAGTAAGGCCCTTGTCGCCGAGCCTC TTCCAGCTGGTCCAGCGAGTCTTCTCGACTAGCTTGCTTGCTTTGGCT GGCATAGGCCCTTGTCCTCGGGGAGCCGTATTCAGGTCTGTGGGCAA GATGGCTCGGCCCATAGACTAGAAAGTACGGCGTGAAGCCCGTGGC TCGGCTCGGCGTTGTCTCAGGCTCCAAACCACCAAGGGGAGTTCTCT CATCCATCGCTTGCCGAACCTTGTTGAGGTTGTTGTAGATCCGAGGCTT GAGCCCTTGTAGAATCATGCCGTTGGCAGCCTCTACTTGCCCATTTGA CATGGGGTGAGCTACGGCGGCCAGTCCACCCGATGTGGTGATCCTC ACAGAAGTCCAGGAACCTTCTGTGCGTGAAGTGGGTGCCGTTGTGCGT GATGATGGAGTTCGGGACCCGTGAATCGATGGATGATGTTGGTGAAGAA CGCCACCGCCTGCTCGGACCTGATGCTGTTTAGGGGTGCGACCTCGAT CCACTTGGAGAATTTGTGATGGCGACCAACAGGTGCGAGTAGCCCCC GGGTGCCCTTCTGCAAGGGGCCGACGAGGTCCAGACCCCATACAGCAAA AGGCTAGGTGATGGGTATCGTCTGCAGAGCCTGAGCGGGCAGGTGGGT CTGCCCTGCGTAGAACTGACACCCCTTCGCAGGCGCGGACAATTCTAGT GGCGTCGGCCACCGCCGTGGGCCAGTAGAAGCCTTGTCGGAAGCATT CCCGACAAGGGGTGAGGGCGCTGCGTGATGGCCGCAAGCCCCGAGTG TATCTCTCGCAGAAGTTCCTGACCTTCGGCGATGGAGATGCATCGCTG GAGGATGCCGGAGGGGCTGCGGTGGTAGAGCTCCTTCTCATCTCCCAG CAGGACGAACGACTTGGCGCGCCGCGCCACCCGCCAAGCTTCGGCTCG GTCGGAGGGTAGCTCTCCTTGGTGGAGATATTGCAGGTACGGGGTCTG CCAGTTTTGATCAGGCGTGACCCGCTTTGCTCCCCCTCGACGCGCAG CGTCTCGCCCTCGGGGGCCGAGGGTACCTCGAGCCGAACCGAGGGTAC CTCGGGCCGAGCCGAGGGTGCCCTCGGGCTGTGCCGAGGGTACCTCGGG CTGGGCCGAGGGTGCCCTCAGGCTCGGGCGTGTGCCGATCTTGACGAG GGTTGATGCAAGTCTCGGGAGAAGACGTCCGGGGGAATCGTTGTTGCG CCCGAGGCTATTTTAGCCAGCTCATCCGAGTCTCGTTGTAGCGTCCG GCGATGTGGTTGAGCTCGAGCCCGTAGAACTTGCTTCCAGGCGCCGA ACCTCATCGCAGTAGGCCCTCCATCTTCGGGTGCGGCGAGTGGGAGTTC TTCATGACTTGGTGCATGACGAGCTGCGAGTACCACGAGCGTCGAGG CGTCAGACCCCTAGCTCGATGGCGATTTGCAACCCGTTGACCAGAGCC TCATACTCGGCCACATTGTTGGACGCCGGGAAATGGAGGCGTAGCACA TAGCGTAGGTGCTTCCCGAGGGGCGAGATGAAGAGTAGGCCTACGCCG GCTCCTATCTTCATCAGCGACCCGTGGAAGAATGAGTCCAGAGCTCC GGCTGGATCGGAGCTGTTGGGAGCTGAGTGTGACCCATTACAGCTACG AAGTCCGCCAAGACCTGGGACTTGATGGCCTTCCGAGGACCGAACGAG ATTGCTTCGCCCATGATTTCCACCGCCTTTCGATTTCTACCCGAGGCC TCTCGGCACTGGATGATCTCCCCAGGGGGAAGGACGACACCATAGTT ACCGGATGAGACTCGAAGTAGTGTGCGAGCTTCCGCCGTGTCAGGATC ACCGCATACAACAGCTTCTAGACTTGTTGGATAGCGGATCTTGGTTTTG GACGTACCTCGCTGACGAAGTAGACTGGCCTCTGAACGAGCAATGCA TGCCCTCTTCTTGCCCTCTGACCAACAATCGCGGCGCTAACCACCTGA GTGGTGCAGGCGACGTAGACCAAGAGGGCTTCTCCGGCAGCTGGGGGC ACCAAGATAGGCGCCTTCGCGAGGAGCGCCTTCAGGTTCCCGAGGGCT TCCTCGGCCCTCAGGGGTCCAAGTGAAGCACTCAGCCTTCTTAAAGAG CGGTACAGAGGTAGACCTCTTTCGCCGAGGCGTGAGATGAAGCGGCTC AGAGCCGCGAGACATCCCATGACCCTCTGCACGCCTTCAAGTCCTTG	TAGACTACTTTTA AAATCTATAAGG CTGAAGTTGAAA ACCAACTAGAGA GAAAGATCAAAC GTCTTAGATCAG ATCGTGGTGGCG AATTCCTTTCCCA AAGCCTTTGATG ATTTTTGTGCAG AACATGGTATTA TTCATGAGAGGA CGCCTCCCTATT CACCCGAATCAA ACGGGATTTGTTG AAAGGAAAAATC GTACGTTGACTG ATCTGGTGAATG CCATGTTAGACA CTTGTGGTTTTA TCTAAGGCATGG TGGGAGGAGGCA GTCCTGACTTCA TGTCATGTTCTG AATAGAATTTCT ATGGGCAAAGAA GAGAAAACCCCT TATGAGAAGTGG GTTGGGAGAAAA CCATCACTTTCA TACTTGCGCACT TGGGGTTGCATG GCGAAAGTCAAT GTACCAATTAAT AAAAAGCGCAAG CTTGGTCCAAGG ACAGTGGATTGT GTCTTCTTTGGA TATGCTTTGTGT AGCATAGCATAT AGATTTTTTAGTA GTTAAATCTGAA GTTTCTGATGTG TATGTTGATACT ATTATGGAATCA CGTGATGCTACT TTCTTTGAACAC ATATTTCTAATG AAAGACATTCAT AGCAATCTAGA TATCTTCTGAG ATAACTTCTGAA CATAGTACACCT GTTGAGATTTTT GAACGCCACAT GAAATTGTCTTA AAGGAGGATGAC AATGATGCTCCT AAAAGGAGCAAG AGACAAAGGGTT GAAAAATCCTTT
--	--

ATGGGCCCCATGCTGGTGATGGCTGCGATCTTCTCCGGGTTGGCTTCG ATGCCCCGCTCGGAGACGATGAACCCCAAGAGCATGCCTCGGGGCACC CCGAAGACACACTTCTCGGGATTGAGCTTGACGCCTTTCGCCTTGAGA CATCGGAATGTCACTTCAAGGTCGGAAAGGAGGTCGGAGGCTTTCCTC GTCTTGACTACGATGTCATCGACGTAGGCCTCGACCGTGCGGCCAATG TGTTTCGCCGAACACATGGTTCATGCACCGCTGGTACGTGCGACCCGCA TTCCTCAAACCGAACGGCATGGTGACATAGCAGTACATGCCGAAGGGT GTGATGAAAGAAGTCACGAGCTGGTCGGACTCTTTCATCCTGATTTGG TGATAACCTGAGTAGGCATCGAGGAAAGACAGGGTTTCGCACCCAGCA GTGGAATCCACGATTTGATCGATGCGAGGCAGAGGGTAGGGAACCTTC GGACATGCTTTGTTGAGACCAGTGTAGTCTACACACATCCGCCATTTCC CCCC- TTTCTTTCTCACAAAGCACAGGGTTGGCAAGCCATTTCGGGATGGAATAC CTCTTTGATGAACTCTGCCGCCATTAGCTTGTGGATCTCCTCGCCTAT GGCTCTGCGCTTCTCCTCGTCGAACCAGCACAGAGGCTGCTTGACGGG TCGGGCTCCGGCTCGGATGTCCAGCGAGTGTCTCGGCGACATCCCTCGG TATGCCGGGCATGTCCGAGGGACTCCACGCGAAGACGTGCGCGTTTCGC GCGGAGAAAGTCGACGAGCACTGCTTCCTATTTGGGATCGAGCCCGGA GCCGATCCGGATCTGCTTGGAGGCGTCGCCGCTGAGGTGAGGGGGACG GACCTTAACCGTCTCCGCTTGTCTCAAAGTTGCCGGCATGACGCTTAC GTCTGGCACCTCCTTAGAGAGGCTCTCTAGGTGCGCGATGAGGGCCTC GGACTCGGCGAGGGCCTCGGCGTACTCCACGCACTCCACGTGCGATTTC GAACGCGTGTTTGTACGTGGGGCCGACGGTGATGACCCCGTTGGGGCC CGACATCTTGAGCTTCAGGTAGGTGTAGTTGGGGACGGCCATGAACTT CGCGTAGCCTGGTCTTCCCAGTACCGCGTGGTAGGTTCTTCGGAACCC GACCACCTCGAACGTGAGGGTCTCCCTTCGGAAGTTGGAGGGTGTTCC GAAGCAGACGGGAAGGTGAGTTGTCCGAGGGGCTGGACGCGCTTCCC GGGAATGATCCCATGGAAGGGCGCAGCGCTGCCCGGACAGAGGACAG ATCGACACGCAGGAGCCCCGAGGGTCTCGGCGTAGATGATGTTGAGGCT GCTGCCTCCGTCCATGAGGACCTTGGTGAGCCTGATGTGCGCGATGAC GGGGTCGACGACGAGCGGGTATTTCCCCGGGCTCGGCACATGGTCGGG GTGGTCGGCTTGGTCGAAAGTGATGGGCTTGTGCGACCACTAGGTA GACTGGCGCCGCCACCTTCACCGAACAGACCTCCCGGCGCTCTTGCTT GTGGTGCCGAGCCGAGGCATCCGCCGCTTGCCACCGTAGATCATGAA GCAGTCGCGGACCTCGGGGAACCTCCTGCTTGGTGATCTTCTTCTT GTCGTGTCGCGGGCCTTGCCACCCCTCCGCGGGTGGCCGACCCTGTG GAAGTGGCGCCGAAGCATGACGCACTCCTCAAGGGTGTGCTTGACGGG CCCCCTTTCGGGGGGTTTCCGAGGGTTCTTGTACTCGGCGGCGGCGAC AAGGTCCGCGTCGCGGCGCTCGCGTTTTCGCTTGCGACTTCTTCTTGCC CTTCTTCTTGCGCGCGAGCTGAGTTGACGCCTCGGGGGCATCTTCCGA TGGGCGACCCCTGGGGCTGCTTGTCTTCTGGAAGATAGCCTCGACCGC CTCCTGGCCGGAGGCGAACTTGGTGGCGATGTCCATCAACTCGCTCGC CCTGGTGGGGGTCTTGCGACCCAGCTTGCTCACCAGGTGCGGGCAGGT GGTGCCAGCGAGGAACGCGCCGATGACATCCGAGTCGGTGATGTTGGG CAGCTCAGTGCCTGCTTCGAGAATCGCCGGATGTAGTCCCGGAGAGA CTCTCCCGGTGCTGTGCGGAGCTTCGGAGGTCCAGAAATTCCCGGG GCGCACGTACGTGCCATGGAATTGCCGGCGAAAGCTTGGACTAGGTTC ATCCCAGTTGGAGATCTGCCTCGGAGGCAGGTGCTCCAACCAGGCGCG GGCGGTGTGCGAGAGAAACAGGGGGAGGTTACGGATGATGAGGTTGTC ATCGTCCGTTCCACCTAGTTGGCAGGCCAGACGGTAGTCCGCGAGCCA CAGTTCCGGTCTCGTCTCCCCCGAGTACTTTGTGATAGTAGTCGGGGG TCAGAACCAGGGTCGGGAACGGCGCCCGTCTGATGGCCCGGCTGAAGGC CTGCGGACCGAGTGGTTCGGGCGAAGGACTCCGATCCTCCCCGCTGTC GTAGCGTCCCCACGCCGTGGGGTGGTAGCCTCGGTGCACCCCTCTCGT GAGGTGGGCTCGACGGTTCGCGGTGATGGTGCTCGTTGCCGAGGCGACC CGGGGCCGAGGCGCTGTGTTGCGCGTGCGCCCGGTGTAGACCGAGGC TTCCCGCATGAATCGGGAAGTCGCGGCATGAGGTTCTGAGGGGTAACC CTGCCCTTCGGGAGGCGGAGCTCTCGGCCCGTCGGACCGCGGCGCCTTC CAGGAGATTCTTGAGCTCTCCCTGGATTGCGCGCCCTCGGTGGTTGA TGGCTCCGGCATCGCGCGGAGGAGCATGCTGCTGCAGCCAGGTTCTG	GGTGATGATTTTC ATTGTGTACCTT TGAGAGCACCTA GAGGGGGGGTGA ATAGGTGATCCT GTAAATCAAAC ACTAAATAGCCA CAAACTTAGTT ATTAAGGTGTCA GCGCGACTAAGT GGTTTTGAAGCG AGTTCCTTGTGGA TAAAACAATCAC AGAGAAGGCAAC ACAAGAGACACG CGATTTTTATCC CGTGGTTCGGCC AAGTAACACTTG CCTACTTCCACG TTGTGGTGTCCT AATGGACATGGG TTGCACTCAACC CCTTTCAGCGGA TCCAATGATCAA CTTGAATACCAC GGCTTCTTTCT TATAGTCTTTTT CCCGTTTTCGAG GAATATCCACAA TTTGGAGTCTCT TGCCTTTACAAT TGAGATCACAAA GAAGGCACAGAG TAAGGTCTTCCA TACTAGAGAATT GGTAGCTGAATT GTTCTCTAGAT CATCTATATGTT GTGCTGCAACTT GGGGGAGTGGCT CAGGAGGCCATT CCTCCCGCCCAT CTGCATCTTGGT CAACAACATTGT TATTACAAGGG ATTGTTTAGGTC AAACCCTGCATT TAAAGTCGGTTG CCCTAGACCTAT AAAATGAAAAGA CAACTCATGTCC ATTACATGATG GGGAGGGATCAG ATCCTCATCTTT AGGGAGTGCACC CAATAGCTCTTG TTCAATTACTTC ACATTGGATTGT CTAAGAGAATCC CTGAAAACCCCT CCGAAACACAGA
--	---

GCCAACCCCACTGGATGCGGGTGGCGGCCTGACCCTGACATCATCGAC GATGCGGTGCTGGAAGCCCTGGGGTAGATGACGTATTTTTCCAGCCGG AGGTTGAGCTGCCATGCCCTGCCGACGTCCCGGCGGATCGGCTCAAG CGCTCCTGCTCCCTCGTCGAGCCTGGCCTGCACCCCGCGGATTTGCTC GAGCTGTGGGTGATGGCCCCCGCCTGAACGGAGACCACAGCTAGCTC CCGTGGGATGTCAACGCGAGGACCCGGCCTAGGGAGATCACCGTCCCC CGGCATGCCGAGATGGTTGCCTTCGGAGGGACCCCTAGATCGACGTG GAAACATTCGTGGCTTGGGCCGAGCCCTCGTCGCCGAGGCTACGGCT ACCGTCGGAACAGTCGGAAGGCAGTAGTCGCATGCGGTGATGAAGTC CCGCATGGCACTGGGGTTGCCAAGTCCGGAGAAATCCCAACAGAAGTT GGGCTCGTCATCTTCTCGGACCCAGAGGGCCCGTAGGTCGAGACGTC CGTCAGCCGGTCCCAAGGCGACTGCATACGAAACCCAGAGGGTTTGG ACTCGCCTCTACGAGAGCGCCCGCCAAAGCGAAACCGCTAGGCGGGTT GAGGCTGAATCCAAATGACGTGGGATGGGAATCGGTGCGTACCTCTTG GTTGACGGGCGGTGATGAAGTCACGTGCGGGACTGACTGCACCGTCGT CTCAGGTACAAGGTTGACACCCAGCAAGCCTTCCGCGAGCGTGCTGGC GTCGTCCGTTTGTCTCAGGATTGGCGTGTTCGCGGGAGACGGCGCTCGT CTTCGTCTCAAGCGCGAGGTCGATGCCCGGTGCGCCCCCTGTTGGGGT GCTGGCGCTATCGGCTCGCTCGACAGCCGACGAGGCGCTGCCTCTGCG TTGGCCTTGGTTGCCCGCCTTCCCTCCATCGGCGGGGAAGAGGGCG GGATGAGCTTGAAGGTTGTTCTTCCACCACGCGGGGAAGACGTGCTCG ATTCCGCCGCCGCGGGCGGGCTGTGCGCCGCCATTGCCGCTGTGCGC CGGCGGTGGAATGAGTATCATGTTGTAGCTGCCGTGAGGGACATGAA CTCAAGACTCCCGAAACGGAGACCGTCCCGGGTTGGAGAGGTTGCTG GAGACTGCCATCTGGAGCTTGACGGGAAGCTGTTCTGCAACACGCAGC AGGCCCCCTACCTGGCGCGCCAAGTGTGCGCGTTTCGAGACCGGGGGT CCCTAAGCCGACGAGTGAATGTGCGCGCGTGCCCCAGCCCAGATGGGT CGACGCGAGGCCGAGCGCGAAGGGG- AAGTGAGGTGGCCGGAGACAGGCGTGAGAGAGGTGGAATCCTGCGGC CTTCGTGTTCTGTCGCCGCGCCAGGTGCGGTGCGCTTGACGTAGGGGGT TACAAGCGTCCACGCGGGAGAGGGAGCGAGCGGCCTCACGCGAGCGCC TGCTCTCCTCCTCGTCTCGCGCGGCCAACCTTCTCTAAGAGGGCCCTG GTCCTTCTTTTATAGGCGTAAGGAGAGGATCCAGGTGTACAATGGGG GG- TATAGCAGAGTGCTACGTGTCTAGCGGAGGAGAGCTAGCGCCCTAAGT ACATGTCGTGTCGTCAGAGGAGTGTGGAGCCTGGCGGAGGAACAGCTGT TGGAGCTGTTGAGTCTTGTGCTGACGTCTTGTGCTTCCGTAAGGGGGC TGAGAGCCGCGCTGCTCACAGAGTATGCAGGGCGCCATCATTGCCTTT CTGGCGGAGCGAGCCAGATGGGACGCCGCTTGTGTTCCCGTGGCCCG AGTCAGCTCGGGGTAGGGTGATGATGGCGTCTCCTGTTGACGTGGCTG GTCTGCGCCCTAGGTTGGGTGATGTGGAAGCTCCTCCGAAGCCGAGGT CGAGTCTGTCTTCCGTGGCCGAGGTGAGTCCGAGCCCTGGTTCCGG CGAGGCCAAGACCGTCCGCTGAGGCCAGGGCGGAGTCCGAGCCCTGGG GTCGGGCGGAGCGGAGTTCGTGCTTCCGGGGCTGAGCCTAAGTCCG AGCCCTGGGTGCGGCGGAGCAGAGTCCGCCGTCTTCCGGGGCTTAGCC CGAGTCCGAGCCCTGGGTGCGGCGGAGCGGAGTTCGTGCTTCTCCTGG GCTTAGCCCGAGTCCGAGCCCTGGGTGCGGCGGAGCGGAGTTCGCCGT CTTCCAGGGCTTAGCCCGAGTCCGAGCCCTGGGTGCGGCGAAGCGGAG TTCGCCGTCTTCCGGGGCTTAGCCCGAGTCCGAGCCCTGGGTGCGGCG GAGCGGAGTTCGCCGTCTTCCGGGGCTTAGCCCGAGTCCGAGCCCTGG GTCGGGCGGAGCGGAGTTCGCCGTCTTCCGGGGCTTAGCCCGAGTCCG AGCCCTGGGTGCGGCGGAGCGGAGTTCGCTGTCTTCCGGGGCTTAGCC CGAGTCCGAGCCCTGGGTGCGGCGGAGCGGAGTTCGCCGTCTTCCGGG GCTTAGCCCGAGTCCGAGCCCTGGGTGCGGCGAAGCGGAGTTCCTAT GGTGCCTGCGGCCGGGCTGACTGCCTGTGAGCCTCACTCTGTCAAGT GGCACCGCAGTCGGAGCGGCGCAGGCGGCGCTTTCTTTCTGTGAGGCC GGTCAGTGAGCGGCGGAGTGACGGCGGTCACTTCGACTCTGTGCGCT GGGGGGGCGGGTCAGGATAAAGGTGTGAGGCCACCTTTGCATTAAATG CTCCTGCAATTTGGTCGGTCGGTGCGGCGATTGGTCAGGGTTGCTTC TTGGCGAAGACAGAGCCTCGGGCGAGCCGGAATATGTTGCGCGCTGG	AGACTATGAATT GCAGGACCTCTT CTAGAGAAATAA CCCTGGCACGCA CTAGCAGTATCG TGAGATTAGCAT TATCCACTTCCC ACATGATAAGTC CCCAAATTATCC AATAGCATTTCTT TATCTATGTGAT TGGTTGGTTGGC CGGCTGCCTCCA GGCCCGCGACGT CTCGTTGACCTT GCGCAGCGTCTG TTTGATTAACTG GAGAGGAGGGTG GCAGCCTGCACG CAGCATAGCAT TTTTTCGCTTGA TTGCCTGCATGC GCAGAAACACAC CTATCATGCGTA CCCAGCGGTCCA GGTTAGCTGGAG ACCGGCGCTCGA CGCCAATCAACC AATCACGCAGCA TGTGTTCTTGCA AATCCCACAATC ATCAGCCAATAC TCCCTGTTGAAA TTCCTGCCCCGA AAGTTTCATGCT TCATTGTGTCTG ATTACAATGAAA GTGAAACCATGT GCTGCCAGCCG GAGATGCCCGAT CAGCAGCAGCTA GAGCAGCAACAA CGTAAG
---	--

	AGGGGG- CCTCGGGCGAGACGAAGATCCTCCGGGGTCCGGCTGCCCTTGTCCGAGG CTAGGCTCGGGCGAGGCGTGATCGAGTCCCTCGAATGGACTGATCTCT GACTTAATCGCACCCATCAGGCCTTTGCAGCTTTATGCTGATGGGGGT TACCAGCTGAGAATTAGGAGCCTTGAGGGTACCCCTAATTATAGTCCC CGACACTTGTAACCTCCACATATAGGTTTGT'TTAACTATTTTATAGAT TAGGGAAAGGTACTCCAGT'TTTGTATTATAGCTCCAAAACTCTATG GAGTTGT'TTTCTCCGAATTTAAGGTGTTTAGATAGAAACATTTTACT CCAACCTGAAGTTCTGCTCCAAAGCTATTTCAAACAAGCCCTAAGAAC CAACCATGTCAGT'TTTAGTTGGGTCAACCCCTCCTCCCATGGGCTCTA AAAAGATTATTTATTTGTTTGTAGCTTAAATAAGATGGTTCTGACATAT TTTCTAACCATTTTAAATAACAAAAAGTACTATAATGTTTATTTATAG TTTGTTCACTTTTAAATCAATAGCATTGCACCCTAATAGCAAGTATAA TAAGAGGTTGTAAACATGCTAAAGCTAAGGTGGAGGAGAGAGAAAGAG CTGGTTGTAAATTTACATTTGGCTTAGACATAAGAACGAAGAAGATTT GCGAGAGAGATAAATAGGATCCTACATTATTAGTGAAGAGCTAACTAC TATATGGATAGGCTAACAGGTATGCTATAAGAGCGTCTCCAACAAAGT GCCCTATAAAAAATGCCCTAAAATTTAAAATAAGCTCCAATATCAAGTG ATTAACCAAAATTTGTAAATCGTTTGGAAAGAAAAGTTTGTGAAACG TATGTTGTTTAGATTTAGATATGGACATCTTAATAAAGTGCAGTATA TTTGTGGCACTTTAGGAGGTGACATATATTATTTGACATTTT'TTGT AATTAGGACACTGGGAGTAGGATTTT'TGTAGTTTGAGCCCTATAATT TAATTTGGGGCACTGAGATAAGTCATTGTTGGAGATGCTCTAAAGGCC CTCACATGTAGCGATCGGCTATATTAATAAGGGCTTTGTTTAGTTTCT TTAGTCCACGGAATAAAGTTAAGTCCCTACGCTGTTTGGTTACAGGAA CTAAAAATATTTAAACGCATTAAACGACTAAGAAGACCAAAATGTTT CTTACTATTCTCATGTCTATTAGTGAATTTGAAATAAATGAGAGGTTAA AAGATGAATATAATATGGTTTAGTCTCTTTTAGTTATCCCTTTAGGAA ATATGAACATAAGCAGTTTAGTCCCTATTTTACAACCTTAGGCACTAA AATAAAGAGGCTAATTTT'TAGTCCATGGAATCAAACAAGCTCTAAAT TGTTTTTAGTAGAGATCAATCCATGGTCGATGGTGCCTCCATTAGGCC GGCTCCAATGTTGGCCGGAATATGTCCCTCGTCCGCACTTTGGCGTG TCAGCGTGTGCAACGTCTCCTTCCCGGTCCAACCGACCCCC- TATCCTTCGTCTCTTCACTACGACTTCGTAGTGAGTGTTGAGTTTCTAG TGACATGGGCGTGCATGGTATGGCAGCCAGCGATGGAGACTGTGCCG GGTCTTCGAGCCACTGCGCCTCTGGTCTGACTCGTACATGCCGTATGC CTCAAAGGACACTGTACCTCCTGGCTCCTGCATGCAGTTTTTCTGT CAGTTTTATTTGGGGGCTAAATTTCTCGCCAGAAATCTAGCTTAGTTC AAATTTGAATTATGAGAGATTGAAGCTCGAGGATCCAAATCTTCGAG GATGTTCTTCCATAATCAATGGAATACAAAATTATGTCAAAAACAAC TTTTTTT'CAGGATTACCGTAGAAGAATTGCGTCTGGCAAAGGAGAACC CCGAGCAGTCGATGAAACGCTAGCTTTGT'TAAAATTGATCTTCTCGAA AGAACGAGGATTGACAATAGGGACGAAGTTGGCTCCCGAAAGAAAGGG CCTCAGGACCACATCGGCGAATGAAGGAGCTCGAAGTTTGGGATTAGT GGCTAGATTCCATGAAGAGGATTCTTCAGCTCGGCAATCCGATTCCGA CCACCAGAATTCCATAGATAGAAATAAGCCTTGAAGGATTTCGATTCA TAGGATCTTGAATTATAAATAGAAAGGCCAAGTTCTTTTGGGAAGACC GAGAATCTGAATACTCTGTCCGAAAGAGAGATAGCCCGATATCGATGT GCGTATCCACCAAAGATAGATTGCAACAGAAAGCCCGTATTTTGTCTCT GAAGGACGAACTTGCATCGACCAATTGAAATAACTTTCAGGAATTCT GAAGGAGAAATGGCCTTGAAAGCGGTTGATAGGTTGTTGGAATCTGCGC CAAATGCCATCTTGAAAGAACAGACCTTGAGAGAAATCCAGAGAAGAA ATCCGATCATCAACGGGAGAAGAGGGGATGGGGAGAGGGAGCACCTCT CGCCAAGGCGTAGGCGGCTCCCTCATTCTCACACCTTTTATGTCTATT TGTTTGATTGTTTTCGGTAGCCTATAAAAAGTTAGCTAATTTCTGTCACT CGTGTATGTCAGCGCCTCAATCTAAACTCCACACGTCTGTCCAAATTA TGTCAGCTCTGTGAGCTAGATTTTGGTACGGAGGGGGTGCAGATTTTA AGTGAGTCTCCCTTGATTTTATATATATCGGCCCTGCTGTCCAGCGC CCCAGCCGGAGATGCAGCA	
Myosin01	insertion: CATG	

Myosin02	insertion SEQ ID NO: 71 GTGGTGTGTTTGGTTTATAGGGACTAATATTTAGTCCATTTATTTTATTC CTTTTTAGTGTATAAATTGCTAAATATAGAACTAAAATAAAGTTTTA GTTCTATATTTGGCAATTTTGGAACTAAAATGGAATAAATCTAGGG ACTAAACATTAGTCCCTAGAAACCAAACACCCTCTATG	-
Myosin03	insertion SEQ ID NO: 74 GTAAGTGCAATCAACCCCAAGTGAGGGTTTTGGTGATTTAATGACAAA ACAAATAAGGGTACTAACAGTTTTTGTCTCAGTGTTTGATTAAAAGG AAACATGAACCTTAAGGAAGCACCAAGGACTTCATGCAATGGACGTATA AAATTTATGTAAAATCTTGTGTTGCATGATGTAATTCTTCCTTATTCT TTTCTGCACAGGTACAGGTGTTTGTATAGCTCAAGTCCTCTAATTCA AAAGCTATTTTTGAAAACCAAAAATCACTTTAACTTTATTTGGCTGAC CATGGTTTGGGTTGAGAAACCTAGAGTGTTCTTGCTGAAAAACAGCTG AACTTCTTCAACTGCAGCTGAGCTTTCCAGCTGAACCTAACTTCAACT GTGATGAGCTTCTTCAGCTGCAGCTGAGCTTTTCAGCTGAGCTGGACT TCAGCTGAGTTGAACCTTCAACTCCAGCTGAACCTCAACTTCAGCTAT GGACCTCTCTGACCATAACCCAGCTGAACCTTGCCCCAGGGCAGCTGAGC TGGGGTTTTCTCTCCAAAACCTCTCTGGTCTAGACCAGCTGAACCTGGTC CTGGGGGGCAGCTCAACTGGTCTCTGACCCTCTGACTTCTGTTCTGCA GTCTGCCAGTCAGACTGGCAGTCAGGTGCCAGGGGTGTCAGGGGGCG GTTCAACCGTCCTAGGGGGGGCGGTTCAACCGGTCCTGGCCTGTCTGA CCCGCTTGCAAGTCACTCTGCCAGTCAGTCTGGCAGTCAGACTGCTGA CCTGTCAAGGGGGCGGTTCAACCGCCCTTGGGGGCGGTTCAACCGGTTT TCAGCGTCGTTTTTGGTCAAACGACTAGTTTTTGGAGCCGTGACTATAA ATAGACCCCTCTCTCTCTCTCTCTCTCTTCTTAATACGGCTGAACACTCA CTCATTTATTGCTCACCTAACTTGAGACAAAGCACTCATCTCTCCATC TCTCTCACACTTAAATCTTCTCAAGATTCAACCTTGTTGGAGGAGCT CTTGGGTGTGAGGTTTGTGCTTGTGCTCACGATTTGGTTTTCTCTCC CATCTCTCTTACAAAGACTTGAGCTTGTGCAAACCCCTCTTGTGCGTT CTTGGTGTCTTGAAACCCCTAGGTTTCAAGATCTTTGATGTTGCTTGG GAGTCTCCAAATTTGTGGACGACCCTAAGAAGTTTGTATCACCCGCTC TTTGAGCTAAGATAAGAAGAGATTGCCTTGACCTTTGTGGTCGGCTTT TGGAGGATTAGGGTTGAAAAAGACCCGGCCCTTTGTGGGCTCCTCAAC GGGGAGTAGGACACCTTTGTGGTGTGGCCGAACCTCGGATTAAATCTT GTATCTTGTGTTCTTGCTTGTGTTTGAAGTATGAGATATTTTACTTGCAA GGTTAACATAAAGATTTTTCCGTAGAGCTTATCTAGAAGTAAATAGC CTTTACATATTATCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT TCAAGAACTTGCGAAATACCTTTTTCAAGTAGATTTTAGTCCAAAGTTT ATAAAACAGTTGGATTGGTTCAAGAGTGGTTCAACTTGCGCACGTAGCT GTTGAACCGGCTGCACCACTGCAACTGTGTATCTAACAATCTTTTGA AGTTTGTGTTGAAAAATTTTCAAGATTAGCCTATTCACCCCTCTAGGC TACTTTCAATTGGTATCAGAGCTACGTACCTCGTTTTACGCTTAACCG CGTGAGGAAACGATCATGTCTACTCAACGGGACCATGTGGATCCTCTC CTCGAGGAAATCCCCATCACATCTTCCGGTGAGGAAGTGGACCCCAAG GTCTTCGACCTCGCCATGAAGATTGCCGAGAGAATGTTCTTCAAAATG AAAGAGGAAGATGCTAAGAAAAAGGCTGAAGAAGAAGAATCAAGAAGA AAGGCCGAAGAAGATAAAGGTAAGGGAACATTTGATTATAATGACGAT CTAGTGGATCTTTTGGTGTCTAAGGTATTAAGCAAGGTAAGTCTCAAC ACCAAAGGATCATCTACCAAAAGCAAAGGTAATGAATTTAGCAAAGTT CAATTTGATTACTCTAGAAATTATATTTCCCAACTTCTCTTCCGCCCCA CTTGGAAAGTTACCAACTCTTAGTGAGTTGAAGTATGATGAGTGGGCC GACAAGATGAAGTCGCATTTAATCGGTGTGCATCCTAGTCTTTGGGAG ATTGTTAATGTAGGTATGTATAAGCCCGCCCAAGGAGAAGAGATGACT CCGGAAATGATGCAAGAGGTTTCATCGAAATGCTCAAGCAGTGAGCATA ATTAAAGGAAGTCTTTGTCCGGAAGAATACCGGAAAGTTCAAGGAAGA GAAGATGCCCCGTGACATTTGGAATATTCTTAAATGTACATGAAGGA GATCCCAAAGCTAAAAGACATAGAGTTGAAGCTTTGGAAAGTGAGCTT GCAAGATATGATTGGACAAAGGGTGAGTCGCTTCAATCACTCTTTGAT CGGTTGATGGTGTGGTCAACAAAATAAGAGTTCTTGGGAGTGAAGAT TGGAGTGACTCCAAGGTCACAAGATTGTTTCATGAGAGCTTATAAAGAA AAGGATAAGAGTCTTGCAAGGATGATTAGGGATCTGTTGGAAGTTGCG	-

CTCAGCAGCAAGTAGACCAACAAGGGGGAACAGTGGTGACAACAGGGT TACGTGCTCGGGGGCAATAGCTCTGTTAATCTAGCCTCTCACGGGCAC TGTGCGGGGTATTTATAGGTATCTGAGTGCCCAACGCCTTGTGTTAA GGACGCATGTGCCCTCAGACACCTAGTTTATCCCCAGAATATTCCCAT AAAGCAGGGTTACAAGCTGTAATTACAAGTATGCTTTTACAAGTTAGG CCCGTAATACAGAGGCGGCCACGCGGGGCCCCGTTACAATGGGCCAGAT CACACGTGGGCCCTTAGGACTGAACGAGGCCGCGCCGTGGGAAGACGTC GCCGCGGGCCCTTCGTCAAAGTGCCGAATGGAGCGAAGGGTGTCCCTGC CCGTTTTGTCTCTGTCTCAGACCAGCGACTGCGGCGAAGGGCTCGAGCGA AGGGTGGCGTCTTTGCCCTTCGCCCCAACATTTGCCCTCCGAGGGACCA GTTTCGACAAAGTCACCTGGTGCCGAAGACGTCGCTAGATGGCGGAGAC GCTGCCCTCGCTCGAAGTGGTTCGCGGGGGTTTTTTTGATGTGACCGTT GATGGACGGCGTACTGTTGGATTGCGGGTTTTCCCGAAGCGTCGCGCCC TGTCGGATTGCGGGTTTTCCCGAAGAGCCGCGCCCTGCCTATAAAAGGG GGCGGGGTTCGGCGCTGTTTGAACTTTGCTTCCGCGCTCCGTAAAAA CCCTAGCCGCCCGCCGTCTTGCTGCTCCTCTTCTCCGCTGCTCCTTT TGCTCGCCGGAGTTCTAGCTCGAGTGAAGCAGACTGCCCGCCGCTGTC GACGGTACGTAGCAATGCCATCCTCTTCTTCTTCCGCCGACGCCACGC CGCCGGCCGCCGCCCTCGTCTGAGGAGACGTTGAGTAGCGTGATGGTGG AGGAGTTGCGCGCCGGCGACTCTGTTGGTTTTTGGTGTGTACGAGTGA CGTCAGCCCGCGTTGAAGATATGCAGCGGTTGGGCTACTTTGGCGGCG GAGTCGCTCGTGCTCCGGGGACGGAGGAAGTCCCCGAGCCGGAGGGCG AGTTGGTTGTTTTCGAGGCGTTCTTCGTGCGCCGCTCTCCGCTTGCTG CGCACCGGTTTCGTTGGCGAAGTCCTGCGTAGATTCAATGTGCAAATAC ATCAGCTGACACCGAACGCCATGGTGGCGCTGGCGAAGTATGTCTGGG CGACGACTTCGTACGGCGGACAGCCGTCAGTTGAAGTTTTTGCGAAGA ATTATTGCTTGCATTGGCAGAAAAGGATGGTTGAGGACGGAGTTGCCC AGTTCGGGTCATGCACATTCACGCCGAAGACCGGCAAGACCACAATGC CAGTAGTGGAGTTGGTTTTCGTGCGCCCGCAACAAGTGGGGCAATTGGA ATGAATTTTGGTTTTACGTTGCTGAGGCTACTGTGCAAGGCCACGAAG GACTCCCCGTTTCTGAGATGTGCTCTCACTATTACTCTGCGTATCCGC CTTTTGAAGTGGCAGAAGAGGATGCAGACGAAGGGGGCCCTTCGGTGCG CCGCCGCTCAGAGCAGTGGTTCGTGATTAGTTGAAGAGTTTCGTGGCGT ACGGGGTCTGGCCCTTGGCGCACGGCTGGGCGCTGGGCGAAGTATGCC CTCGCCAGATGCCTTTTACGGAGGAAGGGTGGTGCGAAGTCTGCCT TCGCACTGAATTTGCGAAACCGTGACCCGGCCGCCCTTTGTGCGTGATG CGGAGGACTTGGCGGTGCGGCTCGTGGGGCGTTACGTGCCGAGGACGG AAGGCCAGCGCAGCTTTGATATCCGCGGGTCAAATGAACGTTTTGAACA GGGTCTTCGAATTAAATCAATTGCCGTACGACGGCTACCCTGGACAAG ACGAGGCGGACCGCCGTGGGAAGAAGCCGGCGGTGGAGACTGGAGGCG ACCTTGCGCTGGCGGCCGCCCTCCTCAAAAAAGAGAAAATTAGGTA CTGCGATGGGAGGACTTGGGGTGTCTGATAGTTTTGCTAGAGAGTTAA TGAGGACGTGTGCGGCCCCGGGGGAAGGATGTCTTCGCCCAGACTCC GGGAGTCTTCGGCTCGGATGCTGAGGGTTACCGAGGGTTGTTGGCCTA GAAATGTTCTATCCCCTGCGCGGCTGAGGAGGACTGTTTTACATCTC GTATGGTTCATCGGTGGAGAGTTTTTCTTACGGGCGAAATATCGGTCT ATGTTGTGTGGGCAGTGATGGACAAGGATCGCCAAGGGGCGGCGCAAA AACGCCAGGCGACTGTGAGGGTTCACAAAGCTCGGTGCAAAATGCAGC GGGGAACGGCGAAGGCTGCGGCCCTCCGGCGGAGGCAAACCGCCGCTGC CGGCGAAGGTGGGCGCCTCTGCACCTAGCAAGGCGCCGAGGCGACGG TGGGCGCCGGAGGCTCCAAGCTGACGAAGGTGGTGCCGCGCGGCCAGCG GAGTGGCGGAGGCTGCGAAGGCGGCCCGAACGTTGCCGCGCGCGGGCA AACGCGTTGCCGACTTCGGCACCGAGATTAATGTGGAGGATTATCTTG GGGGTAAGTCTATATATATATGTATATATATATATATTTAATTTGTT GATGCGTTGCAGGGTCGGACGAGGGGCCGCTTGCTCTAGTTGCGCCTG GCGCGGCGATCGCGACAGCTGCGCTAGCGCCGAAGGCAAGGGGTGAGA CCCTTGCGCTCGGAGGCGAGGTGTCGGCCCTCGCGCTGGTCAGGGACG AGGCGGGCGTGGCGACCCGCCGCTGAAGGGAGTGACGAGGCGCTGA GTCAGGCGACGGAGGCGCTGAGTCAGGTCCGCTGAGCTATGCCCTCTC CGCCCCCTTCCCCCTTTTTTTTTT- AAGCAACAGCCTTACGTGTGCTCTGCAGGTGGCTGACTTCACCAGCCG	
--	--

CACTGCGTCGGGGGGCCCTTACGGCGACTTTGTCTGCCGAAGTCGAGAG ACTTCGGACGCAACATGCTGACGCCGTCCGTGAGAAGTCGGCCTCCGA CAGCAAGTGCCGCAAACCTGGCGGACAAGGTGGCCGCCCTGGAGAAGGA GAAGGCTGACCTCCGGCGCCAGCTGGCGGGGGAGAGGAGGGAGGCCAA CGAAGCCATCGCCAAAGAGCAGGCTGCGCAGGCGGAGGCTAGGCTGGC GCGGGCGGAAGGCAATATCGCCAAGGAGCTCGCCGGAGAGCTGTAGGG GCGGCTTACTGCCCTGGAGAGCCGCGCGGAGGGGGCTGAGGCCGCGAT GCGAGCGGAGGCCGAACGGACGCGCACGCAGCTCATGACTGCATATCG CGAGCTGGGTGCGCGGACCGGTGACTTCGAGGTGCCGGAGCGAGAGGC CGGGCTTCGCTGTCTGGAATGGGTGCAGGAGGAACTGCAGGAGCTCCC CACTGTCTTGAGGGGTAAATGTCGTTACCTCCCTGGTCACCTGCGA GGCGGCGATGAACGCGCTCTCTCGCGAAGGGTGCCGGCACTTTGAGGC CCTCGACCAGGCAGATGAGGATTTTGGGCGAGATATCTTTGAGGTCGA AGACCCCGTGGTGAAGGAATCCGCCGGGGCCCTCTATGACCGGATGTG GGGTTCGTACGGTCGTGAGGTGGTCAGGGAGCGGGCCGAGGCTGCGAG GGCTCAGGTAACGTTGTCGTTTGTGTTGGCGTTTGCTGGATGTGGTCTG TGGTATGTTTGTGAATTTTGTGTGCTTTGTCTTAGGCGGAGCGCG GCGAGCGGGTGGACGACTTCGGGGCACTGAACAGCGCGCTGCCTGACC CGGAGACGACCCCGCGGAGGCCGCGACTGGAGTCGCCCTGGGACCGA CCCTGGAGACTGCGGAAGACGCACACGAGCGCCCCGGCATCTGCTGTG CCGGCGAAGACCTGCCCCCAAAGGTGGCCGAAGGCGCGCCTGATGCCC CCGCAGCTGCTGTGCCGAGCGCTGAAGATCCCGCGGCGGTGGTGGCGG AGGCAGCGGCGGAGGCCCTCGGCGCAGGTGGCGTAGTGGGTGTTTAGGG TAGGGGAATTTTGGATGTTGTACTGAATTTTGGTTGCTGCGTAGGTT CAAGGTCTTGGGCCGTGCGCACGCAACCGCTACTACTGAGTTTTTGGCG GCTTCGACCGCGGAAGGTGGAATGAATATGGTGGGTCTGTATCTGAG TTTTTTGATTTTACTGACTCTGGGAGCTCGGTTGAGTGGACTGAGGAG TCGTCGGAGGAGGACTTGGATTATTTTGCGGCCTTGGACGTGGCTGCG GCGGAGGCTTCGCCACATGCCGTGGACGCGCCAGACGTGTGGGTCTT AGCACCGGGCGCCGACGACGAAGGGCGGTTGCAAAGCGTAGGGTTGGC GCGGAGGAAGGGGCTCGGCTTCGCGTGAGTAGGGTTGGTCGGCAGGAA CTGTTGTCTTTGCCGGCCGCCGCTAGCACAGGGGAAGGGGAAGTGCGA AGTCTTTTTGCGGGTGAGGAGCTTAAGATAATGTTATTTAATTATAGG GAGATGGGAATTAATCCCTAAGGTGGAGCCGATGTAGAGTGTGGTGCCC TCGCTTGATATGTGTAAAGGGTTGTAAATGAAGTTGTGTGCCCTCG CGTGCCCTGTGCTTCCGAGGGCGCTGTGTACTTTGTGTGGCATGTTTGG CTATGCTGTGCTTCTGCTTTTGTGTACTAGTTTGTGCTGCGCCCTTA GGCGGCTTCTGCGCGGAGTCTTCGCGATAGACATCGGAATTTTTCGCA CTTGTTGTGCTAAGTCCGGCTGCACCCTTAGGCGACTTCTGCGCGGGT AGTCTTCGCGATAGACATCGGAATTTTTCGCACTTGTGTGCTAAGTC CGGCTGCACCCTTAGGTGACTTCTGCGCGGATAGTCTTCGCGATACAC ATCGGAGTTTTTTCGCACTTGTGTGCTAAGTCCGGCTGCACCCTTAGG CGACTTCTGCGCGGATAGTCTTCGCGATAGACATCGGAATTTTTCGCA CTTGTTGTGCTAAGTCCGGCTGCACCCTTAGGCGACTTCTGCGCGGAT TAGTCGCACGCGAGGGTGGCTAGCGCTTAGCCTCGCGATGACTTCGAA TTGGTTCGAATGCCGAAGACCGTTGTGCGGGTCTTTTTTTCGACACATGT TTTTGCGGGGGATTTTTCTCATATATATTACATGGCTCCGCCTCGTTA AAAACCTCACCCCCGGGAGGAAAAGAGTGCGGGCCGGTACAAGATTG TTTGCGGTGAATTACAAGGGCGAAGTCAAGCCCTAAATGGATTAAACAA AAAATTTGCGAAGGTGTGCGATGTTCCAGGAGTGCTCCAGGTCTCGC CATTTGGCGTTGTGAGCCTGTAAGCGCTGGGGGAAGCTTTTGTTTTGA CTATGAAGGGGCCCTCCCACTTGGGCTCCAACTTGCCCTTGGACTCCG TCCGAGCTGTTCCGATGAGTACGAGGTCCCCTGCGCTGAACTCCCTTG GGATGACTGCGCGGTGCGGCCATGCTTTGGTCTGGGCTTGGTATTTGT TTAGGGCTGCAAGGGCGAAGACCCGGTCTCCGTCAATGAGATCTTTTG AGGTTGGCTCATCCACGTACGGGACGGCTGACGGAGCTGTACGCGGAG ACCCATGCTTTATTTCTTGCGGGGTCAATTGCCTCCGATCCGTATAGAA GGCGGAAAGGAGTGAACCCGGTGCCTTGCCTCAGTTGTGTTTAGCG CCCAGACTGCCCTCCGGTAACAAGTCGGTCCATCTGCCTTTTTTTTTCA TCGAGGAGCATCTTCTTGACAGCTGTGAAGATTTTCCCATTTGGCACGC TCCACGACCCCGTTGGACTGCGGATGATAGACTGAAGCGAAGGCAAGC	
---	--

TTGGTGCCGATGGAGAAACAAAAATCCTTGAAATCTTGGCTGTCAAAC TGCTTGCCGTTGTCAACTGTTAGCTCGGACGGTACTCCGAAGCGGCAA ACAATGTTTTTGGCAGAAGAATTTTTGGGCAGTCTTTGATGTGATTGTG GAAACAGCCCTCGCTTCAATCCATTTGGTGAAGTATTCGACGGCTACG AAGGCGAACTTAAGGTTCCCCTGAGCAGTGGGCAGGGGCCCAACAATG TCCAGGGCCCCAGCGCTGGAGAGGCCATGTATGGGCGATCAGCTTTGTG TATTGCGAGGGGTTGCCTGACCGAGGAGAGAACTTCTGGCAGGCTTCG CAGGACCTTGTGACTCGATTTGCGGCGCAGATCATTGCAGGCCAGTAA AACCCCTTGTGCGATCACCTTGGCAGTCAGGGCCCTTGGCCCTGCGTGC GAGCCGCACGTGCCACTGTGGACTTCGCGTAGGATCTGCATGCCTTCG GTTTCGGTGACGCATTTGAGCATTGGCTGACTGACCCCTTTCTTGAT AATTGGCCCTCAATCAGTGCGAAGTCCCGGCTTCGATGTCTGAGGCGC TTTGCCTCACTGACGTCGGTTGGATGATAGTACCCCTGTAGGAACAGG GTTATTGGTGCCCGCCAGTCTTCGGTCATGATTAGGTTGACTATACGG TGGCCCTCGCTATCATTGGTTATTTGGAGCCCTTCGGGGCTCCGGACG GCCGGCGTGCCGATGGTGTGAAAGAACACGTCGGAGGGCAGGGGGTCG CCCCTGGCGGGCGGCCCTTGGCTAATGCATCGGCCTCCTCGTTCTTGGCC CTGTCCACATGCTGCAGGGTGAACCCCTTGAAGTGTCTTTCGAGACTT CGGATAGTCGCGAGGTATTCATGAGCGCGGGATCTTTTGCTGCATAG TCTTTCTCGACCTGGCCGGCAACTATCTTGGAGTCTGTTTTGATGATA CATGTGGTGACTCCGAGGGCCCTTAGCTTGCGGAGGCCGAGGATAACC GCTTCATACTCTGCTATGTTGTTTGTACATTTGTCGATTTTCAGAGCG AAGCTGAGGCGTGCTGCATATCTGTGTTTGAAGTCCGGCTGGTGAGGTG ACGACTGCAGCGACGCCTGCCCCCGCATGACACCATGCACCGTCGCAA TGGATAGTCCAGACCTTCTCTGTGGACGGGTCCGGCTGTGTTACTGGC CCAGTCCAGTCGACGACGAAGTCTGCCAGGACTTGCGACTTGATAGCT GTCCTGGGCTCGAAGCTGATGTGGTAGCCGGAGAGTTCCGGCTGCCAC TTGGCAATCCTCACTGATGCCTCCGGGTTTCTGAATAGTTCGCCGAGC CCCCTATCTGAGGTGACTCGGACCTTGAATGCTTCAAAGTAATGGCGC AGCTTGCGCGAGGACATAATAACTGCGTAGGCGATCTTCTCCAGCTCT GTCATGTTGCATTTGGACGGTGTGAGGACTTCGGAGACATAGTAAACA GGGCACTGTCTGACTGCGCCCTTGAAGTGTCTGCTCCTGCACTAGTGCC GCGCTGACCGCGTGCGGCGAAGCCGCGACATAGAGCAATAGGGGGAGC GAGGAGTCAGGGCTTGTGAGGATGGCCAGCTCCGACAGGTACTGTTTT AGTGAGGCGAAGGCCGCTGCTTGCTCCGGTCCCCAGGCGAAGTCTTTC GCACCGCGGAGTGTGTTGAGAGCGGCCAATCTGCCTGTGAGGCGCTGGACG GAGATGAATCTGTTGAGAGCGGCCAATCTGCCTGTGAGGCGCTGGACG TCCCTGGCGGACTACGGAGGCGACATGTGACGATGGCCCTAGATCTTG GTTGGGTTGGCCTCGATGCCGCGGGGACACCAGGTAGCCCAATATC TTGCCCTGGCGAACACCGAAGACGCACCTTTCCGGGTTTAGGCGGAGT CGTGATCTCGCATGTTGCGCAATGTCTCGGCCAGGTGCGCGAGATGG TCCTCCTTATTTCTTGCTGGCGACGACGATGTGCTCCACATATGTAAAT ATATTTCTGCCGACCTGCCCTTCGAGTACTGTTTTGGTGAGTCTGGAG AAGGTGGACCCGGCATTCTTGAGACCCTCCGGCATTCTGACAAAGCAA TAAGTGCCGAAGGGTGTATAAAGCTGGTGCTAGCCTTGTCTTCCTCC TTCATGTATATCTGGTGATAGCCGGAGAAGCAGTCGAGGAGGGACATG ACCTCGCATCCGGCCGCGCTATCGACTATTTTGTGATCCGTGGCAAC GGGAAATTTGTCCTTTGGGCAGGCCCTATTGAGACTGGTGAAGTCTATG CACATTCGCCATTTCCCGCTTTTTTCTGCACCATCACAAACATTGGAG AGCCACGTGGGGTAAGCCACTGGCTCGATGAATTTAGCCTCCAGGAGG TGGTGACCTCCGCCCTTGGCGGCCCTCTATCTTTTCGTGTGACATTTTG CGAAGCTTCTGCTTTTTTGGGTCGCACCGAAGGGTCAATGCCAAGCTG GTTTCGATTATGGATCGGCTGACTCCGACCAGATCGAGGGCGGACCAG CGGAAGACATCTTTATTTCTTGGCCAGGCAGCAGAGAAGTTTCTCTTCT TCAGGTGAGGTGAGGTCTTCGCTGATAGTGACTGTCTGCTTGGGCGTG GCCTGATCGAGGGGGACAGTCTTGGTCCCGTCTTGGCACTGCAACTGT GCCTTGTGCTGCTGCTTGTGCGGTTGGGCTGGCTGGCGCAGGGACCTCG CGCTGGGTCGTGAGACAGTGACGTTCCCTCTGACCAGGCACGAAGTCC CGCTCTATGTTGCGCGCCGTCTGCTGGTTGCCGTAGATTGTGATGGCG CCTAGCGGACCTGGTATCTTCATACATAGGTACAACCCGTGGATGGCG GCTTCGAACCTTATTGATGGAGCCCCGGGCCATGATGGCGTTGTATGGA	
--	--

TATACCATGTCGACAATGTCGAAGGTTACTTGCTCACTTCGGGCATTG GGTGCTACACCGAAGGAGAGGGGCAACTCTATCTTGCCAACTTGAAAG GTGCCCTTGCCGCCGAAACCATAACAAGGGGTTGTCCGAAGGCTTCAGC AAGCTGTGGCTTATACCCATGCGGTGCAAGGCGTGAGGAAGATGATG TCCGCCTGACTGCCGTTGTGACTAGGACTTTGTGTAAGTCCCAGCCT GCCACGCTGCAATTGATGACCATAGCGTCAATGTGGGGGGCGCTGCGC AGATCGACGTCTCGTGATCGAAGGTTAGCGGTATGTGGGACCACTTC GTCTGCACGACTGGGCCAGTGACCGTGACATGGTTGATGCTGCGGTAG TGGTCCCGCTTCTGCCGCTTGGTGTGCAAGTCAGTGCTGGACCCCCCT GTTATCATGTGAATGACTCCGCGATATGGCTGATCGGCGAAATCTTCC TGCTTTGGGGCATGGGGGATGTTTTGGTGTGCTGGTGTGGTGGTGGG GGCGGAGGAGGTACGATTTGTACTTCCTGATGGTGTGGTAAGCGTGG TGCGGTGGATGCGGAGCGGGAGCGTGGATATATGGTGGAGGGGGTGGC TGGTAATTATGCGCGACAATTCTGGGATTGTGCGGCTGGCTGCGCCCCG GCCATTCTGTCTCTGGTGGCCTTCGTCTCGGGGCAATCTTTGGTTTGG TGGGCGCAGTCTTCGCCGTGGAAAAGGCAGTAGAACCGGCGCGCGCGC TGCGCTCGCCCCCGCCCCCTGCCACGAGCTCCATTTCCGCGGCCCTGA GGGGGATATTCCTGGCGACGAGGGGGTTGAGCAGCAGGCTGCTGGTTG GCGATGTTGTGCACTTGCTGTTGACTGCGGCCATCTCGACCGGAGTCT GGTTGCGGAGTCCCTCGTCCACGTGCGGCTGGACTGCGGGGCATCTTTG GGCTTCCGCTGTGACTCAACCTTGCGCTGGTGGAGTTCTTCGATTG GCATATTTTTCAAAGAGCTGATATAGCTCCTGGAGGTTCTTGGGTGGA TCTCTGATACAGTGGCTGTAGAGGACGCCAGCCCGAAGGACACTGATG GCGTAATGGATAGCGATCTGATCATCGACGAGGGCAGTTGTGATTTG AGTGTAAAAAATTTGCGGTAATACTCCCGCAGAGTCTCCTTTTTTCAGC TGTTTGCAAAGCGAGAGCTCGGCCAAAGCGTCGGTGTCTGGACGGTAC CCTTGGAAGTTGAGAAGGAACTTGTCCTGAGACTTCTCCACGAATCA ATGGACAGCGGAGGCAATCTGGTGAACCAGGTGAGTGCTGGGCCCTCT AGAGCGATGATGAAAGACTTCGCCATTGTGGCGTCGTCCCCTCCGGCG GATGCAACGGCGACTTGATAACTCATTATGTACTGTGCCGGGTGCGTG CTGCCGTTGTACTTGGGATATGCCCCCTGCTCGGAAGTTAGCTGGCCAA GGTGTCACTTGCAAGGTGTGGTGCCAGGGGACTTCGCTCGTCGAGGTAG TTGACCCCTTGGAATGGCGCGCTGTGCTGGAAGGCGTAGTCATGCTGG GGGAAACGTGGGTTCTCCATCTGCGCCCCGTGTTGCGGGGGTGGGCCA TGCTGCAGGCCGAGGTGGCCTTCGCGCGTGTCTTCGGTTGTGCGCGC TGCTGGAGGGGTGGCTCTTGCTGTGGACCGAGGTGGCCTTCGCGCTGC ATCAGCGGATCTCGCGCTCGAGGTCTTGGGCTTCTGCTCCTCGTCG CGTATCATTTGCCGCACCTTGCGTAGTGCGGACACACGCTGGCGCTTG GCCTCCAGTATCTCTTTCTGCCTCTGGAGATTGCGGTTCTTGAGGCGC AGGGCACGCAGCTGTAGCTGTTCTTCTGTTGAGACGCCGAGGACCTCG CCGTCTCGGTGAGATCCGCGCCCTCCAGTGGGGCGAAGCCTGGGGGC GGCTGCGATTGTCTTCGAGTCCGCAGGTGCGGAGGGTGTGCTCTTCG CAGGTGCGGGGAATATTGTGCTCGGTGGGCTCTTGGTGGTGGATTGG GTGAGGGCGAGGGCCTTGCCCTTCCTTGCGGCAAGTAGTGCTGCCTTC GCAGCCTCGTCTGCCTTCGGGTTAGCTCTCTTGGGTGCCATCGCGGGT GGTTTGGTGTGACACGAACGGTGGGCGCCAAATGTTGGAAGTTGCGC TCAGCAGCAAGTAGACCAACAAGGGGGAACAGTGGTGACAACAGGGTT ACGTGCTCGGGGGCAATAGCTCTGTTAATCTAGCCTCTCACGGGCACT GTGCGGGGGTATTTATAGGTATCTGAGTGCCCAACGCCTTGTGTTAAG GACGCATGTGCCCCTCAGACACCTAGTTTATTTCCAGAATATTTCCATA AAGCAGGGTTACAAGCTGTAATTACAAGTATGCTTTTACAAGTTAGGC CCGTAATACAGAGGCGGCCACGCGGGGCTGTTACAATGAGCCAGATC ACACGTGGGCCCTTAGGACTGAACGAGGCCGCGTCGCAGGAAGACGTG CCGCGGGCCTTCGTCAAAGTGCCGAATGGAGCGAAGGGTGTCCCTGCC CGTTTTGTCTCTGTGACACCAGCGACTGCGGCGAAGGGCTCGAGCGAA GGGTGGCGTCTTTGCCTTCGCCCCAACAGGATCGTGATGACTATGAGG ATATGACGCCTCATCAATTATTTGCAAAGATTTAACAACACGAGTCCG AAGAAGCCCCATCAAGTCAAGGGACTCTCATGCCTTGATCACTAATG AACAAGACAACCCCAAGAAGAACAAAGACCACAAAGCAAAGAAAGTGG TCGAGACCTCAAGTGATGAAGATAGCTCAAGTGATGAAGACACTGCTA TGTTTCATCAAAACATTTCAAGAAATTTGTAAGGAAAAATGACAAGTTCC	
---	--

AAAGGAAAGGAAAGAAGAGGGCATGCTATGAATGTGGCCAAACCGGTC ATTTTATAGCGGATTGTCCTAACAAAGAAGGAACAAGAAGCTAAGAAGG AATACAAGAAGGACAAGTTCAAAAAGGGGGGCAAGACCAAGGGTACT TCAAGAAGAAGAAATATGGTCAAGCCCATATTGGTGAAGAATGGAAC CCGATGAAGAGAGTTCTAGCTCCGAGGAAGAGGAAGTGGTGGCAAACG TGGCCATCCAATCTACATCAAGCTCGCAACTCTTCACCAACCTACAAG ACGAATCCTACACTCCAACCTGCCTCATGGCAAAGGAGATAAGGTAA CCTTATTTAGTAATGATTTTCCAAATGATGATGATGATGAACAAATTG CCATGAAAAATAAAATGATTAAGGAATTTGGCTTGAATGGGTACAATG TTATCACCAAATTAATGGAGAAGCTAGATAAAAGAAAAGCAACTCTTG ATGCTCAAGAAGACTTGCTTATCCTTGAAAAGGAAAGAAACCTAGAGC TTCAAGAATTGCTTCACAATAAAGATGAAATGCTAGATGTCTTGACTA AGGAAGTATCCTTAGTCAAGATAACTATAGAGAATAAAGATAAAGAAG TAATTAATATGAAAACCTCTATAGCTAATCTTGCAAATGAAAAGAATG CACTTGAAACAAGCATGTTAAGCTTGAATGTTCAAAATCAAGAACCTC AAGTGCAACTTGAAAATTGCAAGAACATCAATGCCTCATCTTTAGTGA TTGAATCTAAGTCTAGCTCCTTAAATGATAATTTTTGCAAACATTGTG CCAAATATCATGCTTCTTGTGTCTAACTAACCATGCAAGGAAGAATA GCCACAGGTGAAGGTCAAAGGAATTTTGAAAAGATGCTCTAGCAATG ATGGGTTAAAGAAAGTTGAACCCAAGTTCAAGTCCCTAAGGCCCAACA ATGGAAGAAGGGGGTTTGGGTTCAACTCATCCAAGGAAAACCTAGCA CAGTGCATAAGGGGTGGAGATCCCCCAAGTTCATAGAGGGAACCACCC TATATGATGCCTTGGGGAGGATTCACTCCTCAAATGACAAGTCACCTC AGGTAAAGGTTAACTTGAGTTCCACAAAGAGTAAGATGAAGGAAGTGG GATCCTCAAGTGGACAAAAATCTAATGCTCCCATTTCCCACTCATATC TTTGTGATTATATGTTGACTTGGGATTTAGGGAAATTGGTTGTGAAAT ATGTGGGTGCCTACACTAAACGAAAAGTCATGAAAAGGAGTGTGTGGG TACCCAAGGCTATAACTAACACTGTAGGACCCAATTCAATTTGGGTAC CTAAAAGCATAGCCTAAACTTGTTTTGCAGGTCTACTCCTCTGGTGGG TCAAGTTGGGTGCTTGACAGTGGATGTACAAATCACATGACCGGGGAG AAAGACATGTTTCATACATTGCAACTAACTCAAGAAGCACAGAATTT GTGTTTGGAGATAGTGGCAAGAGTAAGGTGATTGGTATTGGTAAAATT CCTATCTCTGACCAACAATCACTTTCAAATGTTTTATTGGTAGATTCT TTAAGCTATAATTTGTTGTCCGTTTCACAACTTTGTGGAATGGGTAT AATTGTCATTTTTCGGATGTGGATGTGAAGATCCTTAGAAGGGAGGAC TCCTCAGTTGCCTTTACCGGTCGCTTGAAGGGCAAGCTTTATCTTGTT GATTTACAACAAGTAAAGTGACGCCTGAGACTTGTTTAGTGGCAAAG TCCGACAAGGGTTGGCTATGGCATCGCCGGTTAGCCCATGTCGGTATG AGGAATTTGGCCAAACTTCAAAAGGATAATCACATCATTGGACTAACA AATGTTGTATTTGAGAAAGATAGGGTTTGTGGCGCATGCCAAGCAGGA AAGCAACATGGAGTCCACATCAATCAAAGAATGTGGTCACAACAAAG AGGCCATTGGAGCTTCTTCACATGGACCTCTTCGGACCTGTGGCCTAC ATTAGCATTGGTGGTAGTAAGTATGGTTTAGTCATTGTTGATGATTTT TCTCGATTACCTGGGTTTTCTTTTTGAGTGATAAAGGTGAACTCAA GAAATATTGAAGAAATTCATGAGGAGAGCTCAAATGAATTTGAGCTC AAAATCAAGAAAGTGAGAAGTGATAATGGGACGAAATGCAAGAACACA GGTGTGCAAGAATTCCTTAGGTGAAGAAGGAATCAAACATGAGTTCTCG GTTCCCTTACACTCCACAACAAAGTGGTGTGTGGAAAGAAAGAACCGA ACTCTAATTGAAGCTGCAAGAACCATGTTGGATGAGTACAAGACACCT GACAAATTTTGGGCAGAGGCGGTCAACACCGCCTGTCATGCAATCAAC CGTCTCTATCTTCATAAGATCTACAAAAGACTGCTTATGAGCTTCTC ACCGGTAACGAACCTAAAGTTGATTATTTTAGAGTATTTGGTTGTAAG TGTTTTATTCTTAACAAGAAAGTCAAGAGCTTCAAGTTTGCTCCTAGA GTGGACGAGGGCTTCTTGCTTGGTTATGCATCAAATGCGCATGGATAT CGTGTTTTCAACAATACCACCGGTCTTGTTGAAATAGCGATAGACGTG ACATTTGATGAGTCTAATGGCTCGCAAGGGCATGTTTCTAATATCACT ACAGGAAATGAAGAACTACCTTGTGAAGCCATAAAGAACTTGCAATA GGTGAGGTGAGACCTCAAGAAAAGGATGATGAGGAAGGAACCTTATGG ATGACCAATGAGGTTGTTGATGTGGGTGCAAGGGTGGTGAGTGACAAA GTCTCCACCCAAGCAAACCCATCAACCTCAAGTCATCCAAGCCATGAA GAAAATCATCAAAGAATGCCAACAGTGGTAGAAGATGAACACGAAAGC	
---	--

ATTGATGGTGAAGTGCCTCTTGATCAAGTGAATGATGAGGAGGAGCAA ATACAAAGACAACCATCAGTGCCTCATCCTAGAGTCCATCATACAATT CAAAGGGATCATCCGGTGGACAACATCCTGGGTAGCATCAGGAGAGGG GTAACGACTCGATCTCGTTTAGCTAATTTTTGTGAATTTTACTCGTTT GTTTTCTCTCTTGAGCCACTTAAGGTTGAAGAAGCATTGGGTGATCCG GATTGGATAAATTGTCATGCAAGAGGAGTTGAACAACCTTCACCCGGAAT GAAGTCTGGTCCCTTAGTCCAAAGACCCAAACAAAATGTGATTGGGACT AAATGGGTCTTTAGGAACAAACAAGATGAACATGGCGTGGTTACAAGA AACAAAGCACGGTTGGTTGCCCCAAGGCTACACTCAAGTGGAAGGACTT GATTTTGGCGAAACATATGCGCCAGTAGCAAGGTTAGAATCAATTAGA ATATTAATTGCATATGCTACTAACCATGATTTCAAGCTATATCAAATG GATGTCAAGAGCGCATTTCTAAATGGACCACTACAAGAGAGAGTATAT GTGGAGCAACCACCGGGTTTTGAAGATCCAAAGAAGCCAAATCATGTT TATCTACTTCACAAGGCACCTCTACGGGCTTAAACAAGCCCCCTAGAGCT TGGTATGACTGTCTTAAAGATTTTTTAATTAAGAATGGGTTTACAATA GGAAAAGCTGACTCTACCTTATTACTAGAAAAGTTGACAATGAATTAT TTGTGTGCCAAGTATATGTTGATGACATTATATTTGGTAGTACTAATG AAAAATTTTGTGAAGAGTTTAGCAAAGTAATGACGAACAGGTTTGAGA TGTCTATGATGGGCGAGCTTAAATACTTCCTGGGATTTCAAGTCAAAC AGTCAAGGAAGGTACTTTTCTATGCCAAACCAAATATACACAAGATA TGCTCAAGAAGTTTGGCATGGAAAAAGCAAAACACGCCAAGACTCCAA TGTCATCAAATGGACATCTCGACCTAAATGAGGAAGGTAAACCTGTAG ATCAAAAATTATATAGATCAATGATAGGATCACTGCTTTACTTATGTG CATCTAGACCTGATATAATGTTGAGTGTGTCATGTGTGCACGTTTTTC AAGCTAATCCTAAAGATTGCCATCTGTAGCAGTTAAGAGAATTCTAA GATACTTAGTTTACACCCCAAACCTAGGATTATGGTATCCCAAAGGCT CCCTTTTTCGATTTACTTGGCTATTCTGACTCAGATTATGCCGTTTGA AAGTAGATCGAAAAAGCACTACTGGGACTTGCCAATTCCTTGGGCGAT CCTTAGTATCATGGAGTTCCAAGAAACAAAGTTGTGTTACACTTTCCA CTGCAGAAGCTGAGTACATAGCAGCTGGGGCATGTTGTGCACAGTTGT TATGGATGAAGCAAACCCCTTAGAGACTTTGGTTGTGAGTTTAAACAAA TTCCACTTTTGTGTGACAATGAGAGTGCCATAAACTTGCAAACAACCT CTGTGCAACACTCTAGAACTAAACATATTGACATCAGACACCATTCT TGAGAGACCACGAAGCCAAAGGAGATATCGAATTGTTTCATGTGAGCA CCGAAAATCAACTAGCCGATATCTTCACAAAACCCCTCGATGAGACTA GGTTTTGCTTTCTTAGGAGTGAATTAAATATCTTGGATTCTCGTAACT TAACCTGAAAACGGTCACAAAAATTGTTTGATTCACTATTTTTAATTG ATAGTTTTTAAAGCTTTGTGATGATCTTTCAAAAACCTTGGTGAAAATGT TAAAAATCGAATGAATTTTAGTGCTAAGATTTTTTTTT- GGGCTCAACTGGCTTGACCAGCTGAACCTTCTAGTTTCACTGGAGGAG GCCAGCTGAACCTCGGAGCTGAACCTCAGCTCAGCTGTTTTTAACCAG TTTTTTACCAGGACCATCCTGGTTAAAACCGGTTGAACCGGCATAGTGG ACCGGTTGAACCGGTCTCTGACCGGTCTGTCCGCGCAGGTCTGTCCGT GCAGGCCTGTCCGCGCAGGTCTGTGCTGACCTTTTTTCAGTCGCGC AGCATCCTTTTTTGACCGGCCGCTTCTCTCTGCTGGCGGCCTGGCGTG CAATTTTTTTT- AATCCGGTTCGCATCCCGCAGGGCGCGCAGCACCTTTTTTTTTCTTTTGT CCAGTCCAATCCACTCTGCAGGGCGCATAACATTTTTTTTTCTTTCTTA TTGACTGGTCACTTCCTCTGCTGGGCGCGCAGATTATTTTTTTTTGCTT GCTGCCGCTAAGCAGCGTCAGTGCGCGCCATTTTTTATTTGTTCTGTC TGGTCACTTCTCCTCTGCTGGGCGCGCGCAGAATTTTTTTTGGACGCT TCTGCTCAGCCCGCGTGCCATGCTGCAGACTTTTTCTCTCTGTTCCGGT CGTTCTCCTCAGCCTTGTGAGCCGCGCGCCAGTAACTTTTTTTGGG GCTGTCTGGTCTCAGCATTGCATGCAAATGTGAGGCGCCATGCAGTA TTTTTTGGTCTCCGCTGTGCTCTAGCGTCAGTGCGCGCACAGAAATTT TTTGTGGTGTCCGGTCAGCATCTGCTGTGCATGCAGAAATTTTTTTTCT TCTGATCGGTCACTTCTGCACAGCCGTCTGGGCGCCATGCTGCAGATT TTTTCTCTCTCCTCTTACAGCCATGCATGCTTGTGAGGGGCGCTAGC GGATTTTTTTTTTCCGGTCAACAGCAGCTGCGCGCACAGTTTTTTTTTG TCACAACCCCACTCTACTGCCGTGTGCGCAGCCTTTTGTCTTGTGAC AGGTGCTGTGTCAACCGCGCAGCAAGGAACTATTTTTTTTTGTCACT	
--	--

TCCGTGCAGCAGTGTGAGGGATACAGGGCGCACAGAACTTTTCTTTTG CGCAGTTCTTTTTCTTTTGCCGTCCGGTCTCATCTACGCAGACACTTG GTTTTGTTTTTTTGCACAGAATCTCTGCAGTACACTTTTTTCAGAATCC TGCGTCAGAAATTTGATTTTTTGGCATCTTCCCTCTGCTGTTTCTGGC TCTCTGACCACACTGGTCAGAACCAGCTGAACTGGCCCATCGGGGCGG TTGAACCGGTCTGAGGACCAGCTCAACTGCCTATATATATTTGACAC TCCCCTTCTCTTCTCCCACTTCATTCACTCTCGCCTTAGTCGCTTAC AACCTTTAGAGCCTCTCTTTCACACACCTCAACCGGTTTACAGTTTCA CCACCTTTTCAATCATGGTGCCTCGTGCACACCTTTCAGTGATTGAGT TTAGCAGCGACTCGGACGAGATTGAGGAGGAATCTCCCGTCCCCTCTC CTCCGCCTCGTGCCTCCCTCTCTAAGAGAGGACATGGCTCTGGTCGGA TGGATGGTGGGGCTCTTCCATGCCCTCGCAGCCGACTCGCGGGAGAC GCCAAAAGGTTGGCGCCTCTCGTCCCTCGTCTAGCACACCTTCTCTG GGTATGCTCCCTGCCAGGAGGAGGATGGGGCTTTCGACGACTTCTGTCG ACCTTCCGGCGCCTGATGCTCTCTACGTCACCGGACTGCACATGCATC CGGCGAACGTGAGCCGCGGTCCGCATGAATCTCCCGTCGACTTCACTC TGAAGGGAATAGACACCTTTCAGCGCCTAAGGTTTACCAACCTACTC TCACTCCTCGTCATCCTGGTGTCCAGGACCTCGGTTCTAGAATCTTT TCCAGGCTGACTTCTACAATCTGTCTATTCTTTCCAAGAAACACCCAG TCGTGAGACATAGGTTTATTGACTGGGAGGGTTGTGAGAATATGGGAG ATGCTGACATGTCTGTCAGCTCTCGAAAACCTTAGAGAGAAAGGGGTTAC GAAACATTATGACCATGGAATACCCCTGGAACGATGAGGTGGTAGCTC AGTTCTATGCCACCCTCTGGATAAAGAAAGTAGATGAGGAAGCCGATG GGTATGATTACCCAGTCATGTATTTCTTTCATCCAGGGTAACTGGCACA AAGTCAGTTATCGTCTGTTTTGCCCATATACTGGGCTTCTCTGATGCAG ATATTCTGGGCAGCAACATGCGAGTGCATAACATTTCGGCTGCCATGC GGAAAGAGACGGAATTTATTCATGTCTCTACTGAGCGGGAGTTCTGGA TAACCGCTAACATGCATAGGTACTACAAGTACCTTAATTCTTTGTCCC GGATGACCGTTCTCCCGAAGGGCGGAAATCAAATGAACGTCCTTGGGG AGAGTCGAGTCTTCCCTCCTCCTTGTCTCAAACAGTCTCGCCCGCA TCAATGTGTTTGACATGATCTGGGAAGAGATCGTCCGTGCTTCTCTGGT CTCCTCTCAGAGGGTGTCTTCATGCACCCTTTATCATGAAGATGATCG AAGTGGTCACCCAGACCCACTTCGAAAACCCGTCAAACACTCCTGCT ACGTACCCTACTGGGTTGATTCCTCCAACCCAGCTGCTCGTACGAAGC GGGCTCCCTCACGGTCTGGAGGCCCTGCCTCCTCCACTGAGCCACCTC GCCAGCTCGCCATCATCCATCCTCCTCTCCCCCGCTGCTGCCTCGC GTCCCTCTCCTGGCCGCGGATCTCCCATGCCACGTGGTCCATGGTCTGTG GTCCAGGTCTGTGGCCGCGGGATGGGTGCTCGCTTGGTCCATGGGTTTG TGGCATTTTTCTCCATGTGCCGTAACATTTCTGCTGATGTCCATGAGG TGGCACGGCGTCAGCGGGAGACTGACGACAATCTTCTGTCGTCAAGCTT CCTCCATGGGCATGCCTGTAGGAAACGGGTGACGCCTAAGAGGGGGGG GGGTGAATTAGGACTTCTAAAACCTTTCGCTAAACTAGGCCACAATTAA ATCCCTAGAGCAAACCTATGCAAATAGTCAAACCTAGAATGTGCAAAC TAGGTTTTGTCTAAGTGTGCTATCTCTACCGCAAAGGCTAAGTTTCA ATCTACACTAGATATGTATGAATACAAGATTGAACTTAAATGCTTAA TATAAATGCGGAAACCTTAAAGAGCAAAGTAGAGAAGCAAACCTCTCGTG GATGACGCCGTATTTTTTACCGAGGTATCCGGAACCGTGCAAGGTCCC GACTAATCCTCGTTGGTGCCCTACGCGAAGGGAAGCCACGCGAGGG CCAAGCACCTCGGTGAGTAACCTCCGTATTGAGCCGCGGGCCTTCTCC ACACGCAAGTGGTGTCCGCTTCCGGCTCCTCTCGGACGCTCCCCGCC GTCTCCACTATCGAGCTTCCGGCCGAAACGCCGCGGGCCTCGTTCCCT CCGGTACACGGTGGCGGCCGTGACACAAACGCGGTTGTACGGTCTCG CAAGACTCTCGCCCCACTCGGTACAATTACAATGGCTCACGCAAGAGC CGAGGGGTTGTGAGGTTTATCTAAAACCTCACTCAACAATCTAGGGTTCC CCTAGAGCAAGCGCTAAAGCGGTCTAACTAACCTAAGCACTTCGAAA GCACCTACGCTAATCACCGAATGATTCTATTAAGCACTTGGGTGTTTG AGCACTTGGAATGTCTACAATATGCCTTGGTATGTTGCTTGGACTCC CACACCTCCAAATAGCCGGTTGGGTGAAGTATATATAGGCCCAACTC AAATATAGTCGTTGGAAAGCAGTTCTGCAGCTTCTGCGGCGCACCGG ACAGGTGCACCGGACTGTCCGGTGGTGCACCGGACATCCACGTGTAC TAGCCGTTAAAAGTAGCCGTTTGTAGCTTCCGAAGATGGCGCACCGGAC	
---	--

ATGGCGCACCGGACGGTCCGGTGTGCACCGGACAGCCCATGTCCGCTT GTCATATTTTGGCCATAACTTTTAGCTCCAAACTCCGATTTTCGATGAT CTGTACTTTTTGGAAAGCTTATGAAATTTCTCTACATCCCAGAGTTGA AGCCATATCAGTTTGAGCAAATTTGGCACACCGGACAGGCCAGTGTGC ACCGGACATCCCATATGCTATGCAGTTTTTCAGCAATTCCGACTTCGT ATTGTGGGCATAACTTTTAGCTCCGAACCTCCGATTTTGACGATCTTGG ACTTTTTGGAAAGCTTATAAAATTTGTCTACATCTCAGAGTTAAAGCCA TGTCAGTTTGAATAGATTTGATGTTGTGAAACACTTCCAATTCAGTCA GCCCTTCCTCTCAACTTTATCAACTTCACTTTTGTGTTGTAGCTTTGT ATCTCACTTCTACTTCTTGATGTGTTGGACTCTTAGCATATATAAAGT GTCTTAAATATTACTTTGTAATGTCTTCTATGAGTCTTATAATGTCTT CTTTGAGGTGTTGCATCCTCAGAGCCTCAGTCCAATCCACTTCGCATC CTGTGAACACACACAAATACTTGGGAAATACATTAGTCCACAGG TTATGTTGATCATCAAACACCAAAACCTATTGAGCTAAATGGCCCGGG GTCATTTTCCTTACAATCTCCCCCTTTTGGTGATTGATGACAACACA ACCAAAGCAAGCAAATAATACAATTATTTGAATGATAATATGCAATCT TCTTGCTAAGATGCATGTTTGTCCCCGTAATGTGAGATTATGGACTTA AGCCTCCCCCTAACTCCATAATTCACTTACTTCTTATTTTAGACCAAA ATGCAAAAACCACTTGAAAGATCAAATTTTAAATTTGGTGGTGTGTTGG TGGTTGATAACATTTTCATTGCTTTGGTCCCTACAACCTTCTCCCCCTT TGGCATCAACCACCGAAAAGGAAGACATTAAAAGCATATAGGAAGCAA ATAAGACTTGCCCTCAAAA- GGATTTGTAAAATAATACCAATTGAAACACCAAATTGATACTCCCCCT AAATGAGTGTTCTCCTTTAGAAAGAATTTCTCCCCC- TGTAGAGACGAAGAAATACTCCCCC- TGACAGAGATCTTCTACTTAGGAAAAAGCATGTGAGAAATGGAGGTGC AAGTCATGATTTGAATAGACTAACTTCCCTTATGCATAGGTAACAAAA TATGAGTTAACAATATATGCATTTATAGCAACATGGGAAGTATAAGAT ATGAAATATAGTCTAATCAAATAATGGAGAAGACATGTAAATGTTACT AAGTGATGCTCAAAGATACCAATTGAAAGTCAACTGCGGACCAATTG AATACCAATGGCAGGCTTGCAGACATAAGAATTTCTGTAGATTTGCAG TGGAAGCTTGTGTCCTTCTCCGGGGACTTCATTTTCCTTACAATGAG ACTACTACATGAAATAGACTTGAAAACAGATATTAGTCTCAAAGACTT AGATTATAGAGTAATCTCCCCCTGAAAGTGTGCATACAAGTTTGAAT ACTTGTAGGAGACATGCATTTTATTTGATATTAAGAGGGGCAGATTA TCCATAATCCGAGCTTTGGCACATATAAGTTAAATGATACCAATTGTA GGATATAAGCTTTTGATGCATTAGTGTCACTTACTTAGGGATGTTAAA TAAGCTATGTGTCGTGCTTAATTAAACATTTTAAACCATGCAGGTTTG CTAAGAATATGAATTGAGAACGAGTTACCCTACCATATGATTTACCT AGTATGCATGCATTGTTAATCTAAAGTAAGTACCAAATGAAATACTAG GCATGAAAGGTAAAACCTAGATACAAGAAAATAGATACGCATATTTAAA AATAAGGAATACAATAAATCTAGTTACCTTAGTCATGGGTGGGGAAAT TTGGTTCCAAAATATTCACTAACCCTTGGCAATAAACTTGATCCAA TGTGAAGTATCCCATGATATACATCCTAACTTTGCCAAGTCTCCAAAT TTCTCGTTCCTCCTTTGGTCTACTCACTTCCCTTTGGGATCCAAACC TTCTTGGTGCTTTTCATGGTTGTAATTATTTCTTTTGGCACCCAGATG CGTTTGGCCCCCTTTCTTTTGTGGCTTGCTTGTTGGCCTTGATTGCT ATCACCTTTCCATTCTTTTTCTTTTTGATCAAATATGGTGTAGCAGCC TTTTTGTCCACCTTTTTGATGTAGGTGTTGGAGATTTTGCTTGATGGC TTTTGCTTGAGCTTTTGCCCTTGATTGTCCCCGGTCATCGCCTTGCAAT TGGAAAGACTTGTGGCCTTCCCTGTGGCATAGTCTACAAATCACAGTT TCTCCCTCCATAGGCTTGTCACTCCCGCAGTGGTGTATCCTGTGGA GGTCCGATTTGTTTGGCTTTGCCTTTCTTGTATACAAAGCCTTGCCA AGGCGAGCCACTTCTTGCCTTAATTGTTCACTTTTCTTTGTAACCTCA TCCGAACATGTTTCTGCAACAATATTTCTCAACACGAACTTGGTTGCAT GGGTTAGAATCATCAATTAAATCAAAGCAAGAAGTAGAAGCATCCTTT TTAATAATTTTCAAGGAATTTCAACTAAAGATGCTTCTGGGGTGCTACCA ATGTTTTCTAGCTTAGTAGTTATCTCATTGTTGTGTATGCTAAGAATT TCCATTTTCTTAGCAAATTTTCAATAGTATTTTGGGAGCTAGCTAAC TTAGCCTTGAGTTTTTCTTCTTTTAAATAAGGCCATCATTGGTAGCA GTGGATGGAGCACTAGATTCTAATAGCTCAATTTTGGTTGATAGCTCA	
---	--

CTATTTAAGTTTTCAGAAAGTTTCAAATTTTTCAAGCAAACCTTTGTAA TCATTTTGGGAGCTAATCAACTTATTTTTCAAGGTTTTAAGTTGAGCT TTTTGTTTAGTGCAAACATCTTGAAAAAATTCAACGGCATTACAAAGC TCATCTATAGAGGGCTTTCCCGCACCTTCATCATCACTACCACTTTCA TCACTAGAGTATGGAATGCTTTTGTTACCTCTTGCCATAAGGCATTTG TGTGATGACCGTGATGAGCGTGATGAGGACCGGTGGCTGCGTGTCTT GGAGGTTTCATCTTCACTTGAAGAGTCATCCCAAGTTCTAACACTTGTT AGCGCCTTGCCCTTGCTCCTCTCCTTTTTGGGTTTCGGCCATAGTTGGA CAGTTGTCCCTGAAGTGGCCCTTCTCCC- ACATGCGAAGCATCCTCTCTTCTTTGCTTTTTCTGTCAATGTTAAA GAGAAGATCTTCGACCTGCAGGGGCACACCCATTAGATTAATCTTGCG GATCATCCCCATTACCTTTCCGACGCGTCGGATTGTTTCTTCGTCTC GGAGGATGATGTGCATGGTTGATTATCTTCATCATCATCAATTTCTTC TTCTTCTCTTCTTCACTTGAGGAACCTGGAGTGGGAGCCTTCTTTTT GCCCTTCTTTTCATCACATGCAAAGCATATGGCCTTGAGGAAGTTGG CTCCTCTCCTCGACACATTTTTTCGCGACATCTCAAAGCCGCGATTTT TCCAATTACAATGGTTGGGGTCATGTTACCCAAGTCTCCATGTTGTG AAGGATAGTGATGATGCTCCCATATCTTCGTTGTGGTAGTAGGGAGAT AATCTTCTCACAATGTCCGCATCACCTAGCTTATTAATGCCAATGGA ATTGAGCTCATTGATAATTAGATTCAAACGAGAATACATGTCTCTAAC AAGTTCATCATCTTTTCATAGCAAAGAATTATACTCATTTAAGACTAG GCAATGTTTTTGCTCACGGACATTGGATGTGCCGTGATGGAGCTCATG CAATTTTAGCCAAATCTCATTAGCAGTTTTCAAGGTAAATACTTGATT AAAAATATCCATGCTAAGAGATTATACAAGCAATTTTTGGCTCTAGC ATTTAAGTGAATTTCTTTTTCTCACTCGTGGTTGGTTTCTCGGGATT CTTGAGGGGTTTCATCCCATCACGAGTGAATCTCCAAACACCTAGATC AACGGCCTCTAGGTAGCAAGCCATTCTAGCACTATAATATGGGAAGTT AGTGCCGTAGAAAGTGTGGTGGCCTATGGGTATCCATCCCTTCTCTAA AAAGCGTCGGCTCTTTTAGCAGTGAAGCTAAAACGTTTCAAATGAGCC AAACCAGGCTCTGATACCACTTGTAGGAAACGGGTGACGCCTAAGAGG GGGGGGGGTGAATTAGGACTTCTAAAACTTTTCGCTAAACTAGGCCAC AATTAATCCCTAGAGCAAACCTATGCAAATAGTCAAACCTAGAATGT GCAAACCTAGGTTTTGTCTAAGTGTGCAATCTCTACCGCAAAGGCTAA GTTTCAATCTACACTAGATATGTATGAATACAAGATTGAACTTAAAT ACTTAATATAAATGCGGAACTTAAAGAGCAAAGTAGAGAAGCAAACCT CTCGTGGATGACGCCGGTATTTTTACCGAGGTATCCGGAACCGTGCAA GGTCCCGACTAATCCTCGTTGGTGCCCTACGCGAAGGGAAGCCACGC GAGGGCCAAGCACCTCGGTGAGTAACTCCGTATTGAGCCGCGGGCCT TCTCCACACGCAAGTGGTGCTCCGCTTCCGGCTCCTCTCGGACGCTCC CCGCCGTCTCCACTATCGAGCTTCCGGCCGAAACGCGCGGGCCTCGT TCCCTCCGGTACACGGTGGCGGCCGTGACACAAACGCGGTTGTACGG TCTCGCAAGACTCTCGCCCCACTCGGTACAATTACAATGGCTCACGCA AGAGCCGAGGGGTTGTGAGGTTTATCTAAACTCACTCAACAATCTAGG GTTCCC- TAGAGCAAGCGCTAAAGCGGTCTAACTAACCTAAGCACTTCGCAAAGC ACCTACGCTAATCACCGAATGATTCTATTAAGCACTTGGGTGTTTGAG CACTTGGAATGTCTACAATATGCCTTGGTATGTTGCTTGGAATCCCA CACCTCCAAATAGCCGGTTGGGTGAAGTATATATAGGCCCCAACTCAA ATATAGTCTGTTGGAAAGCAGTTCTGCAGCTTTCTGCGGCGCACCGGAC AGGTGCACCGGACTGTCCGGTGGTGCACCGGACATCCCACGTGTACTA GCCGTTAAAAGTAGCCGTTTGAGCTTCCGAAGATGGCGCACCGGACAT GGCGCACCGGACGGTCCGGTGTGCACCGGACAGCCCATGTCCGCTTGT CATATTTTGGCCATAACTTTTAGCTCCAACTCCGATTTTCGATGATCT TGTACTTTTTTGAAAGCTTATGAAATTTCTTACATCC- AGATTGAAGCCATATCAGTTTGAGCAAATTTGGCACACCGGACAGGC CAGTGTGCACCGGACATCCCATATGCTATGCAGTTTTTTCAGCAATTCC GACTTCGTATTGTGGGCATAACTTTTAGCTCCGAACCTCCGATTTTGAC GATCTTGGACTTTTTTGAAAGCTTATAAAATTGTCTACATCTCAGAGT TAAAGCCATGTCAGTTTGAATAGATTGATGCTGTGAAACACTTCCAA TTCAGTCAGCCCTTCTCTCAACTTTATCAACTTCACTTTTGTGTTTGT AGCTTTGTATCTCACTTCTACTTCTTGATGTGTTGGACTCTTAGCATA	
--	--

TATAAAGTGTCTTAAATATTACTTTGTAATGTCTTCTATGAGTCTTAT AATGTCTTCTTTGAGGTGTTGCATCCTCAGAGCCTCAGTCCAATCCAC TTCGCATCCTGTGAACACACACAAATACTTGGGAAATACATTAG TCTACAGGTTATGTTGATCATCAAACACCAAAACCTATTGAGCTAAAT GGCCCGGGTCATTTTCCTTACAATGCCTTTTGTCCCTCGCTCGCCTG ATGTGCCTCTCCATCCTCCTCCCC- AGAGATTAATGAGTGGCACCAGCAGGCTTACGGG- TGCCTTTTATGTCTGCAGACGATGACGAAGAAGAAGCCTATGTTGATG ATCGCGAGGAGTTTGGCCCGCCTCCCTACCATGGGGATCCGGGTCAAT CATCCTCCCATCCTCCGCCTCCTTACCCAGGTCCGGGACCCATTCCCTG GTGACCCTAGGCCATCCGGCTCTGGTTACCACCACCACCCTCCACCTC CTCCTGATCATCGTACCTCTTCGGCATCTGTTTTTCTGCGGCTGATG AGCCACCTCGGGACTCCACTTTTGAGGAGGACGCGATGAGGACTTTTT TCCCTGACTACACTTCATATCCTCCATCATATCCTCCGTATATCCTC CTCCTGGAGGATATTGATTCTTGTATCCTTGTCTCTCTCTCCGTTTTT GGTACTTGTGCCAAAAGGGGGAGAAAATCCAACCTCTGGTTTGTGG TTCTTTTCGGTTTTCTTTTGGATTATTTTTGTTTTGGCCACTGTTGTG GCTTGTACCTTTTTTTTTTACCATTGTATGAAATAAAACTGTTGGATTGC TATTCAAAGTAATATGATGGTCTTCTGGCCATCATCTTTTGTCTATC AGTTTGTAAATGACTACTCTGCTATGAAGTAAAAATAATGTGTTTATAG TCATGTGCATCACAACCTCCTGTTATGACACTCTTGACCCCTCGGATG ATTGTATTTAGTATGGTTTTTGTCACTATCTCTAAATGTGTTGCTCAC ATGACATATTATGTCAAGACGTTGGATTCACAATCATGCACACATTTA GGG- AGCCATCTATATACAATCATTCAAAAGAAATTTTCTATTGCAAATATA TATCTTTTTGACCTTAATTGTTTTGGCATCAATCACCAAAAAGGGG- AGATTGTAAGTGAATCAACCC- AAGTGAGGGTTTTGGTGATTTAATGACAAAACAAATAAGGGTACTAAC AGCTTTTGTCTCAGTGTTTGATTAAAAGGAAACATGAACCTAAGGAA GCACCAAGGACTTCATGCAATGGACGTATAAAATTTATGTAAATCTT GTGTTGCATGATGTAATTCTTCCTTATTCTTTTCCGTACAGGTACAGG TGTTTGTATAGCTCAAGTCTCTAATTCAAAGCTATTTTTGAAAAC CAAAAATCACTTTAACTTTATTTGGCTGACCATGGTTTGGGTTGAGAA ACCTAGAGTGTTCTTGCTGAAAACAGCTGAACCTCTTCAACTGCAGC TGAGCTTTCCAGCTGAACCTTAACTTCAGCTGTGATGAGCTTCTTCAGC TGCAGCTGAGCTTTTTCAGCTGAGCTGGACTTCAGCTGAGTTGAACCTT CAACTCCAGCTGAACCTTCAACTTCAGCTATGGACCTCTCTGACCATAC CCAGCTGAACCTGCCCCAGGGCAGCTGAGCTGGGGTTTTCTCTCCAAA ACTCTCTGGTCTAGACCAGCTGAACCTGGTCCTGGGG- CAGCTCAACTGGTCTCTGACCCTCTGACTTCTGTTCTGCAGTCTGCCA GTCAGACTGGCAGTCAGGTGCCAGGGGTGTCAGGGG- CGGTTCAACCGCCCTAGGGGG- CGGTTCAACCGGTCTTGCCCTGTCTGACCCGCCTGCAGGTGAGTCTGC CAGTCAGTCTGGCAGTCAGACTGGCAGACAGACTGCTGACCTGTCAGG GG-CGGTTCAACCGCCCTTGGGG- CGGTTCAATCGGTTTTTCAGCGTCGTTTTTTGTCAAACGGCTAGTTTTT GAGCCGTGACTATAAATAGACCCCTCTCTCTCTCTTCTTTAATACGGC TGGACACTCACTCATTTATTGCTCACCTAACTTGATACAAAGCACTCA TCTCTCCATCTCTCTCACACTTAAATCTTCCTCAAGATTCAACCTTGT TGGAGGAGCTCTTGGGTGTGAGGTTTGTGCTTGTGCTCACGATTTGGT TTTCTCTCCCATCTCTCTTACAAAGACTTGAGCTTGTGCAAACCCCT CTTGTGCGTTCTTGGTGTTCTTGAAACCCTAGGTTTCAAGATCTTTGA GGTTGCTTGGGAGTCTCCAAATTTGTGGACGACCCTAAGAAGTTTGT TCAACCGCTCTTTGAGCTAAGATAAGAAGAGATTGCCTTGACCCTTGT GGTCGGCTTTTGGAGGATTAGGGTTGAAAAAGACCCGGCCCTTTGTGG - CTCCTCAACGGGGAGTAGGACACCTTTGTGGTGTGGCCGAACCTCGGA TTAAATCTTGTATCTTGTGTTCTTGCTTGTGTTTACTTGAGATATTTT TACTTGCAAGGTTGACATAAAGATTTTTCCGTAGAGCTTATCTGGAAG TAAATAGCCCTTACATATTCATCTTTTCATCCTCACTTAGCTATATC TTACTTCTCTCAAGAACTTGCGAAATACTTTTTTCGAGTAGATTTTAGT	
--	--

	CCAAAGTTTATAAAACAGTTGGATTGGTTCAGAGTGGTTCAACTTGCG CACGTAGCTGTTGAACCGGCCTGCACCAGTCGAACTGTGTATCTAACA ATCTTTCGAAGTTTGTGTGAAAATTTTCAGATTAGCCTATTCACCCC CCCC-TCTAGGCTACTTTCACCTTCT	
--	---	--

Also envisaged herein is a plant which can be or is produced with the preceding method, including descendants of the plant which have the at least one mutation, or a part of this plant. A part of the plant may be a fertilized or unfertilized seed, an embryo, a pollen, a tissue, an organ, or a plant cell, wherein the fertilized or unfertilized seed, the embryo, or the pollen are generated at the plant obtainable by the preceding method, and the at least one mutation is present in its genome.

A non-limiting example of a plant which can be produced with the preceding method is a plant, preferably *Zea mays* or *Sorghum bicolor*, which:

has at least one mutation in an endogenous DNA sequence that is identical to a nucleic acid that (i) has a sequence selected from SEQ ID NOs: 1, 5, and 12, or a functional fragment thereof, preferably SEQ ID NO: 1; or (ii) has a coding sequence selected from SEQ ID NO: 2, 6 and 13, preferably SEQ ID NO: 2; or (ii) is complementary to the sequence from (i) or (ii); or (iii) is at least 80% identical to the sequence from (i) or (ii); or (iv) encodes for a protein having the amino acid sequence selected from SEQ ID NOs: 3, 7 and 14, preferably SEQ ID NO: 3, or a functional part of the protein; or (v) encodes for a protein having an amino acid sequence that is at least 80% identical to the amino acid sequence selected from SEQ ID Nos: 3, 7 and 14, preferably SEQ ID NO: 3, or a functional part of the protein; or (vi) is identical to a sequence hybridized under stringent conditions with a sequence from (iii), or

has at least one mutation in a regulatory sequence of said endogenous DNA sequence,

wherein said mutation produces an increase in the transcription or expression rate of said endogenous DNA sequence in the identified plant, in comparison to a non-mutagenized wild-type plant, or an increase in the activity or stability of a protein or polypeptide encoded by the endogenous DNA sequence in the identified plant, in comparison to a non-mutagenized
5 wild-type plant,

wherein the at least one mutation causes the property of an enhancer of the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) in the identified plant.

The mutation is preferably an alteration in the coding sequence of SEQ ID No: 1, 5 or 12 (for
10 example, a point mutation), i.e. an alteration in SEQ ID NO: SEQ ID NO: 2, 6 or 13, which causes an amino acid exchange at the amino acid position 325 of SEQ ID No: 3, 7 or 14. This may here involve a mutation according to SEQ ID NO: 10. The mutation caused by TILLING in SEQ ID NO: 10 causes an amino acid exchange in the encoded amino acid at position 325, wherein the alanine is replaced by valine (A325V).

15 Accordingly, also disclosed herein is a plant, preferably a *Zea mays* or *Sorghum bicolor* plant, comprising:

- i) the nucleotide sequence according to SEQ ID NO: 10, or
- ii) a nucleotide sequence encoding the amino acid according to SEQ ID NO: 11, or
- iii) a nucleotide sequence encoding a myosin having at position 325 referenced to
20 SEQ ID NO: 11 or 14 an amino acid different to alanine, preferably valine.

In embodiments, the methods for the production of a transgenic or mutated plant, such as a plant which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) described herein, may further comprise the steps of:

- 25 - crossing the plant, in particular the transgenic or mutated plant, which is carrier of an enhancer for the haploid induction capability of a haploid inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) obtained by the method with a haploid inducer plant preferably comprising a mutated patatin-like phospholipase (PLA1); and
selecting a plant from the progeny which comprises A) the haploid inducer phenotype,
30 preferably the mutated patatin-like phospholipase (PLA1) and B) the introduced nucleic acid in the method for producing a transgenic carrier plant or the at least one mutation created in the method for producing a mutated carrier plant.

The selected plant may be suitable for use as a haploid inductor. Accordingly, the present

invention also relates to a method for producing a plant suitable for use as a haploid inducer comprising the following steps:

- producing a plant, in particular a transgenic plant or a mutated plant, which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) according to a method described herein;
- crossing said plant with a plant preferably comprising a mutated patatin-like phospholipase (PLA1); and
- selecting a plant from the progeny which comprises A) the haploid inducer phenotype, preferably the mutated patatin-like phospholipase (PLA1) and B) the introduced nucleic acid in the method for producing a transgenic carrier plant described herein or the at least one mutation created in the method for producing a mutated carrier plant described herein.

In alternative embodiments for the methods for the production of a plant which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably caused by a mutated patatin-like phospholipase (PLA1) described herein, the nucleic acid may be introduced in a plant cell preferably comprising the mutated patatin-like phospholipase (PLA1) or a plant cell comprising the haploid inducer phenotype, preferably the mutated patatin-like phospholipase (PLA1) may be mutagenized. The plants obtained by such methods may also be suitable for use as haploid inducer.

In a further aspect, the present invention relates to a method for the production of a haploid plant, said method comprising the following steps:

- a) crossing a plant which is suitable for use as a haploid inducer as described herein with a plant of the same genus, preferably a plant of the same species,
- b) selecting a fertilized haploid seed or embryo, and
- c) generating a haploid plant from the seed or embryo selected in step b).

In the method for the production of a haploid plant, the plant which is suitable for use as a haploid inducer is preferably used as a pollen parent, and is crossed with a seed elder of the same genus, preferably of the same species. The plant which is suitable for use as a haploid inducer may also be used as a seed parent and be crossed with a pollen elder of the same genus, preferably of the same species. Both cross partners in step a), thus, seed parent and pollen parent, may also be the same individual. The crossing step then represents a selfing.

In embodiments, the selection of the haploid fertilized seed or embryo encompasses a step of the verification of the haploidy, and the separation of the haploid fertilized seed or embryo of polyploid fertilized seed or embryo. The verification of the haploidy of a fertilized seed or

embryo may take place phenotypically or genotypically, in that, for example, the inductor is provided with an embryo-specific dominant marker that is visible in all diploid descendants, but not in the induced haploid descendants. Furthermore, the ploidy status may be determined via flow cytometry. Moreover, a complete, homozygotic pattern of molecular markers provides an indication of haploid plants. For example, the separation may take place automatically on the basis of data of the verification of the haploidy.

Also envisaged herein is a haploid, fertilized seed, or embryo which is created upon crossing in step a) of the method for production of a haploid plant described herein, as well as a haploid plant which can be or is produced with this method, including its descendant, or a part of this plant.

The invention further relates to a doubled haploid (diploid) plant or a part thereof, wherein the doubled haploid (diploid) plant or a part thereof was generated by chromosome duplication of the haploid plant described herein or of the part thereof via e.g. use of a chromosome doubling agent or via spontaneous doubling.

A related aspect is directed to a method for the production of a double haploid plant, comprising the following steps:

- a) crossing a plant which is suitable for use as a haploid inductor as described herein with a plant of the same genus, preferably a plant of the same species,
- b) selecting a fertilized haploid seed or embryo,
- c) generating a haploid plant from the seed or embryo selected in step b), and
- d) doubling the set of chromosomes of the haploid plant via chromosome doubling agent or via spontaneous doubling. For example, the chromosome doubling agent may be colchicine, pronamide, dithiopyr, trifluralin, or another known anti-microtubule agent.

Another aspect relates to the use of the induction-promoting nucleic acid and the nucleic acid encoding double-stranded RNA described herein, or a vector comprising said induction-promoting nucleic acid or said induction-inhibiting nucleic acid described herein, in a haploid inducer plant, preferably a plant comprising a mutated patatin-like phospholipase (PLA1) to enhance the haploid induction capability of said plant.

Furthermore, the present invention also relates to the use of a plant which is suitable for use as a haploid inductor as described herein, to produce a haploid, fertilized seed or embryo, a haploid plant, or a doubled haploid plant.

Yet another aspect relates to RNA that has a double-stranded portion, wherein at least one strand of the double-stranded portion has a nucleotide sequence which is homologous or

identical to at least 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25, preferably at least 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, or 140, more preferably at least 160, 180, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, or 1000 successive nucleotides of a coding sequence of a nucleic acid that:

- 5 (i) has a sequence selected from the group consisting of SEQ ID Nos: 25, 28 and 33, or a fragment thereof, in a sense or anti-sense orientation; or
- (ii) is complementary to the sequence from (i); or
- (iii) is at least 80%, 82%, 84%, 86%, or 88%, preferably at least 90%, 91%, 92%, 93%, 94%, 95%, or 96%, more preferably at least 97%, 97.5%, 98%, 98.5%, 99%, or 99.5% identical to the
- 10 sequence from (i); or
- (iv) hybridizes with the sequence from (ii) under stringent conditions.

Such RNA is suitable for external application to plants and is suitable for enhancing the induction capability of a haploid inductor plant caused by a mutated patatin-like phospholipase (PLA1). The application preferably occurs at the point in time of the anther

15 formation, pollen formation, or fertilization. The RNA may be used by being sprayed in the form of a spray, or via additional ways of external application that are commonplace to the person skilled in the art, onto the plant tissue, or by spraying or mixing with additional additives before or after the flowering of the plant. For example, additives may be wetting agents, carrier substances, or RNA stabilizers, e.g., liposomes

20 Double-stranded RNA may be produced *in vitro* by means of methods known to the person skilled in the art. For example, the synthesis of double-stranded RNA may take place synthetically, wherein the RNA is formed directly *in vitro*. Starting from a double-stranded DNA, double-stranded RNA may also be synthesized via the formation of an mRNA transcript, which then forms a hairpin structure, for example.

25 In an aspect, the invention relates to a plant, preferably a plant of the genus *Zea* or *Sorghum*, preferably *Zea mays* or *Sorghum bicolor*, comprising

- i) the nucleotide sequence according to SEQ ID NO: 10,
- ii) a nucleotide sequence encoding the protein according to SEQ ID NO: 11, or
- iii) a nucleotide sequence encoding a myosin heavy chain having at position 325 referenced to
- 30 SEQ ID NO: 11 or 14 an amino acid different than alanine, preferably valine.

In an aspect, the invention relates to a plant, preferably a plant of the genus *Zea* or *Sorghum*, preferably *Zea mays* or *Sorghum bicolor*, comprising

- i) the nucleotide sequence according to SEQ ID NOs: 38 or 39, or

ii) a nucleotide sequence encoding the protein according to SEQ ID NO: 40.

The skilled person will understand that SEQ ID NOs: 38-40 are Zea mais sequences, and that corresponding sequences, in particular comprising the corresponding mutations, may be used from different plant species, such as Sorghum.

5 In an aspect, the invention provides a method of editing a plant's genomic DNA or RNA (such as mRNA, microRNA, etc) or otherwise targeting or modifying such (for instance transcriptional or translational activation or repression, epigenomic modification, etc). This is done by taking a first plant— which is a haploid inducing plant and which also has encoded into its DNA the machinery necessary for accomplishing the editing, targeting, or modification (for example, a
10 Cas9 enzyme and a guide RNA)— and using that first plant's pollen to pollinate a second plant. The second plant is the plant to be edited. From that pollination event, progeny (e.g., embryos or seeds) are produced; at least one of which will be a haploid seed. This haploid seed will only contain the chromosomes of the second plant; the first plant's chromosomes have vanished (having been eliminated, lost or degraded), but before doing so, the first plant's chromosomes
15 permitted the gene-editing, targeting, or modification machinery to be expressed. Alternately, and without wishing to be bound by theory, the first plant delivers the already- expressed editing, targeting or modification machinery upon pollination via the pollen tube. Or, in the case that the haploid inducer line is the female in the cross, the haploid inducing plant's egg cell contains the editing machinery that is present and perhaps already being expressed, upon
20 fertilization with the "wild type" or non-haploid inducing pollen grain. Through any of these routes, the haploid progeny obtained by the cross will also have had its genome or transcriptome edited, targeted, or otherwise modified.

Simultaneous editing plus haploid induction can be done in almost any crop via wide cross or de novo haploid induction for instance via CENH3 mutation (i.e., CENH3 -modified haploid
25 inducer; see, e.g., WO 2017/004375, incorporated herein by reference in its entirety) or via lipid spray (see WO 2017/087682, incorporated herein by reference in its entirety).

In one aspect, the editing machinery is any DNA or RNA modification enzyme, but is preferably a site-directed nuclease. The site-directed nuclease is preferably CRISPR-based, but could also be a meganuclease, a transcription-activator like effector nuclease (TALEN), or a zinc finger
30 nuclease. The nuclease used in this invention could be Cas9, CfpI, dCas9- FokI, chimeric FENI-FokI. In one aspect, the DNA or RNA modification enzyme is a site-directed base editing enzyme such as Cas9-cytidine deaminase or Cas9-adenine deaminase, wherein the Cas9 can have one or both of its nuclease activity inactivated, i.e. chimeric Cas9 nickase (nCas9) or deactivated Cas9 (dCas9) fused to cytidine deaminase or adenine deaminase. The optional

guide RNA targets the genome at the specific site intended to be edited. In certain embodiments, the modification enzyme is mutated, such as catalytically inactive or a nickase. In certain embodiments, the modification enzyme is fused to a heterologous domain, preferably a functional heterologous domain. In certain embodiments, the 'functional)

5 heterologous domain comprises base editing activity, nucleotide deaminase activity, methylase activity, demethylase activity, translation activation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, chromatin modifying or remodeling activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity,

10 single-strand DNA cleavage activity, double-strand DNA cleavage activity, and nucleic acid binding activity. In certain embodiments, the modification enzyme is modified. In certain embodiments, the modification enzyme is modified, such as to increase or decrease target recognition specificity, increase or decrease nuclease activity, increase or decrease stability, and/or alter PAM recognition.

15 "Gene editing" or "genome editing" refers to genetic engineering in which in which DNA or RNA is inserted, deleted, modified or replaced in the genome of a living organism. Gene editing may comprise targeted or non-targeted (random) mutagenesis. Targeted mutagenesis may be accomplished for instance with designer nucleases, such as for instance with meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector-based nucleases (TALEN), and

20 the clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) system. These nucleases create site-specific double-strand breaks (DSBs) at desired locations in the genome. The induced double-strand breaks are repaired through nonhomologous end-joining (NHEJ) or homologous recombination (HR), resulting in targeted mutations or nucleic acid modifications. The use of designer nucleases is particularly suitable for generating gene knockouts or

25 knockdowns. In certain embodiments, designer nucleases are developed which specifically induce a mutation in the gene, as described herein elsewhere, such as to generate a mutated gene or a knockout of the gene. In certain embodiments, designer nucleases, in particular RNA-specific CRISPR/Cas systems are developed which specifically target the mRNA, such as to cleave the mRNA and generate a knockdown of the gene/mRNA/protein. Delivery and

30 expression systems of designer nuclease systems are well known in the art.

In certain embodiments, the nuclease or targeted/site-specific/homing nuclease is, comprises, consists essentially of, or consists of a (modified) CRISPR/Cas system or complex, a (modified) Cas protein, a (modified) zinc finger, a (modified) zinc finger nuclease (ZFN), a (modified) transcription factor-like effector (TALE), a (modified) transcription factor-like effector nuclease

(TALEN), or a (modified) meganuclease. In certain embodiments, said (modified) nuclease or targeted/site-specific/homing nuclease is, comprises, consists essentially of, or consists of a (modified) RNA-guided nuclease. It will be understood that in certain embodiments, the nucleases may be codon optimized for expression in plants. As used herein, the term

5 "targeting" of a selected nucleic acid sequence means that a nuclease or nuclease complex is acting in a nucleotide sequence specific manner. For instance, in the context of the CRISPR/Cas system, the guide RNA is capable of hybridizing with a selected nucleic acid sequence. As used herein, "hybridization" or "hybridizing" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the

10 bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogsteen binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PGR, or the

15 cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence.

Gene editing may involve transient, inducible, or constitutive expression of the gene editing components or systems. Gene editing may involve genomic integration or episomal presence of the gene editing components or systems. Gene editing components or systems may be

20 provided on vectors, such as plasmids, which may be delivered by appropriate delivery vehicles, as is known in the art. Preferred vectors are expression vectors.

Gene editing may comprise the provision of recombination templates, to effect homology directed repair (HDR). For instance a genetic element may be replaced by gene editing in which a recombination template is provided. The DNA may be cut upstream and downstream

25 of a sequence which needs to be replaced. As such, the sequence to be replaced is excised from the DNA. Through HDR, the excised sequence is then replaced by the template. In certain embodiments, the genes of the invention as described herein may be provided on/as a template. By designing the system such that double strand breaks are introduced upstream and downstream of the corresponding region in the genome of a plant not comprising the

30 gene, this region is excised and can be replaced with the template comprising the gene of the invention. In this way, introduction of the gene of the invention in a plant need not involve multiple backcrossing, in particular in a plant of specific genetic background. Similarly, the gene of the invention may be provided on/as a template. More advantageously however, the gene of the invention may be generated without the use of a recombination template, but

solely through the endonuclease action leading to a double strand DNA break which is repaired by NHEJ, resulting in the generation of indels.

In certain embodiments, the nucleic acid modification or mutation is effected by a (modified) transcription activator-like effector nuclease (TALEN) system. Transcription activator-like effectors (TALEs) can be engineered to bind practically any desired DNA sequence. Exemplary methods of genome editing using the TALEN system can be found for example in Cermak T. Doyle EL. Christian M. Wang L. Zhang Y. Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 2011;39:e82; Zhang F. Cong L. Lodato S. Kosuri S. Church GM. Arlotta P Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol.* 2011;29:149–153 and US Patent Nos. 8,450,471, 8,440,431 and 8,440,432, all of which are specifically incorporated by reference. By means of further guidance, and without limitation, naturally occurring TALEs or “wild type TALEs” are nucleic acid binding proteins secreted by numerous species of proteobacteria. TALE polypeptides contain a nucleic acid binding domain composed of tandem repeats of highly conserved monomer polypeptides that are predominantly 33, 34 or 35 amino acids in length and that differ from each other mainly in amino acid positions 12 and 13. In advantageous embodiments the nucleic acid is DNA. As used herein, the term “polypeptide monomers”, or “TALE monomers” will be used to refer to the highly conserved repetitive polypeptide sequences within the TALE nucleic acid binding domain and the term “repeat variable di-residues” or “RVD” will be used to refer to the highly variable amino acids at positions 12 and 13 of the polypeptide monomers. As provided throughout the disclosure, the amino acid residues of the RVD are depicted using the IUPAC single letter code for amino acids. A general representation of a TALE monomer which is comprised within the DNA binding domain is X1-11-(X12X13)-X14-33 or 34 or 35, where the subscript indicates the amino acid position and X represents any amino acid. X12X13 indicate the RVDs. In some polypeptide monomers, the variable amino acid at position 13 is missing or absent and in such polypeptide monomers, the RVD consists of a single amino acid. In such cases the RVD may be alternatively represented as X*, where X represents X12 and (*) indicates that X13 is absent. The DNA binding domain comprises several repeats of TALE monomers and this may be represented as (X1-11-(X12X13)-X14-33 or 34 or 35)_z, where in an advantageous embodiment, z is at least 5 to 40. In a further advantageous embodiment, z is at least 10 to 26. The TALE monomers have a nucleotide binding affinity that is determined by the identity of the amino acids in its RVD. For example, polypeptide monomers with an RVD of NI preferentially bind to adenine (A), polypeptide monomers with an RVD of NG preferentially

bind to thymine (T), polypeptide monomers with an RVD of HD preferentially bind to cytosine (C) and polypeptide monomers with an RVD of NN preferentially bind to both adenine (A) and guanine (G). In yet another embodiment of the invention, polypeptide monomers with an RVD of IG preferentially bind to T. Thus, the number and order of the polypeptide monomer repeats in the nucleic acid binding domain of a TALE determines its nucleic acid target specificity. In still further embodiments of the invention, polypeptide monomers with an RVD of NS recognize all four base pairs and may bind to A, T, G or C. The structure and function of TALEs is further described in, for example, Moscou et al., *Science* 326:1501 (2009); Boch et al., *Science* 326:1509-1512 (2009); and Zhang et al., *Nature Biotechnology* 29:149-153 (2011), each of which is incorporated by reference in its entirety.

In certain embodiments, the nucleic acid modification or mutation is effected by a (modified) zinc-finger nuclease (ZFN) system. The ZFN system uses artificial restriction enzymes generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain that can be engineered to target desired DNA sequences. Exemplary methods of genome editing using ZFNs can be found for example in U.S. Patent Nos. 6,534,261, 6,607,882, 6,746,838, 6,794,136, 6,824,978, 6,866,997, 6,933,113, 6,979,539, 7,013,219, 7,030,215, 7,220,719, 7,241,573, 7,241,574, 7,585,849, 7,595,376, 6,903,185, and 6,479,626, all of which are specifically incorporated by reference. By means of further guidance, and without limitation, artificial zinc-finger (ZF) technology involves arrays of ZF modules to target new DNA-binding sites in the genome. Each finger module in a ZF array targets three DNA bases. A customized array of individual zinc finger domains is assembled into a ZF protein (ZFP). ZFPs can comprise a functional domain. The first synthetic zinc finger nucleases (ZFNs) were developed by fusing a ZF protein to the catalytic domain of the Type IIS restriction enzyme FokI. (Kim, Y. G. et al., 1994, Chimeric restriction endonuclease, *Proc. Natl. Acad. Sci. U.S.A.* 91, 883-887; Kim, Y. G. et al., 1996, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.* 93, 1156-1160). Increased cleavage specificity can be attained with decreased off target activity by use of paired ZFN heterodimers, each targeting different nucleotide sequences separated by a short spacer. (Doyon, Y. et al., 2011, Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* 8, 74-79). ZFPs can also be designed as transcription activators and repressors and have been used to target many genes in a wide variety of organisms.

In certain embodiments, the nucleic acid modification is effected by a (modified) meganuclease, which are endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs). Exemplary method for using

meganucleases can be found in US Patent Nos: 8,163,514; 8,133,697; 8,021,867; 8,119,361; 8,119,381; 8,124,369; and 8,129,134, which are specifically incorporated by reference.

In certain embodiments, the nucleic acid modification is effected by a (modified) CRISPR/Cas complex or system. With respect to general information on CRISPR/Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, and making and using thereof, including as to amounts and formulations, as well as Cas9CRISPR/Cas-expressing eukaryotic cells, Cas-9 CRISPR/Cas expressing eukaryotes, such as a mouse, reference is made to: US Patents Nos. 8,999,641, 8,993,233, 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418, 8,895,308, 8,906,616, 8,932,814, 8,945,839, 8,993,233 and 8,999,641; US Patent Publications US 2014-0310830 (US App. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); US 2015-0184139 (U.S. App. Ser. No. 14/324,960); 14/054,414 European Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), and EP 2 784 162 (EP14170383.5); and PCT Patent Publications WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO 2014/093701 (PCT/US2013/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US2014/041790), WO 2014/204724 (PCT/US2014/041800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US2014/041804), WO 2014/204727 (PCT/US2014/041806), WO 2014/204728 (PCT/US2014/041808), WO 2014/204729 (PCT/US2014/041809), WO 2015/089351 (PCT/US2014/069897), WO 2015/089354 (PCT/US2014/069902), WO 2015/089364 (PCT/US2014/069925), WO 2015/089427

(PCT/US2014/070068), WO 2015/089462 (PCT/US2014/070127), WO 2015/089419 (PCT/US2014/070057), WO 2015/089465 (PCT/US2014/070135), WO 2015/089486 (PCT/US2014/070175), PCT/US2015/051691, PCT/US2015/051830. Reference is also made to US provisional patent applications 61/758,468; 61/802,174; 61/806,375; 61/814,263; 5 61/819,803 and 61/828,130, filed on January 30, 2013; March 15, 2013; March 28, 2013; April 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to US provisional patent application 61/836,123, filed on June 17, 2013. Reference is additionally made to US provisional patent applications 61/835,931, 61/835,936, 61/835,973, 61/836,080, 61/836,101, and 61/836,127, each filed June 17, 2013. Further reference is made to US provisional patent 10 applications 61/862,468 and 61/862,355 filed on August 5, 2013; 61/871,301 filed on August 28, 2013; 61/960,777 filed on September 25, 2013 and 61/961,980 filed on October 28, 2013. Reference is yet further made to: PCT/US2014/62558 filed October 28, 2014, and US Provisional Patent Applications Serial Nos.: 61/915,148, 61/915,150, 61/915,153, 61/915,203, 61/915,251, 61/915,301, 61/915,267, 61/915,260, and 61/915,397, each filed December 12, 15 2013; 61/757,972 and 61/768,959, filed on January 29, 2013 and February 25, 2013; 62/010,888 and 62/010,879, both filed June 11, 2014; 62/010,329, 62/010,439 and 62/010,441, each filed June 10, 2014; 61/939,228 and 61/939,242, each filed February 12, 2014; 61/980,012, filed April 15, 2014; 62/038,358, filed August 17, 2014; 62/055,484, 62/055,460 and 62/055,487, each filed September 25, 2014; and 62/069,243, filed October 27, 2014. 20 Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014. Reference is made to US provisional patent application 61/930,214 filed on January 22, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014. Mention is also made of US application 62/180,709, 17-Jun-15, PROTECTED GUIDE RNAS (PGRNAS); US 25 application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US applications 62/091,462, 12-Dec-14, 62/096,324, 23-Dec-14, 62/180,681, 17-Jun-2015, and 62/237,496, 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14 and 62/180,692, 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; 30 US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, 35 METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION;

US application 62/098,059, 30-Dec-14, 62/181,641, 18-Jun-2015, and 62/181,667, 18-Jun-2015, RNA-TARGETING SYSTEM; US application 62/096,656, 24-Dec-14 and 62/181,151, 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 61/939,154, 12-Feb-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US applications 62/054,675, 24-Sep-14 and 62/181,002, 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14 and 62/181,690, 18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14 and 62/181,687, 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS. Mention is made of US applications 62/181,659, 18-Jun-2015 and 62/207,318, 19-Aug-2015,

ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION. Mention is made of US applications 62/181,663, 18-Jun-2015 and 62/245,264, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US applications 62/181,675, 18-Jun-2015, and Attorney Docket No. 5 46783.01.2128, filed 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US application 62/232,067, 24-Sep-2015, US application 62/205,733, 16-Aug-2015, US application 62/201,542, 5-Aug-2015, US application 62/193,507, 16-Jul-2015, and US application 62/181,739, 18-Jun-2015, each entitled NOVEL CRISPR ENZYMES AND SYSTEMS and of US application 62/245,270, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. Mention is also made of US application 10 61/939,256, 12-Feb-2014, and WO 2015/089473 (PCT/US2014/070152), 12-Dec-2014, each entitled ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. Mention is also made of PCT/US2015/045504, 15-Aug-2015, US application 62/180,699, 17-Jun-2015, and US application 62/038,358, 17-Aug-2014, each entitled GENOME EDITING USING CAS9 NICKASES.

15 European patent application EP3009511. Reference is further made to Multiplex genome engineering using CRISPR/Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. *Science* Feb 15;339(6121):819-23 (2013); RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. *Nat Biotechnol* Mar;31(3):233-9 (2013); One-Step Generation of 20 Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. *Cell* May 9;153(4):910-8 (2013); Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. *Nature*. 2013 Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23; Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing 25 Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. *Cell* Aug 28. pii: S0092-8674(13)01015-5. (2013); DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., 30 Bao, G., & Zhang, F. *Nat Biotechnol* doi:10.1038/nbt.2647 (2013); Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. *Nature Protocols* Nov;8(11):2281-308. (2013); Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelsen, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. *Science* Dec 12. (2013). [Epub ahead of print]; 35 Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA.,

Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27. (2014). 156(5):935-49; Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. (2014) Apr 20. doi: 10.1038/nbt.2889; CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling, Platt et al., Cell 159(2): 440-455 (2014) DOI: 10.1016/j.cell.2014.09.014; Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu et al, Cell 157, 1262-1278 (June 5, 2014) (Hsu 2014); Genetic screens in human cells using the CRISPR/Cas9 system, Wang et al., Science. 2014 January 3; 343(6166): 80–84. doi:10.1126/science.1246981; Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench et al., Nature Biotechnology 32(12):1262-7 (2014) published online 3 September 2014; doi:10.1038/nbt.3026, and In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech et al, Nature Biotechnology 33, 102–106 (2015) published online 19 October 2014; doi:10.1038/nbt.3055, Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System, Zetsche et al., Cell 163, 1-13 (2015); Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems, Shmakov et al., Mol Cell 60(3): 385-397 (2015); C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector, Abudayyeh et al, Science (2016) published online June 2, 2016 doi: 10.1126/science.aaf5573. Each of these publications, patents, patent publications, and applications, and all documents cited therein or during their prosecution ("apln cited documents") and all documents cited or referenced in the apln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the apln cited documents) are incorporated herein by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

In certain embodiments, the CRISPR/Cas system or complex is a class 2 CRISPR/Cas system. In certain embodiments, said CRISPR/Cas system or complex is a type II, type V, or type VI CRISPR/Cas system or complex. The CRISPR/Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas protein can be programmed by an RNA guide (gRNA) to recognize a specific nucleic acid target, in other words the Cas enzyme protein can be recruited to a specific nucleic acid target locus (which may comprise or consist of RNA and/or DNA) of interest using said short RNA guide.

In general, the CRISPR/Cas or CRISPR system is as used herein foregoing documents refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene and one or more of, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and, where applicable, transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides.

In certain embodiments, the gRNA is a chimeric guide RNA or single guide RNA (sgRNA). In certain embodiments, the gRNA comprises a guide sequence and a tracr mate sequence (or direct repeat). In certain embodiments, the gRNA comprises a guide sequence, a tracr mate sequence (or direct repeat), and a tracr sequence. In certain embodiments, the CRISPR/Cas system or complex as described herein does not comprise and/or does not rely on the presence of a tracr sequence (e.g. if the Cas protein is Cpf1).

As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a CRISPR/Cas locus effector protein, as applicable, comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at

soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid -targeting complex to a target nucleic acid sequence may be assessed by any suitable assay.

- 5 A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be genomic DNA. The target sequence may be mitochondrial DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA,
- 10 ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scrRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred
- 15 embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

- In certain embodiments, the gRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem
- 20 loop. In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt,
- 25 from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer. In particular embodiments, the CRISPR/Cas system requires a tracrRNA. The “tracrRNA” sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when
- 30 optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and gRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a

secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop may correspond to the tracr mate sequence, and the portion of the sequence 3' of the loop then corresponds to the tracr sequence. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop may alternatively correspond to the tracr sequence, and the portion of the sequence 3' of the loop corresponds to the tracr mate sequence. In alternative embodiments, the CRISPR/Cas system does not require a tracrRNA, as is known by the skilled person.

In certain embodiments, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a target locus and (2) a tracr mate or direct repeat sequence (in 5' to 3' orientation, or alternatively in 3' to 5' orientation, depending on the type of Cas protein, as is known by the skilled person). In particular embodiments, the CRISPR/Cas protein is characterized in that it makes use of a guide RNA comprising a guide sequence capable of hybridizing to a target locus and a direct repeat sequence, and does not require a tracrRNA. In particular embodiments, where the CRISPR/Cas protein is characterized in that it makes use of a tracrRNA, the guide sequence, tracr mate, and tracr sequence may reside in a single RNA, i.e. an sgRNA (arranged in a 5' to 3' orientation or alternatively arranged in a 3' to 5' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr mate sequence. In these embodiments, the tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence.

Typically, in the context of an endogenous nucleic acid-targeting system, formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence and complexed with one or more nucleic acid-targeting effector proteins) results in modification (such as cleavage) of one or both DNA or RNA strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest). The skilled person will be aware of specific cut sites for selected CRISPR/Cas systems, relative to the target sequence, which as is known in the art may be within the target sequence or alternatively 3' or 5' of the target sequence.

In some embodiments, the unmodified nucleic acid-targeting effector protein may have

nucleic acid cleavage activity. In some embodiments, the nuclease as described herein may direct cleavage of one or both nucleic acid (DNA, RNA, or hybrids, which may be single or double stranded) strands at the location of or near a target sequence, such as within the target sequence and/or within the complement of the target sequence or at sequences associated with the target sequence. In some embodiments, the nucleic acid-targeting effector protein may direct cleavage of one or both DNA or RNA strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, the cleavage may be blunt (e.g. for Cas9, such as SaCas9 or SpCas9). In some embodiments, the cleavage may be staggered (e.g. for Cpf1), i.e. generating sticky ends. In some embodiments, the cleavage is a staggered cut with a 5' overhang. In some embodiments, the cleavage is a staggered cut with a 5' overhang of 1 to 5 nucleotides, preferably of 4 or 5 nucleotides. In some embodiments, the cleavage site is upstream of the PAM. In some embodiments, the cleavage site is downstream of the PAM. In some embodiments, the nucleic acid-targeting effector protein that may be mutated with respect to a corresponding wild-type enzyme such that the mutated nucleic acid-targeting effector protein lacks the ability to cleave one or both DNA or RNA strands of a target polynucleotide containing a target sequence. As a further example, two or more catalytic domains of a Cas protein (e.g. RuvC I, RuvC II, and RuvC III or the HNH domain of a Cas9 protein) may be mutated to produce a mutated Cas protein substantially lacking all DNA cleavage activity. In some embodiments, a nucleic acid-targeting effector protein may be considered to substantially lack all DNA and/or RNA cleavage activity when the cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the nucleic acid cleavage activity of the non-mutated form of the enzyme; an example can be when the nucleic acid cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. As used herein, the term "modified" Cas generally refers to a Cas protein having one or more modifications or mutations (including point mutations, truncations, insertions, deletions, chimeras, fusion proteins, etc.) compared to the wild type Cas protein from which it is derived. By derived is meant that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as known in the art or as described herein.

In certain embodiments, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair

sequences adjacent the protospacer (that is, the target sequence). Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme. Further, engineering of the PAM Interacting (PI) domain may allow programming of PAM specificity, improve target site recognition fidelity, and increase the versatility of the Cas, e.g. Cas9, genome engineering platform. Cas proteins, such as Cas9 proteins may be engineered to alter their PAM specificity, for example as described in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. The skilled person will understand that other Cas proteins may be modified analogously.

The Cas protein as referred to herein, such as without limitation Cas9, Cpf1 (Cas12a), C2c1 (Cas12b), C2c2 (Cas13a), C2c3, Cas13b protein, may originate from any suitable source, and hence may include different orthologues, originating from a variety of (prokaryotic) organisms, as is well documented in the art. In certain embodiments, the Cas protein is (modified) Cas9, preferably (modified) *Staphylococcus aureus* Cas9 (SaCas9) or (modified) *Streptococcus pyogenes* Cas9 (SpCas9). In certain embodiments, the Cas protein is (modified) Cpf1, preferably *Acidaminococcus* sp., such as *Acidaminococcus* sp. BV3L6 Cpf1 (AsCpf1) or *Lachnospiraceae* bacterium Cpf1, such as *Lachnospiraceae* bacterium MA2020 or *Lachnospiraceae* bacterium MD2006 (LbCpf1). In certain embodiments, the Cas protein is (modified) C2c2, preferably *Leptotrichia wadei* C2c2 (LwC2c2) or *Listeria newyorkensis* FSL M6-0635 C2c2 (LbFSLC2c2). In certain embodiments, the (modified) Cas protein is C2c1. In certain embodiments, the (modified) Cas protein is C2c3. In certain embodiments, the (modified) Cas protein is Cas13b.

In certain embodiments, the nucleic acid modification is effected by random mutagenesis. Cells or organisms may be exposed to mutagens such as UV radiation or mutagenic chemicals (such as for instance such as ethyl methanesulfonate (EMS)), and mutants with desired characteristics are then selected. Mutants can for instance be identified by TILLING (Targeting Induced Local Lesions in Genomes). The method combines mutagenesis, such as mutagenesis using a chemical mutagen such as ethyl methanesulfonate (EMS) with a sensitive DNA

screening-technique that identifies single base mutations/point mutations in a target gene. The TILLING method relies on the formation of DNA heteroduplexes that are formed when multiple alleles are amplified by PCR and are then heated and slowly cooled. A “bubble” forms at the mismatch of the two DNA strands, which is then cleaved by a single stranded nucleases.

- 5 The products are then separated by size, such as by HPLC. See also McCallum et al. “Targeted screening for induced mutations”; Nat Biotechnol. 2000 Apr;18(4):455-7 and McCallum et al. “Targeting induced local lesions IN genomes (TILLING) for plant functional genomics”; Plant Physiol. 2000 Jun;123(2):439-42.

- 10 RNA interference (RNAi) is a biological process in which RNA molecules inhibit gene expression or translation, by neutralizing targeted mRNA molecules. Two types of small ribonucleic acid (RNA) molecules – microRNA (miRNA) and small interfering RNA (siRNA) – are central to RNA interference. RNAs are the direct products of genes, and these small RNAs can bind to other specific messenger RNA (mRNA) molecules and either increase or decrease their activity, for example by preventing an mRNA from being translated into a protein. The RNAi pathway is
- 15 found in many eukaryotes, including animals, and is initiated by the enzyme Dicer, which cleaves long double-stranded RNA (dsRNA) molecules into short double-stranded fragments of about 21 nucleotide siRNAs (small interfering RNAs). Each siRNA is unwound into two single-stranded RNAs (ssRNAs), the passenger strand and the guide strand. The passenger strand is degraded and the guide strand is incorporated into the RNA-induced silencing complex (RISC).
- 20 Mature miRNAs are structurally similar to siRNAs produced from exogenous dsRNA, but before reaching maturity, miRNAs must first undergo extensive post-transcriptional modification. A miRNA is expressed from a much longer RNA-coding gene as a primary transcript known as a pri-miRNA which is processed, in the cell nucleus, to a 70-nucleotide stem-loop structure called a pre-miRNA by the microprocessor complex. This complex consists of an RNase III enzyme
- 25 called Drosha and a dsRNA-binding protein DGCR8. The dsRNA portion of this pre-miRNA is bound and cleaved by Dicer to produce the mature miRNA molecule that can be integrated into the RISC complex; thus, miRNA and siRNA share the same downstream cellular machinery. A short hairpin RNA or small hairpin RNA (shRNA/Hairpin Vector) is an artificial RNA molecule with a tight hairpin turn that can be used to silence target gene expression via RNA
- 30 interference. The most well-studied outcome is post-transcriptional gene silencing, which occurs when the guide strand pairs with a complementary sequence in a messenger RNA molecule and induces cleavage by Argonaute 2 (Ago2), the catalytic component of the RISC. As used herein, an RNAi molecule may be an siRNA, shRNA, or a miRNA. It will be understood that the RNAi molecules can be applied as such to/in the plant, or can be encoded by

appropriate vectors, from which the RNAi molecule is expressed. Delivery and expression systems of RNAi molecules, such as siRNAs, shRNAs or miRNAs are well known in the art.

In an aspect, the invention relates to a polynucleic acid, preferably an isolated polynucleic acid, capable of specifically hybridizing with any of the nucleic acid molecules of the invention as described herein, or the complement or reverse complement thereof. In certain embodiments, such polynucleic acid is capable of specifically hybridizing with a nucleotide sequence molecule of SEQ ID NO: 1, 2, 4, 10, 15, 16, 24, 25, 30, 35, 36, 38, or 39; or the complement or the reverse complement thereof. It will be understood that such polynucleic acid hybridizes specifically with the recited sequences if it does not hybridize with related sequences (e.g. mutated genes versus wild type genes). Such polynucleic acids can therefore be used to discriminate between the haploid inducer enhancers according to the invention and other non-haploid inducer enhancer plants. In certain embodiments, the polynucleic acid comprises less than 500 nucleotides, such as less than 400 nucleotides, such as less than 300 nucleotides, such as less than 200 nucleotides, such as less than 100, nucleotides, such as preferably less than 80 nucleotides, more preferably less than 60 nucleotides, most preferably less than 40 nucleotides. In certain embodiments, such polynucleic acids comprise at least 5 nucleotides, preferably at least 10 nucleotides, more preferably at least 15 nucleotides. In certain embodiments, polynucleic acids can therefore be used to discriminate between the haploid inducer enhancers according to the invention and other non-haploid inducer enhancer plants. In certain embodiments, the polynucleic acid comprises less than 500 nucleotides, such as less than 400 nucleotides, such as less than 300 nucleotides, such as less than 200 nucleotides, such as less than 100, nucleotides, such as preferably less than 80 nucleotides, more preferably less than 60 nucleotides, most preferably less than 40 nucleotides; and at least 5 nucleotides, preferably at least 10 nucleotides, more preferably at least 15 nucleotides. In certain embodiments, such polynucleic acids comprise from 10 to 60 nucleotides, preferably from 15 to 40 nucleotides. In certain embodiments, such polynucleic acids are primers or probes, as described herein elsewhere.

The invention further relates to a kit comprising one or more of the above polynucleic acid.

The aspects and embodiments of the invention are further supported by the following non-limiting examples.

EXAMPLES**Example 1: QTL analysis and identification of candidate genes**

The functionality of the QTL HIR8 on chromosome 9 was shown by comparing the induction rate of recombinants carrying only the functional mutation (HIR1 carrying the PLA1 gene), only the QTL on chromosome 9 (HIR8) or both:

HIR1	HIR8	# lines	haploid induction rate
no	no	31	0.4%
no	yes	36	0.2%
yes	no	22	4.1%
yes	yes	11	10.6%

In order to identify the underlying genes and to molecularly describe this new locus, a detailed molecular analysis was performed.

All genes in this locus were analyzed for differential expression in pollen in RWS, a haploid inducer line based on Stock 6 (Coe, 1959), compared to three control genotypes (B73, A188 and IG (IG type haploid induction, which is a result of a mutation in the INDETERMINATE GAMETOPHYTE1 gene on chromosome 3)) using RNASeq data.

In addition, for all pollen expressed genes, BAC clones were isolated from RWS, and the gene and amino acid sequences were compared to known reference sequences.

This data analysis resulted in the following list of candidate gene:

Gene ID	Gene	Expression in Pollen	Diff-Expression	Annotation	Seq analysis	Potential function in haploid induction
1	GRMZ M2G4 35294	yes	2x over-expression in RWS	Myosin heavy chain	same amino acid sequence in RWS as in Ref	Myosin involved in cell movement and transport processes in the pollen tube
2	GRMZ M2G0 94586	yes	no	PPR repeat	exchange of 2 amino acids at 5' end, specific to RWS	PPR proteins important for mitochondrial function, known as restorers for cytoplasmic male sterility. The gene could be important for maintaining the energy supply in the pollen tube
3	GRMZ M2G0	yes	missing in RWS and	Aquaporin	no data yet	regulation of water content in the

	81239		IG			pollen tube
4	GRMZ M2G1 24276	yes	only expressed in RWS	SUR2, Membrane lipid metabolism; sphingolipid biosynthesis	amino acid exchange P90R	lipid synthesis in the fast growing pollen tube
5	GRMZ M5G8 62101	yes	no	TB2/DP1 and HVA22 related proteins (ABA, stress related)	amino acid exchange, Frameshift at 3'end	protein involved in membrane traffic

By fine-mapping of the region in an F2 population with the fixed RWS-allele of HIR1 and segregation of HIR8, the region could be reduced to a genomic fragment containing gene 1 (myosin; RWS and W22: genomic DNA=SEQ ID NO: 1; cDNA=SEQ ID NO: 2; protein: SEQ ID NO: 3; B73 control: genomic DNA=SEQ ID NO: 5; cDNA=SEQ ID NO: 6; protein: SEQ ID NO: 7), gene 2 (PPR; RWS and W22: genomic DNA=SEQ ID NO: 15; cDNA=SEQ ID NO: 16; protein: SEQ ID NO: 17; B73 control: genomic DNA=SEQ ID NO: 18; cDNA=SEQ ID NO: 19; protein: SEQ ID NO: 20) and gene 3 (aquaporin; RWS and W22: genomic DNA=SEQ ID NO: 24; cDNA=SEQ ID NO: 25; protein: SEQ ID NO: 26; B73 control: genomic DNA=SEQ ID NO: 27; cDNA=SEQ ID NO: 28; protein: SEQ ID NO: 29).

The haplotype of the region of interest, based on 600k markers, is identical between RWS, KW5526 and the public line W22. In order to control if the locus in RWS contains any additional mutation in comparison to the wildtype line KW5526, KW5526 was sequenced (40x coverage, Illumina) and the genic sequences were compared. No differences could be identified. The functionality of the HIR8 allele in KW5526 was shown in an F2 line containing HIR1 of RWS and HIR8 of KW5526. The induction rate increased from 1.6% (control lines) to 4%.

The complete genomic sequence of W22 was determined by NRGene and published in 2017. This sequence was compared to the BAC sequences in the promoter regions of myosin and aquaporin, both being differentially expressed in RWS compared to control lines. Again, no difference between RWS and W22 was observed. Anyhow both promoters showed extreme differences to the control line B73.

Directly upstream of the myosin gene (upstream region: SEQ ID NO: 4), a fragment of 31.7 kb length is inserted in W22. Annotation showed two additional genes, one zinc ion binding gene with two exons and UTR, i.e. all characteristics of a functional gene. The second gene seems to

be a transposable element. The RNASeq data of RWS pollen, which were mapped to W22, did not show any expression of these genes. Myosin itself is expressed 2-3 fold higher than in control lines. The introgression in the promotor may be responsible for this effect.

The line PH207 is not carrier of the HIR8 QTL and has been used for generation of a TILLING
 5 population. The TILLING population of PH207 was screened for myosin mutants. These were crossed to a mutant of HIR1, which has an induction rate of 1-2% on its own. Double homozygous plants were tested for induction using a glossy tester and subsequent flowcytometry. The myosin mutant PH207m021m showed an induction rate of 6.9% after combination with the HIR1 mutant PH207m007ad (sequences of mutated PLA1: genomic
 10 DNA=SEQ ID NO: 38; cDNA=SEQ ID NO: 39; protein=SEQ ID NO: 40). PH207m021m carries an alanine instead of a valine at position 325 (mutated fragment of the genomic DNA=SEQ ID NO: 10, mutated fragment of the protein=SEQ ID NO: 11; for comparison wildtype fragment of the genomic DNA=SEQ ID NO: 8, wildtype fragment of the protein=SEQ ID NO: 9. Without wishing to be bound by theory, this mutation in the myosin gene may disrupt a helix structure around
 15 amino acid position 325 (Fig. 1) leading to either an increase of the stability of the protein or its functionality, thus mimicking an effect of overexpression.

In the region upstream of aquaporin, a fragment of 3.7 kb (in B73) is exchanged by a fragment of 18.3 kb (W22 and RWS) (upstream region in RWS and W22=SEQ ID NO: 30; upstream region in B73=SEQ ID NO: 31). The inserted fragment contains one expressed gene of unknown
 20 function (genomic DNA=SEQ ID NO: 35; cDNA=SEQ ID NO: 36; protein: SEQ ID NO: 37). It appears to be the reason for suppression of expression of aquaporin in RWS pollen.

Overall, four genes could be identified which are important for the in vivo haploid induction or the in vivo haploid induction capability in maize:

- myosin, overexpressed by a modified promoter
- 25 - PPR, by exchange of amino acid sequence
- aquaporin, by suppression of expression (caused by modified promoter)
- gene of unknown function upstream aquaporin

Any of these genes individually, or any combination of the genes, may be responsible for the effect of haploid induction.

CLAIMS

1. A nucleic acid molecule, comprising a nucleotide sequence, which:
- (i) is a sequence selected from the group consisting of SEQ ID NOs: 1, 4, 5, 8, 10, 12, 15, 18, 21, 30, and 35, or a functional fragment thereof;
 - (ii) has a coding sequence selected from the group consisting of SEQ ID NOs: 2, 6, 13, 16, 19, 22 and 36; or
 - (iii) is complementary to the sequence from (i) or (ii); or
 - (iv) is at least 80% identical to the sequence from (i) or (ii); or
 - (v) encodes a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 3, 7, 9, 11, 14, 17, 20, 23 and 37, or a functional part of the protein; or
 - (vi) encodes a protein comprising an amino acid sequence which is at least 80% identical to a sequence selected from the group consisting of SEQ ID NOs: 3, 7, 9, 11, 14, 17, 20, 23 and 37, or a functional part of the protein; or
 - (vii) hybridizes with a sequence from (iii) under stringent conditions.,
2. A nucleic acid molecule, comprising a nucleotide sequence, which:
- (i) is a sequence selected from the group consisting of SEQ ID NOs: 24, 27, and 32, or a functional fragment thereof; or
 - (ii) has a coding sequence selected from the group consisting of SEQ ID NOs: 25, 28 and 33; or
 - (iii) is complementary to the sequence from (i) or (ii); or
 - (iv) is at least 80% identical to the sequence from (i) or (ii); or
 - (v) encodes a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 26, 29 and 34, or a functional part of the protein; or
 - (vi) encodes a protein comprising an amino acid sequence which is at least 80% identical to a sequence selected from the group consisting of SEQ ID NOs: 26, 29 and 34, or a functional part of the protein; or
 - (vii) hybridizes with a sequence from (iii) under stringent conditions.
3. A nucleic acid molecule encoding an RNA that has a double-stranded portion, wherein at least one strand of the double-stranded portion has a nucleotide sequence which is homologous or identical to at least 14 successive nucleotides of a coding sequence of a nucleic acid that:

- (i) has a sequence selected from the group consisting of SEQ ID Nos: 25, 28, and 33, or a fragment thereof, in a sense or anti-sense orientation; or
- (ii) is complementary to the sequence from (i); or
- (iii) is at least 80% identical to the sequence from (i); or
- 5 (vi) hybridizes with the sequence from (ii) under stringent conditions.
4. The nucleic acid molecule according to claim 1 or 3, wherein said nucleic acid, after transcription or expression in a plant, or the nucleic acid molecule according to claim 2, wherein said nucleotide, after silencing in a plant, is suitable for enhancing the haploid
- 10 induction capability of a haploid inducer plant.
5. The nucleic acid molecule according to claim 4, wherein said haploid inducer plant comprises the HIR1 QTL, preferably wherein said haploid inducer plant is selected from or derived from Stock 6, RWK, RWS, UH400, AX5707RS, NP2222-mai/.
- 15
6. The nucleic acid molecule according to claim 4 or 5, wherein said haploid inducer plant comprises a mutated patatin-like phospholipase (PLA1), preferably wherein said PLA1 has a sequence identical to the PLA1 sequence of Stock 6, RWK, RWS, UH400, AX5707RS, or NP2222-mai/, more preferably wherein said PLA1
- 20 (i) having the nucleotide sequence of SEQ ID NO: 41 or 47 or a sequence which is at least 80% identical to SEQ ID NO: 41 or 47,
- (ii) having the coding sequence of SEQ ID NO: 42 or 48 or a sequence which is at least 80% identical to SEQ ID NO: 42 or 48, or
- (iii) encoding the amino acid sequence of SEQ ID NO: 43, 49 or 50 or a sequence which is at
- 25 least 80% identical to SEQ ID NO: 43, 49 or 50,
- comprises one or more mutation, preferably producing a change in the transcription or expression rate of the PLA1 sequence in a plant, in comparison to a non-mutagenized wild-type plant, or a change in the activity or stability of the encoded amino acid sequence in a plant, in comparison to a non-mutagenized wild-type plant, preferably wherein the mutation causes the
- 30 property of a haploid inducer plant or increases the induction capability of a haploid inducer plant,
- preferably wherein the mutation is
- (A) a deletion or mutation in SEQ ID NO: 41, 42, 47 or 48, preferably which leads to an early stop codon causing a shortening of the encoded amino acid sequence, wherein

- preferably the amino acid sequence at the C-terminal end is shortened by 5 or more amino acids, and/or
- (B) a deletion or mutation in SEQ ID NO: 47 or 48 preferably resulting in an amino acid sequence having SEQ ID NO: 46, or a deletion in SEQ ID NO: 41 or 42 preferably resulting
- 5 in an amino acid sequence having SEQ ID NO: 40;
- (C) an alteration in the coding sequence of SEQ ID NO: 42 or 48 which causes an amino acid exchange between amino acid positions 57 and 289 of SEQ ID NO: 43, 49, or 50;
- (D) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 58, wherein arginine is replaced by glutamine (R58Q);
- 10 (E) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 74, wherein aspartate is replaced by asparagine (D74N);
- (F) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 78, wherein glycine is replaced by arginine (G78R);
- (G) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or
- 15 50 at position 160, wherein valine is replaced by isoleucine (V160I);
- (H) causes an amino acid exchange in the amino acid sequence of SEQ ID NO: 43, 49 or 50 at position 288, wherein serine is replaced by leucine (S288L); and/or
- (I) is a knock-out mutation.
- 20 7. The nucleic acid according to claim 6, wherein said PLA1 has a sequence as set forth in any of SEQ ID NO: 38-40 or 44-46.
8. A vector comprising the nucleic acid according to any of claims 1 to 7.
- 25 9. A plant cell comprising the nucleic acid or vector according to any of claims 1 to 8.
10. A plant or a part thereof comprising the plant cell according to claim 9.
11. A method for producing a transgenic plant which is carrier of an enhancer for the haploid
- 30 induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of claims 4 to 7 comprising:
- a1) transforming at least one plant cell with a nucleic acid molecule according to claim 1 or 3 or a vector according to claim 8 for overexpression of myosin heavy chain,
- a2) transforming at least one plant cell with a nucleic acid molecule according to claim 2 or a
- 35 vector according to any of claims 8 for repression of aquaporin, or

a3) transforming at least one plant cell with a nucleic acid molecule according to claim 1 or 3 or a vector according to claim 8 for expression of PPR corresponding to RWS PPR, or

a4) transforming at least one plant cell with a nucleic acid molecule according to claim 1 or 3 or a vector according to any of claims 8 for expression of the gene set forth in SEQ ID NO: 35;

5 and

b) regenerating a plant from the at least one transgenic plant cell of a) or a transgenic progeny thereof.

12. A method for producing a mutated plant which is carrier of an enhancer for the haploid
10 induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of claims 4 to 7

comprising the following steps:

a) mutagenizing plant cells and subsequent regenerating of plants from the mutagenized plant cells or mutagenizing plants, and

15 b) identifying a regenerated or mutagenized plant which:

has at least one mutation in an endogenous DNA sequence that is identical to a nucleic acid according to claim 1 or 2, or

has at least one mutation in a regulatory sequence of said endogenous DNA sequence, wherein said mutation produces an increase in the transcription or expression rate of the
20 endogenous DNA sequence that is identical to a nucleic acid according to claim 1 in the identified plant or a decrease in the transcription or expression rate of the endogenous DNA sequence that is identical to a nucleic acid according to claim 2 in the identified plant, in comparison to a non-mutagenized wild-type plant, or wherein said mutation produces an increased activity or stability of a protein or polypeptide encoded by the endogenous DNA
25 sequence that is identical to a nucleic acid according to claim 1 in the identified plant or an decreased activity or stability of a protein or polypeptide encoded by the endogenous DNA sequence that is identical to a nucleic acid according to claim 2 in the identified plant, in comparison to a non-mutagenized wild-type plant,

wherein said at least one mutation causes the property of an enhancer of the haploid
30 induction capability of a haploid inducer plant, preferably a haploid inducer plant as defined in any of claims 4 to 7.

13. A method for producing a plant which is carrier of an enhancer for the haploid induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of claims 4
35 to 7 comprising

- overexpressing the myosin heavy chain gene or protein, or increasing the myosin heavy chain mRNA or protein stability or functionality; or
- repressing the aquaporin gene or protein, or decreasing the aquaporin mRNA or protein stability or functionality; or
- 5 - mutating the PPR gene, preferably resulting in a PPR protein having the sequence as set forth in SEQ ID NO: 17.

14. The method according to any of claims 11 to 13, wherein the plant cell of step a) is a cell comprising a mutated patatin-like phospholipase (PLA1).

10

15. The method according to any of claims 11 to 14, wherein the method further comprises a step of crossing the plant with a plant comprising a mutated patatin-like phospholipase (PLA1) and selecting a plant from the progeny which comprises A) the mutated patatin-like phospholipase (PLA1) and B) the nucleic acid or vector introduced in step a) of the method of
- 15 claim 11, the at least one mutation created in the method of claim 12, or the manipulation of claim 13.

16. A plant produced by the method according to any of claims 11 to 14, or a part thereof.

- 20 17. A plant, preferably a Zea mays or Sorghum bicolor plant, comprising
- i) the nucleotide sequence according to SEQ ID NO: 10, or
 - ii) a nucleotide sequence encoding the amino acid according to SEQ ID NO: 11, or
 - iii) a nucleotide sequence encoding a myosin heavy chain having at position 325 referenced to SEQ ID NO: 11 or 14 an amino acid different to alanine, preferably valine.

25

18. A plant, preferably a Zea mays or Sorghum bicolor plant, comprising
- i) the nucleotide sequence according to SEQ ID NO: 38 or 39, or
 - ii) a nucleotide sequence encoding the amino acid according to SEQ ID NO: 40.

- 30 19. Method for identification/selection of an enhancer for the haploid induction capability of an inducer plant, preferably a haploid inducer plant as defined in any of claims 4 to 7, comprising steps of

- a) isolating DNA from at least one cell of a plant, preferably a maize plant or sorghum plant,
- 35 b) determining

- (i) the presence of the nucleic acid molecule according to SEQ ID NOs: 1, 10, 15, or 35
 - (ii) the presence of the nucleic acid molecule of SEQ ID NO: 4 or 30, or
 - (iii) the increased stability or activity or overexpression of the nucleic acid molecule according to claim 1 or the myosin heavy chain protein, or
 - 5 (iv) the decreased stability or activity or suppression of the nucleic acid molecule according to claim 2 or the aquaporin protein,
 - c) identify/select the plant from which the at least one cell had been isolated; preferably wherein step b)(iii) or step b)(iv) comprises comparing the expression level and/or activity of said one or more gene with a predetermined threshold.
- 10
20. A method for the production of a haploid plant or doubled haploid plant, comprising the following steps:
- a) crossing a plant according to claim 10, or 16 to 18 with a plant of the same genus,
 - b) selecting a fertilized haploid seed or embryo, and
 - 15 c) generating a haploid plant from the seed or embryo selected in step b),
 - d) optionally, doubling the set of chromosomes via a chromosome doubling agent or via spontaneous doubling.
21. An isolated polynucleic acid specifically hybridising with a nucleotide sequence molecule of
- 20 SEQ ID NO: 1, 2, 4, 10, 15, 16, 17, 24, 25, 30, 35, 36, 38, or 39; or the complement or the reverse complement thereof.
22. A method of editing or modifying plant genomic DNA or RNA, comprising:
- a) obtaining or providing a first plant, wherein the first plant is a haploid inducer plant as
 - 25 obtained by a method according to any of claims 11 to 15 or 20, or a plant according to any of claims 10, 16 to 18, and wherein said first plant is capable of expressing a DNA or RNA modification enzyme and optionally at least one guide nucleic acid;
 - b) obtaining a second plant, wherein the second plant comprises the plant genomic DNA or RNA which is to be edited or modified;
 - 30 c) pollinating the second plant with pollen from the first plant; and
 - d) selecting at least one haploid progeny produced by the pollination of step (c) wherein the haploid progeny comprises the genome of the second plant but not the first plant, and the genome of the haploid progeny has been edited or modified by the DNA or RNA editing or modification enzyme and optional at least one guide nucleic acid delivered by the first plant.

23. The method of claim 22, wherein the edited or modified haploid progeny is treated with a chromosome doubling agent, thereby creating an edited doubled haploid progeny.
24. A molecular marker which is able to hybridize to nucleic acid molecule as defined in claim 19 b) (i) or (ii).
25. Use of the molecular marker according to claim 24 or the isolated polynucleic acid according to claim 21 for the identification of the plant of any of claims 10 or 16 to 18.
26. Use of a nucleic acid molecule according to any of claims 1 or 3-7 for enhancing the haploid induction capability of an inducer plant, preferably a haplotype inducer plant as defined in any of claims 4 to 7.

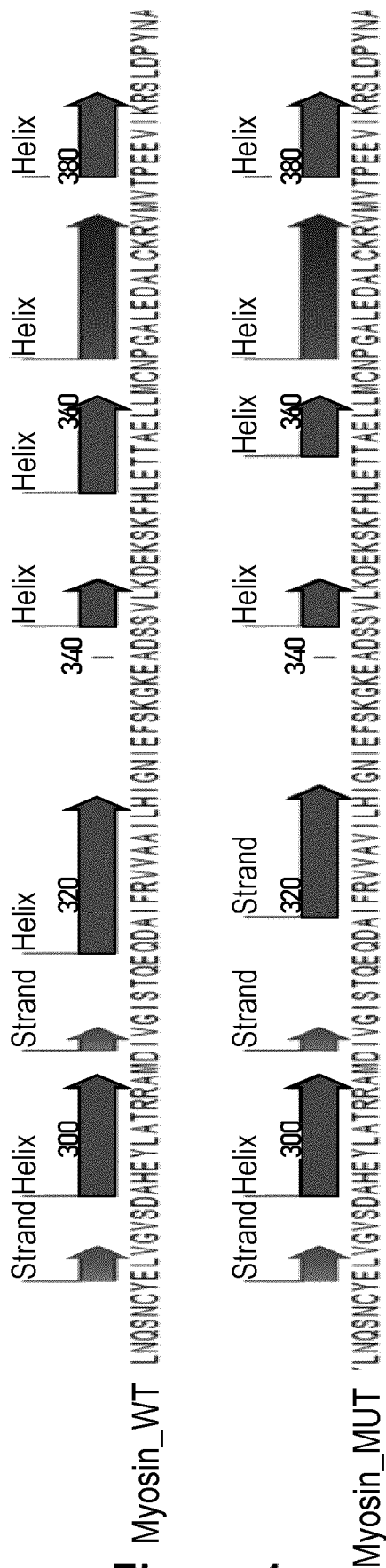


Figure 1