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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 August 2016 (11.08.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/124918 Al

(51) International Patent Classification:
A01H 5/00 (2006.01) *C07K 14/415* (2006.01)
A01H 5/10 (2006.01) *C12N 15/82* (2006.01)

(21) International Application Number:
PCT/GB20 16/050245

(22) International Filing Date:
3 February 2016 (03.02.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/CN20 15/072 143
3 February 2015 (03.02.2015) CN

(71) Applicant: **THE INSTITUTE OF GENETICS AND DEVELOPMENTAL BIOLOGY** [CN/CN]; Chinese Academy of Sciences, No.l BeiChen West Road, Chaoyang District, Beijing 100101 (CN).

(71) Applicant *(for SD only):* **WILLIAMS, Andrea** [GB/GB]; 62-68 Hills Road, Cambridge Cambridgeshire CB2 1LA (GB).

(72) Inventor: **YUNHAI, Li;** The Institute of Genetics and Developmental Biology,Chinese Academy of Sciences, No.l BeiChen West Road, Chaoyang District, Beijing 100101 (CN).

(74) Agent: **MARKS & CLERK LLP;** 62-68 Hills Road, Cambridge Cambridgeshire CB2 1LA (GB).

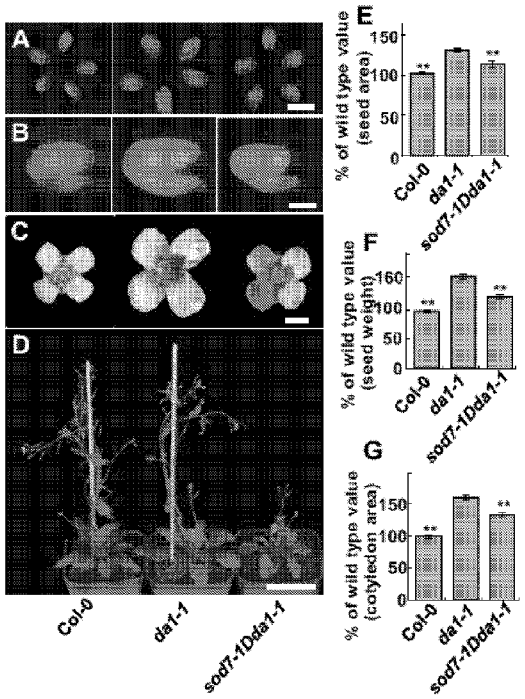
(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP,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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

[Continued on next page]

(54) Title: PLANTS WITH INCREASED SEED SIZE

Figure 1



(57) Abstract: The invention relates to genetically modified plants with an altered seed phenotype, in particular increased seed size. The invention relates to a plant that does not produce a functional NGAL2 polypeptide or functional NGAL2 and NGAL3 polypeptides. NGAL2 and NGAL3 are members of the RAV family and comprise a B3 DNA- binding domain and a transcriptional repression motif.

WO 2016/124918 A1

DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

PLANTS WITH INCREASED SEED SIZE

Field of the invention

The invention relates to transgenic plants with improved growth and yield-related traits, in particular increased seed size. Also within the scope of the invention are related methods, uses, isolated nucleic acids and vector constructs.

Introduction

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

A trait of particular economic interest is increased seed size. Seed size is an important agronomic trait which increased crop yield, and is also a key ecological trait that influences many aspects of a species' regeneration strategy, such as seedling survival rates and seed dispersal syndrome (Harper et al., 1970; Westoby et al., 2002; Moles et al., 2005; Fan et al., 2006; Orsi and Tanksley, 2009; Gegas et al., 2010). Although the size of seeds is one of the most important agronomic traits in plants, the genetic and molecular mechanisms that set the final size of seeds are almost unknown. In higher plants, seed development starts with a double fertilization process, in which one of the two haploid pollen nuclei fuses with the haploid egg cell to produce the diploid embryo,

while the other sperm nucleus fuses with the diploid central cell to form the triploid endosperm (Lopes and Larkins, 1993). The integuments surrounding the ovule are maternal tissues and form the seed coat after fertilization. Therefore, the size of the seed is the result of the growth of the embryo, the endosperm and the maternal tissues. However, the genetic and molecular mechanisms setting the limits of seed growth are almost unknown in plants.

Several factors that function maternally to regulate seed size have been identified in Arabidopsis. For example, TRANSPARENT TESTA GLABRA 2 (TTG2) influences seed growth by increasing cell elongation in the maternal integuments (Garcia et al., 2005; Ohto et al., 2009), while APETALA2 (AP2) may control seed growth by limiting cell elongation in the maternal integuments (Jofuku et al., 2005; Ohto et al., 2005; Ohto et al., 2009). By contrast, AUXIN RESPONSE FACTOR 2 (ARF2) acts maternally to control seed growth by restricting cell proliferation (Schruff et al., 2006). Similarly, the ubiquitin receptor DA1 acts synergistically with the E3 ubiquitin ligases DA2 and EOD1/BB to control seed size by limiting cell proliferation in the maternal integuments (Li et al., 2008; Xia et al., 2013). Mutations in the suppressor of *da1-1* (SOD2), which encodes the ubiquitin-specific protease (UBP15), suppress the large seed phenotype of *da1-1* (Du et al., 2014). DA1 physically associates with UBP15/SOD2 and modulates the stability of UBP15. These studies show that the ubiquitin pathway plays an important part in the maternal control of seed size. KLU/CYTOCHROME P450 78A5 (CYP78A5) regulates seed size by increasing cell proliferation in the maternal integuments of ovules (Adamski et al., 2009). KLU has also been suggested to generate mobile plant-growth substances that promote cell proliferation (Anastasiou et al., 2007; Adamski et al., 2009). By contrast, overexpression of *CYP78A6/EOD3* increases both cell proliferation and cell elongation in the integuments, resulting in large seeds (Fang et al., 2012). Seed size is also determined by zygotic tissues. Several factors have been described to influence seed size via the zygotic tissues in Arabidopsis, including HAIKU1(IKU1), IKU2, MINISEED3 (MINI3) and SHORT HYPOCOTYL UNDER BLUE1 (SHB1) (Garcia et al., 2003; Luo et al., 2005; Zhou et al., 2009; Wang et al., 2010; Kang et al., 2013). iku and mini3 mutants form small seeds due to precocious cellularization of the endosperm (Garcia et al., 2003; Luo et al., 2005; Wang et al., 2010). SHB1 associates with MINI3 and IKU2 promoters and regulates expression of *MINI3* and *IKU2* (Zhou et al., 2009; Kang et al., 2013). ABA INSENSITIVE5 (ABI5) has been recently described to repress the expression of *SHB1* (Cheng et al., 2014), and MINI3 has been reported to activate expression of the

cytokinin oxidase (CKX2) (Li et al., 2013), suggesting the roles of phytohormones in regulating endosperm growth. In addition, the endosperm growth is influenced by parent of-origin effects (Scott et al., 1998; Xiao et al., 2006).

The invention is aimed at providing plants with improved yield traits that are beneficial to agriculture.

Summary of the invention

In a first aspect, the invention relates to a plant generated that does not produce a functional NGAL2 polypeptide or does not produce functional NGAL2 and NGAL3 polypeptides.

In another aspect, the invention relates to a method for altering a plant phenotype comprising reducing or abolishing the expression of a nucleic acid sequence encoding a NGAL2 polypeptide or reducing or abolishing the activity of a NGAL2 or reducing or abolishing the expression of a nucleic acid sequences encoding NGAL2 and NGAL3 polypeptides or reducing or abolishing the activity of a NGAL2 and NGAL3 polypeptide relative to a control plant.

In another aspect, the invention relates to a method for making a plant with an altered phenotype comprising reducing or abolishing the expression of a nucleic acid sequence encoding a NGAL2 polypeptide or reducing or abolishing the activity of a NGAL2 or reducing or abolishing the expression of a nucleic acid sequences encoding NGAL2 and NGAL3 polypeptides or reducing or abolishing the activity of a NGAL2 and NGAL3 polypeptide relative to a control plant..

In another aspect, the invention relates to a plant obtained or obtainable any method described above.

In another aspect, the invention relates to an isolated nucleic acid comprising a sequence comprising or consisting of SEQ ID NO: 1 or 2 or a functional variant or homologue thereof.

In another aspect, the invention relates to a vector comprising an isolated nucleic acid described above.

In another aspect, the invention relates to a silencing nucleic acid construct targeting sequence comprising or consisting of SEQ ID NO: 1, 2 or 3 or a functional variant, part or homologue thereof.

Figures

The invention is further described in the following non-limiting figures.

Figure 1. Isolation of a suppressor of *da1-1* (sod7-1D).

(A) Seeds from wild-type, *da1-1* and *sod7-1D da1-1* plants (from left to right). (B) Mature embryos of the wild type, *da1-1* and *sod7-1D da1-1* (from left to right). (C) Flowers from wild-type, *da1-1* and *sod7-1D da1-1* plants (from left to right). (D) 30-day-old plants of the wild type, *da1-1* and *sod7-1D da1-1* (from left to right). (E) Projective area of wild-type, *da1-1* and *sod7-1D da1-1* seeds. (F) Weight of wild-type, *da1-1* and *sod7-1D da1-1* seeds. (G) Cotyledon area of 10-d-old wild-type, *da1-1* and *sod7-1D da1-1* seedlings. Values (E-G) are given as mean ± SD relative to the respective wild-type values, set at 100%. **, P<0.01 compared with *da1-1* (Student's t-test). Bars = 0.5 mm in (A), 0.2 mm in (B), 1 mm in (C) and 5 cm in (D).

Figure 2. Seed and organ size in the *sod7-1D* mutant.

(A and B) Seeds of Col-0 (A) and *sod7-1D* (B). (C and D) Mature embryos of Col-0 (C) and *sod7-1D* (D). (E and F) 10-day-old seedlings of Col-0 (E) and *sod7-1D* (F). (G) Projective area of Col-0 and *sod7-1D* seeds. (H) Weight of Col-0 and *sod7-1D* seeds. (I) Cotyledon area of 10-day-old Col-0 and *sod7-1D* seedlings. Values (G-I) are given as mean ± SD relative to the respective wild-type values, set at 100%. **, P<0.01 compared with the wild type (Student's t-test). Bars = 0.5 mm in (A) and (B), 0.2 mm in (C) and (D), and 1 mm in (E) and (F).

Figure 3. Cloning of the *SOD7* gene.

(A) Structure of the T-DNA insertion in the *sod7-1D* mutant. (B) Expression levels of At3g1 1580 (SOD7) and At3g1 1590 in *da1-1* and *sod7-1D da1* seedlings.

(C) The SOD7 protein contains a B3 DNA binding domain (second domain in lighter shading) and a transcriptional repression motif (small light box in darker shading, marked with an arrow). (D) Projective area of Col-0, *35S:GFP-SOD7#3* and *35S:GFP-SOD7#5* seeds. (E) Cotyledon area of 10-day-old Col-0, *35S:GFP-SOD7#3* and *35S:GFP-SOD7#5* seedlings. (F) Expression levels of SOD7 in Col-0, *35S:GFP-SOD7#3* and *35S:GFP-SOD7#5* seedlings. Values (D-F) are given as mean ± SD relative to the respective wild-type values, set at 100%. **, P<0.01 compared with the wild type (Student's t-test).

Figure 4. Expression pattern and subcellular localization of SOD7.

(A-K) *SOD7* expression activity was monitored by pSOD7:GUS transgene expression. Histochemical analysis of GUS activity in the developing leaves (A, B and C), the developing sepals (D, E), the developing petals (F, G), the developing stamens (H, I), and the developing carpels (J, K). (L) GFP florescence of SOD7-GFP in a young ovule of *pSOD7:SOD7-GFP* transgenic plants. (M-O) GFP fluorescence of SOD7-GFP (M),

DAPI staining (N), and merged (O) images are shown. Epidermal cells in *pSOD7:SOD7-GFP* leaves were used to observe GFP signal. (P-R) GFP fluorescence of GFP-SOD7 (P), DAPI staining (Q), and merged (R) images are shown. Epidermal cells in *35S:GFP-SOD7* leaves were used to observe GFP signal. Bars = 100 μm in (A-K), 10 μm in (L), and 2 μm in (M-R).

Figure 5. SOD7 acts redundantly with NGAL3 to control seed size.

(A) The *SOD7* gene structure. The start codon (ATG) and the stop codon (TGA) are shown. Closed boxes indicate the coding sequence, and the line between boxes indicates intron. The T-DNA insertion site *(sod7-ko1)* in the SOD7 gene was indicated. (B) The *NGAL3* gene structure. The start codon (ATG) and the stop codon (TGA) are shown. Closed boxes indicate the coding sequence, and the line between boxes indicates intron. The T-DNA insertion site *(ngal3-ko1)* in the NGAL3 gene was indicated. (C) Seeds from Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1* plants (from left to right). (D) Mature embryos of Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1* (from left to right). (E) 25-day-old plants of Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1* (from left to right). (F) Flowers of Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1* (from left to right). (G) Projective area of Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1* seeds. (H) Weight of Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1* seeds. (I) Cotyledon area of Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1* seedlings. Values (G-I) are given as mean ± SD relative to the respective wild-type values, set at 100%. **, P<0.01 compared with the wild type (Col-0) (Student's t-test). Bars = 0.5 mm in (C), 0.2 mm in (D), 5 cm in (E), and 1 mm in (F).

Figure 6. SOD7 acts maternally to determine seed size.

(A) Projective area of Col-OxCol-0 (C/C) F1, **Col-0**×*sod7-ko1 ngal3-ko1* (C/d) F1, *sod7-ko1 ngal3-ko1*×*Col-0* (d/C) F1 and *sod7-ko1 ngal3-ko Usod7-ko 1 ngal3-ko1* (d/d) F1 seeds. Values are given as mean ± SD relative to the respective wild-type values, set at 100%. (B) Projective area of Col-OxCol-0 (C/C) F2, *Col-0***sod7-ko1 ngal3-ko1* (C/d) F2, *sod7-ko1 ngal3-ko1***Col-0* (d/C) F2 and *sod7-ko1 ngal3-ko1***sod7-ko1 ngal3-ko1* (d/d) F2 seeds. Values are given as mean ± SD relative to the respective wild-type values, set at 100%. (C and D) Mature ovules of Col-0 (C) and *sod7-ko1 ngal3-ko1* (D). (E) Outer integument length of mature Col-0 (lighter bar to the left) and *sod7-ko1 ngal3-ko1* (darker bar to the right) ovules. Values are given as mean ± SD. (F) The number of cells in the outer integuments of Col-0 and *sod7-ko1 ngal3-ko1* at 0, 6 and 8 DAP. Values are given as mean ± SD. (F) The length of cells in the outer integuments of Col-0 and *sod7-ko1 ngal3-ko1* at 0, 6 and 8 DAP. Values are given as

mean ± SD. **, P<0.01 compared with the wild type (Col-0) (Student's t-test). Bars = 50 μm in (C) and (D).

Figure 7. *klu-4* is epistatic to *sod7-ko1 ngal3-ko1* with respect to seed size.

(A) Seed area of Col-0, *klu-4, sod7-ko1 ngal3-ko1* and *klu-4 sod7-ko1 ngal3-ko1* (from left to right). Values are given as mean ± SD relative to the respective wild-type values, set at 100%. (B) Seed weight of Col-0, *klu-4, sod7-ko1 ngal3-ko1* and *klu-4 sod7-ko1 ngal3-ko1* (from left to right). Values are given as mean ± SD relative to the respective wild-type values, set at 100%. (C) The outer integument length of Col-0, *klu-4, sod7-ko1 ngal3-ko1* and *klu-4 sod7-ko1 ngal3-ko1* (from left to right). *ngal3-ko1* at 0 and 8 DAP. Values are given as mean ± SD. (D) The number of cells in the outer integuments of Col-0, *klu-4, sod7-ko1 ngal3-ko1* and *klu-4 sod7-ko1 ngal3-ko1* (from left to right)at 0 and 8 DAP. Values are given as mean ± SD. **, P<0.01 compared with their respective controls (Student's t-test).

Figure 8. SOD7 directly binds to the promoter of KLU and represses the expression of KLU.

(A) Expression dynamics of *SOD7* and *KLU* in *pER8-SOD7* transgenic plants treated with β -estradiol for 0, 4 and 8 hours. Means were calculated from three biological samples. Values are given as mean ± SD. **, P<0.01, compared with the expression levelof *KLU* and *SOD7* at 0 hour, respectively (Student's t-test). (B) A 2-kb promoter region of *KLU* upstream of its ATG codon contains a CACTTG sequence. PF1 and PF2 represent PCR fragments used for ChIP-quantitative PCR analysis. A and *A-m* indicate the wild-type probe and the mutated probe used in the EMSA essay, respectively. (C) ChIP-qPCR analysis shows that SOD7 binds to the promoter fragment PF1 of *KLU.* Chromatin from *35S:GFP* and *35S:GFP-SOD7* transgenic plants was immunoprecipitated by anti-GFP, and the enrichment of the fragments was determined by quantitative real-time PCR. The ACTIN7 promoter was used as a negative control. The fold enrichment was normalized to the ACTIN7 amplicon, set at 1. Means were calculated from three biological samples. Values are given as mean ± SD. **, P<0.01, compared with *35S:GFP* transgenic plants (Student's t-test). (D) Direct interaction between SOD7 and the *KLU* promoter determined by EMSA. The biotin-labeled probe *A* and MBP-SOD7 formed the DNA-protein complex, but the mutated probe *A-m* and MBP-SOD7 did not form the DNA-protein complex. The retarded DNA-protein complex was reduced by competition using the unlabeled probe *A*.

Figure 9. The organ size phenotype of 35S:GFP-SOD7 transgenic plants.

Overexpression of SOD7 results in small plants compared with the wild type. Bar = 5 cm.

Figure 10. Phylogenetic tree of the RAV family members in Arabidopsis.

Figure 11. SOD7 acts redundantly with NGAL3 to influence organ size.

Petal area of *Col-0, sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1.* (B) The seventh leaf area of Col-0, *sod7-ko1, ngal3-ko1* and *sod7-ko1 ngal3-ko1.* Values (A and B) are given as meant SD relative to the respective wild-type values, set at 100%. **, P<0.01 and *, P<0.05 compared with the wild type (Col-0).

Figure 12: Conserved domains in NGAL2, NGAL3 and homologs. a) B box motif. b) Repressor motif

Figure 13: Alignment of sequences. The following sequences are shown (from top to bottom): RMZM2G053008, HvMLOC_57250, Os12g0157000, GmLoc100778733, Bra004501, Bra000434, Bra040478, Bra014415, Bra003482, Bra007646, GmLoc100781489, GRMZM2G024948_T01, os02g0683500, HvMLOC_66387, os04g0581400, GRMZM2G102059_T01, Os10g0537100, GRMZM2G142999_T01, GRMZM2G125095_T01, os03g0120900, GRMZM2G098443_T01 , GRMZM2G082227_T01 , Os1 1g01 56000, GRMZM2G328742_T01 , GmLoc100802734 GmLod 00795470, GmLod 0081 8 164, Bra017262, At2g36080/NGAL1, Bra005301, At3g1 1580/SOD7, BraLOCI 03849927, Bra034828, At5g06250/NGAL3, Bra005886, GmLod 02660503, HvMLOC_38822, os01g0693400, HvMLOC44012, HvMLOC_7940 HvMLOC_75135, TRAECDM81004, HvMLOC_56567, TRAES3BF098300010CFD_t1 HvMLOC_63261, TRAES3BF062700040CFD_t1, TRAES3BF062600010CFD_t1, Bra038346, GmLoc732601 , GmLod 00789009, GmLod 00776987, GmLod 00801 107. Conserved B3 domain and repressor motif are boxed.

Figure 14: Genome editing experiments to knock out rice genes Os1 1g01560000 and Os12g01 57000 in rice. gRNA stands for guide RNA, target site linked with gRNA scaffold will recruit CAS9 enzyme to target site in the genome and cause gene-editing.

Detailed description

The present invention will now be further described. 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.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of botany, microbiology, tissue culture, molecular biology,

chemistry, biochemistry and recombinant DNA technology, bioinformatics which are within the skill of the art. Such techniques are explained fully in the literature.

As used herein, the words "nucleic acid", "nucleic acid sequence", "nucleotide", "nucleic acid molecule" or "polynucleotide" are intended to include DNA molecules (e.g., cDNA or genomic DNA), RNA molecules (e.g., mRNA), naturally occurring, mutated, synthetic DNA or RNA molecules, and analogues of the DNA or RNA generated using nucleotide analogues. It can be single-stranded or double-stranded. Such nucleic acids or polynucleotides include, but are not limited to, coding sequences of structural genes, anti-sense sequences, and non-coding regulatory sequences that do not encode mRNAs or protein products. These terms also encompass a gene. The term "gene" or "gene sequence" is used broadly to refer to a DNA nucleic acid associated with a biological function. Thus, genes may include introns and exons as in the genomic sequence, or may comprise only a coding sequence as in cDNAs, and/or may include cDNAs in combination with regulatory sequences.

The terms "peptide", "polypeptide" and "protein" are used interchangeably herein and refer to amino acids in a polymeric form of any length, linked together by peptide bonds.

For the purposes of the invention, "transgenic", "transgene" or "recombinant" means with regard to, for example, a nucleic acid sequence, an expression cassette, gene construct or a vector comprising the nucleic acid sequence or an organism transformed with the nucleic acid sequences, expression cassettes or vectors according to the invention, all those constructions brought about by recombinant methods in which either

(a) the nucleic acid sequences encoding proteins useful in the methods of the invention, or

(b) genetic control sequence(s) which is operably linked with the nucleic acid sequence according to the invention, for example a promoter, or

(c) both (a) and (b)

are not located in their natural genetic environment or have been modified by genetic intervention techniques, it being possible for the modification to take the form of, for example, a substitution, addition, deletion, inversion or insertion of one or more nucleotide residues. The natural genetic environment is understood as meaning the natural genomic or chromosomal locus in the original plant or the presence in a

genomic library. In the case of a genomic library, the natural genetic environment of the nucleic acid sequence is preferably retained, at least in part. The environment flanks the nucleic acid sequence at least on one side and has a sequence length of at least 50 bp, preferably at least 500 bp, especially preferably at least 1000 bp, most preferably at least 5000 bp. A naturally occurring expression cassette - for example the naturally occurring combination of the natural promoter of the nucleic acid sequences with the corresponding nucleic acid sequence encoding a polypeptide useful in the methods of the present invention, as defined above - becomes a transgenic expression cassette when this expression cassette is modified by non-natural, synthetic ("artificial") methods such as, for example, mutagenic treatment. Suitable methods are described, for example, in US 5,565,350 or WO 00/15815 both incorporated by reference.

In certain embodiments, a transgenic plant for the purposes of the invention is thus understood as meaning, as above, that the nucleic acids used in the method of the invention are not at their natural locus in the genome of said plant, it being possible for the nucleic acids to be expressed homologously or heterologously. Thus, the plant can express a silencing construct transgene. However, as mentioned, in certain embodiments, transgenic also means that, while the nucleic acids according to the different embodiments of the invention are at their natural position in the genome of a plant, the sequence has been modified with regard to the natural sequence, and/or that the regulatory sequences of the natural sequences have been modified, for example by mutagenesis.

Transgenic is preferably understood as meaning the expression of the nucleic acids according to the invention at an unnatural locus in the genome, i.e. homologous or, preferably, heterologous expression of the nucleic acids takes place. According to the invention, the transgene is stably integrated into the plant and the plant is preferably homozygous for the transgene.

The various aspects of the invention use genetic engineering methods. Thus, the plants have been generated using genetic engineering methods, for example transgene expression, mutagenesis, gene targeting, gene silencing or genome editing as detailed below. Thus, the various aspects of the invention can involve recombinant DNA technology. The plants of the invention are thus mutant plants which have been genetically engineered, that is manipulated by human intervention. The plants of the

various aspects of the invention do not relate to natural variants which have not been manipulated by genetic engineering methods. The plant may be a transgenic plant in some embodiments, for example a plant which comprises a nucleic acid construct expressing a silencing construct.

In preferred embodiments exclude embodiments that are solely based on generating plants by traditional breeding methods.

The inventor has identified a B3 domain transcriptional repressor termed AtNGAL2, encoded by the suppressor of *Atda1-1 (AtSOD7),* which acts maternally to control seed size by restricting cell proliferation in the integuments of ovules and developing seeds.

The inventor previously identified the ubiquitin receptor DA1 as a negative regulator of seed size in Arabidopsis (Li et al., 2008). The *da1-1* mutant formed large seeds due to increased cell proliferation in the maternal integuments (Li et al., 2008; Xia et al., 2013). To identify novel components in the DA1 pathway or other seed size regulators, the inventor initiated a T-DNA activation tagging screen for modifiers of *da1-1* (Fang et al., 2012). A dominant suppressor of *da1-1 (sod7-1D)* was isolated from seeds produced from approximate 16,000 T1 plants (Fig.lA). Seeds of the *sod7-1D da1-1* double mutant were significantly smaller and lighter than *da1-1* seeds (Figures 1A, E and F). The results show that the *sod7-1D* mutation suppressed the seed and organ size phenotypes of *da1-1.* The *SOD7* gene was isolated and found to encode a NGATHA like protein (NGAL2) containing a B3 DNA-binding domain and a transcriptional repression motif (Figure 3C) (Alvarez et al., 2009; Ikeda and Ohme-Takagi, 2009; Trigueros et al., 2009). SOD7 belongs to the RAV gene family that consists of 13 members in Arabidopsis (Figure 10) (Swaminathan et al., 2008). Several members of the RAV family contain the putative transcriptional repression motifs, including NGA1, NGA2, NGA3, NGA4, NGAL1 , NGAL2/SOD7 and NGAL3 (Figure 10) (Ikeda and Ohme-Takagi, 2009). The transcriptional repression motifs in NGA1, NGAL1 and NGAL2/SOD7 have been known to possess the repressive activity (Ikeda and Ohme-Takagi, 2009), indicating that they are transcriptional repressors. *SOD7* exhibits the highest similarity to Arabidopsis NGAL3/DEVELOPMENT-RELATED PcG TARGET IN THE APEX 4 (DPA4) (Figure 10), which has known roles in the regulation of leaf serrations (Engelhorn et al., 2012), but no previously identified function in seed size control.

The inventor has shown that overexpression of *AtSOD7* significantly decreases seed size of wild-type plants, while the disruption of *AtSOD7* increases seed size. The inventors have shown that disruption of *AtNGAL3,* a close homolog of *AtSOD7* also increases seed size. Moreover, the simultaneous disruption of *AtSOD7* and *AtNGAL3* further increases seed size in a synergistic manner. Genetic analyses carried out by the inventor indicate that *AtSOD7* acts in a common pathway with the seed size regulator *AtKLU* to control seed growth, but does so independently of *AtDA1.* Further results show that *AtSOD7* directly binds to the promoter of *AtKLU in vitro* and in *vivo* and represses expression of *AtKLU.* Therefore, the inventor's findings show that *AtSOD7* (aka *AtNGAL2)* is a target for seed size improvement in crops. The plants of the invention are characterised by increased organ size, for example increased seed size, and also increased petal size, increased embryo size, for example. Increased seed size leads to an increase in seed yield and the plants of the invention are thus characterised by increased seed yield.

Thus, the invention relates to a plant wherein said plant does not produce a functional NGAL2 and/or NGAL3 polypeptide. For example, the plant does not produce a full length transcript of a nucleic acid sequence encoding a NGAL2 and/or NGAL3 protein. In another embodiment, the plant produces a full length transcript of a nucleic acid sequence encoding a NGAL2 and/or NGAL3, but the resulting protein is not functional. In a preferred embodiment, said plant does not produce a functional NGAL2 polypeptide and also does not produce a functional NGAL3 polypeptide. Such plants are double knock-out or knock-down mutants (loss of function mutants) and methods according to the invention as described below relate to making such double mutants.

The plants of the invention are mutant plants which have been genetically modified and are not naturally occurring varieties. Thus, the plants have been generated using genetic engineering methods, for example mutagenesis, gene targeting, gene silencing or genome editing as detailed below. Thus, the various aspects of the invention can involve recombinant DNA technology. The plant may be a transgenic plant in some embodiments, for example a plant which comprises a transgene to silence gene expression of *SOD7* and/or *NGAL3.* In other embodiments, the plant does not carry a transgene, but is a mutant plant wherein the endogenous nucleic acid sequence encoding a NGAL2 and/or NGAL3 polypeptide or the endogenous *SOD7* and/or *NGAL3* promoter sequence has been manipulated to either reduce or abolish expression of a nucleic acid sequence encoding a NGAL2 and/or NGAL3 polypeptide

or reduce or abolish the activity of a NGAL2 and/or NGAL3 polypeptide. The plants of the various aspects of the invention do not relate to natural variants which have not been manipulated by genetic engineering methods.

In one aspect, the invention relates to a plant generated by genetic engineering methods wherein the expression of a nucleic acid sequence encoding a NGAL2 and/or NGAL3 polypeptide and/or the activity of a NGAL2 and/or NGAL3 polypeptide is reduced or abolished relative to a control plant. In one embodiment, expression of a nucleic acid sequence encoding a NGAL2 polypeptide or the activity of a NGAL2 polypeptide is reduced or abolished. In another embodiment, expression of a nucleic acid sequence encoding a NGAL3 polypeptide or the activity of a NGAL3 polypeptide is reduced or abolished. In a preferred embodiment the presence of function of both proteins is affected, in other words, the plant is characterised in that expression of a nucleic acid sequence encoding a NGAL2 polypeptide or the activity of a NGAL2 polypeptide is reduced or abolished and also expression of a nucleic acid sequence encoding a NGAL3 polypeptide or the activity of a NGAL3 polypeptide is reduced or abolished in said plant.

For example, said plant can have reduced or abolished expression of a nucleic acid sequence encoding a NGAL2 polypeptide and reduced or abolished expression of a nucleic acid sequence encoding a NGAL3 polypeptide. In another embodiment, said plant can have reduced or abolished activity of a NGAL2 polypeptide and reduced or abolished activity of a NGAL3 polypeptide. In another embodiment, said plant can have reduced or abolished expression of a nucleic acid sequence encoding a NGAL2 polypeptide and reduced or abolished activity of a NGAL3 polypeptide. In another embodiment, said plant can have reduced or abolished expression of a nucleic acid sequence encoding a NGAL3 polypeptide and reduced or abolished activity of a NGAL2 polypeptide.

A NGAL2 or NGAL3 polypeptide as described in the various aspects of the invention has a characteristic domain structure as explained below.

A NGAL2 OR NGLA3 polypeptide as described in the various aspects of the invention comprises a B3 DNA binding domain which has the structure shown in figure 12 .

In one embodiment, the domain is: SNNNNNNGGSGDDVACHFQRFDLHRLFIGWRGE

(SEQ ID NO:6) or a domain with at least 80%, at least 95% or at least 95% sequence identity thereto.

A NGAL2 OR NGAL3 polypeptide as described in the various aspects of the invention also comprises a transcriptional repression motif shown in figure 12.

In one embodiment, the domain is: VRLFGVNLE (SEQ ID NO:7) or a domain with at least 95% sequence identity thereto.

In one embodiment, the NGAL2 protein is AtNGAL2, a functional variant, part or homologue thereof. AtNGAL2 is encoded by AtSOD7. The term AtSOD7 refers to the wild type *AtSOD7* nucleic acid sequence comprising or consisting of SEQ ID NO. 1 (CDNA) or SEQ ID NO 2 (genomic DNA). The protein encoded by *AtSOD7* is termed *At*NGAL2 SEQ ID NO.3. In one embodiment, said functional homologue is not AtNGAL3.

In one embodiment, the NGAL3 protein is AtNGAL3, a functional variant, part or homologue thereof. The term *AtNGAL3* refers to the wild type *AtNGAL3* nucleic acid sequence comprising or consisting of SEQ ID NO. 4. The protein encoded by *AtNGAL3* is termed *At*NGAL3 SEQ ID NO.5.

The term "functional" refers to the biological function of the NGAL2 or NGAL3, that is their function in controlling organ size, in particular seed size. The terms "functional variant" or "functional part "as used herein, for example with reference to SEQ ID NOs: 1, 2 or 3, or SEQ ID NOs: 4 or 5 refers to a variant gene or polypeptide sequence or part of the gene or polypeptide sequence which retains the biological function of the full non-variant SOD7/NGAL2 or NGAL2/NGAL3 sequence, that is regulation of seed size. Such sequences complement the *Atsod7-1D* mutant or *Atngal3* mutant respectively.

Thus, it is understood, as those skilled in the art will appreciate, that the aspects of the invention, encompass not only targeting a *AtSOD7* and/or *AtNGAL3* nucleic acid, for example a nucleic acid sequence comprising or consisting of SEQ ID NO: 1 or SEQ ID NO: 2, or SEQ ID NO: 4 respectively or a polypeptide comprising or consisting of SEQ ID NO: 3, or SEQ ID NO: 5, or a promoter of a AtSOD7 and/or *AtNGAL3* nucleic acid. The aspects of the invention encompass also functional variants of AtNGAL2 or AtNGAL3 that do not affect the biological activity and function of the resulting protein.

Alterations in a nucleic acid sequence which result in the production of a different amino acid at a given site that do however not affect the functional properties of the encoded polypeptide, are well known in the art. For example, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also produce a functionally equivalent product. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Also encompassed is a variant that is substantially identical, i.e. has only some sequence variations, for example in non-conserved residues, to the wild type sequences as shown herein and is biologically active.

Generally, variants of a particular *SOD7/NGAL3* nucleotide sequence or NGAL2/NGAL3 polypeptide as described herein will have at least about 60%, preferably at least about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 92%, 94%, 95%, 96%, 97%, 98% or 99% or more sequence identity to that particular non-variant nucleotide sequence, as determined by sequence alignment programs described elsewhere herein.

Furthermore, the various the aspects of the invention encompass not only a *AtSOD7* and/or *AtNGAL3* nucleic acid, for example a nucleic acid sequence comprising or consisting of SEQ ID NO: 1 or SEQ ID NO: 2, or SEQ ID NO: 4 respectively or a polypeptide comprising or consisting of SEQ ID NO: 3, or SEQ ID NO: 5, or their functional variants but also homologues of *AtSOD7* and/or *AtNGAL3* in *Arabidopsis* or other plants. Also within the scope of the invention are functional variants of such homologues as defined above.

The term homologue as used herein also designates an *AtSOD7* and/or *AtNGAL3* orthologue from other plant species. A homologue of AtNGAL2 or AtNGAL3 polypeptide respectively has, in increasing order of preference, at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%,

87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least 99% overall sequence identity to the amino acid represented by SEQ ID NO: 3 or 5 respectively. Preferably, overall sequence identity is at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, most preferably 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least 99%.

In another embodiment, the homologue of a *AtSOD7* or *AtNGAL3* nucleic acid sequence respectively has, in increasing order of preference, at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least 99% overall sequence identity to the nucleic acid represented by SEQ ID NO: 1 or 2 or 4 respectively. Preferably, overall sequence identity is at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, most preferably 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least 99%. The overall sequence identity is determined using a global alignment algorithm known in the art, such as the Needleman Wunsch algorithm in the program GAP (GCG Wisconsin Package, Accelrys).

In a preferred embodiment, the NGAL2 or NGAL3 homologue is from a plant that is not Arabidopsis.

In one embodiment, an AtNGAL2 or a homologue thereof or AtNGAL3 or a homologue thereof comprises a B3 domain having the sequence as defined above

In one embodiment, an AtNGAL2 or a homologue thereof or AtNGAL3 or a homologue thereof comprises a transcriptional repression motif having the sequence as defined above

Examples of homologues are shown in figure 13 and in SEQ ID NO: 49-145. In certain embodiments, if a plant has more than one AtNGAL2 and/or AtNGAL3 homologue, then all homologues are knocked out or knocked down. Suitable homologues can be

identified by sequence comparisons and identifications of conserved domains. There are predictors in the art that can be used to identify such sequences. The function of the homologue can be identified as described herein and a skilled person would thus be able to confirm the function, for example when overexpressed in a plant or knocked out in a plant or when expressed in a plant or by expressing the homologous nucleic acid sequence in an Arabidopsis gain of function mutant.

Thus, the nucleotide sequences of the invention and described herein can also be used to isolate corresponding sequences from other organisms, particularly other plants, for example crop plants. In this manner, methods such as PCR, hybridization, and the like can be used to identify such sequences based on their sequence homology to the sequences described herein. Topology of the sequences and the characteristic domains structure can also be considered when identifying and isolating homologues. Sequences may be isolated based on their sequence identity to the entire sequence or to fragments thereof. In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen plant. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labelled with a detectable group, or any other detectable marker. Thus, for example, probes for hybridization can be made by labelling synthetic oligonucleotides based on the ABA-associated sequences of the invention. Methods for preparation of probes for hybridization and for construction of cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook, et al., (1989) Molecular Cloning: A Library Manual (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

Hybridization of such sequences may be carried out under stringent conditions. By "stringent conditions" or "stringent hybridization conditions" is intended conditions under which a probe will hybridize to its target sequence to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences that are 100% complementary to the probe can be identified (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing).

Generally, a probe is less than about 1000 nucleotides in length, preferably less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Duration of hybridization is generally less than about 24 hours, usually about 4 to 12. Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

According to the invention, preferred homologues of *AtSOD7* and *AtNGAL3* peptides are selected from crop plants, for example cereal crops. Preferred homologues of AtNGAL2 and AtNGAL3 and their polypeptide sequences are also shown in Fig. 13.

A plant according to the various aspects of the invention, including the transgenic plants, methods and uses described herein may be a monocot or a dicot plant.

A dicot plant may be selected from the families including, but not limited to *Asteraceae, Brassicaceae (e.g. Brassica napus), Chenopodiaceae, Cucurbitaceae, Leguminosae (Caesalpiniaceae, Aesalpiniaceae Mimosaceae, Papilionaceae or Fabaceae), Malvaceae, Rosaceae or Solanaceae.* For example, the plant may be selected from lettuce, sunflower, *Arabidopsis,* broccoli, spinach, water melon, squash, cabbage, tomato, potato, yam, capsicum, tobacco, cotton, okra, apple, rose, strawberry, alfalfa, bean, soybean, field (fava) bean, pea, lentil, peanut, chickpea, apricots, pears, peach, grape vine, bell pepper, chilli or citrus species.

A monocot plant may, for example, be selected from the families *Arecaceae, Amaryllidaceae* or *Poaceae.* For example, the plant may be a cereal crop, such as maize, wheat, rice, barley, oat, sorghum, rye, millet, buckwheat, or a grass crop such as Lolium species or Festuca species, or a crop such as sugar cane, onion, leek, yam or banana.

Also included are biofuel and bioenergy crops such as rape/canola, sugar cane, sweet sorghum, *Panicum virgatum* (switchgrass), linseed, lupin and willow, poplar, poplar hybrids, Miscanthus or gymnosperms, such as loblolly pine. Also included are crops for

silage (maize), grazing or fodder (grasses, clover, sanfoin, alfalfa), fibres (e.g. cotton, flax), building materials (e.g. pine, oak), pulping (e.g. poplar), feeder stocks for the chemical industry (e.g. high erucic acid oil seed rape, linseed) and for amenity purposes (e.g. turf grasses for golf courses), ornamentals for public and private gardens (e.g. snapdragon, petunia, roses, geranium, Nicotiana sp.) and plants and cut flowers for the home (African violets, Begonias, chrysanthemums, geraniums, Coleus spider plants, Dracaena, rubber plant).

Preferably, the plant is a crop plant. By crop plant is meant any plant which is grown on a commercial scale for human or animal consumption or use. In a preferred embodiment, the plant is a cereal.

Most preferred plants are maize, rice, wheat, oilseed rape/canola, sorghum, soybean, sunflower, alfalfa, potato, tomato, tobacco, grape, barley, pea, bean, field bean, lettuce, cotton, sugar cane, sugar beet, broccoli or other vegetable brassicas or poplar.

The term "plant" as used herein encompasses whole plants, ancestors and progeny of the plants and plant parts, including seeds, fruit, shoots, stems, leaves, roots (including tubers), flowers, and tissues and organs, wherein each of the aforementioned comprise the gene/nucleic acid of interest. The term "plant" also encompasses plant cells, suspension cultures, callus tissue, embryos, meristematic regions, gametophytes, sporophytes, pollen and microspores, again wherein each of the aforementioned comprises the gene/nucleic acid of interest.

According to the various aspects of the invention, including the plants and methods of the invention, abolishing, inactivating, repressing, reducing or down-regulating the activity of a NGAL2 and/or NGAL3 polypeptide can be achieved through different means. Such means that are within the scope of the various aspects of the invention are methods for abolishing or reducing translation or transcription of the *SOD7* and/or *NGAL3* gene, destabilizing *SOD7* and/or *NGAL3* transcript stability, destabilizing NGAL2 and/or NGAL3 polypeptide stability or abolishing or reducing the activation or activity of the NGAL2 and/or NGAL3 or polypeptide. Thus, in one embodiment, endogenous *SOD7* and/or *NGAL3* gene or its promoter carry a functional mutation so that no full length transcript is made. In another embodiment, the *SOD7* and/or *NGAL3* gene is silenced in said plant using gene silencing techniques. In another embodiment,

the *SOD7* and/or *NGAL3* nucleic acid sequence has been altered to introduce a mutation which results in a NGAL2/NGAL3 protein with reduced or abolished activity. These embodiments and the techniques used are described in more detail below.

In another aspect, the invention relates to a method for altering a plant phenotype comprising reducing or abolishing the expression of a nucleic acid sequence encoding a NGAL2 and/or NGAL3 polypeptide and/or reducing or abolishing the activity of a NGAL2 and/or NGAL3 polypeptide relative to a control plant.

In another aspect, the invention relates to a method for making a plant with an altered phenotype comprising reducing or abolishing the expression of a nucleic acid sequence encoding a NGAL2 and/or NGAL3 polypeptide and/or reducing or abolishing the activity of a NGAL2 and/or NGAL3 polypeptide relative to a control plant.

As previously described, such methods above use genetic engineering methods.

In this aspect, a wild type plant may be targeted to simultaneously knock out or down both *SOD7* and *NGAL3* function. Alternatively, the method may comprise the following steps

a) Knocking out or down *SOD7* function in a first plant;
b) knocking out or down *NGAL3* function in a second plant and
c) crossing plants regenerated from said first plant with plants regenerated from said second plant.

In one embodiment of these methods, expression of a nucleic acid sequence encoding a NGAL2 polypeptide or the activity of a NGAL2 polypeptide is reduced or abolished. In another embodiment, expression of a nucleic acid sequence encoding a NGAL3 polypeptide or the activity of a NGAL3 polypeptide is reduced or abolished. In a preferred embodiment, the method comprises reducing or abolishing expression of a nucleic acid sequence encoding a NGAL2 polypeptide or the activity of a NGAL2 polypeptide and reducing or abolishing expression of a nucleic acid sequence encoding a NGAL3 polypeptide or the activity of a NGAL3 polypeptide to create a double loss of function mutant.

For example, the method comprises reducing or abolishing expression of a nucleic acid sequence encoding a NGAL2 polypeptide and reducing or abolishing expression of a

nucleic acid sequence encoding a NGAL3 polypeptide. In another embodiment, the method comprises reducing or abolishing activity of a NGAL2 polypeptide and reducing or abolishing activity of a NGAL3 polypeptide. In another embodiment, the method comprises reducing or abolishing expression of a nucleic acid sequence encoding a NGAL2 polypeptide and reducing or abolishing activity of a NGAL3 polypeptide. In another embodiment the method comprises reducing or abolishing expression of a nucleic acid sequence encoding a NGAL3 polypeptide or reducing or abolishing activity of a NGAL2 polypeptide.

According to these methods, the phenotype is preferably selected from increased organ size, for example increased seed size or increased seed weight. Increased seed size leads to an increase in yield and the methods of the invention also increased yield.

The term "yield" in general means a measurable produce of economic value, typically related to a specified crop, to an area, and to a period of time. Individual plant parts directly contribute to yield based on their number, size and/or weight, or the actual yield is the yield per square meter for a crop and year, which is determined by dividing total production (includes both harvested and appraised production) by planted square meters. The term "yield" as described herein relates to yield-related traits and may relate to vegetative biomass (root and/or shoot biomass), to reproductive organs, and/or to propagules (such as seeds) of that plant. Thus, according to the invention, the term yield refers to organ size, in particular seed size and can be measured by assessing seed size or seed weight or cotyledon size.

The terms "increase", "improve" or "enhance" are interchangeable. Yield or seed size for example is increased by at least a 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9% or 10%, preferably at least 15% or 20%, more preferably 25%, 30%, 35%, 40% or 50% or more in comparison to a control plant.

A control plant as used herein according to all of the aspects of the invention is a plant which has not been modified according to the methods of the invention. Accordingly, the control plant has not been genetically modified to alter either expression of a nucleic acid encoding a NGAL2 or NGAL3 polypeptide or to alter the activity of a NGAL2 or NGAL3 polypeptide as described herein. In one embodiment, the control plant is a wild type plant that has not been genetically altered. In another embodiment, the control plant is a transgenic plant that does not have altered expression of a nucleic

acid encoding a NGAL2 or NGAL3 polypeptide or altered activity of a NGAL2 or NGAL3 polypeptide, but has been genetically altered in other ways, for example by expressing a desirable transgene to confer certain traits.

The reduction, decrease, down-regulation or repression of the activity of the NGAL2 and/or NGAL3 polypeptide or corresponding *SOD7* and/or *NGAL3* nucleic acid sequences according to the aspects of the invention is at least 10%, 20%, 30%, 40% or 50% in comparison to the control plant.

For example, the plant is a reduction (knock down) or loss of function (knock out) mutant wherein the function of the *SOD7* and/or *NGAL3* nucleic acid sequence is reduced or lost compared to a wild type control plant. To this end, a mutation is introduced into the *SOD7* and/or *NGAL3* nucleic acid sequence or the corresponding promoter sequence which disrupts the transcription of the gene leading to a gene product which is not functional or has a reduced function. The mutation may be a deletion, insertion or substitution. The expression of active protein may thus be abolished by mutating the nucleic acid sequences in the plant cell which encode the NGAL2 or NGAL3 polypeptide and regenerating a plant from the mutated cell. The nucleic acids may be mutated by insertion or deletion of one or more nucleotides. Techniques for the inactivation or knockout of target genes are well-known in the art. These techniques include gene target using vectors that target the gene of interest and which allow integration allows for integration of transgene at a specific site. The targeting construct is engineered to recombine with the target gene, which is accomplished by incorporating sequences from the gene itself into the construct. Recombination then occurs in the region of that sequence within the gene, resulting in the insertion of a foreign sequence to disrupt the gene. With its sequence interrupted, the altered gene will be translated into a nonfunctional protein, if it is translated at all. Other techniques include genome editing (targeted genome engineering) as described below. Using either of these techniques, in preferred embodiment, conserved domains which confer function of NGAL2 or NGAL3 respectively are modified.

A skilled person will know further approaches can be used to generate such mutants. In one embodiment, insertional mutagenesis is used, for example using T-DNA mutagenesis (which inserts pieces of the T-DNA from the *Agrobacterium tumefaciens* T-Plasmid into DNA causing either loss of gene function or gain of gene function mutations), site-directed nucleases (SDNs) or transposons as mutagens. Insertional

mutagenesis is an alternative means of disrupting gene function and is based on the insertion of foreign DNA into the gene of interest (see Krysan et al, The Plant Cell, Vol. 11, 2283-2290, December 1999).

In one embodiment, as discussed in the examples, T-DNA may be used as an insertional mutagen which disrupts *SOD7* and/or *NGAL3* gene expression. T-DNA not only disrupts the expression of the gene into which it is inserted, but also acts as a marker for subsequent identification of the mutation. Since the sequence of the inserted element is known, the gene in which the insertion has occurred can be recovered, using various cloning or PCR-based strategies. The insertion of a piece of T- DNA on the order of 5 to 25 kb in length generally produces a disruption of gene function. If a large enough population of T-DNA transformed lines is generated, there are reasonably good chances of finding a transgenic plant carrying a T-DNA insert within any gene of interest. Transformation of spores with T-DNA is achieved by an *Agrobacterium*-mediated method which involves exposing plant cells and tissues to a suspension of *Agrobacterium* cells.

The details of this method are well known to a skilled person. In short, plant transformation by Agrobacterium results in the integration into the nuclear genome of a sequence called T-DNA, which is carried on a bacterial plasmid. The use of T-DNA transformation leads to stable single insertions. Further mutant analysis of the resultant transformed lines is straightforward and each individual insertion line can be rapidly characterized by direct sequencing and analysis of DNA flanking the insertion. Gene expression in the mutant is compared to expression of the *SOD7* and/or *NGAL3* nucleic acid sequence in a wild type plant and phenotypic analysis is also carried out. Other techniques for insertional mutagenesis include the use of transposons.

In another embodiment, mutagenesis is physical mutagenesis, such as application of ultraviolet radiation, X-rays, gamma rays, fast or thermal neutrons or protons. The targeted population can then be screened to identify a *SOD7* or *NGAL3* loss of function mutant.

In another embodiment of the various aspects of the invention, the plant is a mutant plant derived from a plant population mutagenised with a mutagen. The mutagen may be fast neutron irradiation or a chemical mutagen, for example selected from the following non-limiting list: ethyl methanesulfonate (EMS), methylmethane sulfonate

(MMS), N-ethyl-N-nitrosurea (ENU), triethylmelamine (TEM), N-methyl-N-nitrosourea (MNU), procarbazine, chlorambucil, cyclophosphamide, diethyl sulfate, acrylamide monomer, melphalan, nitrogen mustard, vincristine, dimethylnitosamine, N-methyl-N'-nitro-Nitrosoguanidine (MNNG), nitrosoguanidine, 2-aminopurine, 7,12 dimethyl-benz(a)anthracene (DMBA), ethylene oxide, hexamethylphosphoramide, bisulfan, diepoxyalkanes (diepoxyoctane (DEO), diepoxybutane (BEB), and the like), 2-methoxy-6-chloro-9 [3-(ethyl-2-chloroethyl)aminopropylamino]acridine dihydrochloride (ICR-170) or formaldehyde.

In one embodiment, the method used to create and analyse mutations is targeting induced local lesions in genomes (TLLING), reviewed in Henikoff et al, 2004. In this method, seeds are mutagenised with a chemical mutagen, for example EMS. The resulting M1 plants are self-fertilised and the M2 generation of individuals is used to prepare DNA samples for mutational screening. DNA samples are pooled and arrayed on microtiter plates and subjected to gene specific PCR. The PCR amplification products may be screened for mutations in the *SOD7* and/or *NGAL3* target gene using any method that identifies heteroduplexes between wild type and mutant genes. For example, but not limited to, denaturing high pressure liquid chromatography (dHPLC), constant denaturant capillary electrophoresis (CDCE), temperature gradient capillary electrophoresis (TGCE), or by fragmentation using chemical cleavage. Preferably the PCR amplification products are incubated with an endonuclease that preferentially cleaves mismatches in heteroduplexes between wild type and mutant sequences. Cleavage products are electrophoresed using an automated sequencing gel apparatus, and gel images are analyzed with the aid of a standard commercial image-processing program. Any primer specific to the *SOD7* or *NGAL3* nucleic acid sequence may be utilized to amplify the *SOD7* or *NGAL3* nucleic acid sequence within the pooled DNA sample. Preferably, the primer is designed to amplify the regions of the *SOD7* and/or *NGAL3* gene where useful mutations are most likely to arise, specifically in the areas of the *SOD7* and/or *NGAL3* gene that are highly conserved and/or confer activity as explained elsewhere. To facilitate detection of PCR products on a gel, the PCR primer may be labelled using any conventional labelling method.

Rapid high-throughput screening procedures thus allow the analysis of amplification products for identifying a mutation conferring the reduction or inactivation of the expression of the *SOD7* and/or *NGAL3* gene as compared to a corresponding non-mutagenised wild type plant. Once a mutation is identified in a gene of interest, the

seeds of the M2 plant carrying that mutation are grown into adult M3 plants and screened for the phenotypic characteristics associated with the target gene *SOD7* or *NGAL3.* Loss of function or reduced function mutants with increased seed size compared to a control can thus be identified.

Plants obtained or obtainable by such method which carry a functional mutation in the endogenous *SOD7* and/or *NGAL3* locus are also within the scope of the invention

In another embodiment, RNA-mediated gene suppression or RNA silencing may be used to achieve silencing of the *SOD7* and/or *NGAL3* nucleic acid sequence. "Gene silencing" is a term generally used to refer to suppression of expression of a gene via sequence-specific interactions that are mediated by RNA molecules. The degree of reduction may be so as to totally abolish production of the encoded gene product, but more usually the abolition of expression is partial, with some degree of expression remaining. The term should not therefore be taken to require complete "silencing" of expression.

Transgenes may be used to suppress endogenous plant genes. This was discovered originally when chalcone synthase transgenes in petunia caused suppression of the endogenous chalcone synthase genes and indicated by easily visible pigmentation changes. Subsequently it has been described how many, if not all plant genes can be "silenced" by transgenes. Gene silencing requires sequence similarity between the transgene and the gene that becomes silenced. This sequence homology may involve promoter regions or coding regions of the silenced target gene. When coding regions are involved, the transgene able to cause gene silencing may have been constructed with a promoter that would transcribe either the sense or the antisense orientation of the coding sequence RNA. It is likely that the various examples of gene silencing involve different mechanisms that are not well understood. In different examples there may be transcriptional or post-transcriptional gene silencing and both may be used according to the methods of the invention.

The mechanisms of gene silencing and their application in genetic engineering, which were first discovered in plants in the early 1990s and then shown in *Caenorhabditis elegans* are extensively described in the literature.

RNA-mediated gene suppression or RNA silencing according to the methods of the invention includes co-suppression wherein over-expression of the target sense RNA or mRNA, that is the *SOD7* and/or *NGAL3* sense RNA or mRNA, leads to a reduction in the level of expression of the genes concerned. RNAs of the transgene and homologous endogenous gene are co-ordinately suppressed. Other techniques used in the methods of the invention include antisense RNA to reduce transcript levels of the endogenous target gene in a plant. In this method, RNA silencing does not affect the transcription of a gene locus, but only causes sequence-specific degradation of target mRNAs. An "antisense" nucleic acid sequence comprises a nucleotide sequence that is complementary to a "sense" nucleic acid sequence encoding a NGAL2 and/or NGAL3 protein, or a part of the protein, i.e. complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA transcript sequence. The antisense nucleic acid sequence is preferably complementary to the endogenous *SOD7* and/or *NGAL3* gene to be silenced. The complementarity may be located in the "coding region" and/or in the "non-coding region" of a gene. The term "coding region" refers to a region of the nucleotide sequence comprising codons that are translated into amino acid residues. The term "non-coding region" refers to 5' and 3' sequences that flank the coding region that are transcribed but not translated into amino acids (also referred to as 5' and 3' untranslated regions).

Antisense nucleic acid sequences can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid sequence may be complementary to the entire *SOD7* and/or *NGAL3* nucleic acid sequence, but may also be an oligonucleotide that is antisense to only a part of the nucleic acid sequence (including the mRNA 5' and 3' UTR). For example, the antisense oligonucleotide sequence may be complementary to the region surrounding the translation start site of an mRNA transcript encoding a polypeptide. The length of a suitable antisense oligonucleotide sequence is known in the art and may start from about 50, 45, 40, 35, 30, 25, 20, 15 or 10 nucleotides in length or less. An antisense nucleic acid sequence according to the invention may be constructed using chemical synthesis and enzymatic ligation reactions using methods known in the art. For example, an antisense nucleic acid sequence (e.g., an antisense oligonucleotide sequence) may be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acid sequences, e.g., phosphorothioate derivatives and acridine-substituted nucleotides

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

The nucleic acid molecules used for silencing in the methods of the invention hybridize with or bind to mRNA transcripts and/or insert into genomic DNA encoding a polypeptide to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid sequence which binds to DNA duplexes, through specific interactions in the major groove of the double helix. Antisense nucleic acid sequences may be introduced into a plant by transformation or direct injection at a specific tissue site. Alternatively, antisense nucleic acid sequences can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense nucleic acid sequences can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid sequence to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid sequences can also be delivered to cells using vectors.

RNA interference (RNAi) is another post-transcriptional gene-silencing phenomenon which may be used according to the methods of the invention. This is induced by double-stranded RNA in which mRNA that is homologous to the dsRNA is specifically degraded. It refers to the process of sequence-specific post-transcriptional gene silencing mediated by short interfering RNAs (siRNA). The process of RNAi begins when the enzyme, DICER, encounters dsRNA and chops it into pieces called small-interfering RNAs (siRNA). This enzyme belongs to the RNase III nuclease family. A complex of proteins gathers up these RNA remains and uses their code as a guide to search out and destroy any RNAs in the cell with a matching sequence, such as target mRNA.

Artificial and/or natural microRNAs (miRNAs) may be used to knock out gene expression and/or mRNA translation. MicroRNAs (miRNAs) miRNAs are typically single stranded small RNAs typically 19-24 nucleotides long. Most plant miRNAs have perfect or near-perfect complementarity with their target sequences. However, there are natural targets with up to five mismatches. They are processed from longer non-coding RNAs with characteristic fold-back structures by double-strand specific RNases of the Dicer family. Upon processing, they are incorporated in the RNA-induced silencing complex (RISC) by binding to its main component, an Argonaute protein. miRNAs serve as the specificity components of RISC, since they base-pair to target nucleic acids, mostly mRNAs, in the cytoplasm. Subsequent regulatory events include target mRNA cleavage and destruction and/or translational inhibition. Effects of miRNA overexpression are thus often reflected in decreased mRNA levels of target genes. Artificial microRNA (amiRNA) technology has been applied in *Arabidopsis thaliana* and other plants to efficiently silence target genes of interest. The design principles for amiRNAs have been generalized and integrated into a Web-based tool (http://wmd.weigelworld.org).

Thus, according to the various aspects of the invention a plant may be transformed to introduce a RNAi, shRNA, snRNA, dsRNA, siRNA, miRNA, ta-siRNA, amiRNA or cosuppression molecule that has been designed to target the expression of an *SOD7* and/or *NGAL3* nucleic acid sequence and selectively decreases or inhibits the expression of the gene or stability of its transcript. Preferably, the RNAi, snRNA, dsRNA, shRNA siRNA, miRNA, amiRNA, ta-siRNA or cosuppression molecule used according to the various aspects of the invention comprises a fragment of at least 17 nt, preferably 22 to 26 nt and can be designed on the basis of the information shown in SEQ ID NO: 1. Guidelines for designing effective siRNAs are known to the skilled person. Briefly, a short fragment of the target gene sequence (e.g., 19-40 nucleotides in length) is chosen as the target sequence of the siRNA of the invention. The short fragment of target gene sequence is a fragment of the target gene mRNA. In preferred embodiments, the criteria for choosing a sequence fragment from the target gene mRNA to be a candidate siRNA molecule include 1) a sequence from the target gene mRNA that is at least 50-100 nucleotides from the 5' or 3' end of the native mRNA molecule, 2) a sequence from the target gene mRNA that has a G/C content of between 30% and 70%, most preferably around 50%, 3) a sequence from the target gene mRNA that does not contain repetitive sequences (e.g., AAA, CCC, GGG, TTT, AAAA, CCCC, GGGG, TTTT), 4) a sequence from the target gene mRNA that is

accessible in the mRNA, 5) a sequence from the target gene mRNA that is unique to the target gene, 6) avoids regions within 75 bases of a start codon. The sequence fragment from the target gene mRNA may meet one or more of the criteria identified above. The selected gene is introduced as a nucleotide sequence in a prediction program that takes into account all the variables described above for the design of optimal oligonucleotides. This program scans any mRNA nucleotide sequence for regions susceptible to be targeted by siRNAs. The output of this analysis is a score of possible siRNA oligonucleotides. The highest scores are used to design double stranded RNA oligonucleotides that are typically made by chemical synthesis. In addition to siRNA which is complementary to the mRNA target region, degenerate siRNA sequences may be used to target homologous regions. siRNAs according to the invention can be synthesized by any method known in the art. RNAs are preferably chemically synthesized using appropriately protected ribonucleoside phosphoramidites and a conventional DNA/RNA synthesizer. Additionally, siRNAs can be obtained from commercial RNA oligonucleotide synthesis suppliers.

siRNA molecules according to the aspects of the invention may be double stranded. In one embodiment, double stranded siRNA molecules comprise blunt ends. In another embodiment, double stranded siRNA molecules comprise overhanging nucleotides (e.g., 1-5 nucleotide overhangs, preferably 2 nucleotide overhangs). In some embodiments, the siRNA is a short hairpin RNA (shRNA); and the two strands of the siRNA molecule may be connected by a linker region (e.g., a nucleotide linker or a non-nucleotide linker). The siRNAs of the invention may contain one or more modified nucleotides and/or non-phosphodiester linkages. Chemical modifications well known in the art are capable of increasing stability, availability, and/or cell uptake of the siRNA. The skilled person will be aware of other types of chemical modification which may be incorporated into RNA molecules.

In one embodiment, recombinant DNA constructs as described in US 6635805, incorporated herein by reference, may be used.

The silencing RNA molecule is introduced into the plant using conventional methods, for example a vector and Agrobacterium-mediated transformation. Stably transformed plants are generated and expression of the *SOD7* and/or *NGAL3* gene compared to a wild type control plant is analysed.

Silencing of the *SOD7* and/or *NGAL3* nucleic acid sequence may also be achieved using virus-induced gene silencing.

Thus, in one embodiment of the invention, the plant expresses a nucleic acid construct comprising a RNAi, shRNA snRNA, dsRNA, siRNA, miRNA, ta-siRNA, amiRNA or co-suppression molecule that targets the *SOD7* or *NGAL3* nucleic acid sequence as described herein and reduces expression of the endogenous *SOD7* or *NGAL3* nucleic acid sequence. A gene is targeted when, for example, the RNAi, snRNA, dsRNA, siRNA, shRNA miRNA, ta-siRNA, amiRNA or cosuppression molecule selectively decreases or inhibits the expression of the gene compared to a control plant. Alternatively, a RNAi, snRNA, dsRNA, siRNA, miRNA, ta-siRNA, amiRNA or cosuppression molecule targets A *SOD7* or *NGAL3* nucleic acid sequence when the RNAi, shRNA snRNA, dsRNA, siRNA, miRNA, ta-siRNA, amiRNA or cosuppression molecule hybridises under stringent conditions to the gene transcript.

Gene silencing may also occur if there is a mutation on an endogenous gene and/or a mutation on an isolated gene/nucleic acid subsequently introduced into a plant. The reduction or substantial elimination may be caused by a non-functional polypeptide. For example, the polypeptide may bind to various interacting proteins; one or more mutation(s) and/or truncation(s) may therefore provide for a polypeptide that is still able to bind interacting proteins (such as receptor proteins) but that cannot exhibit its normal function (such as signalling ligand).

A further approach to gene silencing is by targeting nucleic acid sequences complementary to the regulatory region of the gene (e.g., the promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. Other methods, such as the use of antibodies directed to an endogenous polypeptide for inhibiting its function *in planta,* or interference in the signalling pathway in which a polypeptide is involved, will be well known to the skilled man. In particular, it can be envisaged that manmade molecules may be useful for inhibiting the biological function of a target polypeptide, or for interfering with the signalling pathway in which the target polypeptide is involved.

In one embodiment, the suppressor nucleic acids may be anti-sense suppressors of expression of the NGAL2 or NGAL3 polypeptides. In using anti-sense sequences to down-regulate gene expression, a nucleotide sequence is placed under the control of a

promoter in a "reverse orientation" such that transcription yields RNA which is complementary to normal mRNA transcribed from the "sense" strand of the target gene.

An anti-sense suppressor nucleic acid may comprise an anti-sense sequence of at least 10 nucleotides from the target nucleotide sequence. It may be preferable that there is complete sequence identity in the sequence used for down-regulation of expression of a target sequence, and the target sequence, although total complementarity or similarity of sequence is not essential. One or more nucleotides may differ in the sequence used from the target gene. Thus, a sequence employed in a down-regulation of gene expression in accordance with the present invention may be a wild-type sequence (e.g. gene) selected from those available, or a variant of such a sequence.

The sequence need not include an open reading frame or specify an RNA that would be translatable. It may be preferred for there to be sufficient homology for the respective anti-sense and sense RNA molecules to hybridise. There may be down regulation of gene expression even where there is about 5%, 10%, 15% or 20% or more mismatch between the sequence used and the target gene. Effectively, the homology should be sufficient for the down-regulation of gene expression to take place.

Suppressor nucleic acids may be operably linked to tissue-specific or inducible promoters. For example, integument and seed specific promoters can be used to specifically down-regulate a *SOD7* or *NGAL3* nucleic acids in developing ovules and seeds to increase final seed size.

Nucleic acid which suppresses expression of a NGAL2 or NGAL3 polypeptide as described herein may be operably linked to a heterologous regulatory sequence, such as a promoter, for example a constitutive, inducible, tissue-specific or developmental specific promoter. The construct or vector may be transformed into plant cells and expressed as described herein. Plant cells comprising such vectors are also within the scope of the invention.

In another aspect, the invention relates to a silencing construct to silence expression of *NGAL2* or *NGAL3* obtainable or obtained by a method as described herein and to a plant cell comprising such construct. Accordingly, the invention also relates to the use

of a nucleic acid sequence comprising or consisting of SEQ ID NO: 1, 2 or 3 or a part thereof or a homologue of SEQ ID NO: 1, 2 or 3 or a part thereof in silencing expression of *NGAL2* or *NGAL3.* Host cells transformed with such construct are also within the scope of the invention.

Recently, genome editing techniques have emerged as alternative methods to conventional mutagenesis methods (such as physical and chemical mutagenesis) or methods using the expression of transgenes in plants to produce mutant plants with improved phenotypes that are important in agriculture. These techniques employ sequence-specific nucleases (SSNs) including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the RNA-guided nuclease Cas9 (CRISPR/Cas9), which generate targeted DNA double-strand breaks (DSBs), which are then repaired mainly by either error-prone non-homologous end joining (NHEJ) or high-fidelity homologous recombination (HR). The SSNs have been used to create targeted knockout plants in various species ranging from the model plants, *Arabidopsis* and tobacco, to important crops, such as barley, soybean, rice and maize. Heritable gene modification has been demonstrated in *Arabidopsis* and rice using the CRISPR/Cas9 system and TALENs.

Targeted genome modification or targeted genome editing is a genome engineering technique that uses targeted DNA double-strand breaks (DSBs) to stimulate genome editing through homologous recombination (HR)-mediated recombination events. To achieve effective genome editing via introduction of site-specific DNA DSBs, four major classes of customizable DNA binding proteins can be used: meganucleases derived from microbial mobile genetic elements, ZF nucleases based on eukaryotic transcription factors, transcription activator-like effectors (TALEs) from Xanthomonas bacteria, and the RNA-guided DNA endonuclease Cas9 from the type II bacterial adaptive immune system CRISPR (clustered regularly interspaced short palindromic repeats). Meganuclease, ZF, and TALE proteins all recognize specific DNA sequences through protein-DNA interactions. Although meganucleases integrate its nuclease and DNA-binding domains, ZF and TALE proteins consist of individual modules targeting 3 or 1 nucleotides (nt) of DNA, respectively. ZFs and TALEs can be assembled in desired combinations and attached to the nuclease domain of FokI to direct nucleolytic activity toward specific genomic loci.

Upon delivery into host cells via the bacterial type III secretion system, TAL effectors enter the nucleus, bind to effector-specific sequences in host gene promoters and activate transcription. Their targeting specificity is determined by a central domain of tandem, 33-35 amino acid repeats. This is followed by a single truncated repeat of 20 amino acids. The majority of naturally occurring TAL effectors examined have between 12 and 27 full repeats.

These repeats only differ from each other by two adjacent amino acids, their repeat-variable di-residue (RVD). The RVD that determines which single nucleotide the TAL effector will recognize: one RVD corresponds to one nucleotide, with the four most common RVDs each preferentially associating with one of the four bases. Naturally occurring recognition sites are uniformly preceded by a T that is required for TAL effector activity. TAL effectors can be fused to the catalytic domain of the FokI nuclease to create a TAL effector nuclease (TALEN) which makes targeted DNA double-strand breaks (DSBs) *in vivo* for genome editing. The use of this technology in genome editing is well described in the art, for example in US 8,440,431, US 8,440,432 and US 8,450,471. Reference 30 describes a set of customized plasmids that can be used with the Golden Gate cloning method to assemble multiple DNA fragments. As described therein, the Golden Gate method uses Type IIS restriction endonucleases, which cleave outside their recognition sites to create unique 4 bp overhangs. Cloning is expedited by digesting and ligating in the same reaction mixture because correct assembly eliminates the enzyme recognition site. Assembly of a custom TALEN or TAL effector construct and involves two steps: (i) assembly of repeat modules into intermediary arrays of 1-10 repeats and (ii) joining of the intermediary arrays into a backbone to make the final construct.

Another genome editing method that can be used according to the various aspects of the invention is CRISPR. The use of this technology in genome editing is well described in the art, for example in US 8,697,359 and references cited herein. In short, CRISPR is a microbial nuclease system involved in defense against invading phages and plasmids. CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage (sgRNA). Three types (I-III) of CRISPR systems have been identified across a wide range of bacterial hosts. One key feature of each CRISPR locus is the presence of an array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences

(spacers). The non-coding CRISPR array is transcribed and cleaved within direct repeats into short crRNAs containing individual spacer sequences, which direct *Cas* nucleases to the target site (protospacer). The Type II CRISPR is one of the most well characterized systems and carries out targeted DNA double-strand break in four sequential steps. First, two non-coding RNA, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the target DNA via Watson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the protospacer.

Cas9 is thus the hallmark protein of the type II CRISPR-Cas system, and a large monomeric DNA nuclease guided to a DNA target sequence adjacent to the PAM (protospacer adjacent motif) sequence motif by a complex of two noncoding RNAs: CRIPSR RNA (crRNA) and trans-activating crRNA (tracrRNA). The Cas9 protein contains two nuclease domains homologous to RuvC and HNH nucleases. The HNH nuclease domain cleaves the complementary DNA strand whereas the RuvC-like domain cleaves the non-complementary strand and, as a result, a blunt cut is introduced in the target DNA. Heterologous expression of Cas9 together with an sgRNA can introduce site-specific double strand breaks (DSBs) into genomic DNA of live cells from various organisms. For applications in eukaryotic organisms, codon optimized versions of Cas9, which is originally from the bacterium Streptococcus pyogenes, have been used.

The single guide RNA (sgRNA) is the second component of the CRISPR/Cas system that forms a complex with the Cas9 nuclease. sgRNA is a synthetic RNA chimera created by fusing crRNA with tracrRNA. The sgRNA guide sequence located at its 5' end confers DNA target specificity. Therefore, by modifying the guide sequence, it is possible to create sgRNAs with different target specificities. The canonical length of the guide sequence is 20 bp. In plants, sgRNAs have been expressed using plant RNA polymerase III promoters, such as U6 and U3.

Using these techniques, it is possible to specifically target conserved domains to abolish the function of the NGAL2 and/or NGAL3 polypeptide.

For example, the conserved B3 domain or repression motif may be targeted.

Thus, in another embodiment of the invention directed to a mutant plant, plant cell, plant or a part thereof characterised in that the activity of a NGAL2 polypeptide is altered and said plant expresses a nucleic acid comprising a mutant SEQ ID NO. 1 or 2 and encoding a mutant NGAL2 polypeptide, a functional homologue or variant thereof, for example one which carries a mutation in the B3 or repressor domain.

Thus, in another embodiment of the invention directed to a mutant plant, plant cell, plant or a part thereof characterised in that the activity of a NGAL3 polypeptide is altered and said plant expresses a nucleic acid comprising a mutant SEQ ID NO. 4 and encoding a mutant NGAL3 polypeptide, a functional homologue or variant thereof which carries a mutation in the B3 or repressor domain.

In a preferred embodiment, the invention directed to a mutant plant, plant cell, plant or a part thereof characterised in that the activity of a NGAL2 and a NGAL3 polypeptide is altered and said plant expresses a nucleic acid comprising a mutant SEQ ID NO. 1 or 2 and encoding a mutant NGAL2 polypeptide, a functional homologue or variant thereof, for example one which carries a mutation in the B3 or repressor domain and said plant expresses a nucleic acid comprising a mutant SEQ ID NO. 4 and encoding a mutant NGAL3 polypeptide which carries a mutation in the B3 or repressor domain.

Mutations in the promoter region of *SOD7* and/or *NGAL3* resulting in a loss of function are also within the scope of the invention.

Constructs designed using the genome editing technologies to knock out or knock down *NGAL2* or *NGAL3,* for example as shown herein, are also within the scope of the invention as well as host cells comprising these constructs. In one embodiment, the constructs comprise or consist of a sequence selected from SEQ ID NO: 155, 156, 157 or 158. Accordingly, in a further aspect of the invention, there is provided a nucleic acid construct comprising a sequence selected from SEQ ID NO: 155, 156, 157 or 158. In a further aspect of the invention, there is provided a nucleic acid construct comprising at least one CRISPR target sequence, wherein the target sequence is selected from SEQ ID Nos 150, 160, 161 , 162 and 163. Preferably, the target sequence comprises at least

two CRISPR target sequences, preferably SEQ ID No 159 and 160 or SEQ ID No 161 and 162, or SEQ ID No 161 and 163 or SEQ ID No 159 and 163.

In another embodiment of the methods of the invention, inactivating, repressing or down-regulating the activity of NGAL2 and/or NGAL3 can be achieved by manipulating the expression of *SOD7* and/or *NGAL3* inhibitors in a plant, for example transgenic plant. For example, a gene expressing a protein that inhibits the expression of the *SOD7* and/or *NGAL3* gene or activity of the *SOD7* and/or *NGAL3* protein can be introduced into a plant and over-expressed. The inhibitor may interact with the regulatory sequences that direct *SOD7* and/or *NGAL3* gene expression to down-regulate or repress *SOD7* and/or *NGAL3* gene expression. For example, the inhibitor may be a transcriptional repressor. Alternatively, it may interact and repress transcriptional regulators, for example transcription factors, that positively regulate expression of the *SOD7* and/or *NGAL3* gene. Alternatively, the inhibitor it may directly interact with the NGAL2 and/or NGAL3 protein to inhibit its activity or interact with modulators of the NGAL2 and/or NGAL3 protein. For example, the activity of the NGAL2 and/or NGAL3 protein may be inactivated, repressed or down-regulated by manipulating post-transcriptional modifications, of the NGAL2 and/or NGAL3 protein resulting in a reduced or lost activity.

In one embodiment, the methods of the invention comprise comparing the activity of the NGAL2 and/or NGAL3 polypeptide and/or expression of the *SOD7* and/or *NGAL3* gene with the activity of the NGAL2 and/or NGAL3 polypeptide and/or expression of the *SOD7* and/or *NGAL3* gene in a control plant.

In another aspect, the invention relates to a plant obtainable or obtained by a method as described herein.

In another aspect, the invention relates to an expression cassette comprising an isolated nucleic acid sequence comprising or consisting of a sequence as shown in SEQ ID NO: 1 or 2 a functional part, variant, homologue or orthologue thereof operably linked to a regulatory element. In another aspect, the invention relates to an expression cassette comprising an isolated nucleic acid sequence comprising or consisting of a sequence as shown in SEQ ID NO: 4 or a functional part, variant, homologue or orthologue thereof operably linked to a regulatory element. The regulatory element may be a promoter. The invention also relates to a vector comprising such expression

cassette. The invention also relates to a composition comprising the two expression cassettes above.

In the methods described here, plants can be regenerated from plants transformed or genetically altered as described above and the phenotype, specifically the seed phenotype is analysed by known methods.

Transformation methods are known in the art. The nucleic acid sequence is introduced into said plant through a process called transformation. The term "introduction" or "transformation" as referred to herein encompasses the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for transfer. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a genetic construct of the present invention and a whole plant regenerated there from. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem). The polynucleotide may be transiently or stably introduced into a host cell and may be maintained non-integrated, for example, as a plasmid. Alternatively, it may be integrated into the host genome. The resulting transformed plant cell may then be used to regenerate a transformed plant in a manner known to persons skilled in the art.

The transfer of foreign genes into the genome of a plant is called transformation. Transformation of plants is now a routine technique in many species. Advantageously, any of several transformation methods may be used to introduce the gene of interest into a suitable ancestor cell. The methods described for the transformation and regeneration of plants from plant tissues or plant cells may be utilized for transient or for stable transformation. Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts, electroporation of protoplasts, microinjection into plant material, DNA or RNA-coated particle bombardment, infection with (non-integrative) viruses and

the like. Transgenic plants, including transgenic crop plants, are preferably produced via *Agrobacterium tumefaciens* mediated transformation.

To select transformed plants, the plant material obtained in the transformation is, as a rule, subjected to selective conditions so that transformed plants can be distinguished from untransformed plants. For example, the seeds obtained in the above-described manner can be planted and, after an initial growing period, subjected to a suitable selection by spraying. A further possibility is growing the seeds, if appropriate after sterilization, on agar plates using a suitable selection agent so that only the transformed seeds can grow into plants. Alternatively, the transformed plants are screened for the presence of a selectable marker such as the ones described above. Following DNA transfer and regeneration, putatively transformed plants may also be evaluated, for instance using Southern analysis, for the presence of the gene of interest, copy number and/or genomic organisation. Alternatively or additionally, expression levels of the newly introduced DNA may be monitored using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

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

The various aspects of the invention described herein clearly extend to any plant cell or any plant produced, obtained or obtainable by any of the methods described herein, and to all plant parts and propagules thereof unless otherwise specified. The present invention extends further to encompass the progeny of a primary transformed or transfected cell, tissue, organ or whole plant that has been produced by any of the aforementioned methods, the only requirement being that progeny exhibit the same genotypic and/or phenotypic characteristic(s) as those produced by the parent in the methods according to the invention.

The invention also extends to harvestable parts of a plant of the invention as described above such as, but not limited to seeds, leaves, fruits, flowers, stems, roots, rhizomes, tubers and bulbs. The invention furthermore relates to products derived, preferably directly derived, from a harvestable part of such a plant, such as dry pellets or powders, oil, fat and fatty acids, starch or proteins. The invention also relates to food products and food supplements comprising the plant of the invention or parts thereof.

While the foregoing disclosure provides a general description of the subject matter encompassed within the scope of the present invention, including methods, as well as the best mode thereof, of making and using this invention, the following examples are provided to further enable those skilled in the art to practice this invention and to provide a complete written description thereof. However, those skilled in the art will appreciate that the specifics of these examples should not be read as limiting on the invention, the scope of which should be apprehended from the claims and equivalents thereof appended to this disclosure. Various further aspects and embodiments of the present invention will be apparent to those skilled in the art in view of the present disclosure.

All documents mentioned in this specification are incorporated herein by reference in their entirety, including references to gene and protein accession numbers.

"and/or" where used herein is to be taken as specific disclosure of each of the multiple specified features or components with or without the other at each combination unless otherwise dictated. For example "A, B and/or C" is to be taken as specific disclosure of each of (i) A, (ii) B, (iii) C, (iv) A and B, (v) B and C or (vi) A and B and C, just as if each is set out individually herein.

Unless context dictates otherwise, the descriptions and definitions of the features set out above are not limited to any particular aspect or embodiment of the invention and apply equally to all aspects and embodiments which are described.

The invention is further described in the following non-limiting examples.

Examples

METHODS

Plant materials and growth conditions

Arabidopsis thaliana Columbia (Col-0) was used as wild-type line. The *da1-1,* sod7-1D, *sod7-ko1* and *ngal3-ko1* were in the Col-0 background. *sod7-1D* was identified as a suppressor of *da1-1* by using T-DNA activation tagging method. The *sod7-ko1* (SM_3_34191) and *ngal3-ko1* (SM_3_36641) were identified in AtIDB (www.atidb.org) and obtained from Arabidopsis Stock Centre NASC collection. T-DNA insertions were confirmed by PCR and sequencing by using the primers described in Table 1. Arabidopsis plants were grown under long-day conditions (16 h light/8 h dark) at 22°C. Activation tagging screening The activation tagging plasmid pJFAT260 was introduced into the *da1-1* mutant plants using Agrobacterium tumefaciens strain GV3101 (Fan et al., 2009; Fang et al., 2012), and T1 plants were selected by using the herbicide Basta. Seeds produced from T1 plants were used to isolate modifiers of *da1-1.*

Morphological and cellular analysis

To measure seed size, we photographed dry seeds of the wild type and mutants under a Leica microscope (LEICA S8APO) using Leica CCD (DFC420). The projective area of wild-type and mutant seeds was measured by using Image J software. Average seed weight was determined by weighing mature dry seeds in batches of 100 using an electronic analytical balance (METTLER TOLEDO AL104, China). The weights of five sample batches were measured for each seed lot. Fully expanded cotyledons, petals (stage 14) and leaves were scanned to produce digital images for area measurement. To measure cell number and cell size, petals, leaves, ovules and seeds were placed in a drop of clearing solution [30ml H20, 80g Chloral hydrate (Sigma, C8383), 10 ml 100% Glycerol (Sigma, G6279)]. Cleared Samples were imaged under a Leica microscope (LEICA DM2500) with differential interference contrast (DIC) optics and photographed with a SPOT FLEX Cooled CCD Digital Imaging System. Area measurement was made by using Image J software.

Cloning of the SOD7 gene

The flanking sequences of the T-DNA insertion of the *sod7-1D* mutant were identified by the thermal asymmetric interlaced PCR (TAIL-PCR) according to a previously reported method (Liu et al., 1995). Briefly, TAIL-PCR utilizes three nested specific primers (OJF22, OJF23 and OJF24) within the T-DNA region of the pJFAT260 vector and a shorter arbitrary degenerate primer (AD1). Thus, the relative amplification efficiencies of specific and non-specific products can be thermally controlled. TAIL-

PCR products were sequenced using the primer OJF24. The specific primers OJF22, OJF23 and OJF24 and an arbitrary degenerate (AD1) primer are described in Table 1.

Constructs and plant transformation

The *35S:GFP-SOD7, pSOD7:SOD7-GFP* and *pSOD7:GUS* constructs were made using a PCR-based Gateway system. The coding sequence (CDS) of SOD7 was amplified using the primers SOD7CDS-F and SOD7CDS-R (Table 1). PCR products were cloned into pCR8/TOPO TA cloning vector. The SOD7 CDS was then subcloned into the binary vector pMDC43 with the GFP gene to generate the transformation plasmid 35S:GFP-SOD7. The SOD7 genomic sequence containing 2040-bp promoter sequence and 2104-bp SOD7 gene was amplified using the primers SOD7G-F and SOD7G-R (Table 1). PCR products were cloned into pCR8/TOPO TA cloning vector. The SOD7 genomic sequence was then subcloned into the binary vectors pMDC107 with the GFP gene to generate the transformation plasmid *pSOD7:SOD7-GFP.* The 2262-bp SOD7 promoter sequence was amplified using the primers SOD7P-F and SOD7P-R (Table 1). PCR products were cloned into pCR8/TOPO TA cloning vector. The SOD7 promoter was then subcloned into the binary vectors pGWB3 with the GUS gene to generate the transformation plasmid pSOD7:GUS. The plasmids *35S:GFP-SOD7, pSOD7:SOD7-GFP* and *pSOD7:GUS* were introduced into Col-0 or *sod7-ko1 ngal3ko1* plants using Agrobacterium tumefaciens GV3101 , respectively, and transformants were selected on hygromycin (30μg/ml)-containing medium. The SOD7 cDNA was cloned into the ApaI and SpeI sites of the binary vector pER8 to generate a chemically inducible construct pER8-SOD7. The specific primers for the pER8-SOD7 construct were SOP7ER-F and SOD7ER-R. The plasmid pER8-SOD7 was introduced into Col-0 plants using Agrobacterium tumefaciens GV3101 , and transformants were selected on hygromycin (30μg/ml)-containing medium. GUS staining Samples (pSOD7:GUS) were stained in a GUS staining solution (1 mM X-gluc, 50 Mm NaP04 buffer, 0.4 mM each K3Fe(CN)6/K4Fe(CN)6, and 0.1% (v/v) Triton X-100) and incubated at 37°C for 3 hours. After GUS staining, chlorophyll was removed by 70% ethanol. RT-PCR and quantitative real-time RT-PCR. Total RNA was extracted from Arabidopsis seedlings using an RNAprep pure Plant kit (TIANGEN). mRNA was reverse transcribed into cDNA using SuperScriptIII reverse transcriptase (Invitrogen). cDNA samples were standardized on ACTIN2 transcript amount using the primers ACTIN2-F and ACTIN2-R (Table 1). Quantitative real-time RT-PCR analysis was performed with a Lightcycler 480 machine (Roche) using the Lightcycler 480 SYBR Green I Master (Roche). ACTIN2 mRNA was used as an internal control, and relative

amounts of mRNA were calculated using the comparative threshold cycle method. The primers used for RT-PCR and quantitative real-time RT-PCR are described in Table 1.

The chromatin immunoprecipitation (ChIP) assay

The chromatin immunoprecipitation (ChIP) assay was performed as described previously with minor modifications (Gendrel et al., 2005). Briefly, 35S:GFP and *35S:GFP-SOD7* transgenic seeds were grown on 1/2 MS plates for 10 days. The seedlings were cross-linked by 1% formaldehyde for 15 min in vacuum and stopped by 0.125 M Glycine. Samples were ground in liquid nitrogen, and nuclei were isolated. Chromatin was immunoprecipitated by anti-GFP (Roche, 11814460001) and protein A+G beads (Millpore Magna ChIP Protein A+G Magnetic Beads, 16-663). DNA was precipitated by glycogen, NaOAc and ethanol, washed by 70% ethanol, and dissolved in 60 μl of water. Gene-specific primers (PF1-F, PF1-R, PF-2F, PF2-R, ACTIN7-ChIP-F, and ACTIN7-ChIP-R) were used to quantify the enrichment of each fragment (Table 1).

The DNA electrophoretic mobility shift assay (EMSA)

The coding sequence of SOD7 was cloned into the NdeI and BamHI sites of the pMAL-C2 vector to generate the construct MBP-SOD7. MBP-SOD7 fusion proteins were expressed in Escherichia coli BL21 (DE3) (Biomed) and purified by Amylose resins(New England Biolabs). The biotin-labeled and unlabeled probes were synthesized as forward and reverse strands. The forward and reverse strands were then incubated in a solution (50mM Tris-HCl, 5mM EDTA and 250mM NaCl) at 95 °c for 10min and renatured to double stranded probes at room temperature. The gel-shift assay was performed according to the method described previously (Smaczniak et al., 2012).

Results

sod7-1D suppresses the seed size phenotype of *da1-1*

We previously identified the ubiquitin receptor DA1 as a negative regulator of seed size in Arabidopsis (Li et al., 2008). The *da1-1* mutant formed large seeds due to increased cell proliferation in the maternal integuments (Li et al., 2008; Xia et al., 2013). To identify novel components in the DA1 pathway or other seed size regulators, we initiated a T-DNA activation tagging screen for modifiers of *da1-1* (Fang et al., 2012). A dominant suppressor of *da1-1* (sod7-1 D) was isolated from seeds produced from

approximate 16,000 T1 plants (Fig.lA). Seeds of the *sod7-1D da1-1* double mutant were significantly smaller and lighter than *da1-1* seeds (Figures 1A, E and F). The embryo constitutes the major volume of a mature seed in Arabidopsis. *sod7-1D da1-1* embryos were smaller than *da1-1* embryos (Figure 1B). The size of *sod7-1D da1-1* cotyledons was significantly reduced, compared with that of *da1-1* cotyledons (Figure 1G). In addition, *sod7-1D da1-1* double mutant formed smaller leaves and flowers than *da1-1* (Figures 1C and 1D). Thus, these results show that the *sod7-1D* mutation suppressed the seed and organ size phenotypes of *da1-1.*

sod7-1D produces small seeds

We isolated the single *sod7-1D* mutant among F2 progeny derived from a cross between the wild type (Col-0) and *sod7-1D da1-1.* The *sod7-1D* seeds were significantly smaller and lighter than wild-type seeds (Figures 2A, B, G and H). We further isolated and visualized embryos from mature wild-type and *sod7-1D* seeds. The *sod7-1D* embryos were obviously smaller than wild-type embryos (Figures 2C and D). The changes in seed size were also reflected in the size of seedlings (Figures 2E and F). The 10-d old *sod7-1D* cotyledons were significantly smaller than wild-type cotyledons (Figure. 2E, F and I). In addition, the *sod7-1D* mutants exhibited small leaves and flowers compared with the wild type. The decreased size of *sod7-1D* leaves and petals was not caused by smaller cells, indicating that the *sod7-1D* mutation results in a decrease in cell number. In fact, the average area of epidermal cells in *sod7-1D* petals was larger than that in wild-type petals, suggesting a possible compensation mechanism between cell number and cell size.

SOD7 encodes a B3 domain transcriptional repressor NGAL2

To determine whether the seed and organ size phenotypes of *sod7-1D* was caused by the T-DNA insertion, we firstly analyzed the genetic linkage of the mutant phenotypes with Basta resistance, which is conferred by the selectable marker of the activation tagging vector (Fan et al., 2009). In a T2 population, 181 plants with *sod7-1D da1-1* phenotypes were resistant, whereas 55 plants with *da1-1* phenotypes were sensitive, indicating that the insertion is cosegregated with the *sod7-1D* phenotypes. To clone the SOD7 gene, we isolated the T-DNA flanking sequences using thermal asymmetric interlaced PCR (Liu et al., 1995). DNA sequencing revealed that the T-DNA had inserted approximately 5.6 kb upstream of the At3g1 1580 and about 3.7 kb upstream of the At3g1 1590 gene (Figure 3A). To determine which gene is responsible for the *sod7-1D* phenotypes, we examined the mRNA levels of these two genes. The mRNA

of the At3g1 1590 gene accumulated at a similar level in *sod7-1D da1-1* and *da1-1,* suggesting that At3g1 1590 is not the *SOD7* gene (Figure 3B). By contrast, expression level of the At3g1 1580 gene in *sod7-1D da1-1* plants was dramatically higher than that in *da1-1* plants, suggesting that At3g1 1580 is the *SOD7* gene (Figure 3B). To further confirm whether the *sod7-1D* phenotypes were caused by ectopic At3g1 1580 expression, we overexpressed the At3g1 1580 gene (35S:GFP-SOD7) in wild-type plants (Col-0) and isolated 37 transgenic plants. Most transgenic lines showed small seeds and organs (Figures 3D-F), similar to those observed in the *sod7-1D* single mutant, indicating that At3g1 1580 is the SOD7 gene. The *SOD7* gene encodes a NGATHA like protein (NGAL2) containing a B3 DNA-binding domain and a transcriptional repression motif (Figure 3C) (Alvarez et al., 2009; Ikeda and Ohme-Takagi, 2009; Trigueros et al., 2009). SOD7 belongs to the RAV gene family that consists of 13 members in Arabidopsis (Figure 10) (Swaminathan et al., 2008). Several members of the RAV family contain the putative transcriptional repression motifs, including NGA1, NGA2, NGA3, NGA4, NGAL1 , NGAL2/SOD7 and NGAL3 (Figure 10) (Ikeda and Ohme-Takagi, 2009). The transcriptional repression motifs in NGA1, NGAL1 and NGAL2/SOD7 have been known to possess the repressive activity (Ikeda and Ohme-Takagi, 2009), indicating that they are transcriptional repressors. *SOD7* exhibits the highest similarity to Arabidopsis NGAL3/DEVELOPMENT-RELATED PcG TARGET IN THE APEX 4 (DPA4) (Figure 10), which has known roles in the regulation of leaf serrations (Engelhorn et al., 2012), but no previously identified function in seed size control.

Expression pattern and subcellular localization of SOD7

To monitor *SOD7* expression pattern during development, the *pSOD7:GUS* and *pSOD7:SOD7-GFP* vectors were constructed and transformed to wild-type plants, respectively. The tissue-specific expression patterns of *SOD7* were examined using a histochemical assay for GUS activity. In seedlings, relatively higher GUS activity was detected in younger leaves than in older leaves (Figures 4A -C). In flowers, GUS activity was observed in sepals, petals, stamens and carpels (Figures 4D-K). GUS activity was stronger in younger floral organs than in older ones (Figures 4D-K). Expression of *SOD7* was also detected in ovules (Figure. 4L). Thus, these analyses indicate that *SOD7* is a temporally and spatially expressed gene. As SOD7 encodes a B3 domain transcriptional repressor, we speculated that SOD7 is localized in the nucleus. To determine subcellular localization of SOD7, we observed GFP inflorescence in *pSOD7:SOD7-GFP* transgenic plants. As shown in Figures 4M-0,

GFP signal was only detected in nuclei. We also expressed a GFP-SOD7 fusion protein under the control of the 35S promoter in wild-type plants. Transgenic lines overexpressing *GFP-SOD7* formed smaller seeds than the wild type (Figure 3D), indicating that the *GFP-SOD7* fusion protein is functional. As shown in Figures 4P-R, GFP fluorescence in *35S:GFP-SOD7* transgenic plants was exclusively observed in nuclei. Thus, these results show that SOD7 is a nuclear-localized protein.

SOD7/NGAL2 acts redundantly with NGAL3 to control seed size

In order to further investigate the function of SOD7 in seed size control, we isolated T-DNA inserted loss-of-function mutants for *SOD7* and *NGAL3,* the most closely related family member. *sod7-ko1* (SM_3_34191) was identified with T-DNA insertion in the first exon of the SOD7 gene (Figure 5A). *ngal3-ko1* (SM_3_36641) had T-DNA insertion in the first exon of the NGAL3 gene (Figure 5B). The T-DNA insertion sites were confirmed by PCR using T-DNA specific and flanking primers and sequencing PCR products. *sod7-ko1* and *ngal3-ko1* mutants had no detectable full-length transcripts of *SOD7* and *NGAL3,* respectively. Seeds from *sod7-ko1* and *ngal3-ko1* mutants were slightly larger and heavier than seeds from wild-type plants (Figures 5C, G and H). The cotyledon area of *sod7-ko1* and *ngal3-ko1* mutants was increased, compared with that of the wild type (Figure 5I). Considering that SOD7 shares the highest similarity with NGAL3, we speculated that SOD7 may act redundantly with NGAL3 to influence seed size. To test this, we generated the *sod7-ko1 ngal3-ko1* double mutant. As shown in Figures 5C, D, G and H, the seed size and weight phenotypes of *sod7-ko1* mutant were synergistically enhanced by the disruption of NGAL3, indicating that SOD7 functions redundantly with NGAL3 to control seed size. We further measured the cotyledon area of 10-d-old seedlings. A synergistic enhancement of cotyledon size of *sod7-ko1* by the *ngal3-ko1* mutation was also observed (Figure 5I). In addition, the *sod7-ko1 ngal3-ko1* double mutant formed larger leaves and flowers than their parental lines (Figures 5E and F; 11). Thus, these results indicate that SOD7 and NGAL3 act redundantly to control seed and organ growth.

SOD7 acts maternally to control seed size

As the size of a seed is determined by the zygotic and/or maternal tissues (Garcia et al., 2005; Xia et al., 2013; Du et al., 2014), we asked whether SOD7 functions maternally or zygotically. We therefore performed reciprocal cross experiments between the wild type and *sod7-ko1 ngal3-ko1.* The effect of *sod7-ko1 ngal3-ko1* on seed size was observed only when *sod7-ko1 ngal3-ko1* was used as maternal plants

(Figure 6A). The size of seeds from *sod7-ko1 ngal3-ko1* plants pollinated with wild-type pollen was similar to that from the self-pollinated *sod7-ko1 ngal3-ko1* plants (Figure 6A). By contrast, the size of seeds from wild-type plants pollinated with *sod7-ko1 ngal3-ko1* mutant pollen was similar to that from the self-pollinated wild-type plants (Figure 6A). These results indicate that *sod7-ko1 ngal3-ko1* acts maternally to influence seed size. We further investigated the size of Col-O/Col-0 F2, *Co\-0/sod7-ko1 ngal3-ko1* F2, *sod7-ko1 ngal3-ko* 7/Col-O F2 and *sod7-ko1 ngal3-ko1/ sod7-ko1 ngal3-ko1* F2 seeds. As shown in Figure 6B, *sod7-ko1 ngal3-ko1/ sod7-ko1 ngal3-ko1* F2 seeds were larger than wild-type seeds, while the size of *Co\-0/sod7-ko1 ngal3-ko1* F2 and *sod7-ko1 ngal3-ko* 7/Col-O F2 seeds was similar to that of wild-type seeds. Thus, these results indicate that the embryo and endosperm genotypes for SOD7 do not determine seed size, and SOD7 is required in the sporophytic tissue of the mother plant to control seed growth.

SOD7 regulates cell proliferation in the maternal integuments

The reciprocal crosses showed that SOD7 functions maternally to influence seed size. The integuments surrounding the ovule are maternal tissues, which could set the growth potential of the seed coat after fertilization. Consistent with this idea, several studies showed that the integument size influences the final size of seeds in Arabidopsis (Garcia et al., 2005; Schruff et al., 2006; Adamski et al., 2009; Xia et al., 2013; Du et al., 2014). We therefore asked whether SOD7 acts through the maternal integuments to determine seed size. To test this, we characterized mature ovules of the wild type and *sod7-ko1 ngal3-ko1.* As shown in Figures 6C and D, the *sod7-ko1 ngal3-ko1* ovules were obviously larger than wild-type ovules. The outer integument length of *sod7-ko1 ngal3-ko1* ovules was significantly increased, compared with that of wild-type ovules (Figure 6E). As the size of the integument is determined by cell proliferation and cell expansion, we examined the number and size of outer integument cells in wild-type and *sod7-ko1 ngal3-ko1* ovules. As shown in Figure 6F, the number of outer integument cells in *sod7-ko1 ngal3-ko1* ovules was increased, compared with that in wild-type ovules. By contrast, the length of outer integument cells in *sod7-ko1 ngal3-ko1* ovules was similar to that in wild-type ovules (Figure 6G). These results showed that SOD7 is required for cell proliferation in the maternal integuments of ovules. After fertilization, cells in the integument mainly undergo expansion but still have division. We further examined the number and size of outer integument cells in wild-type and *sod7-ko1 ngal3-ko1* seeds at 6 and 8 day after pollination (DAP). In wild-type seeds, the number of outer integument cells at 6 DAP was comparable with that at

8 DAP (Figure 6F), indicating that cells in the outer integuments of wild-type seeds completely stop dividing by 6 DAP. Similarly, cells in the outer integuments of *sod7-ko1 ngal3-ko1* seeds also cease division by 6 DAP. The number of outer integument cells in *sod7-ko1 ngal3-ko1* seeds was significantly increased, compared with that in wild-type seeds (Figure 6F). By contrast, the length of outer integument cells in *sod7-ko1 ngal3-ko1* seeds was not increased in comparison to that in wild-type seeds (Figure 6G). Thus, these analyses indicate that *SOD7* is required for cell proliferation in the maternal integuments of ovules and developing seeds.

SOD7 acts in a common pathway with *KLU* to control seed size, but does so independently of *DM*

The Arabidopsis *klu* mutants formed small seeds due to the decreased cell proliferation in the integuments, while plants overexpressing *KLU/CYP78A5* produced large seeds as a result of the increased cell proliferation in the integuments (Adamski et al., 2009), suggesting that *SOD7* and *KLU* could function antagonistically in a common pathway to control seed growth. To test for genetic interactions between SOD7 and KLU, we generated the klu-4 *sod7-ko1 ngal3-ko1* triple mutant and measured the size of seeds from wild-type, klu-4, *sod7-ko1 ngal3-ko1* and klu-4 *sod7-ko1 ngal3-ko1* plants. As shown in Figures 7A and B, the average size and weight of klu-4 *sod7-ko1 ngal3-ko1* seeds were similar to those of the klu-4 single mutant, indicating that klu-4 is epistatic to *sod7-ko1 ngal3-ko1* with respect to seed size and weight. We further investigated the mature ovules from wild-type, klu-4, *sod7-ko1 ngal3-ko1* and klu-4 *sod7-ko1 ngal3-ko1* plants. The outer integument length of klu-4 *sod7-ko1 ngal3-ko1* ovules was comparable with that of klu-4 ovules (Figure 7C). Similarly, the outer integument length of klu-4 *sod7-ko1 ngal3-ko1* seeds was indistinguishable from that of klu-4 seeds at 8 DAP (Figure 7C). In addition, the size of klu-4 *sod7-ko1 ngal3-ko1* petals was similar to that of klu-4 petals).

Thus, these genetic analyses show that klu-4 is epistatic to *sod7-ko1 ngal3-ko1* with respect to seed and organ size, indicating that SOD7 and KLU act antagonistically in a common pathway to control seed and organ growth. To further understand the cellular basis of epistatic interactions between *SOD7* and *KLU,* we investigated the outer integument cell number of ovules and developing seeds from wild-type, klu-4, *sod7-ko1 ngal3-ko1* and klu-4 *sod7-ko1 ngal3-ko1* plants. The number of outer integument cells in klu-4 *sod7-ko1 ngal3-ko1* ovules was similar to that in klu-4 ovules (Figure 7D). Similarly, the number of outer integument cells in klu-4 *sod7-ko1 ngal3-ko1* seeds was

comparable with that in klu-4 seeds (Figure 7D). These results indicate that klu-4 is epistatic to *sod7-ko1 ngal3-ko1* with respect to the number of outer integument cells. We also observed that cells in the outer integuments of klu-4 and klu-4 *sod7-ko1 ngal3-ko1* seeds were slightly longer than those in wild-type seeds, suggesting a possible compensation mechanism between cell proliferation and cell expansion. Together, these findings show that SOD7 functions antagonistically in a common pathway with KLU to control cell proliferation in the maternal integuments.

Considering that *sod7-1D* was identified as a suppressor of *da1-1* in seed size, we further asked whether SOD7 and DA1 could act in the same genetic pathway. To test this, we measured the size of wild-type, *da1-1, sod7-1D* and *sod7-1D da1-1* seeds. The genetic interaction between *sod7-1D* and *da1-1* was essentially additive for seed size, compared with that of *sod7-1D* and *da1-1* single mutants, indicating that *SOD7* might function independently of *DM* to control seed size. We further crossed *sod7-ko1 ngal3-ko1* with *da1-1* and generated the *sod7-ko1 ngal3-ko1 da1-1* triple mutant and measured its seed size. The genetic interaction between *sod7-ko1 ngal3-ko1* and *da1-1* was also additive for seed size, compared with their parental lines, further supporting that *SOD7* functions to control seed growth separately from *DM.*

SOD7 directly binds to the promoter of *KLU* and represses the expression of *KLU*

Considering that *SOD7* acts antagonistically in a common pathway with *KLU* to control seed size, we asked whether the transcription repressor *SOD7* could repress the expression of *KLU.* We therefore investigated the expression of *KLU* in the chemically-inducible SOD7 *{pER8-SOD7)* transgenic plants. After the *pER8-SOD7* transgenic plants were treated with the inducer (β -estradiol), the expression of *SOD7* was strongly induced at 4 and 8 hours (Figure 8A). As expected, the expression of *KLU* was dramatically repressed at 4 and 8 hours (Figure 8A). Thus, these results indicate that SOD7 represses the expression of *KLU* and also suggest that *KLU* might be a direct target of SOD7.

To determine whether SOD7 can directly bind to the promoter of the *KLU* gene, we performed a chromatin immunoprecipitation (ChIP) assay with *35S:GFP* and *35.GFP-SOD7* transgenic plants. It has been reported that the CACCTG sequence is recognized by the B3 domain of RAV1 , one member of the RAV family (Kagaya et al., 1999; Yamasaki et al., 2004). We therefore analyzed the promoter sequence of KLU and did not find an intact CACCTG sequence within 2 kb promoter region of KLU.

However, we found a similar sequence (CACTTG) in the promoter region of *KLU* (Figure 8B), which could be the potential SOD7-binding site. To test this, we examined the enrichment of a *KLU* promoter fragment (PF1) containing the CACTTG sequence by ChIP analyses and found that the fragment PF1 was strongly enriched in the chromatin-immunoprecipitated DNA with anti-GFP antibody (Figures 8B and C). By contrast, we did not detect significant enrichment of an ACTIN7 promoter sequence and the KLU promoter fragment PF2, which do not contain the CACTTG sequence (Figures 8B and C). This result shows that SOD7 associates with the promoter of *KLU* in vivo. We further expressed SOD7 as a MBP fusion protein (MBP-SOD7) and performed the DNA electrophoretic mobility shift assays (EMSA). As shown in Figures 8B and D, MBP-SOD7 was able to bind to the biotin-labeled probe A containing the CACTTG sequence, and the binding was reduced by the addition of an unlabeled probe A. By contrast, MBP-SOD7 failed to bind to a probe A-m with mutations in the CACTTG sequence (Figures 8B and D). Taken together, these results show that SOD7 directly binds to the promoter of *KLU* and represses *KLU* expression.

Discussion

Seed size is crucial for plant fitness and agricultural purposes, but little is known about the genetic and molecular mechanisms that set the final size of seeds in plants. In this study, we show that SOD7 acts maternally to control seed size by restricting cell proliferation in the integuments of ovules and developing seeds. *SOD7* encodes a B3 domain transcriptional repressor NGAL2 and acts redundantly with its closest homolog NGAL3 to control seed size. Genetic analyses indicate that SOD7 functions in a common pathway with the maternal factor *KLU* to control seed growth, but does so independently of DA1. Further results reveal that SOD7 directly binds to the promoter region of *KLU* and represses *KLU* expression. Thus, our findings identify SOD7 as a negative factor for seed size and define the genetic and molecular mechanisms of *SOD7* and KLU in seed size control.

SOD7 acts maternally to regulate seed size

The *sod7-1D* gain-of-function mutant was identified as a suppressor of the large seed phenotype of *da1-1.* However, genetic analyses showed that *SOD7* functions independently of DA1 to control seed growth. The *sod7-1D* single mutant produced small seeds and organs (Figure 2), while the simultaneous disruption of *SOD7* and the closely related family member *NGAL3* resulted in large seeds and organs (Figure 5), indicating that *SOD7* is a negative regulator of seed and organ size. Several previous

studies suggest that there is a possible link between seed size and organ growth. For instance, *ar†2, da1-1, da2A* and *eod3-1D* mutants produced large seeds and organs (Schruff et al., 2006; Li et al., 2008; Fang et al., 2012; Xia et al., 2013), whereas klu and *sod2/ubp15* mutants formed small seeds and organs (Anastasiou et al., 2007; Adamski et al., 2009; Du et al., 2014). However, seed size is not invariably associated with organ size. For example, *eod8/med25* mutants with large organs formed normal-sized seeds (Xu and Li, 2011), while ap2 mutants with normal-sized organs produced large seeds (Jofuku et al., 2005; Ohto et al., 2005). Thus, these findings suggest that seeds and organs not only share common mechanisms but also possess distinct pathways to control their respective size.

Reciprocal cross experiments showed that *SOD7* acts maternally to restrict seed growth, and the endosperm and embryo genotypes for *SOD7* do not determine seed size (Figure 6). The integuments surrounding the ovule are maternal tissues and form the seed coat after fertilization. Arabidopsis *ar†2, ap2, da1-1, da2-1* and *eod3-1D* mutants with large integuments formed large seeds (Jofuku et al., 2005; Ohto et al., 2005; Schruff et al., 2006; Li et al., 2008; Fang et al., 2012; Xia et al., 2013), while *klu-4* and *ubp15/sod2* mutants with small integuments produced small seeds (Adamski et al., 2009; Du et al., 2014), indicating that the maternal integuments are crucial for determining seed size in Arabidopsis. Consistent with this notion, mature *eod7-ko1 ngal3-ko1* ovules were larger than wild-type ovules (Figures 6C and D). The outer integument length of *eod7-ko1 ngal3-ko1* ovules and developing seeds was significantly increased, compared with that of wild-type ovules and seeds (Figures 6E and 7C). Considering that the maternal integument or seed coat not only acts as a protective structure but also restricts seed growth, the regulation of maternal integument size is one of important mechanisms for seed size control. The size of the integument is determined by cell proliferation and cell expansion; these two processes are assumed to be coordinated. The number of outer integument cells in *sod7-ko1 ngal3-ko1* ovules and seeds was significantly increased, compared with that in wild-type ovules and seeds (Figure 6F), indicating that *SOD7* controls seed growth by limiting cell proliferation in the maternal integuments. Similarly, several mutants with the increased number of cells in the maternal integuments produced large seeds in Arabidopsis (Schruff et al., 2006; Li et al., 2008; Xia et al., 2013). By contrast, several other mutants with the decreased number of cells in the maternal integuments formed small seeds in Arabidopsis (Adamski et al., 2009; Du et al., 2014). Considering that cells in the integuments mainly undergo expansion after fertilization (Garcia et al.,

2005), it is possible that the number of cells in the integuments determines the growth potential of the seed coat after fertilization.

The genetic and molecular mechanisms of *SOD7* and *KLU* in seed size control

The *sod7-1D* mutant had small seeds and organs (Figure 2), as had been seen in *klu* mutants (Anastasiou et al., 2007; Adamski et al., 2009). KLU encodes a cytochrome P450 CYP78A5 that has been proposed to generate mobile plant-growth substances (Anastasiou et al., 2007; Adamski et al., 2009). KLU regulates seed size by promoting cell proliferation in the maternal integuments of ovules (Anastasiou et al., 2007; Adamski et al., 2009). By contrast, *SOD7* acts maternally to control seed size by limiting cell proliferation in the integuments of ovules and developing seeds (Figure 6). These results suggest that SOD7could function antagonistically in a common pathway with *KLU* to control seed size. In our growth conditions, klu-4 formed slightly smaller seeds than the wild type due to the decreased cell number and the slightly increased cell length in the integuments of developing seeds (Figures 7A and D), suggesting a possible compensation mechanism between cell proliferation and cell expansion in *klu-4* integuments. Importantly, our genetic analyses showed that *klu-4* is epistatic to *sod7-ko1 ngal3-ko1* with respect to seed and organ size (Figures 7A and B). *klu-4\s* also epistatic to *sod7-ko1 ngal3-ko1* for the outer integument length (Figure 7C). Further results revealed that the number of cells in the outer integuments of *klu-4 sod7-ko1 ngal3-ko1* ovules and developing seeds was similar to that of *klu-4ovu\es* and developing seeds (Figure 7D). Thus, these genetic results demonstrate that *SOD7* act in a common pathway with KLU to control seed size by regulating cell proliferation in the maternal integuments.

SOD7 encodes a B3 domain transcriptional repressor NGAL2 that is localized in nuclei of Arabidopsis cells (Figures 4M-R). Thus, it is possible that *SOD7* could directly bind to the promoter of *KLU* and repress *KLU* expression. Supporting this idea, the inducible expression of *SOD7* resulted in a strong reduction of *KLU* expression (Figure 8A). Our ChIP-qPCR data showed that SOD7 associates with the promoter region of *KLU* in vivo (Figures 8B and C). EMSA experiments revealed that SOD7 directly binds to the CACTTG sequence in the promoter of the KLU gene (Figures 8B and D). Thus, these results illustrate that SOD7 directly targets the promoter region of KLU and represses the expression of *KLU,* thereby determining seed size. Taken together, these findings

reveal the genetic and molecular mechanisms of SOD7 and KLU in regulating Arabidopsis seed size.

For many plants, the seeds are the main product to be harvested, and an increase in seed size would be beneficial for growers. In this study, we identify *SOD7* as a negative regulator of seed size, and demonstrate that *SOD7* acts in a common genetic pathway with *KLU* to control seed size. Our current knowledge of *SOD7* functions suggests that the *SOD7* gene (and its homologs in other plant species) could be used to engineer large seed size in crops. Considering that crop plants have undergone selection for large seed size during domestication (Fan et al., 2006; Song et al., 2007; Gegas et al., 2010), it will be a worthwhile challenge to know whether beneficial alleles of the *SOD7* gene have already been utilized by plant breeders.

Knockout experiments in rice using genome editing

Genome editing experiments to knock out os1 1g01560000 and/or Os12g0157000 in rice are being carried out using the crisper-cas9 system. Four vectors, each with two recognition (CRISPR target) sites, have been constructed, to achieve these knock outs, as described in Fig. 14. In summary, the vectors were obtained as follows:

1. The target sites were identified. The target site should be (or approximately so) 20 nucleotides before a NGG sequence, N being for any nucleotide. The target sequence was then evaluated using the website: http://cbi.hzau.edu.cn/crispr/help.php (incorporated herein by reference). Of note, the target site should be unique in the genome.
2. Using overlap PCR, the target sequence is linked with the U6 sequence, as shown in Figure 14. U6 is for transcriptional activity.
3. Using infusion technology we connected the U6-guide-gRNA scaffold fragment to the vector pMDC99-cas9 to obtain the pMDC99-cas9- U6-guide-gRNA scaffold constructs. These constructs were named zyy1,zyy2, zyy3 ,zyy4. The full sequences of these constructs are represented in SEQ ID NO: 155, 156, 157 and 158 respectively. Each construct contains two recognition sites, which are highlighted in the sequence information, and are represented separately as SEQ ID Nos 159, 160, 161 , 162 and 163.
4. We then transformed these constructs into Agrobacteria and used an Agrobacteria mediated method to transform rice and obtain gene-edited rice.

Transformation of plants is a routine technique that is well known to the skilled person. Nonetheless, a brief outline of transformation techniques is provided above.

Knock out lines are being analysed to assess the phenotype.

Table 1. Primers used in this study

Primer Name	Primer Sequences
Primers for T-DNA identification	
SM_3_34191-LP	ACCATGACATTCGAGGTTCAC (SEQ ID NO. 8)
SM_3_34191-RP	ATCACCACCAAAACGACGTAG (SEQ ID NO. 9)
SM_3_36641-RP	TACGTCATGCTTCAAATCGTG (SEQ ID NO. 10)
SM_3_36641-RP	AGGACACGAACAATTCATTCG (SEQ ID NO. 11)
Spm32	TACGAATAAGAGCGTCCATTTTAGAGTGA (SEQ ID NO. 12)
SM_3_39145- LP	ACCCAAAGAACAGCAATCATG (SEQ ID NO. 13)
SM_3_39145- RP	AAAACACTCCGCCATTAAACC (SEQ ID NO. 14)
Primers for TAIL-PCR	
OJF22	CGAGTATCAATGGAAACTTAACCG (SEQ ID NO. 15)
OJF23	AACGGAGAGTGGCTTGAGAT (SEQ ID NO. 16)
OJF24	TGGCCCTTATGGTTTCTGCA (SEQ ID NO. 17)
AD1	NTCGA(G/C)T(A/T)T(G/C)G(A/T)GTT (SEQ ID NO. 18)
Primers for Constructs	
SOD7CDS-F	ATGTCAGTCAACCATTACCAC (SEQ ID NO. 19)
SOD7CDS-R	CAGGTAGGAGATGGACGAGGTTGA (SEQ ID NO. 20)
SOD7G-F	TGAGAGGAACCATTTCTTAGAGG (SEQ ID NO. 21)
SOD7G-R	ACCTCGTCCATCTCCTACCTGC (SEQ ID NO. 22)
SOD7P-F	AAACACGTCAAATATAACGAAT (SEQ ID NO. 23)
SOD7P-R	CTTTTTTTTGGTTTCTTGGAGTGAGAGAGAGAG (SEQ ID NO. 24)
SOD7-ER-F	AGTCTGGGCCCATGTCAGTCAACCATTAC (SEQ ID NO. 25)
SOD7-ER-R	GCGACTAGTTTATAAAAGAGTTAAAATTA (SEQ ID NO. 25)
MBP-SOD7-FP	CGGGATCCTCAGTCAACCATTACC (SEQ ID NO. 27)
MBP-SOD7-RP	ACTAGTCGACTCAACCTCGTCCATCTCC (SEQ ID NO. 28)
Primers for RT-PCR and qRT-PCR	
ACTIN2-F	GAAATCACAGCACTTGCACC (SEQ ID NO. 29)
ACTIN2-R	AAGCCTTTGATCTTGAGAGC (SEQ ID NO. 30)
SOD7-EX-F	GCGACGACGGAGAAAGGG (SEQ ID NO. 31)
SOD7-EX-R	ACGACGGCGCCATAGTGT (SEQ ID NO. 32)

NGAL3-EX-F TTTGAAGACGAGTCAGGCAAGT (SEQ ID NO. 33)

NGAL3-EX-R TACGGCGGCTCCATAGTGGG (SEQ ID NO. 34)

SOD7-q-FP GTATTGGAGCGGCTTGACTACACC (SEQ ID NO. 35)

SOD7-q-RP GACGGCATCACCATGACATTCG (SEQ ID NO. 36)

KLU-q-FP TGATTCTGACATGATTGCTGTTCT (SEQ ID NO. 37)

KLU-q-RP TCGCAACTGTATCTGTCCCTCTA (SEQ ID NO. 38)

Primers for ChIP assay

ACTIN7-ChIP-FP CGTTTCGCTTTCCTTAGTGTTAGCT (SEQ ID NO. 29)

ACTI N7-Chl P-RP AGCGAACGGATCTAGAGACTCACCTTG (SEQ ID NO. 40)

PF1-F CAGGCCTAAGCCTAACAGTAGAC (SEQ ID NO. 41)

PF1-R TGTACTAGGATTTATTTACGTAG (SEQ ID NO. 42)

PF2-F TATTGTTCATAGAAACCCTGCAAA (SEQ ID NO. 43)

PF2-R AGTCAATGGTTTAATGGCGGAGTG (SEQ ID NO. 44)

Probes for EMSA

A-Biotin-FP TTCTACTACACTTGCTCTCTGTA (SEQ ID NO. 45)

A-Biotin-RP TACAGAGAGCAAGTGTAGTAGAA (SEQ ID NO. 46)

A-Biotin-m-FP TTCTACTAACACCTCTCTCTGTA (SEQ ID NO. 47)

A-Biotin-m-RP TACAGAGAGAGGTGTTAGTAGAA (SEQ ID NO. 48)

References

Adamski, N.M., Anastasiou, E., Eriksson, S., O'Neill, CM., and Lenhard, M. (2009). mLocal maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. Proceedings of the National Academy of Sciences of the United States of America 106, 201 15-20120.

Alvarez, J.P., Goldshmidt, A., Efroni, I., Bowman, J.L, and Eshed, Y. (2009). Th NGATHA distal organ development genes are essential for style specification in Arabidopsis. Plant Cell 21, 1373-1393.

Anastasiou, E., Kenz, S., Gerstung, M., MacLean, D., Timmer, J., Fleck, C , and Lenhard, M. (2007). Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. Developmental cell 13, 843-856.

Cheng, Z.J., Zhao, X.Y., Shao, X.X., Wang, F., Zhou, C, Liu, Y.G., Zhang, Y., and Zhang, X.S. (2014). Abscisic Acid Regulates Early Seed Development in Arabidopsis by ABI5-Mediated Transcription of SHORT HYPOCOTYL UNDER BLUE1. Plant Cell 26, 1053-1068.

Du, L , Li, N., Chen, L , Xu, Y., Li, Y., Zhang, Y., and Li, C. (2014). The Ubiquitin Receptor DA1 Regulates Seed and Organ Size by Modulating the Stability of the Ubiquitin-Specific Protease UBP15/SOD2 in Arabidopsis. Plant Cell 26, 665-677.

Engelhorn, J., Reimer, J.J., Leuz, I., Gobel, U., Huettel, B., Farrona, S., and Turck, F. (2012). Development-related PcG target in the apex 4 controls leaf margin architecture in Arabidopsis thaliana. Development 139, 2566-2575.

Fan, C, Xing, Y., Mao, H., Lu, T., Han, B., Xu, C, Li, X., and Zhang, Q. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112, 1164-1171.

Fan, J., Hill, L., Crooks, C , Doerner, P., and Lamb, C. (2009). Abscisic acid has a key role in modulating diverse plant-pathogen interactions. Plant physiology 150, 1750-1761 .

Fang, W., Wang, Z., Cui, R., Li, J., and Li, Y. (2012). Maternal control of seed size by EOD3/CYP78A6 in Arabidopsis thaliana. Plant J 70, 929-939.

Garcia, D., Fitz Gerald, J.N., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. Plant Cell 17, 52-60.

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jurgens, G., and Berger, F. (2003). Arabidopsis haiku mutants reveal new controls of seed size by endosperm. Plant physiology 131, 1661-1670.

Gegas, V.C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., Sayers, L., Doonan, J.H., and Snape, J.W. (2010). A genetic framework for grain size and shape variation in wheat. Plant Cell 22, 1046-1056.

Gendrel, A.V., Lippman, Z., Martienssen, R., and Colot, V. (2005). Profiling histone modification patterns in plants using genomic tiling microarrays. Nat Methods 2, 213-218.

Harper, J.L., Lovell, P.H., and Moore, K.G. (1970). The Shapes and Sizes of Seeds. Annual Review of Ecology and Systematics 1, 327-356

Ikeda, M., and Ohme-Takagi, M. (2009). A novel group of transcriptional repressors in Arabidopsis. Plant & cell physiology 50, 970-975.

Jofuku, K.D., Omidyar, P.K., Gee, Z., and Okamuro, J.K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proceedings of the National Academy of Sciences of the United States of America 102, 3117-3122.

Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucleic Acids Res 27, 470-478.

Kang, X., Li, W., Zhou, Y., and Ni, M. (2013). A WRKY transcription factor recruits the SYG1-like protein SHB1 to activate gene expression and seed cavity enlargement. PLoS Genet 9, e1003347.

Li, J., Nie, X., Tan, J.L., and Berger, F. (2013). Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 110, 15479-15484.

Li, Y., Zheng, L , Corke, F., Smith, C , and Bevan, M.W. (2008). Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. Genes Dev 22, 1331-1336.

Liu, Y.G., Mitsukawa, N., Oosumi, T., and Whittier, R.F. (1995). Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J 8, 457-463.

Lopes, M.A., and Larkins, B.A. (1993). Endosperm origin, development, and function. Plant Cell 5, 1383-1399.

Luo, M., Dennis, E.S., Berger, F., Peacock, W.J., and Chaudhury, A. (2005). MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 102, 17531-17536.

Moles, AT., Ackerly, D.D., Webb, CO., Tweddle, J.C., Dickie, J.B., and Westoby, M. (2005). A brief history of seed size. Science 307, 576-580.

Ohto, M.A., Fischer, R.L., Goldberg, R.B., Nakamura, K., and Harada, J.J. (2005). Control of seed mass by APETALA2. Proceedings of the National Academy of Sciences of the United States of America 102, 3123-3128.

Ohto, M.A., Floyd, S.K., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (2009). Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. Sex Plant Reprod 22, 277-289.

Orsi, C.H., and Tanksley, S.D. (2009). Natural variation in an ABC transporter gene associated with seed size evolution in tomato species. PLoS Genet 5, e1000347.

Schruff, M.C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R.J. (2006). The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. Development 133, 251-261 . Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. Development 125, 3329-3341 .

Smaczniak, C , Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F.,

Xu, L , Carles, C.C., Angenent, G.C., and Kaufmann, K. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proceedings of the National Academy of Sciences of the United States of America 109, 1560-1565.

Song, X.J., Huang, W., Shi, M., Zhu, M.Z., and Lin, H.X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39, 623-630.

Swaminathan, K., Peterson, K., and Jack, T. (2008). The plant B3 superfamily. Trends Plant Sci 13, 647-655.

Trigueros, M., Navarrete-Gomez, M., Sato, S., Christensen, S.K., Pelaz, S., Weigel, D., Yanofsky, M.F., and Ferrandiz, C. (2009). The NGATHA genes direct style development in the Arabidopsis gynoecium. Plant Cell 21, 1394-1409.

Wang, A., Garcia, D., Zhang, H., Feng, K., Chaudhury, A., Berger, F., Peacock, W.J., Dennis, E.S., and Luo, M. (2010). The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. Plant J 64, 670-679.

Westoby, M., Falster, D.S., Moles, AT., Vesk, P.A., and Wright, I.J. (2002). PLANT ECOLOGICAL STRATEGIES: Some Leading Dimensions of Variation Between Species. Annual Review of Ecology and Systematics 33, 125-159.

Xia, T., Li, N., Dumenil, J., Li, J., Kamenski, A., Bevan, M.W., Gao, F., and Li, Y. (2013). The Ubiquitin Receptor DA1 Interacts with the E3 Ubiquitin Ligase DA2 to Regulate Seed and Organ Size in Arabidopsis. Plant Cell 25, 3347-3359.

Xiao, W., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). Regulation of seed size by hypomethylation of maternal and paternal genomes. Plant physiology 142, 1160-1 168.

Xu, R., and Li, Y. (201 1). Control of final organ size by Mediator complex subunit 25 in Arabidopsis thaliana. Development 138, 4545-4554.

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Tomo, Y., Hayami, N., Terada, T., Shirouzu, M., Osanai, T., Tanaka, A., Seki, M., Shinozaki, K., and Yokoyama, S. (2004). Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. Plant Cell 16, 3448-3459.

Zhou, Y., Zhang, X., Kang, X., Zhao, X., and Ni, M. (2009). SHORT HYPOCOTYL UNDER BLUE1 associates with MINISEED3 and HAIKU2 promoters in vivo to regulate Arabidopsis seed development. Plant Cell 21, 106-1 17.

Sequence information

Identity of homologs to NGAL2 is indicated

AtSOD7 nucleic acid SEQ ID NO. 1 (cDNA) At3g1 1580

ATGTCAGTCAACCATTACCACAACACTCTCTCGTTGCATCATCACCACCAAAACGA
CGTAGCTATAGCACAACGAGAGTCTTTGTTCGAGAAATCACTCACACCAAGCGAC
GTCGGAAAGCTAAACCGCTTAGTCATACCAAAACAACACGCCGAGAAATACTTCC
CTCTCAATAATAATAATAATAATGGCGGCAGCGGAGATGACGTGGCGACGACGGA
GAAAGGGATGCTTCTTAGCTTCGAGGATGAGTCAGGCAAGTGTTGGAAATTCAGA
TACTCTTATTGGAACAGTAGCCAAAGCTACGTGTTGACCAAAGGATGGAGCAGGT
ACGTCAAAGACAAACACCTCGACGCAGGCGACGTTGTTTTCTTTCAACGTCACCG
TTTTGATCTCCATAGACTCTTCATTGGCTGGCGGAGACGCGGTGAAGCTTCTTCCT
CTCCCGCTGTCTCCGTTGTGTCTCAAGAAGCTCTAGTTAATACGACGGCGTATTG
GAGCGGCTTGACTACACCTTATCGTCAAGTACACGCGTCAACTACTTACCCTAATA
TTCACCAAGAGTATTCACACTATGGCGCCGTCGTTGATCATGCTCAGTCGATACCA
CCGGTGGTCGCAGGTAGCTCGAGGACGGTGAGGCTTTTTGGCGTGAACCTCGAA
TGTCATGGTGATGCCGTCGAGCCACCACCGCGTCCTGATGTCTATAATGACCAAC
ACATTTACTATTACTCAACTCCTCATCCCATGAATATATCATTTGCTGGGGAAGCAT
TGGAGCAGGTAGGAGATGGACGAGGTTGA

AtSOD7 nucleic acid SEQ ID NO. 2 (genomic DNA).

ttgtttcggctatttgttatactattgttataacagtcacaagacttgacctcaacgaaaacttttacaaaacgtgaattggaaa
tttttacaaaatatgctcttaatcgttaatgcttcccaattaggtgagttaaattgtgagaggaaccatttcttagaggaaatggt
tcatgaaaacaaatatgaaatagtatcactagtcttagttttgcgagaaaattaggaaaaatagaaacgtgtaagcacca
atgatattcctgaaagcacgtgacagatatttcatgatcctataattaacaagtgataaagatattaaataaaattaacgata
cttgagaaattcgtcaaataaaatagaagaggaccactcacgtaaccatttgcacgtcccattgatttttgtggtagacttgg
tatgttatattacttatattcacagaattatatacgaaactcacgacttaagatgcacggtaataactacagatggaaatttac
ccatcaaacaagaaaacaacatttactcaagcatctagctagaccaaaatgtttgtttacttgttgacttgcgatccatagat
atattagttagaacttttttcttctacaattgatcaaatgtttcacactgttctcaatttctcatctagattcatgacttatatgtttggtc
aaatatcacagcttgatgagcattaaatagcgtcgaagtataggatggttacgttgttcaatattgtaaaggaaaaaaaga
gaaagagtgccaaaaggtcaagtcgatttcacaaataaatcttgaagtctttatccctctcgattataaaatgattaggaaa
agaaaaagagagaataaaatgtagataaagagaaagagaaagagagagaggaacataagggatggtatgaagta
gaagtgaagatgcatgcgatggtgtgtcggaaaggcaaagcacatgctacacaacttgagcttctcacttgcgtcaggg
ataagtatcctctgtaccttcttacttttgcgtaatatgtaccacctcacttctcaaccgtttgatctttaatccttcattatttcttcatt

accttctcttttgtttttgttttcgttttcaatttctcatagattcatttacaaactaaatatcataggaaggtgttatctctagttaatttc
ttatcctactttaacaaaatttaattgtcaaaagattatttttacgtttatagacaaaagatactgacacatcaattccacgaac
caaatggttgagaaaaacaaaacgactatctttgtcttgcaaataaattaatggcagttagtaagattctcagctgaaaatt
catacaagagtaaatgatcaaataaccatttatgagagaaatttaatccttcagaaaccaatgaggatctgatcaagtaat
tgcaaaccacatgagtccatgataaaggattgtttgacttacgcaatccacatatttatggctgcttgatatgtaaggtttatct
gctttgacagtctatagaatcttgctaatcaatacgtcatatccggtgaatactgaaacttttttaattaagaaaacacaaatc
atcttttctccggaggatttcgaatttagttccggcaatgctgaaataacatatgttgaacttataacattccaagacatcaaat
tttactaatatataaataattacatattcttcttctacatgatcaaaacctttcaactttaattaaagggttacgtcgcggcgttttg
tgtggcttactctttttttacactataactatagaacactcgtggatccaatgccgtttaggacaagattttatcagacgagaaa
aaaaaaaacaataccacatttttaaatatatatggattatggactgcaacaacaatatagaaaagaagagaaaaaaat
aaaaataatgattgaaaggaaatatcatcacgcaaaaccttaaaagtactatcggtatcgtgtcgtcctctcctcatcaaat
agttcccacagttttcacatcaatttaaccattttcaatttttttcactctctgtctctctcctttgtataatactatattagtaccattac
ccatctctctttcaccaccaaaccaacacctgcaaatcctctctctctctctcactccaagaaaccaaaaaaaaagATG
TCAGTCAACCATTACCACAACACTCTCTCGTTGCATCATCACCACCAAAACGACGT
AGCTATAGCACAACGAGAGTCTTTGTTCGAGAAATCACTCACACCAAGCGACGTC
GGAAAGCTAAACCGCTTAGTCATACCAAAACAACACGCCGAGAAATACTTCCCTCT
CAATAATAATAATAATAATGGCGGCAGCGGAGATGACGTGGCGACGACGGAGAAA
GGGATGCTTCTTAGCTTCGAGGATGAGTCAGGCAAGTGTTGGAAATTCAGATACT
CTTATTGGAACAGTAGCCAAAGCTACGTGTTGACCAAAGGATGGAGCAGGTACGT
CAAAGACAAACACCTCGACGCAGGCGACGTTGTTTTCTTTCAACGTCACCGTTTTG
ATCTCCATAGACTCTTCATTGGCTGGCGGAGACGCGGTGAAGCTTCTTCCTCTCC
CGCTGTCTCCGTTGTGTCTCAAGAAGCTCTAGTTAATACGACGGCGTATTGGAGC
GGCTTGACTACACCTTATCGTCAAGTACACGCGTCAACTACTTACCCTAATATTCA
CCAAGAGTATTCACACTATGgtaaattcaaacccttttatttcctcttttgttttttctttctctcttatctatatgtcagatt
tatactcctctctgttctcttttaagatttgtctttttcataaaaatagatgattcgtaatttgtattgcatatttacatgttctcttaaaa
aaagtaatagagattaatattttatgcatggtattttagattatctgcctactttatatggtagtaaacaagaacattcatctttatt
tggttttataaacaaaatatgagaatttttaaaggttagggcaagcacttggaaagctcaaccattttagttagctggtggaa
tatctttcttataaaaagcaaatgagttatctaaaactatatgacaattattttagttgcgtgtgtaatgtatataaaataacaac
atgaaataacattttgtcttttatttttgtcattcttattatttaattttggacccgacaatttcaaataatcttctccaagttgtaacta
atccgttacatgcgcgtgaggagaaccgtccaatccacttagactaacgtgccctttatttcttcctttaattctatgttaaaaa
aacaatttaactaaaagatgcgcacgtgtcttgacggtggaaaaaaattgtagGCGCCGTCGTTGATCATG
CTCAGTCGATACCACCGGTGGTCGCAGGTAGCTCGAGGACGGTGAGGCTTTTTG
GCGTGAACCTCGAATGTCATGGTGATGCCGTCGAGCCACCACCGCGTCCTGATG
TCTATAATGACCAACACATTTACTATTACTCAACTCCTCATCCCATGgtaaatatttttttttttt
acatttttgtcagattcaaatttttgcttacgtatgatataattattaaacagatgtcgtggctgtttctcgagacgagacagatg
aaaattagtaattttaaaatagacctgaaagagatttttatgtttaataaattatataaaggaggaatcagagagaataata

ctatacacttgactgtaaaaccacatggccaatttggtttttatttgattactttgatttgttttgtttactcttttgtctctgtagcctcct
tttgttcattaattaatatcagccgtaagtatatagtttcctgtgaaaacagtctctattttggttttactattctaatttgttaggcac
cgtcagtttttttttgtgaaaccaaattattgactaataagctggaaagcaaaactgactaaaagcattacaaacttatcaatg
acataagttttgaatttattaccatgttttgtaatgttcagatataatttgaaatgcttagaattatatatttgtatacttaaattaatg
aaataaagtgaatactaaagatagttttatttttcatattattctatacaattcggtgtacaatttgtttttgatgataataaaaata
ataaaattgcgtgttggaattgtgaaacagAATATATCATTTGCTGGGGAAGCATTGGAGCAGGT
AGGAGATGGACGAGGT

AtNGAL2 SEQ ID N0.3 (protein encoded by *AtSOD7)..*

MSVNHYHNTLSLHHHHQNDVAIAQRESLFEKSLTPSDVGKLNRLVIPKQHAEKYFPLN
NNNNNGGSGDDVATTEKGMLLSFEDESGKCWKFRYSYWNSSQSYVLTKGWSRYVK
DKHLDAGDWFFQRHRFDLHRLFIGWRRRGEASSSPAVSVVSQEALVNTTAYWSGL
TTPYRQVHASTTYPNIHQEYSHYGAVVDHAQSIPPWAGSSRTVRLFGVNLECHGDA
VEPPPRPDVYNDQHIYYYSTPHPMNISFAGEALEQVGDGRG

AtNGAL3 nucleic acid sequence SEQ ID NO. 4 (cDNA) at5g06250

ATGTCAGTCAACCATTACTCCACAGACCACCACCACACTCTCTTGTGGCAGCAAC
AGCAACACCGCCACACCACCGACACATCGGAGACAACCACCACCGCCACATGGC
TCCACGACGACCTAAAAGAGTCACTCTTCGAGAAGTCTCTCACACCAAGCGACGT
CGGGAAACTCAACCGCCTCGTCATACCAAAACAACACGCAGAGAAATACTTCCCT
CTCAATGCCGTCCTAGTCTCCTCTGCTGCTGCTGACACGTCATCTTCGGAGAAAG
GGATGCTTCTAAGCTTTGAAGACGAGTCAGGCAAGTCATGGAGGTTCAGATACTC
TTACTGGAACAGCAGTCAAAGCTATGTCTTGACTAAAGGATGGAGCAGATTTGTCA
AAGACAAACAGCTCGATCCAGGCGACGTTGTTTTCTTCCAACGACACCGTTCTGA
TTCTAGGAGACTCTTCATTGGCTGGCGCAGACGTGGACAAGGCTCCTCATCCTCC
GTCGCGGCCACTAACTCCGCCGTGAATACGAGTTCTATGGGAGCTCTTTCTTATC
ATCAAATCCACGCCACTAGTAATTACTCTAATCCTCCCTCTCACTCAGAGTATTCC
CACTATGGAGCCGCCGTAGCAACAGCGGCTGAGACTCACAGCACACCGTCGTCT
TCCGTCGTCGGGAGCTCAAGGACGGTGAGGCTTTTCGGTGTGAATCTGGAGTGT
CAAATGGATGAAAACGACGGAGATGATTCTGTTGCAGTTGCCACCACCGTTGAAT
CTCCCGACGGTTACTACGGCCAAAACATGTACTATTATTACTCTCATCCTCATAAC
ATGGTAATTTTAACTCTTTTATAA

AtNGAL3 amino acid SEQ ID NO.5

MSVNHYSTDHHHTLLWQQQQHRHTTDTSETTTTATWLHDDLKESLFEKSLTPSDVG
KLNRLVIPKQHAEKYFPLNAVLVSSAAADTSSSEKGMLLSFEDESGKSWRFRYSYWN
SSQSYVLTKGWSRFVKDKQLDPGDWFFQRHRSDSRRLFIGWRRRGQGSSSSVAAT
NSAVNTSSMGALSYHQIHATSNYSNPPSHSEYSHYGAAVATAAETHSTPSSSWGSS
RTVRLFGVNLECQMDENDGDDSVAVATTVESPDGYYGQNMYYYYSHPHNMVILTLL

Oryza sativa

Osl2g0157000 LOC_Osl2g06080.1
Cover 73% identity 53%

SEQ ID NO: 49
MAMHAGHAWWGVAMYTNHYHHHYRHKTSDVGKNRVKHARYGGGDSGKGSDSGKWRRYSYWTSSSYVTKG
WSRYVKKRDAGDVVHRVRGGAADRGCRRRGSAAAVRVTANGGWSMCYSTSGSSYDTSANSYAYHRSVDDHSD
HAGSRADAKSSSAASASRRRGVN DCGADATAMYGYMHHSYAAVSTVNYWSV

CDS SEQ ID NO: 50

ATGGCCATGCACCCTCTCGCCCAGGGGCACCCCCAGGCGTGGCCATGGGGTGTAGCCATG
TACACCAACCTGCACTACCACCACCACTACGAGAGGGAGCACCTGTTCGAGAAGCCGCTG
ACGCCGAGCGACGTCGGCAAGCTCAACAGGCTGGTGATCCCCAAGCAGCACGCCGAGAGG
TACTTCCCGCTCGGCGGCGGCGACTCCGGTGAGAAGGGCCTCCTCCTCTCCTTCGAGGAC
GAGTCCGGCAAGCCATGGCGGTTCCGCTACTCCTACTGGACCAGCAGCCAGAGCTACGTG
CTCACCAAGGGCTGGAGCCGCTACGTCAAGGAGAAGCGCCTCGACGCCGGCGACGTCGTC
CACTTCGAGCGCGTCCGCGGCCTCGGCGCCGCCGACCGCCTCTTCATCGGCTGCAGGCGC
CGCGGCGAGAGCGCGCCCGCGCCGCCGCCCGCCGTTCGCGTCACGCCGCAGCCGCCTGCC
CTCAACGGCGGCGAGCAGCAGCCGTGGAGCCCAATGTGTTACAGCACGTCGGGCTCGTCC
TACGACCCTACCAGCCCTGCCAATTCATATGCCTACCATCGCTCCGTAGACCAAGATCAC
AGCGACATACTACACGCAGGAGAGTCGCAGAGAGAAGCAGACGCCAAGAGCAGCAGCGCG
GCGTCGGCGCCGCCGCCGTCGAGGCGGCTCAGGCTGTTCGGCGTTAACCTCGACTGCGGC
CCGGAGCCGGAGGCGGATCAGGCGACGGCAATGTACGGCTACATGCACCACCAGAGCCCC
TACGCCGCAGTGTCTACAGTGCCAAATTACTGGTCAGTATTTTTTCAGTTTTAA

Osllg0156000
LOC_Osllg05740.1
Cover 81% identity 47%

SEQ ID NO: 51
MAM NHPLFSQEQPQSWPWGVAMYAN FHYHH HYEKEHMFEKPLTPSDVGKLNRLVIPKQHA
ERYFPLGAGDAADKGLILSFEDEAGAPWRFRYSYWTSSQSYVLTKGWSRYVKEKRLDAGD
VVHFERVRGSFGVGDRLFIGCRRRGDAAAAQTPAPPPAVRVAPAAQNAGEQQPWSPMCYS
TSGGGSYPTSPANSYAYRRAADHDHGDMHHADESPRDTDSPSFSAGSAPSRRLRLFGVNL
DCGPEPEADTTAAATMYGYMHQQSSYAAMSAVPSYWGNS

CDS SEQ ID NO: 52
ATGGCCATGAACCACCCTCTCTTCTCCCAGGAGCAACCCCAGTCCTGGCCATGGGGTGTG
GCCATGTACGCCAACTTCCACTACCACCACCACTACGAGAAGGAGCACATGTTTGAGAAG

CCCCTGACGCCCAGTGACGTGGGGAAGCTGAACCGGCTGGTGATCCCCAAGCAGCACGCC
GAGAGGTACTTCCCCCTCGGCGCCGGCGACGCCGCCGACAAGGGCCTGATCCTGTCGTTC
GAGGACGAGGCCGGCGCGCCGTGGCGGTTCAGGTACTCCTACTGGACGAGCAGCCAGAGC
TACGTGCTCACCAAGGGCTGGAGCCGCTACGTCAAGGAGAAGCGCCTCGACGCCGGCGAC
GTCGTGCACTTCGAGAGGGTGCGCGGCTCCTTCGGCGTCGGCGACCGTCTCTTCATCGGC
TGCAGGCGCCGCGGCGACGCCGCCGCCGCGCAAACACCCGCACCGCCGCCCGCCGTGCGC
GTCGCCCCGGCTGCACAGAACGCCGGCGAGCAGCAGCCGTGGAGCCCAATGTGTTACAGC
ACGTCGGGCGGCGGCTCATACCCTACCAGCCCAGCCAACTCCTACGCCTACCGCCGCGCA
GCAGATCATGATCACGGGGACATGCACCATGCAGACGAGTCTCCGCGCGACACGGACAGC
CCAAGCTTCAGTGCAGGCTCGGCGCCATCGAGGCGGCTCAGGCTGTTCGGCGTCAACCTC
GACTGCGGGCCAGAGCCGGAGGCAGACACCACGGCAGCGGCAACAATGTACGGCTACATG
CACCAGCAGAGCTCCTATGCTGCCATGTCTGCAGTACCCAGTTACTGGGGCAATTCATAA

os02g0683500 LOC_Os02g45850
Cover 47% identity 62%

SEQ ID NO: 53
MEFTTSSRFSKEEEDEEQDEAGRREI PFMTATAEAAPAPTSSSSSPAHHAASASASASAS
GSSTPFRSDDGAGASGSGGGGGGGGEAEVVEKEHM FDKVVTPSDVGKLNRLVI PKQYAEK
YFPLDAAAN EKGLLLNFEDRAGKPWRFRYSYWNSSQSYVMTKGWSRFVKEKRLDAGDTVS
FSRGIGDEAARHRLFI DWKRRADTRDPLRLPRGLPLPMPLTSHYAPWGIGGGGGFFVQPS
PPATLYEHRLRQGLDFRAFNPAAAMGRQVLLFGSARI PPQAPLLARAPSPLHHHYTLQPS
GDGVRAAGSPVVLDSVPVI ESPTTAAKRVRLFGVNLDNPHAGGGGGAAAGESSN HGNALS
LQTPAWM RRDPTLRLLELPPH HHHGAESSAASSPSSSSSSKRDAHSALDLDL

CDS SEQ ID NO: 54
ATGGAGTTCACTACAAGCAGTAGGTTTTCTAAAGAAGAGGAGGACGAGGAGCAGGATGAG
GCGGGAAGGCGAGAGATCCCCTTCATGACGGCCACGGCCGAAGCCGCGCCTGCGCCCACG
TCGTCGTCGTCGTCTCCTGCTCATCACGCGGCTTCCGCGTCGGCGTCGGCGTCTGCGTCA
GGGAGCAGCACTCCCTTTCGCTCCGACGATGGCGCCGGGGCGTCTGGGAGCGGCGGCGGC
GGCGGCGGCGGCGGAGAAGCGGAGGTGGTGGAGAAGGAGCACATGTTCGACAAGGTGGTG
ACGCCGAGCGACGTTGGGAAGCTGAACCGGCTGGTGATCCCGAAGCAGTACGCCGAGAAG
TACTTCCCGCTGGACGCGGCGGCGAACGAGAAGGGCCTCCTGCTCAACTTCGAGGACCGC
GCGGGGAAGCCATGGCGGTTCCGCTACTCCTACTGGAACAGCAGCCAGAGCTACGTGATG
ACCAAGGGGTGGAGCCGCTTCGTCAAGGAGAAGCGCCTCGACGCCGGGGACACCGTCTCC
TTCTCCCGCGGCATCGGCGACGAGGCGGCGCGGCACCGCCTCTTCATCGACTGGAAGCGC
CGCGCCGACACCCGCGACCCGCTCCGGCTGCCCCGCGGGCTGCCGCTCCCGATGCCGCTC
ACGTCGCACTACGCCCCGTGGGGGATCGGCGGCGGAGGGGGATTCTTCGTGCAGCCCTCG
CCGCCGGCCACGCTCTACGAGCACCGCCTCAGGCAAGGCCTCGACTTCCGCGCCTTCAAC
CCCGCCGCCGCGATGGGGAGGCAGGTCCTCCTGTTCGGCTCGGCGAGGATTCCTCCGCAA
GCACCACTGCTGGCGCGCGCCGTCGCCGCTGCACCACCACTACACGCTGCAGCCGAGC
GGCGATGGTGTAAGGGCGGCGGGCTCACCGGTGGTGCTCGACTCGGTTCCGGTCATCGAG
AGCCCCACGACGGCCGCGAAGCGCGTGCGGCTGTTCGGCGTGAACCTCGACAACCCGCAT
GCCGGCGGCGGCGGCGGCGCCGCCGCCGGCGAGTCGAGCAATCATGGCAATGCACTGTCA
TTGCAGACGCCCGCGTGGATGAGGAGGGATCCAACACTGCGGCTGCTGGAATTGCCTCCT
CACCACCACCATGGCGCCGAGTCGTCCGCTGCATCGTCTCCGTCGTCGTCGTCTTCCTCC
AAGAGGGACGCGCATTCGGCCTTGGATCTCGATCTGTAG

os04g0581400 LOC_Os04g49230
Cover 46% identity 64%

CDS SEQ ID NO: 55
ATGGAGTTTGCTACAACGAGTAGTAGGTTTTCCAAGGAAGAGGAGGAGGAGGAGGAAGGG
GAACAGGAGATGGAGCAGGAGCAGGATGAAGAGGAGGAGGAGGCGGAGGCCTCGCCCCGC
GAGATCCCCTTCATGACGTCGGCGGCGGCGGCGGCCACCGCCTCATCGTCCTCCCCGACA
TCGGTCTCCCCTTCCGCCACCGCTTCCGCGGCGGCGTCCACGTCGGCGTCGGGCTCTCCC
TTCCGGTCGAGCGACGGTGCGGGAGCGTCGGGGAGTGGCGGCGGCGGTGGCGGCGAGGAC
GTGGAGGTGATCGAGAAGGAGCACATGTTCGACAAGGTGGTGACGCCGAGCGACGTGGGG
AAGCTGAACCGGCTGGTGATCCCGAAGCAGCACGCCGAGAAGTACTTCCCGCTGGACTCG
GCGGCGAACGAGAAGGGCCTTCTCCTCAGCTTCGAGGACCGAACCGGCAAGCTATGGCGC
TTCCGCTACTCCTACTGGAACAGCAGCCAGAGCTACGTCATGACCAAGGGTTGGAGCCGC
TTCGTCAAGGAGAAGCGCCTCGACGCCGGGGACACCGTCTCCTTCTGCCGCGGCGCCGCC
GAGGCCACCCGCGACCGCCTCTTCATCGACTGGAAGCGCCGCGCCGACGTCCGCGACCCG
CACCGCTTCCAGCGCCTACCGCTCCCCATGACCTCGCCCTACGGCCCGTGGGGCGGCGGC
GCGGGCGCTTCTTCATGCCGCCCGCGCCGCCCGCCACGCTCTACGAGCATCACCGCTTTC
GCCAGGGCTTCGACTTCCGCAACATCAACCCCGCTGTGCCGGCGAGGCAGCTCGTCTTCT
TCGGCTCCCCAGGGACGGGGATTCATCAGCACCCGCCCTTGCCACCGCCGCCGTCGCCAC
CTCCGCCTCCTCACCAACTCCACATTACGGTGCACCACCCGAGCCCCGTAG

SEQ ID NO: 56
M EFATTSSRFSKEEEEEEEGEQEMEQEQDEEEEEAEASPREIPFMTSAAAAATASSSSPT
SVSPSATASAAASTSASGSPFRSSDGAGASGSGGGGGGEDVEVI EKEH MFDKVVTPSDVG
KLNRLVIPKQHAEKYFPLDSAANEKGLLLSFEDRTGKLWRFRYSYWNSSQSYVMTKGWSR
FVKEKRLDAGDTVSFCRGAAEATRDRLFI DWKRRADVRDPHRFQRLPLPMTSPYGPWGGG
AGASSCRPRRPPRSTSITAFARASTSATSTPLCRRGSSSSSAPQGRGFISTRPCHRRRRH
LRLLTNSTLRCTTRAP

os03g0120900 LOC_Os03g02900
Cover 47% identity 63%

CDS SEQ ID NO: 57
ATGGAGTTCATCACGCCAATCGTGAGGCCGGCATCGGCGGCGGCGGGCGGCGGCGAGGTG
CAGGAGAGTGGTGGGAGGAGCTTGGCGGCGGTGGAGAAGGAGCACATGTTCGACAAGGTG
GTGACGCCGAGCGACGTGGGGAAGCTGAACCGGCTGGTGATCCCGAAGCAGCACGCGGAG
AAGTACTTCCCGCTGGACGCGGCGTCCAACGAGAAGGGGCTCCTGCTCAGCTTCGAGGAC
CGCACGGGGAAGCCATGGCGGTTCCGCTACTCCTACTGGAACAGCAGCCAGAGCTACGTG
ATGACCAAGGGGTGGAGCCGCTTCGTCAAGGAGAAGCGACTCGACGCCGGGGACACCGTC
TCCTTCGGCCGCGGCGTCGGCGAGGCCGCGCGCGGGAGGCTCTTCATCGACTGGCGCCGC
CGCCCCGACGTCGTCGCCGCGCTCCAGCCGCCCACGCACCGCTTCGCCCACCACCTCCCT
TCCTCCATCCCCTTCGCTCCCTGGGCGCACCACCACGGACACGGAGCCGCCGCCGCCGCC
GCCGCCGCCGCCGGCGCCAGGTTTCTCCTGCCTCCCTCCTCGACTCCCATCTACGACCAC
CACCGCCGACACGCCCACGCCGTCGGGTACGACGCGTACGCCGCGGCCACCAGCAGGCAG
GTGCTGTTCTACCGGCCGTTGCCGCCGCAGCAGCAGCATCATCCCGCGGTGGTGCTGGAG
TCGGTGCCGGTGCGCATGACGGCGGGGCACGCGGAGCCGCCGTCGGCTCCGTCGAAGCGA
GTTCGGCTGTTCGGGGTGAACCTCGACTGCGCGAATTCCGAACAAGACCACGCCGGCGTG
GTCGGGAAGACGGCGCCGCCGCCGCTGCCATCGCCGCCGTCATCATCGTCATCTTCCTCC
GGGAAAGCGAGGTGCTCCTTGAACCTTGACTTGTGA

SEQ ID NO: 58

M EFITPIVRPASAAAGGGEVQESGGRSLAAVEKEHM FDKVVTPSDVGKLNRLVI PKQHAE
KYFPLDAASNEKGLLLSFEDRTGKPWRFRYSYWNSSQSYVMTKGWSRFVKEKRLDAGDTV
SFGRGVGEAARGRLFIDWRRRPDVVAALQPPTHRFAHHLPSSI PFAPWAHHHGHGAAAAA
AAAAGARFLLPPSSTPIYDH HRRHAHAVGYDAYAAATSRQVLFYRPLPPQQQH HPAVVLE
SVPVRMTAGHAEPPSAPSKRVRLFGVNLDCANSEQDHAGVVGKTAPPPLPSPPSSSSSSS
GKARCSLN LDL

os01g0693400
Cover 47% identity 63%

CDS SEQ ID NO: 59
ATGGACAGCTCCAGCTGCCTGGTGGATGATACCAACAGCGGCGGCTCGTCCACGGACAAG
CTGAGGGCGTTGGCCGCCGCGGCGGCGGAGACGGCGCCGCTGGAGCGCATGGGGAGCGGG
GCGAGCGCGGTGGTGGACGCGGCCGAGCCTGGCGCGGAGGCGGACTCCGGGTCCGGGGGA
CGTGTGTGCGGCGGCGGCGGCGGCGGTGCCGGCGGTGCGGGAGGGAAGCTGCCGTCGTCC
AAGTTCAAGGGCGTCGTGCCGCAGCCCAACGGGAGGTGGGGCGCGCAGATCTACGAGCGG
CACCAGCGGGTGTGGCTCGGCACGTTCGCCGGGGAGGACGACGCCGCGCGCGCCTACGAC
GTCGCCGCGCAGCGCTTCCGCGGCCGCGACGCCGTCACCAACTTCCGCCCGCTCGCCGAG
GCCGACCCGGACGCCGCCGCCGAGCTTCGCTTCCTCGCCACGCGCTCCAAGGCCGAGGTC
GTCGACATGCTCCGCAAGCACACCTACTTCGACGAGCTCGCGCAGAGCAAGCGCACCTTC
GCCGCCTCCACGCCGTCGGCCGCGACCACCACCGCCTCCCTCTCCAACGGCCACCTCTCG
TCGCCCCGCTCCCCCTTCGCGCCCGCCGCGGCGCGCGACCACCTGTTCGACAAGACGGTC
ACCCCGAGCGACGTGGGCAAGCTGAACAGGCTCGTCATACCGAAGCAGCACGCCGAGAAG
CACTTCCCGCTACAGCTCCCGTCCGCCGGCGGCGAGAGCAAGGGTGTCCTCCTCAACTTC
GAGGACGCCGCCGGCAAGGTGTGGCGGTTCCGGTACTCGTACTGGAACAGCAGCCAGAGC
TACGTGCTAACCAAGGGCTGGAGCCGCTTCGTCAAGGAGAAGGGTCTCCACGCCGGCGAC
GTCGTCGGCTTCTACCGCTCCGCCGCCAGTGCCGGCGACGACGGCAAGCTCTTCATCGAC
TGCAAGTTAGTACGGTCGACCGGCGCCGCCCTCGCGTCGCCCGCTGATCAGCCAGCGCCG
TCGCCGGTGAAGGCCGTCAGGCTCTTCGGCGTGGACCTGCTCACGGCGCCGGCGCCGGTC
GAACAGATGGCCGGGTGCAAGAGAGCCAGGGACTTGGCGGCGACGACGCCTCCACAAGCG
GCGGCGTTCAAGAAGCAATGCATAGAGCTGGCACTAGTATAG

SEQ ID NO: 49
60M DSSSCLVDDTNSGGSSTDKLRALAAAAAETAPLERMGSGASAVVDAAEPGAEADSGSGG
RVCGGGGGGAGGAGGKLPSSKFKGVVPQPNGRWGAQIYERHQRVWLGTFAGEDDAARAYD
VAAQRFRGRDAVTNFRPLAEADPDAAAELRFLATRSKAEVVDM LRKHTYFDELAQSKRTF
AASTPSAATTTASLSNGHLSSPRSPFAPAAARDHLFDKTVTPSDVGKLN RLVI PKQHAEK
HFPLQLPSAGGESKGVLLNFEDAAGKVWRFRYSYWNSSQSYVLTKGWSRFVKEKGLHAGD
VVGFYRSAASAGDDGKLFI DCKLVRSTGAALASPADQPAPSPVKAVRLFGVDLLTAPAPV
EQMAGCKRARDLAATTPPQAAAFKKQCI ELALV

Osl0g0537100 LOC_Osl0g39190
Cover 47% identity 60%

CDS SEQ ID NO: 61
ATGGAGTTCACCCCAATTTCGCCGCCGACGAGGGTCGCCGGCGGTGAGGAGGATTCCGAG

AGGGGGGCGGCGGCGTGGGCGGTGGTGGAGAAGGAGCACATGTTTGAGAAGGTCGTGACG
CCGAGCGACGTGGGGAAGCTGAACCGATTGGTCATCCCCAAGCAGCACGCCGAGAGGTAC
TTCCCGCTCGACGCCGCGGCGGGCGCCGGCGGCGGCGGTGGTGGCGGCGGTGGCGGCGGC
GGGGGGAAGGGGCTGGTGCTGAGCTTCGAGGACAGGACGGGGAAGGCGTGGAGGTTCCGG
TACTCGTACTGGAACAGCAGCCAGAGCTACGTGATGACCAAAGGGTGGAGCCGCTTCGTC
AAGGAGAAGCGCCTCGGCGCCGGCGACACCGTGTCGTTCGGCCGCGGCCTCGGCGACGCC
GCCCGCGGCCGCCTCTTCATCGACTTCCGCCGCCGCCGCCAGGACGCCGGCAGCTTCATG
TTCCCGCCGACGGCGGCGCCGCCGTCGCACTCGCACCACCATCATCAGCGACACCACCCG
CCGCTCCCGTCCGTGCCCCTTTGCCCGTGGCGAGACTACACCACCGCCTATGGCGGCGGC
TACGGCTACGGCTACGGCGGCGGCTCCACCCCGGCGTCCAGCCGCCACGTGCTGTTCCTC
CGGCCGCAGGTGCCGGCCGCTGTGGTGCTCAAGTCGGTGCCGGTGCACGTCGCGGCCACC
TCGGCGGTGCAGGAGGCGGCGACGACGACAAGGCCGAAGCGTGTCCGGCTGTTCGGGGTG
AACCTCGACTGCCCGGCGGCCATGGACGACGACGACGACATCGCCGGAGCGGCGAGCCGG
ACGGCAGCGTCGTCTCTCCTGCAGCTCCCCTCGCCGTCGTCCTCGACGTCGTCGTCGACG
GCGGGGAAGAAGATGTGCTCCTTGGATCTTGGGTTGTGA

SEQ ID NO: 62

MEFTPISPPTRVAGGEEDSERGAAAWAVVEKEH MFEKVVTPSDVGKLNRLVI PKQHAERY
FPLDAAAGAGGGGGGGGGGGGKGLVLSFEDRTGKAWRFRYSYWNSSQSYVMTKGWSRFV
KEKRLGAGDTVSFGRGLGDAARGRLFIDFRRRRQDAGSFMFPPTAAPPSHSHHHHQRHHP
PLPSVPLCPWRDYTTAYGGGYGYGYGGGSTPASSRHVLFLRPQVPAAVVLKSVPVHVAAT
SAVQEAATTTRPKRVRLFGVNLDCPAAMDDDDDIAGAASRTAASSLLQLPSPSSSTSSST
AGKKMCSLDLGL

Glycine max

L.OC100795470

Cover 75% identity 53%

SEQ ID NO: 63

Msinhysmdlpeptlwwphphhqqqqltlmdpdplrlnlnsddgngndndndenqttttggeqeilddkepmf^ kpltpsdvgklnr
Ivipkqhaekyfplsgdsggseckgllsfedesgkcwrfrysywnssqsyvltkgwsryvkdkrldagdvvlferhrvdaqrlfigwrrrrqsd
aalppahvssrksgggdgnsnknegwtrgfysahhpypthhlhhhqp spyqqqhdclhagrgsqgqnqrmrpvgnnsssssssrvlrl
fgvdmecqpehddsgpstpqcsynsnnmlpstqgtdhshhnfyqqqpsnsnpsphhmmvhhqpyyy

CDS SEQ ID NO: 64

ATGTCCATAAACCACTACTCCATGGACCTTCCCGAACCGACACTCTGGTGGCCACACCCA
CACCACCAACAACAACAACTAACCTTAATGGATCCTGACCCTCTCCGTCTCAACCTCAAT
AGCGACGATGGCAATGGCAATGACAACGACAACGACGAAAATCAAACAACCACAACAGGA
GGAGAACAAGAAATATTAGACGATAAAGAACCGATGTTCGAGAAGCCCTTAACCCCGAGC
GACGTGGGGAAGCTGAACCGTCTCGTAATCCCGAAGCAGCACGCGGAGAAGTACTTCCCA
CTGAGTGGTGACTCGGGCGGGAGCGAGTGCAAGGGGCTGTTACTGAGTTTCGAGGACGAG
TCG GGGAAGTGTTG GCG CTTCCG CTACTCGTACTG GAACAG CAG CCAG A G CTACGTGCTC
ACCAAAGGGTGGAGCCGCTACGTCAAGGACAAGCGCCTTGACGCGGGCGACGTCGTTTTG
TTCGAGCGTCACCGCGTCGACGCGCAGCGCCTCTTCATCGGGTGGAGGCGCAGGCGGCAG
AGCGATGCCGCCTTGCCGCCTGCGCACGTTAGCAGTAGGAAGAGTGGTGGTGGTGATGGG
AATAGTAATAAGAATGAGGGGTGGACCAGAGGGTTCTATTCTGCGCATCATCCTTATCCT
ACGCATCATCTTCATCATCATCAGCCCTCGCCATACCAACAACAACATGACTGTCTTCAT

GCAGGTAGAGGGTCCCAAGGTCAGAACCAAAGGATGAGACCAGTGGGAAACAACAGTTCT
AGCTCTAGTTCGAGTTCAAGGGTACTTAGGCTGTTCGGGGTCGACATGGAATGCCAACCC
GAACATGATGATTCTGGTCCCTCCACACCCCAATGCTCCTACAATAGTAACAACATGTTG
CCATCAACACAGGGCACAGATCATTCCCATCACAATTTCTACCAACAGCAACCTTCTAAT
TCCAATCCTTCCCCTCATCACATGATGGTACATCACCAACCATACTACTACTAG

LOC100818164
Cover 50% identity 73%

SEQ ID NO: 65
MSTNHYTM DLPEPTLWWPHPHQQQLTLI DPDPLPLNLN NDDNDNGDDNDNDENQTVTTTT
TGGEEEI INNKEPMFEKPLTPSDVGKLNRLVI PKQHAEKYFPLSGGDSGSSECKGLLLSF
EDESGKCWRFRYSYWNSSQSYVLTKGWSRYVKDKRLDAGDVVLFQRHRADAQRLFIGWRR
RRQSDALPPPAHVSSRKSGGDGNSSKNEGDVGVGWTRGFYPAHHPYPTHHHHPSPYHHQQ
DDSLHAVRGSQGQNQRTRPVGNSSSSSSSSSRVLRLFGVNMECQPEHDDSGPSTPQCSYN
TNNI LPSTQGTDI HSHLNFYQQQQTSNSKPPPHH MMI RHQPYYY

SEQ ID NO: 66
ATGTCGACAAACCACTACACCATGGACCTTCCCGAACCAACACTCTGGTGGCCACACCCA
CACCAACAACAACTAACCTTAATAGATCCAGACCCTCTCCCTCTGAACCTCAACAACGAC
GACAACGACAATGGCGACGACAACGACAACGACGAAAACCAAACAGTTACAACAACCACA
ACAGGAGGAGAAGAAGAAATAATAAACAATAAAGAACCGATGTTCGAGAAGCCGCTAACC
CCGAGCGACGTGGGGAAGCTGAACCGCCTCGTAATCCCGAAGCAGCACGCTGAGAAGTAC
TTTCCACTGAGTGGTGGTGACTCGGGCAGTAGCGAGTGCAAGGGGCTGTTACTGAGTTTC
GAGGACGAGTCGGGGAAGTGCTGGCGCTTCCGCTACTCGTACTGGAACAGCAGCCAGAGC
TACGTGCTCACCAAAGGGTGGAGCCGTTACGTGAAGGACAAGCGCCTCGATGCGGGAGAT
GTCGTTTTATTCCAGCGCCACCGCGCCGACGCGCAGCGCCTCTTCATCGGCTGGAGGCGC
AGGCGGCAGAGCGACGCCCTGCCGCCGCCTGCGCACGTTAGCAGCAGGAAGAGTGGTGGT
GATGGGAATAGTAGTAAGAATGAGGGTGATGTGGGCGTGGGCTGGACCAGAGGGTTCTAT
CCTGCGCATCATCCTTATCCTACGCATCATCATCATCCCTCGCCATACCATCACCAACAA
GATGACTCTCTTCATGCAGTTAGAGGGTCCCAAGGTCAGAACCAAAGGACGAGACCAGTG
GGAAACAG CAGTTCTAGTTCGAGTTCGAGTTCAAGG GTACTTAGGCTATTCGGGGTCAAC
ATGGAATGCCAACCCGAACATGATGATTCTGGACCCTCCACACCCCAATGCTCCTACAAT
ACTAACAACATATTGCCATCCACACAGGGCACAGATATTCATTCCCATCTCAATTTCTAC
CAACAACAACAAACTTCTAATTCCAAGCCTCCCCCTCATCACATGATGATACGTCACCAA
CCATACTACTACTAG

LOC100802734
Cover 77% identity 53%

SEQ ID NO: 67
MSSI N HYSPETTLYWTNDQQQQAAMWLSNSHTPRFN LNDEEEEEEDDVIVSDKATNNLTQ
EEEKVAMFEKPLTPSDVGKLNRLVIPKQHAEKH FPLDSSAAKGLLLSFEDESGKCWRFRY
SYWNSSQSYVLTKGWSRYVKDKRLHAGDVVLFHRH RSLPQRFFISCSRRQPNPVPAHVST
TRSSASFYSAHPPYPAH HFPFPYQPHSLHAPGGGSQGQNETTPGGNSSSSGSGRVLRLFG
VNMECQPDNHN DSQNSTPECSYTHLYHHQTSSYSSSSNPHHHMVPQQP

SEQ ID NO: 68

ATGTCATCGATAAACCACTATTCACCGGAAACAACACTATACTGGACCAACGACCAACAG
CAACAAGCCGCCATGTGGCTGAGTAATTCCCACACCCCGCGTTTCAATCTGAACGACGAG
GAGGAGGAGGAGGAAGACGACGTTATCGTTTCGGACAAGGCTACTAATAACTTGACGCAA
GAGGAGGAGAAGGTAGCCATGTTCGAGAAGCCGTTGACGCCGAGCGACGTCGGGAAGCTG
AACCGGCTCGTGATTCCGAAACAGCACGCGGAGAAGCACTTCCCTCTCGACTCGTCGGCG
GCGAAGGGGCTGTTGCTGAGTTTCGAGGACGAGTCCGGGAAGTGTTGGCGCTTCCGTTAC
TCTTATTGGAACAGTAG CCA GAGTTACGTTTTGACCAAAG GATGGAG CCGTTACGTCAAA
GACAAACGCCTCCACGCTGGCGACGTCGTTTTGTTCCACAGACACCGCTCCCTCCCTCAA
CGCTTCTTCATCTCCTGCAGCCGCCGCCAACCCAACCCGGTCCCCGCTCACGTTAGCACC
ACCAGATCCTCCGCTTCCTTCTACTCTGCGCACCCACCTTATCCTGCGCACCACTTCCCC
TTCCCATACCAACCTCACTCTCTTCATGCACCAGGTGGAGGGTCCCAAGGACAGAACGAA
ACGACACCGGGAGGGAACAGTAGTTCAAGTGGCAGTGGCAGGGTGCTGAGGCTCTTTGGT
GTGAACATGGAATGCCAACCTGATAATCATAATGATTCCCAGAACTCCACACCAGAATGC
TCCTACACCCACTTATACCACCATCAAACCTCTTCTTATTCTTCTTCTTCAAACCCTCAC
CATCACATGGTACCTCAACAACCATAA

L OC100781489

Cover 49% identity 64%

SEQ ID NO: 69

M ELMQQVKGNYSDSREEEEEEEAAAITRESESSRLHQQDTASNFGKKLDLMDLSLGSSKE
EEEEGNLQQGGGGVVHHAHQVVEKEHMFEKVATPSDVGKLNRLVIPKQHAEKYFPLDSST
NEKGLLLNFEDRNGKVWRFRYSYWNSSQSYVMTKGWSRFVKEKKLDAGDIVSFQRGLGDL
YRHRLYI DWKRRPDHAHAHPPH HHDPLFLPSI RLYSLPPTMPPRYHHDHH FH HHLNYNNL
FTFQQHQYQQLGAATTTH HNNYGYQNSGSGSLYYLRSSMSMGGGDQN LQGRGSNIVPMI I
DSVPVNVAHHNN NRHGNGGITSGGTNCSGKRLRLFGVN MECASSAEDSKELSSGSAAHVT
TAASSSSLHHQRLRVPVPVPLEDPLSSSAAAAARFGDHKGASTGTSLLFDLDPSLQYHRH

CDS SEQ ID NO: 70

ATGGAGTTGATGCAACAAGTTAAAGGTAATTATTCTGATAGCAGGGAGGAAGAGGAGGAA
GAGGAAGCTGCAGCAATCACAAGGGAATCAGAAAGCAGCAGGTTACACCAACAAGATACA
GCATCCAATTTTGGAAAGAAGCTAGACTTGATGGACTTGTCACTAGGGAGCAGCAAGGAA
GAGGAAGAGGAAGGGAATTTGCAACAAGGAGGAGGAGGAGTGGTTCATCATGCTCACCAA
GTAGTGGAGAAAGAACACATGTTTGAGAAAGTGGCGACACCGAGCGACGTAGGGAAGCTG
AACAGGCTGGTGATACCGAAGCAGCACGCGGAGAAGTACTTCCCCCTTGACTCCTCAACC
AACGAGAAGGGTCTGCTCCTGAATTTCGAGGACAGGAATGGGAAGGTGTGGCGATTCAGG
TATTCCTATTGGAACAGCAGCCAGAGCTATGTGATGACAAAAGGGTGGAGCCGCTTTGTT
AAGGAGAAGAAGCTGGATGCCGGTGACATTGTCTCCTTCCAGCGTGGCCTTGGGGATTTG
TATAGACATCGGTTGTATATAGATTGGAAGAGAAGGCCCGATCATGCTCATGCTCATCCA
CCTCATCATCACGATCCTTTGTTTCTTCCCTCTATCAGATTGTACTCTCTCCCTCCCACC
ATGCCACCTCGCTACCACCACGATCATCACTTTCACCACCATCTCAATTACAACAACCTC
TTCACTTTTCAGCAACACCAGTACCAGCAGCTTGGTGCTGCCACTACCACTCATCACAAC
AACTATGGTTACCAGAATTCGGGATCTGGTTCACTCTATTACCTAAGGTCCTCTATGTCA
ATGGGTGGTGGTGATCAAAACTTGCAAGGGAGAGGGAGCAACATTGTCCCCATGATCATT
GATTCTGTGCCGGTTAACGTTGCTCATCACAACAACAATCGCCATGGGAATGGGGGCATC
ACGAGTGGTGGTACTAATTGTAGTGGAAAACGACTAAGGCTATTTGGGGTGAACATGGAA
TGCGCTTCTTCGGCAGAAGATTCCAAAGAATTGTCCTCGGGTTCGGCAGCACACGTGACG
ACAGCTGCTTCTTCTTCTTCTCTTCATCATCAGCGCTTGAGGGTGCCAGTGCCAGTGCCA
CTTGAAGATCCACTTTCGTCGTCAGCAGCAGCAGCAGCAAGGTTTGGGGATCACAAAGGG
GCCAGTACTGGGACTTCGCTGCTGTTTGATTTGGATCCCTCTTTGCAGTATCATCGCCAC
TGA

LOC100776987
Cover 46% identity 62%

SEQ ID NO: 71
MDAISCLDESTTTESLSISQAKPSSTI MSSEKASPSPPPPNRLCRVGSGASAVVDSDGGG
GGGSTEVESRKLPSSKYKGVVPQPNGRWGSQIYEKHQRVWLGTFN EEDEAARAYDVAVQR
FRGKDAVTNFKPLSGTDDDDGESEFLNSHSKSEIVDMLRKHTYNDELEQSKRSRGFVRRR
GSAAGAGNGNSISGACVM KAREQLFQKAVTPSDVGKLNRLVIPKQHAEKHFPLQSAANGV
SATATAAKGVLLNFEDVGGKVWRFRYSYWNSSQSYVLTKGWSRFVKEKNLKAGDTVCFQR
STGPDRQLYIDWKTRNVVNEVALFGPVVEPIQMVRLFGVNI LKLPGSDSIANNNNASGCC
NGKRREM ELFSLECSKKPKI IGAL

CDS SEQ ID NO: 72
ATGGATGCAATTAGTTGCCTGGATGAGAGCACCACCACCGAGTCACTCTCCATAAGTCAG
GCGAAGCCTTCTTCGACGATTATGTCGTCCGAGAAGGCTTCTCCTTCCCCGCCGCCGCCG
AACAGGCTGTGCCGCGTCGGTAGCGGTGCTAGCGCAGTCGTGGATTCCGACGGCGGCGGC
GGGGGTGGCAGCACCGAGGTGGAGTCGCGGAAGCTCCCCTCGTCCAAGTATAAGGGCGTC
GTGCCCCAGCCCAACGGCCGCTGGGGCTCGCAGATTTACGAGAAGCACCAGCGCGTGTGG
CTGGGAACGTTCAACGAGGAAGACGAGGCGGCGCGTGCGTACGACGTCGCCGTGCAGCGA
TTCCGCGGCAAGGACGCCGTCACAAACTTCAAGCCGCTCTCCGGCACCGACGACGACGAC
GGGGAATCGGAGTTTCTCAACTCGCATTCGAAATCCGAGATCGTCGACATGCTGCGTAAG
CATACGTACAATGACGAGCTGGAACAAAGCAAGCGCAGCCGCGGCTTCGTACGTCGGCGC
GGCTCCGCCGCCGGCGCCGGAAACGGAAACTCAATCTCCGGCGCGTGTGTTATGAAGGCG
CGTGAGCAGCTATTCCAGAAGGCCGTTACGCCGAGCGACGTTGGGAAACTGAACCGTTTG
GTGATACCGAAGCAGCACGCGGAGAAGCACTTTCCTTTACAGAGCGCTGCTAACGGCGTT
AGCGCGACGGCGACGGCGGCGAAGGGCGTTTTGTTGAACTTCGAAGACGTTGGAGGGAAA
GTGTGGCGGTTTCGTTACTCGTATTGGAACAGTAGCCAGAGTTACGTCTTGACCAAAGGT
TGGAGCCGGTTCGTTAAGGAGAAGAATCTGAAAGCCGGTGACACGGTTTGTTTTCAACGG
TCCACTGGACCGGACAGGCAGCTTTACATCGATTGGAAGACGAGGAATGTTGTTAACGAG
GTCGCGTTGTTCGGACCGGTTGTCGAACCGATCCAGATGGTTCGGCTCTTTGGTGTTAAC
ATTTTGAAACTACCCGGTTCAGATTCTATCGCCAATAACAATAATGCAAGTGGGTGCTGC
AATGGCAAGAGAAGAGAAATGGAACTCTTTTCATTAGAGTGTAGCAAGAAACCTAAGATT
ATTGGTGCTTTGTAG

Locl00778733
Cover 44% identity 64%

SEQ ID NO: 73
MELMQEVKGYSDGREEEEEEEEAAEEI ITREESSRLLHQHQEAAGSNFI INNN HHHHQHH
HHHTTKQLDFMDLSLGSSKDEGNLQGSSSSVYAHH HHAASASSSANGN NNNSSSSNLQQQ
QQQPAEKEHM FDKVVTPSDVGKLNRLVI PKQHAEKYFPLDSSANEKGLLLNFEDRNGKLW
RFRYSYWNSSQSYVMTKGWSRFVKEKKLDAGDMVSFQRGVGELYRHRLYI DWWRRPDH HH
HHH HGPDHSTTLFTPFLI PNQPHH LMSI RWGATGRLYSLPSPTPPRHHEHLNYNNNAMYH
PFHHHGAGSGI NATTH HYNNYHEMSSTTTSGSAGSVFYHRSTPPISMPLADHQTLNTRQQ
QQQQQQQEGAGNVSLSPMI IDSVPVAHHLHHQQH HGGKSSGPSSTSTSPSTAGKRLRLFG
VNMECASSTSEDPKCFSLLSSSSMANSNSQPPLQLLREDTLSSSSARFGDQRGVGEPSM L
FDLDPSLQYRQ

SEQ ID NO: 74

ATGGAGTTGATGCAAGAAGTGAAAGGGTATTCTGATGGCAGAGAGGAGGAGGAGGAGGAA
GAGGAAGCAGCAGAAGAAATCATCACAAGAGAAGAAAGCAGCAGGTTGTTACACCAGCAC
CAGGAGGCAGCAGGTTCCAATTTCATCATCAACAATAATCATCATCATCATCAACATCAC
CACCACCACACAACAAAGCAGCTAGACTTCATGGACTTGTCACTTGGTAGCAGCAAGGAT
GAAGGGAATTTGCAAGGATCATCTTCTTCTGTCTATGCTCATCATCATCATGCAGCAAGT
GCTAGTTCTTCTGCCAATGGTAACAACAACAACAGCAGCAGCAGCAACTTGCAGCAACAG
CAGCAGCAGCCTGCTGAGAAGGAGCACATGTTTGATAAAGTAGTGACACCAAGTGATGTG
GGGAAGCTGAACCGGTTGGTGATACCAAAGCAGCATGCTGAGAAGTATTTCCCTCTTGAT
TCCTCAGCCAATGAGAAGGGTCTGTTGCTGAATTTTGAGGACAGGAATGGTAAGTTGTGG
AGGTTCAGGTACTCCTATTGGAACAGCAGCCAGAGCTATGTGATGACCAAAGGTTGGAGC
CGTTTTGTTAAGGAGAAGAAGCTTGATGCTGGTGACATGGTGTCCTTCCAGCGTGGTGTT
GGGGAGTTGTATAGGCATAGGTTGTACATAGATTGGTGGAGAAGGCCTGATCATCATCAC
CATCACCATCATGGCCCTGACCATTCAACCACACTCTTCACACCTTTCTTAATTCCCAAT
CAGCCTCATCACTTAATGTCCATCAGATGGGGTGCCACTGGCAGATTGTACTCCCTCCCT
TCCCCAACCCCACCACGCCACCATGAACACCTCAATTACAACAATAACGCCATGTATCAT
CCCTTTCATCACCATGGTGCTGGAAGTGGAATTAATGCTACTACTCATCACTACAACAAC
TATCATGAGATGAGTAGTACTACTACTTCAGGATCTGCAGGCTCAGTCTTTTACCACAGG
TCAACACCCCCAATATCAATGCCATTGGCTGACCACCAAACCTTGAACACAAGGCAGCAG
CAACAACAACAACAACAACAAGAGGGAGCTGGCAATGTTTCTCTTTCCCCTATGATCATT
GATTCTGTTCCAGTTGCTCACCACCTCCATCATCAACAACACCATGGTGGCAAGAGTAGT
GGTCCTAGTAGTACTAGTACTAGTCCTAGCACTGCAGGGAAAAGACTAAGGCTATTTGGG
GTCAACATGGAATGTGCTTCTTCAACATCAGAAGACCCCAAATGCTTCAGCTTGTTGTCC
TCATCTTCAATGGCTAATTCCAATTCACAACCACCACTTCAGCTTTTGAGGGAAGATACA
CTTTCGTCATCATCGGCAAGGTTTGGGGATCAGAGAGGAGTAGGGGAACCTTCAATGCTT
TTTGATCTGGACCCTTCTTTGCAATACCGGCAGTGA

LOC732601
Cover 44% identity 62%

SEQ ID NO: 75
MDGGCVTDETTTSSDSLSVPPPSRVGSVASAVVDPDGCCVSGEAESRKLPSSKYKGVVPQ
PNGRWGAQIYEKHQRVWLGTFNEEDEAARAYDIAALRFRGPDAVTN FKPPAASDDAESEF
LNSHSKFEIVDM LRKHTYDDELQQSTRGGRRRLDADTASSGVFDAKAREQLFEKTVTPSD
VGKLNRLVIPKQHAEKHFPLSGSGDESSPCVAGASAAKGMLLNFEDVGGKVWRFRYSYWN
SSQSYVLTKGWSRFVKEKNLRAGDAVQFFKSTGPDRQLYDCKARSGEVNNNAGGLFVPI
GPVVEPVQMVRLFGVNLLKLPVPGSDGVGKRKEMELFAFECCKKLKVIGAL

CDS SEQ ID NO: 76
ATGGATGGAGGCTGTGTCACAGACGAAACCACCACATCCAGCGACTCTCTTTCCGTTCCG
CCGCCCAGCCGCGTCGGCAGCGTTGCAAGCGCCGTCGTCGACCCCGACGGTTGTTGCGTT
TCCGGCGAGGCCGAATCCCGGAAACTCCCTTCGTCGAAATACAAAGGCGTGGTGCCGCAA
CCGAACGGTCGCTGGGGAGCTCAGATTTACGAGAAGCACCAGCGCGTGTGGCTCGGCACT
TTCAACGAGGAAGACGAAGCCGCCAGAGCCTACGACATCGCCGCGCTGCGCTTCCGCGGC
CCCGACGCCGTCACCAACTTCAAGCCTCCCGCCGCCTCCGACGACGCCGAGTCCGAGTTC
CTCAACTCGCATTCCAAGTTCGAGATCGTCGACATGCTCCGCAAGCACACCTACGACGAC
GAGCTCCAGCAGAGCACGCGCGGTGGTAGGCGCCGCCTCGACGCTGACACCGCGTCGAGC
GGTGTGTTCGACGCGAAAGCGCGTGAGCAGCTGTTCGAGAAAACGGTTACGCCGAGCGAC
GTCGGGAAGCTGAATCGATTAGTGATACCGAAGCAGCACGCGGAGAAGCACTTTCCGTTA
AGCGGATCCGGCGACGAAAGCTCGCCGTGCGTGGCGGGGGCTTCGGCGGCGAAGGGAATG
TTGTTGAACTTTGAGGACGTTGGAGGGAAAGTGTGGCGGTTTCGTTACTCTTATTGGAAC
AGTAGCCAGAGCTACGTGCTTACCAAAG GATGGAGCCG GTTCGTTAAGGA GAAG AATCTT
CGAGCCGGTGACGCGGTTCAGTTCTTCAAGTCGACCGGACCGGACCGGCAGCTATATATA

GACTGCAAGGCGAGGAGTGGTGAGGTTAACAATAATGCTGGCGGTTTGTTTGTTCCGATT
GGACCGGTCGTTGAGCCGGTTCAGATGGTTCGGCTTTTCGGGGTCAACCTTTTGAAACTA
CCCGTACCCGGTTCGGATGGTGTAGGGAAGAGAAAAGAGATGGAACTGTTTGCATTTGAA
TGTTGCAAGAAGTTAAAAGTAATTGGAGCTTTGTAA

Locl00801107
Cover 44% identity 61%

SEQ ID NO: 77
MDAISCMDESTTTESLSISLSPTSSSEKAKPSSM ITSSEKVSLSPPPSNRLCRVGSGASA
VVDPDGGGSGAEVESRKLPSSKYKGVVPQPNGRWGAQIYEKHQRVWLGTFNEEDEAARAY
DIAAQRFRGKDAVTNFKPLAGADDDDGESEFLNSHSKPEIVDM LRKHTYNDELEQSKRSR
GVVRRRGSAAAGTANSISGACFTKAREQLFEKAVTPSDVGKLNRLVIPKQHAEKHFPLQS
SNGVSATTIAAVTATPTAAKGVLLN FEDVGGKVWRFRYSYWNSSQSYVLTKGWSRFVKEK
NLKAGDTVCFH RSTGPDKQLYIDWKTRNVVNNEVALFGPVGPVVEPIQMVRLFGVNI LKL
PGSDTIVGN NNNASGCCNGKRREMELFSLECSKKPKI IGAL

CDS SEQ ID NO: 78
ATGGATGCAATTAGTTGCATGGATGAGAGCACCACCACTGAGTCACTCTCTATAAGTCTT
TCTCCGACGTCATCGTCGGAGAAAGCGAAGCCTTCTTCGATGATTACATCGTCGGAGAAG
GTTTCTCTGTCCCCGCCGCCGTCAAACAGACTATGCCGTGTTGGAAGCGGCGCGAGCGCA
GTCGTGGATCCTGATGGCGGCGGCAGCGGCGCTGAGGTAGAGTCGCGGAAACTCCCCTCG
TCGAAGTACAAGGGCGTGGTGCCCCAGCCCAACGGCCGCTGGGGTGCGCAGATTTACGAG
AAGCACCAGCGCGTGTGGCTTGGAACGTTCAACGAGGAAGACGAGGCGGCGCGTGCGTAC
GACATCGCCGCGCAGCGGTTCCGCGGCAAGGACGCCGTCACGAACTTCAAGCCGCTCGCC
GGCGCCGACGACGACGACGGAGAATCGGAGTTTCTCAACTCGCATTCCAAACCCGAGATC
GTCGACATGCTGCGAAAGCACACGTACAATGACGAGCTGGAGCAGAGCAAGCGCAGCCGC
GGCGTCGTCCGGCGGCGAGGCTCCGCCGCCGCCGGCACCGCAAACTCAATTTCCGGCGCG
TGCTTTACTAAGGCACGTGAGCAGCTATTCGAGAAGGCTGTTACGCCGAGCGACGTTGGG
AAATTG AACCGTTTG GTGATACCGAAG CAG CACG CG GA GAAG CACTTTCCGTTACAGAGC
TCTAACGGCGTTAGCGCGACGACGATAGCGGCGGTGACGGCGACGCCGACGGCGGCGAAG
GGCGTTTTGTTGAACTTCGAAGACGTTGGAGGGAAAGTGTGGCGGTTTCGTTACTCGTAT
TGGAACAGTAGCCAGAGTTACGTCTTAACCAAAGGTTGGAGCCGGTTCGTTAAGGAGAAG
AATCTGAAAGCTGGTGACACGGTTTGTTTTCACCGGTCCACTGGACCGGACAAGCAGCTT
TACATCGATTGGAAGACGAGGAATGTTGTTAACAACGAGGTCGCGTTGTTCGGACCGGTC
GGACCGGTTGTCGAACCGATCCAGATGGTTCGGCTCTTTGGGGTTAACATTTTGAAACTA
CCCGGTTCAGATACTATTGTTGGCAATAACAATAATGCAAGTGGGTGCTGCAATGGCAAG
AGAAGAGAAATGGAACTGTTCTCGTTAGAGTGTAGCAAGAAACCTAAGATTATTGGTGCT
TTGTAA

LOC100789009
Cover 44% identity 62%

SEQ ID NO: 79
M DGGSVTDETTTTSNSLSVPANLSPPPLSLVGSGATAVVYPDGCCVSGEAESRKLPSSKY
KGVVPQPNGRWGAQIYEKHQRVWLGTFNEEDEAARAYDIAAH RFRGRDAVTNFKPLAGAD
DAEAEFLSTHSKSEIVDM LRKHTYDNELQQSTRGGRRRRDAETASSGAFDAKAREQLFEK
TVTQSDVGKLNRLVIPKQHAEKHFPLSGSGGGALPCMAAAAGAKGMLLNFEDVGGKVWRF
RYSYWNSSQSYVLTKGWSRFVKEKNLRAGDAVQFFKSTGLDRQLYDCKARSGKVN NNAA
GLFI PVGPVVEPVQMVRLFGVDLLKLPVPGSDGIGVGCDGKRKEMELFAFECSKKLKVIG
A L

SEQ ID NO: 80

ATGGATGGAGGCAGTGTCACAGACGAAACCACCACAACCAGCAACTCTCTTTCGGTTCCG
GCGAATCTATCTCCGCCGCCTCTCAGCCTTGTCGGCAGCGGCGCAACCGCCGTCGTCTAC
CCCGACGGTTGTTGCGTCTCCGGCGAAGCCGAATCCCGGAAACTCCCGTCCTCGAAATAC
AAAGGCGTGGTGCCGCAACCGAACGGTCGTTGGGGAGCTCAGATTTACGAGAAGCACCAG
CGCGTGTGGCTCGGCACCTTCAACGAGGAAGACGAAGCCGCCAGAGCCTACGACATCGCC
GCGCATCGCTTCCGCGGCCGCGACGCCGTCACTAACTTCAAGCCTCTCGCCGGCGCCGAC
GACGCCGAAGCCGAGTTCCTCAGCACGCATTCCAAGTCCGAGATCGTCGACATGCTCCGC
AAGCACACCTACGACAACGAGCTCCAGCAGAGCACCCGCGGCGGCAGGCGCCGCCGGGAC
GCCGAAACCGCGTCGAGCGGCGCGTTCGACGCGAAGGCGCGTGAGCAGCTGTTCGAGAAA
ACCGTTACGCAGAGCGACGTCGGGAAGCTGAACCGATTAGTGATACCAAAGCAGCACGCG
GAGAAGCACTTTCCGTTAAGCGGATCCGGCGGCGGAGCCTTGCCGTGCATGGCGGCGGCT
GCGGGGGCGAAGGGAATGTTGCTGAACTTTGAGGACGTTGGAGGGAAAGTGTGGCGGTTC
CGTTACTCGTATTGGAACAGTAGCCAGAGCTACGTGCTTACCAAAGGATGGAGCCGGTTC
GTTAAGGAGAAGAATCTTCGAGCTGGTGACGCGGTTCAGTTCTTCAAGTCGACCGGACTG
GACCGGCAACTATATATAGACTGCAAGGCGAGGAGTGGTAAGGTTAACAATAATGCTGCC
GGTTTGTTTATTCCGGTTGGACCGGTTGTTGAGCCGGTTCAGATGGTACGGCTTTTCGGG
GTCGACCTTTTGAAACTACCCGTACCCGGTTCGGATGGTATTGGGGTTGGCTGTGACGGG
AAGAGAAAAGAGATGGAGCTGTTTGCATTTGAATGTAGCAAGAAGTTAAAAGTAATTGGA
GCTTTGTAA

Locl02660503
Cover 36% identity 57%

SEQ ID NO: 81

migvekvticmrievntekgrralmdcwqisgvhessdcseikfafdavvkrarheennaaaqkfkgvvsqqngnwgaqi yahqqriwl
gtfksereaamaydsasiklrsgechrnfpwndqtvqepqfqshysaetvlnmirdgtypskfatflktrqtqkgvakhiglkgddeeqfcct
qlfqkeltpsdvgklnrlvipkkhavsyfpyvggsadesgsvdveavfydklmrlwkfrycywkssqsyvftrgwnrfvkdkklkakdviafft
wgksggegeafalidviynnnaeedskgdtkqvlgnqlqlagseegededanigkdfnaqkglrlfgvcit

CDS SEQ ID NO: 82

atgattggagttgagaaagtgacaatttgtatgagaatagaggtgaatactgaaaagggaagaagggctttaatggactgttggcaaatatcag
gagttcatgaaagttcagattgtagcgaaatcaaatttgcattcgacgcagtagtaaaacgcgcgaggcatgaagagaataatgcagcagcac
agaagttcaaaggcgttgtgtctcaacaaaatgggaactggggtgcacagatatatgcacaccagcagagaatctggttggggaccttcaaat
ctgaaagagaggctgcaatggcttatgacagcgccagcataaaacttagaagcggagagtgccacagaaactttccatggaacgaccaaaca
gttcaagagcctcagttccaaagccattacagcgcagaaacagtgctaaacatgattagagatggcacctatccatcaaaatttgctacatttctc
aaaactcgtcaaacccaaaaaggcgttgcgaaacacataggtctgaagggtgatgacgaggaacagttttgttgcacccaacttttcagaagg
aattaacaccaagtgatgtgggcaagctcaacaggcttgtcatcccaaagaagcatgcagttagctattttccttacgttggtggcagtgctgatg
agagtggtagtgttgacgtggaggctgtgttttatgacaaactcatgcgattgtggaagttccgatactgctattggaagagcagccaaagttacg
tgttcaccagaggctggaatcggtttgtgaaggataagaagttgaaggctaaagatgtcattgcgttttttacgtggggaaaaagtggaggaga
gggagaagcttttgcattgatcgatgtaatttataataataatgcagaagaagacagcaagggagacaccaaacaagttttgggaaaccaatta
caattagctggcagtgaagaaggtgaagatgaagatgcaaacattggaaaggatttcaatgcacaaaagggtctgaggctctttggtgtgtgta
tcacctaa

Hordeum vulgare

MLOC_66387

Cover 47% identity 64%

SEQ ID NO: 83

M EFTATSSRFSKGEEEVEEEQEEASMREIPFMTPAAATCAAAPPSASASASTPASASGSS
PPFRSGDDAGASGSGAGDGSRSNVAEAVEKEHM FDKVVTPSDVGKLNRLVI PKQYAEKYF
PLDSAANEKGLLLNFEDSAGKPWRFRYSYWNSSQSYVMTKGWSRFVKEKRLDAGDTVSFS
RGAGEAARHRLFI DWKRRADTRDPLRLPRLPLPM PLTSHYSPWGLGAGARGFFM PPSPPA
TLYEHRLRQGFDFRGMNPSYPTMGRQVI LFGSAARM PPHGPAPLLVPRPPPPLHFTVQQQ
GSDAGGSVTAGSPVVLDSVPVI ESPTTATKKRVRLFGVNLDNPQHPGDGGGESSNYGSAL
PLQMPASAWRPRDHTLRLLEFPSHGAEASSPSSSSSSKREAHSGLDLDL

SEQ ID NO: 84

ATGGAGTTTACTGCGACAAGCAGTAGGTTTTCTAAAGGAGAGGAGGAGGTGGAGGAGGAG
CAGGAGGAGGCGTCGATGCGCGAGATCCCTTTCATGACGCCCGCGGCCGCCACCTGCGCC
GCGGCGCCGCCTTCTGCTTCTGCGTCGGCCTCGACACCCGCGTCAGCGTCTGGAAGTAGC
CCTCCCTTTCGATCTGGGGATGACGCCGGAGCGTCGGGGAGCGGGGCCGGCGACGGCAGC
CGCAGCAACGTGGCGGAGGCCGTGGAGAAGGAGCACATGTTCGACAAAGTGGTGACGCCG
AGCGACGTGGGGAAGCTTAACCGGCTGGTCATCCCCAAGCAGTACGCCGAGAAGTACTTC
CCGCTGGACTCGGCGGCCAACGAGAAGGGCCTTCTGCTCAACTTCGAGGACAGCGCCGGG
AAG CCATGGCG CTTCCG CTATTCCTACTGGAAC A GCAG CCAGA G CTACGTCATGACCAAA
GGCTGGAGCCGCTTCGTCAAGGAGAAGCGCCTCGACGCTGGGGACACCGTCTCCTTCTCC
CGCGGCGCCGGTGAGGCCGCGCGCCACCGCCTCTTCATCGACTGGAAGCGCCGAGCCGAC
ACCAGAGACCCGCTCCGCTTGCCCCGCCTCCCGCTCCCGATGCCGCTGACGTCGCACTAC
AGCCCGTGGGGCCTCGGCGCCGGCGCCAGAGGATTCTTCATGCCTCCCTCGCCGCCAGCC
ACGCTCTACGAGCACCGTCTCCGTCAAGGCTTCGACTTCCGCGGCATGAACCCCAGTTAC
CCCACAATGGGGAGACAGGTCATCCTTTTCGGCTCGGCCGCCAGGATGCCTCCGCACGGA
CCAGCACCACTCCTCGTGCCGCGCCCGCCGCCGCCGCTGCACTTCACGGTGCAGCAACAA
GGCAGCGACGCCGGCGGAAGTGTAACCGCAGGATCCCCAGTGGTGCTCGACTCAGTGCCG
GTAATCGAAAGCCCCACGACGGCAACGAAGAAGCGCGTGCGCTTGTTCGGCGTGAACTTG
GACAACCCGCAGCATCCCGGTGATGGCGGGGGCGAATCGAGCAATTATGGCAGTGCACTG
CCATTGCAGATGCCCGCATCAGCATGGCGGCCAAGGGACCATACGCTGAGGCTGCTCGAA
TTCCCCTCGCACGGTGCCGAGGCGTCGTCTCCATCGTCGTCGTCGTCTTCCAAGAGGGAG
GCGCATTCGGGCTTGGATCTCGATCTGTGA

M LOC44012

Cover 55% identity 63%

SEQ ID NO: 85

M LRKHTYFDELAQSKRAFAASAALSAPTTSGDAGGSASPPSPAAVREHLFDKTVTPSDVG
KLNRLVIPKQNAEKHFPLQLPAGGGESKGLLLNFEDDAGKVWRFRYSYWNSSQSYVLTKG
WSRFVKEKGLGAGDVVGFYRSAAGRTGEDSKFFI DCRLRPNTNTAAEADPVDQSSAPVQK
AVRLFGVDLLAAPEQGM PGGCKRARDLVKPPPPKVAFKKQCI ELALA

SEQ ID NO: 86

ATGCTCCGCAAGCACACCTACTTCGACGAGCTCGCCCAGAGCAAGCGCGCCTTCGCCGCG
TCGGCCGCGCTCTCCGCGCCCACCACCTCGGGCGACGCCGGCGGCAGCGCCTCGCCGCCC
TCCCCGGCCGCCGTGCGCGAGCACCTCTTCGACAAGACCGTCACGCCCAGCGACGTCGGC
AAGCTGAACAGGCTGGTGATACCGAAGCAGAACGCCGAGAAGCACTTCCCGCTGCAGCTC

CCGGCCGGCGGCGGCGAGAGCAAGGGCCTGCTCCTCAACTTCGAGGACGATGCGGGCAAG
GTGTGGCGGTTCCGCTACTCGTACTGGAACAGCAGCCAGAGCTACGTCCTCACCAAGGGC
TGGAGCCGCTTCGTGAAGGAGAAGGGCCTCGGCGCCGGAGACGTCGTCGGGTTCTACCGC
TCCGCCGCCGGGAGGACCGGCGAAGACAGCAAGTTCTTCATTGACTGCAGGCTGCGGCCG
AACACCAACACCGCCGCCGAAGCAGACCCCGTGGACCAGTCGTCGGCGCCCGTGCAGAAG
GCCGTGAGACTCTTCGGCGTCGATCTTCTCGCGGCGCCGGAGCAGGGCATGCCGGGCGGG
TGCAAGAGGGCCAGAGACTTGGTGAAGCCGCCGCCTCCGAAAGTGGCGTTCAAGAAGCAA
TGCATAGAGCTGGCGCTAGCGTAG

MLOC_57250
Cover 50% identity 57%

SEQ ID NO: 87
MYCSRGRI DPAEEGQVMGGLGVRDASWALFKVLEQSDVQVGQNRLLLTKEAVWGGPI PKL
FPELEELRGDGLNAENRVAVKILDADGCEGDANFRYLNSSKAYRVMGPQWSRLVKETGMC
KGDRLDLYAATATAASSCSGARAAVAPAPPGAIVKAAGF

CDS SEQ ID NO: 88
ATGTATTGTTCCCGCGGCCGCATCGATCCCGCGGAAGAAGGGCAGGTGATGGGCGGCCTC
GGCGTGCGCGACGCCAGCTGGGCGCTGTTCAAGGTGTTGGAGCAGTCCGACGTCCAGGTG
GGGCAGAACCGGCTGCTCCTCACCAAGGAGGCGGTGTGGGGCGGCCCTATCCCCAAGCTT
TTCCCGGAGCTGGAGGAGCTCCGCGGCGACGGCCTCAACGCCGAGAACAGGGTCGCGGTC
AAGATCCTCGACGCCGACGGCTGCGAGGGGGACGCCAACTTCCGCTACCTCAACTCCAGC
AAGGCGTACCGGGTCATGGGGCCTCAGTGGAGCCGGCTCGTGAAGGAGACCGGCATGTGC
AAGGGAGACCGCCTCGATCTGTACGCGGCAACGGCGACCGCTGCCTCTTCGTGTTCTGGA
GCCAGGGCGGCTGTGGCGCCGGCGATACCTCCCGGAGCAATCGTGAAGGCAGCCGGGTTC
TAA

MLOC_38822
Cover 47% identity 56%

SEQ ID NO: 89
MLRKH IYPDELAQHKRAFFFAAASSPTSSSSPLASPAPSAAAARREHLFDKTVTPSDVGK
LNRLVI PKQHAEKHFPLQLPSASAAVPGECKGVLLNFDDATGKVWRFRYSYWNSSQSYVL
TKGWSRFVKEKGLHAGDAVEFYRAASGN NQLFI DCKLRSKSTTTTTSVNSEAAPSPAPVT
RTVRLFGVDLLIAPAARHAHEHEDYGMAKTNKRTMEASVAAPTPAHAVWKKRCVDFALTY
RLATTPQCPRSRDQLEGVQAAGSTFAL

CDS SEQ ID NO: 90
ATGCTGCGCAAGCACATCTATCCCGACGAGCTCGCGCAGCACAAGCGCGCCTTCTTCTTC
GCCGCGGCGTCGTCCCCTACGTCGTCGTCGTCACCTCTCGCCTCGCCGGCTCCTTCAGCC
GCGGCGGCGCGGCGCGAGCACCTGTTCGACAAGACGGTCACGCCCAGCGACGTGGGGAAG
CTGAACCGGCTGGTGATCCCCAAGCAGCACGCCGAGAAGCACTTCCCGCTGCAGCTCCCT
TCTGCCAGCGCCGCCGTGCCAGGCGAGTGCAAGGGCGTGCTGCTCAACTTCGATGACGCG
ACCGGCAAGGTGTGGAGGTTCCGGTACTCCTACTGGAACAGCAGCCAGAGCTACGTGCTC
ACCAAGGGGTGGAGCCGCTTCGTGAAGGAGAAGGGCCTTCACGCCGGCGACGCCGTCGAG
TTCTACCGCGCCGCCTCCGGCAACAACCAGCTCTTCATCGACTGCAAGCTCCGGTCCAAG
AGCACCACGACGACGACCTCCGTCAACTCGGAGGCCGCCCCATCGCCGGCACCCGTGACG
AGGACAGTGCGACTCTTCGGGGTCGACCTTCTCATCGCGCCGGCGGCGAGGCACGCGCAT
GAGCACGAGGACTACGGCATGGCCAAGACAAACAAGAGAACCATGGAGGCCAGCGTAGCG
GCGCCTACTCCGGCGCACGCGGTGTGGAAGAAGCGGTGCGTAGACTTCGCGCTGACCTAC

CGACTTGCCACCACCCCACAGTGCCCGAGGTCAAGAGATCAACTAGAAGGAGTACAAGCA
GCTGGGAGTACATTTGCTCTATAG

MLOC_7940
Cover 49% identity 52%

SEQ ID NO: 91
MGVEI LSSTGEHSSQYSSGAASTATTESGVGGRPPTAPSLPVSIADESATSRSASAQSTS
SRFKGVVPQPNGRWGAQIYERHARVWLGTFPDEDSAARAYDVAALRYRGREAATNFPCAA
AEAELAFLAAHSKAEIVDM LRKHTYTDELRQGLRRGRGMGARAQPTPSWAREPLFEKAVT
PSDVGKLNRLVVPKQHAEKH FPLKRTPETTTTTGKGVLLNFEDGEGKVWRFRYSYWNSSQ
SYVLTKGWSRFVREKGLGAGDSIVFSCSAYGQEKQFFIDCKKNKTMTSCPADDRGAATAS
PPVSEPTKGEQVRVVRLFGVDIAGEKRGRAAPVEQELFKRQCVAHSQHSPALGAFVL

CDS SEQ ID NO: 92
ATGGGGGTGGAGATCCTGAGCTCAACGGGGGAACACTCCTCCCAGTACTCTTCCGGAGCC
GCGTCCACGGCGACGACGGAGTCAGGCGTGGGCGGACGGCCGCCGACTGCGCCGAGCCTA
CCTGTTTCCATCGCCGACGAGTCGGCGACCTCGCGGTCGGCATCGGCGCAGTCGACGTCG
TCGCGGTTCAAGGGCGTGGTGCCGCAGCCCAACGGGCGGTGGGGCGCCCAGATCTACGAG
CGCCACGCCCGCGTCTGGCTCGGCACGTTCCCGGACGAAGACTCTGCGGCGCGCGCCTAC
GACGTGGCCGCGCTCCGGTACCGGGGCCGCGAGGCCGCCACCAACTTCCCGTGCGCGGCC
GCCGAGGCGGAGCTCGCCTTCCTGGCGGCACACTCCAAGGCCGAGATCGTCGACATGCTC
CGGAAGCACACCTACACCGACGAGCTCCGCCAGGGCCTGCGGCGCGGCCGCGGCATGGGG
GCGCGCGCGCAGCCGACGCCGTCGTGGGCGCGGGAGCCCCTTTTCGAGAAGGCCGTGACC
CCGAGCGACGTGGGCAAGCTCAACCGCCTCGTTGTGCCGAAGCAGCACGCCGAGAAGCAC
TTCCCCCTGAAACGCACGCCGGAGACGACAACGACCACCGGCAAGGGGGTGCTTCTCAAC
TTCGAGGATGGCGAGGGGAAAGTGTGGAGGTTCCGGTACTCGTATTGGAACAGCAGCCAG
AGCTACGTGCTCACCAAGGGATGGAGCCGCTTCGTTCGGGAGAAGGGCCTCGGTGCCGGC
GACTCCATCGTGTTCTCCTGCTCGGCGTACGGTCAGGAGAAGCAGTTCTTCATCGACTGC
AAGAAGAACAAGACGATGACGAGCTGCCCCGCCGATGACCGCGGCGCCGCAACAGCGTCG
CCGCCAGTGTCAGAGCCAACAAAAGGAGAACAAGTCCGTGTTGTGAGGCTGTTCGGCGTC
GACATCGCCGGAGAGAAGAGGGGGCGAGCGGCGCCGGTGGAGCAGGAGTTGTTCAAGAGG
CAATGCGTGGCACACAGCCAGCACTCTCCAGCCCTAGGTGCCTTCGTCTTATAG

MLOC_56567
Cover 42% identity 59%

SEQ ID NO: 93
MGVEI LSSMVEHSFQYSSGASSATAESGAVGTPPRHLSLPVAIADESLTSRSASSRFKGV
VPQPNGRWGAQIYERHARVWLGTFPDQDSAARAYDVASLRYRGGDAAFNFPCVVVEAELA
FLAAHSKAEIVDM LRKQTYADELRQGLRRGRGMGVRAQPMPSWARVPLFEKAVTPSDVGK
LNRLVVPKQHAEKHFPLKRSPETTTTTGNGVLLNFEDGQGKVWRFRYSYWNSSQSYVLTK
GWSRFVREKGLGAGDSI MFSCSAYGQEKQFFI DCKKNTTVNGGKSASPLQVMEIAKAEQV
RVVRLFGVDIAGVKRERAATAEQGPQGWFKRQCMAHGQHSPALGDFAL

SEQ ID NO: 94
ATGGGGGTGGAGATCCTGAGCTCCATGGTGGAGCACTCCTTCCAGTACTCTTCGGGCGCG
TCCTCGGCCACCGCGGAGTCAGGCGCCGTCGGAACACCGCCGAGGCATCTGAGCCTACCT
GTCGCCATCGCCGACGAGTCCCTGACCTCACGGTCGGCGTCGTCTCGGTTCAAGGGCGTG

GTGCCGCAGCCCAACGGGCGGTGGGGCGCCCAGATCTACGAGCGCCACGCTCGCGTCTGG
CTCGGCACGTTCCCAGACCAGGACTCGGCGGCGCGCGCCTACGACGTTGCCTCGCTCAGG
TACCGCGGCGGCGACGCCGCCTTCAACTTCCCGTGCGTGGTGGTGGAGGCGGAGCTCGCC
TTCCTGGCGGCGCACTCCAAGGCTGAGATCGTTGACATGCTCCGGAAGCAGACCTACGCC
GATGAACTCCGCCAGGGACTACGGCGCGGCCGTGGCATGGGGGTGCGCGCGCAGCCGATG
CCGTCGTGGGCGCGGGTTCCCCTTTTCGAGAAGGCCGTGACCCCTAGCGATGTCGGCAAG
CTCAATCGCCTGGTGGTGCCGAAGCAGCACGCCGAGAAGCACTTCCCCCTGAAGCGCAGC
CCGGAGACGACGACCACCACCGGCAACGGCGTACTGCTCAACTTTGAGGACGGCCAGGGA
AAAGTGTG GAG GTTCCG GTACTCATATTG GAACAGC AGCCA GA GCTACGTGCTCACCAAA
GGCTGGAGCCGCTTCGTCCGGGAGAAGGGCCTCGGCGCCGGTGACTCCATCATGTTCTCC
TGCTCGGCGTACGGGCAGGAGAAGCAGTTCTTCATCGACTGCAAGAAGAACACGACCGTG
AACGGAGGCAAATCGGCGTCGCCGCTGCAGGTGATGGAGATTGCCAAAGCAGAACAAGTC
CGCGTCGTTAGACTGTTCGGTGTCGACATCGCCGGGGTGAAGAGGGAGCGAGCGGCGACG
GCGGAGCAAGGCCCGCAGGGGTGGTTCAAGAGGCAATGCATGGCACACGGCCAGCACTCT
CCTGCCCTAGGTGACTTCGCCTTATAG

MLOC_75135
Cover 43% identity 57%

SEQ ID NO: 95
MGMEILSSTVEHCSQYSSSASTATTESGAAGRSTTALSLPVAITDESVTSRSASAQPASS
RFKGVVPQPNGRWGSQIYERHARVWLGTFPDQDSAARAYDVASLRYRGRDAATNFPCAAA
EAELAFLTAHSKAEIVDMLRKHTYADELRQGLRRGRGMGARAQPTPSWARVPLFEKAVTP
SDVGKLNRLVVPKQHAEKH FPLKCTAETTTTTGNGVLLNFEDGEGKVWRFRYSYWNSSQS
YVLTKGWSSFVREKGLGAGDSIVFSSSAYGQEKQLFINCKKNTTMNGGKTALPLPVVETA
KGEQDHVVKLFGVDIAGVKRVRAATGELGPPELFKRQSVAHGCGRM NYICYSIGTIGPLM
LN

SEQ ID NO: 96
ATGGGGATGGAAATCCTGAGCTCCACGGTGGAGCACTGCTCCCAGTACTCTTCCAGCGCG
TCCACGGCCACAACGGAGTCAGGCGCCGCCGGAAGATCGACGACGGCTCTGAGCCTACCA
GTTGCCATCACCGACGAGTCCGTTACCTCGCGGTCGGCATCGGCGCAGCCGGCGTCATCA
CGGTTCAAGGGCGTGGTGCCGCAGCCCAACGGGCGGTGGGGCTCCCAGATCTACGAGCGC
CACGCTCGCGTCTGGCTCGGCACCTTCCCGGATCAGGACTCGGCGGCGCGTGCCTACGAC
GTTGCCTCGCTCAGGTACCGGGGCCGCGATGCCGCCACCAACTTCCCGTGCGCCGCTGCG
GAAGCGGAGCTCGCCTTCCTGACCGCGCACTCCAAGGCCGAGATCGTCGACATGCTCCGG
AAGCACACCTACGCCGACGAACTCCGCCAGGGCCTGCGGCGCGGCCGCGGCATGGGTGCG
CGCGCGCAGCCGACGCCGTCGTGGGCGCGGGTTCCCCTTTTCGAGAAGGCTGTGACCCCT
AGCGATGTCGGCAAGCTCAATCGCCTGGTGGTGCCGAAGCAGCACGCCGAGAAGCACTTC
CCCCTGAAGTGCACCGCAGAGACGACGACCACCACCGGCAACGGCGTGCTGCTAAACTTC
GAGGATGGTGAGGGGAAGGTGTGGAGGTTCCGGTACTCGTATTGGAACAGTAGCCAGAGC
TACGTGCTCACCAAAGGCTGGAGCAGCTTCGTCCGGGAGAAGGGCCTCGGCGCAGGCGAC
TCCATCGTCTTCTCCTCCTCGGCGTACGGGCAGGAGAAGCAGTTATTCATCAACTGCAAA
AAGAACACGACTATGAACGGCGGCAAAACAGCGTTGCCGCTGCCAGTGGTGGAGACTGCC
AAAGGAGAACAAGACCACGTCGTTAAGTTGTTCGGTGTTGACATCGCCGGTGTGAAGAGG
GTGCGAGCGGCGACGGGGGAGCTAGGCCCGCCGGAGTTGTTCAAGAGACAATCCGTGGCA
CACGGATGCGGAAGGATGAACTACATTTGCTACTCCATAGGGACAATAGGACCTCTTATG
CTCAACTGA

MLOC_63261
Cover 49% identity 51%

SEQ ID NO: 97

MASSKPTNPEVDNDMECSSPESGAEDAVESSSPVAAPSSRFKGVVPQPNGRWGAQIYEKH
SRVWLGTFGDEEAAACAYDVAALRFRGRDAVTNHQRLPAAEGAGWSSTSELAFLADHSKA
EIVDMLRKHTYDDELRQGLRRGHGRAQPTPAWAREFLFEKALTPSDVGKLNRLVVPKQHA
EKH FPPTTAAAAGSDGKGLLLNFEDGQGKVWRFRYSYWNSSQSYVLTKGWSRFVQEKGLC
AGDTVTFSRSAYVMNDTDEQLFI DYKQSSKNDEAADVATADENEAGHVAVKLFGVDIGWA
GMAGSSGG

SEQ ID NO: 98

ATGGCGTCTAGCAAGCCGACAAACCCCGAGGTAGACAATGACATGGAGTGCTCCTCCCCG
GAATCGGGTGCCGAGGACGCCGTGGAGTCGTCGTCGCCGGTGGCAGCGCCATCTTCGCGG
TTCAAGGGCGTCGTGCCGCAGCCTAACGGGCGCTGGGGCGCGCAGATCTACGAGAAGCAC
TCGCGGGTGTGGCTTGGCACGTTCGGGGACGAGGAAGCCGCCGCGTGCGCCTACGACGTG
GCCGCGCTCCGCTTCCGCGGCCGCGACGCCGTCACCAACCACCAGCGCCTGCCGGCGGCG
GAGGGGGCCGGCTGGTCGTCCACGAGCGAGCTCGCCTTCCTCGCCGACCACTCCAAGGCC
GAGATCGTCGACATGCTCCGGAAGCACACCTACGACGACGAGCTCCGGCAGGGCCTGCGC
CGCGGCCACGGGCGCGCGCAGCCCACGCCGGCGTGGGCGCGAGAGTTCCTCTTCGAGAAG
GCCCTGACCCCGAGCGACGTCGGCAAGCTCAACCGCCTGGTCGTTCCGAAGCAGCACGCC
GAGAAGCACTTCCCCCCGACGACGGCGGCGGCCGCCGGAAGCGACGGCAAGGGCTTGCTG
CTCAACTTCGAGGACGGCCAAGGGAAGGTGTGGAGGTTCCGGTACTCATACTGGAACAGC
AGCCAGAGCTACGTGCTCACCAAGGGCTGGAGCCGCTTCGTCCAAGAAAAGGGCCTCTGC
GCCGGCGACACCGTGACGTTCTCCCGGTCGGCGTACGTGATGAATGACACGGATGAGCAG
CTCTTCATCGACTACAAGCAGAGTAGCAAGAACGACGAAGCGGCCGACGTAGCCACTGCC
GATGAGAATGAGGCCGGCCATGTCGCCGTGAAGCTCTTCGGGGTCGACATTGGCTGGGCT
GGGATGGCGGGATCATCAGGTGGGTGA

MLOC_64708

Cover 49% identity 51%

SEQ ID NO: 99

MLFDSSVSASLGTM RPLVKKLDMLLAPARGYSTLCKRI KEVM HLLKHDVEEISSYLDELT
EVEDPPPMAKCWM NEARDLSYDMEDYI DSLLFVPPGHFI KKKKKKKKKGKKKMVI KKRLK
WCKQIVFTKQVSDHGI KTSKI IHVNVPRLPNKPKVAKI ILQFRIYVQEAI ERYDKYRLHH
CSTLRRRLLSTGSMLSVPIPYEEAAQIVTDGRMNEFISSLAANNAADQQQLKVVSVLGSG
CLGKTTLANVLYDRIGMQFECRAFI RVSKKPDMKRLFRDLLSQFHQKQPLPTSCNELGIS
DNI IKHLQDKRYLIVI DDLWDLSVWDII KYAFPKGNHGSRII ITTQI EDVALTCCCDHSE
HVFEM KPLN IGHSRELFFNRLFGSESDCLEEFKRVSNEIVDICGGLPLATI NIASHLANQ
ETEVSLDLLTDTRDLLRSCLWSNSTSERTKQVLNLSYSNLPDYLKTCLLYLHMYPVGSI I
WKDDLVKQLVAEGFIATREGKDQDQEM IEKAAGLCFDALI DRRFIQPIYTKYNNKVLSCT
VHEVVHDLIAQKSAEENFIVVADHNRKN IALSHKVRRLSLIFGDTIYAKTPAN ITKSQI R
SFRFFGLFECM PCITEFKVLRVLNLQLSGHRGDN DPI DLTGISELFQLRYLKITSDVCI K
LPNQMQKLQYLETLDI MDAPRVTAVPWDI IN LPHLLHLTLPVDTYLLDWISSMTDSVISL
WTLGKLNYLQH LHLTSSSTRPSYHLERSVEALGYLIGGHGKLKTIVVAHVSSAQNTVVRG
APEVTISWDRMSPPPLLQRFECPHSCFIFYRI PKWVTELGN LCI LKIAVKELHM ICLGTL
RGLHALTDLSLYVETAPIDKII FDKAGFSVLKYCKLRFAAGIAWLKFEADAM PSLWKLM L
VFNAIPRMDQNLVFFHHSRPAM HQRGGAVI IVEHMPGLRVISAKFGGAASDLEYASRTVV
SNH PSN PTI NMQLVCYSSNGKRSRKRKQQPYDVVKGQPDEYAKRLERPAEKRISTPTKSS
LRLHVPEITPKPMQITDNNVQRREH MFDTVLTRGDVGMLNRLVVPKKHAEKYFPLDSSST
RTSKAIVLSFEDPAGKSWFFHYSYRSSSQNYVMFKGWTGFVKEKFLEAGDTVSFSRGVGE
ATRGRLFI DCQNEQRYM FERVLTASDMESDGCSLMVPVNLVWPHPGLRKTI KGRHAVLQF
EDGSGNGKVWPFQFEASGQYYLMKGLNYFVNDRDLAAGYTVSFYRAGTRLFVDSGRKDDK

VALGTRSRERIYPKIVRSQ

Brassica rapa

LOC103849927

Cover 99% ident 80%

CDS SEQ ID NO: 100

ATGTTGTTTGATAGTTCAGTGAGTGCTTCGTTGGGCACCATGAGACCACTTGTCAAGAAG
CTCGACATGCTGCTAGCTCCTGCTCGGGGATACAGTACCTTGTGCAAGAGGATCAAGGAA
GTGATGCACCTTCTCAAACATGATGTTGAAGAGATAAGCTCCTACCTTGATGAACTTACA
GAGGTGGAGGACCCTCCACCAATGGCCAAGTGCTGGATGAACGAGGCACGCGACCTGTCT
TATGATATGGAGGATTACATTGATAGCTTGTTATTTGTGCCACCTGGCCATTTCATCAAG
AAGAAGAAGAAGAAGAAGAAGAAGGGAAAGAAGAAGATGGTGATAAAGAAGAGGCTCAAG
TGGTGCAAACAGATCGTATTCACAAAGCAAGTGTCAGACCATGGTATCAAGACCAGTAAA
ATCATTCATGTTAATGTCCCTCGTCTTCCCAATAAGCCCAAGGTTGCAAAAATAATATTA
CAGTTCAGGATCTATGTCCAGGAGGCTATTGAACGGTATGACAAGTATAGGCTTCACCAT
TGCAGCACCTTGAGGCGTAGATTGTTGTCCACTGGTAGTATGCTTTCAGTGCCAATACCC
TATGAAGAAGCTGCCCAAATTGTAACTGATGGCCGGATGAATGAGTTTATCAGCTCACTG
GCTGCTAATAATGCAGCAGATCAGCAGCAGCTCAAGGTGGTATCTGTTCTTGGATCTGGG
TGTCTAGGTAAAACTACGCTTGCGAATGTGTTGTACGACAGAATTGGGATGCAATTCGAA
TGCAGAGCTTTCATTCGAGTGTCCAAAAAGCCTGATATGAAGAGACTTTTCCGTGACTTG
CTCTCGCAATTCCACCAGAAGCAGCCACTGCCTACCAGTTGTAATGAGCTTGGCATAAGT
GACAATATCATCAAACATCTGCAAGATAAAAGGTATCTAATTGTTATTGATGATTTGTGG
GATTTATCAGTATGGGATATTATTAAATATGCTTTTCCAAAGGGAAACCATGGAAGCAGA
ATAATAATAACTACACAGATTGAAGATGTTGCATTAACTTGTTGCTGTGATCACTCGGAG
CATGTTTTCGAGATGAAACCTCTCAACATTGGTCACTCAAGAGAGCTATTTTTTAATAGA
CTTTTTGGTTCTGAAAGTGACTGTCTTGAAGAATTCAAACGAGTTTCAAACGAAATTGTT
GATATATGTGGTGGTTTACCGCTAGCAACAATCAACATAGCTAGTCATTTGGCAAACCAG
GAGACAGAAGTATCATTGGATTTGCTAACAGACACACGTGATTTGTTGAGGTCCTGTTTG
TGGTCAAATTCTACTTCAGAAAGAACAAAACAAGTACTGAACCTCAGCTACAGTAATCTT
CCTGATTATCTGAAGACATGTTTGCTGTATCTTCATATGTATCCAGTGGGCTCCATAATC
TGGAAGGATGATCTGGTGAAGCAATTGGTGGCTGAAGGGTTTATTGCTACAAGAGAAGGG
AAAGACCAAGACCAAGAAATGATAGAGAAAGCTGCAGGACTCTGTTTCGATGCACTTATT
GATAGAAGATTCATCCAGCCTATATATACCAAGTACAACAATAAGGTGTTGTCCTGCACG
GTTCATGAGGTGGTACATGATCTTATTGCCCAAAAGTCTGCTGAAGAGAATTTCATTGTG
GTAGCAGACCACAATCGAAAGAATATAGCACTTTCTCATAAGGTTCGTCGACTATCTCTC
ATCTTTGGCGACACAATATATGCCAAGACACCAGCAAACATCACAAAGTCACAAATTCGG
TCATTCAGATTTTTTGGATTATTCGAGTGTATGCCTTGTATTACAGAGTTCAAGGTTCTC
CGTGTTCTAAACCTTCAACTATCTGGTCATCGTGGGGACAATGACCCTATAGACCTCACT
GGGATTTCAGAACTGTTTCAGCTGAGATATTTAAAGATTACAAGTGATGTGTGCATAAAA
CTACCAAATCAAATGCAAAAACTGCAATATTTGGAAACGTTGGACATTATGGATGCACCA
AGAGTCACTGCTGTTCCATGGGATATTATAAATCTCCCACACCTGTTGCACCTGACTCTT
CCTGTTGATACATATCTGCTGGATTGGATTAGCAGCATGACTGACTCCGTCATCAGTCTG
TGGACCCTTGGCAAGCTGAACTACCTGCAGCATCTTCATCTTACTAGTTCTTCTACACGT
CCTTCATACCATCTGGAGAGAAGTGTGGAGGCTCTGGGTTATTTGATCGGAGGACATGGC
AAGCTGAAAACTATAGTAGTCGCTCATGTCTCCTCTGCTCAAAATACTGTGGTTCGTGGC
GCCCCAGAAGTAACCATTTCATGGGATCGTATGTCACCTCCCCCCCTTCTCCAGAGATTC
GAATGCCCACACAGCTGCTTCATATTTTACCGAATTCCTAAGTGGGTTACAGAACTTGGC
AACCTGTGCATTTTGAAGATTGCAGTGAAGGAGCTTCATATGATTTGTCTTGGTACTCTC
AGAGGATTGCATGCCCTCACTGATCTGTCGCTGTATGTGGAGACAGCGCCCATTGACAAG
ATCATCTTTGACAAGGCCGGGTTCTCAGTTCTCAAGTACTGCAAATTGCGCTTCGCGGCT

GGTATAGCTTGGCTGAAATTTGAGGCTGATGCAATGCCTAGTCTATGGAAACTGATGCTA
GTTTTCAACGCCATCCCACGAATGGACCAAAATCTTGTTTTCTTTCACCACAGCCGACCG
GCGATGCATCAACGTGGTGGTGCAGTAATCATTGTCGAGCATATGCCAGGGCTTAGAGTG
ATCTCCGCAAAATTTGGGGGCGCAGCTTCTGATCTAGAGTATGCTTCGAGGACCGTCGTT
AGTAACCATCCAAG CAATCCTACAATCAACATGCAATTGGTGTGTTATAGTTCCAATGGT
AAGAGAAGCAGAAAAAGGAAACAACAACCTTACGACGTTGTGAAGGGACAACCAGATGAA
TACGCCAAGAGATTGGAGAGACCAGCTGAGAAAAGGATTTCAACGCCGACAAAGTCTTCT
TTGCGTCTGCATGTTCCAGAAATTACACCAAAACCTATGCAGATTACAGACAACAATGTT
CAGAGGAGGGAGCACATGTTCGATACGGTTCTGACTCGGGGGGACGTGGGGATGCTGAAC
CGGCTGGTGGTACCGAAGAAGCACGCGGAGAAGTACTTCCCGCTGGACAGTTCCTCCACC
CGCACCAGCAAGGCCATCGTACTCAGCTTTGAGGACCCTGCTGGGAAGTCATGGTTCTTC
CACTACTCCTACCGGAGCAGCAGCCAGAACTACGTCATGTTCAAGGGGTGGACTGGCTTC
GTCAAGGAGAAGTTTCTCGAAGCCGGCGACACCGTCTCCTTCAGCCGCGGCGTCGGGGAG
GCCACGAGGGGGAGGCTCTTCATCGACTGTCAAAATGAGCAGAGGTACATGTTCGAGCGA
GTG CTGACG GCG AGTG ATATGGAGTCG GATGGCTG CTCG CTGATGGTCCCAGTGAACTTG
GTGTGGCCGCACCCCGGCCTCCGCAAGACGATCAAGGGGAGGCACGCCGTGCTGCAGTTT
GAGGACGGCAGCGGCAACGGGAAGGTGTGGCCATTTCAGTTTGAGGCCTCCGGCCAATAC
TATCTCATGAAGGGCTTGAACTACTTTGTTAACGACCGCGACCTTGCGGCTGGCTATACC
GTCTCCTTCTACCGCGCCGGCACGCGGTTGTTCGTCGACTCCGGGCGTAAAGATGACAAA
GTAG CCTTGG GAACCAG AAG CCG CG AAAG GATCTATCCTAAG ATCGTGCGGTCG CAGTAG

LOC103849927

SEQ ID NO: 101

msgnhysrdihhntpsvhhhqnyavvdreylfeksltpsdv gklnrlvipkqhaekhfplnnagddvaaaettekgmlltfedesgkcwkf
rysywnssqsyvltkgwsryvkdkhlhagdvvffqrhrfdlhrvfigwrkrgevssptavsvvsqearvnttaywsglttpyrqvhastssyp
nihqeyshygavaeiptvvtgssrtvrlfgvnlechgdvvetppcpdgyngqhfyyystpdpmnisfageameqvgdgrr

Bra034828

Cover 100% identity 79%

SEQ ID NO: 102

MSVNHYSNTLSSHNH HNEH KESLFEKSLTPSDVGKLNRLVIPKQHAERYLPLNNCGGGGD
VTAESTEKGVLLSFEDESGKSWKFRYSYWNSSQSYVLTKGWSRYVKDKHLNAGDVVLFQR
HRFDI HRLFIGWRRRGEASSSSAVSAVTQDPRANTTAYWNGLTTPYRQVHASTSSYPNNI
HQEYSHYGPVAETPTVAAGSSKTVRLFGVNLECHSDVVEPPPCPDAYNGQH IYYYSTPHP
M NISFAGEAM EQVG DG RG

CDS SEQ ID NO: 103

ATGTCAGTCAACCATTACTCAAACACTCTCTCGTCGCACAATCACCACAACGAACATAAA
GAGTCTTTGTTCGAGAAGTCACTCACGCCAAGCGATGTTGGAAAGCTAAACCGTTTAGTC
ATACCAAAACAACACGCCGAGAGATACCTCCCTCTCAATAATTGCGGCGGCGGCGGCGAC
GTGACGGCGGAGTCGACGGAGAAAGGGGTGCTTCTCAGCTTCGAGGACGAGTCGGGAAAA
TCTTGGAAATTCAGATACTCATATTGGAACAGTAGTCAAAGCTACGTGTTGACCAAAGGA
TGGAGCAGGTACGTCAAAGACAAGCACCTCAACGCAGGGGACGTCGTTTTATTTCAACGG
CACCGTTTTGATATTCATAGACTCTTCATTGGCTGGAGGAGACGCGGAGAGGCTTCTTCC
TCTTCCGCCGTTTCCGCCGTGACTCAAGATCCTCGAGCTAACACGACGGCGTACTGGAAC
GGTTTGACTACACCTTATCGTCAAGTACACGCGTCAACTAGTTCTTACCCTAACAACATC
CACCAAGAGTATTCACATTATGGCCCTGTTGCTGAGACACCGACGGTAGCTGCAGGGAGC
TCGAAGACGGTGAGGCTATTTGGAGTTAACCTCGAATGTCACAGTGACGTTGTGGAGCCA
CCACCGTGTCCTGACGCCTACAACGGCCAACACATTTACTATTACTCAACTCCACATCCC

ATGAATATCTCATTTGCTGGAGAAGCAATGGAGCAGGTAGGAGATGGACGAGGTTGA

Bra005886
Cover 100% identity 79%

SEQ ID NO: 104
MSVN HYSTDH HQVHH HHTLFLQNLHTTDTSEPTTTAATSLREDQKEYLFEKSLTPSDVGK
LNRLVI PKQHAEKYFPLNTI ISNNAEEKGMLLSFEDESGKCWRFRYSYWNSSQSYVLTKG
WSRYVKDKQLDPADVVFFQRQRSDSRRLFIGWRRRGQGSSSAANTTSYSSSMTAPPYSNY
SNRPAHSEYSHYGAAVATATETHHPSSSAVGSSRTVRLFGVNLECQMDEDEGDDSVATA
AAAECPRQDSYYDQN MYNYYTPHSSAS

CDS 105
ATGTCAGTCAACCATTACTCCACGGACCACCACCAGGTCCACCACCACCACACTCTCTTC
TTGCAGAACCTCCACACCACCGACACATCGGAGCCAACCACAACCGCCGCCACATCACTC
CGCGAAGACCAGAAAGAGTATCTCTTCGAGAAATCTCTCACACCAAGCGACGTTGGCAAA
CTCAACCGTCTCGTTATACCAAAACAGCACGCGGAGAAGTACTTCCCTCTCAACACCATC
ATCTCCAATAATGCTGAGGAGAAAGGGATGCTTCTAAGCTTCGAAGACGAGTCAGGCAAG
TGCTGGAGGTTCAGATACTCTTACTGGAACAGCAGTCAAAGCTACGTGTTGACTAAAGGA
TGGAGCAGATACGTCAAAGACAAACAGCTCGACCCAGCCGATGTTGTTTTCTTCCAACGT
CAACGTTCTGATTCCCGGAGACTCTTTATTGGCTGGCGTAGACGCGGTCAAGGCTCCTCC
TCCGCCGCGAATACGACGTCGTATTCTAGTTCCATGACTGCTCCACCGTATAGTAATTAC
TCTAATCGTCCTGCTCACTCAGAGTATTCCCACTATGGCGCCGCCGTAGCAACAGCGACG
GAGACGCACTTCATACCATCGTCTTCCGCCGTCGGGAGCTCGAGGACGGTGAGGCTTTTT
GGTGTGAATTTGGAGTGTCAAATGGATGAAGACGAAGGAGATGATTCGGTTGCCACGGCA
GCCGCCGCTGAGTGTCCTCGTCAGGACAGCTACTACGACCAAAACATGTACAATTATTAC
ACTCCTCACTCCTCAGCCTCATAA

Bra005301
Cover 100% identity 58%

SEQ ID NO: 106
MSI NQYSSDFNYHSLMWQQQQHRHHHHQNDVAEEKEALFEKPLTPSDVGKLNRLVIPKQH
AERYFPLAAAAADAMEKGLLLCFEDEEGKPWRFRYSYWNSSQSYVLTKGWSRYVKEKQLD
AGDVI LFHRHRVDGGRFFIGWRRRGNSSSSSDSYRH LQSNASLQYYPHAGVQAVESQRGN
SKTLRLFGVNMECQLDSDLPDPSTPDGSTICPTSHDQFH LYPQQHYPPPYYMDISFTGDV
HQTRSPQG

CDS SEQ ID NO: 107
ATGTCAATAAACCAATACTCAAGCGATTTCAACTACCACTCTCTCATGTGGCAACAACAG
CAGCACCGCCACCACCACCATCAAAACGACGTCGCGGAGGAAAAAGAAGCTCTTTTCGAG
AAACCCTTAACCCCAAGTGACGTCGGAAAACTCAACCGCCTCGTCATCCCAAAACAGCAC
GCCGAGAGATACTTCCCTCTCGCAGCAGCCGCCGCAGACGCGATGGAGAAGGGATTACTT
CTCTGCTTCGAGGACGAGGAAGGTAAGCCATGGAGATTCAGATACTCGTATTGGAACAGT
AGCCAGAGTTATGTCTTGACCAAAGGATGGAGCAGATACGTCAAGGAGAAGCAGCTCGAC
GCCGGTGACGTCATTCTCTTCCACCGCCACCGTGTTGACGGAGGAAGATTCTTCATTGGC
TGGAGAAGACGCGGCAACTCTTCCTCCTCTTCCGACTCTTATCGCCATCTTCAGTCCAAT
GCCTCGCTCCAATATTATCCTCATGCAGGAGTTCAAGCGGTGGAGAGCCAGAGAGGGAAT

TCGAAGACATTAAGACTGTTCGGAGTGAACATGGAGTGTCAGCTAGACTCCGACTTGCCC
GATCCATCTACACCAGACGGTTCCACCATATGTCCGACCAGTCACGACCAGTTTCATCTC
TACCCTCAACAACACTATCCTCCTCCGTACTACATGGACATAAGTTTCACAGGAGATGTG
CACCAGACGAGAAGCCCACAAGGATAA

Bra017262
Cover 92% identity 56%

SEQ ID NO: 108
MSI NQYSSEFYYHSLMWQQQQQHHHQNEVVEEKEALFEKPLTPSDVGKLNRLVIPKQHAE
RYFPLAAAAVDAVEKGLLLCFEDEEGKPWRFRYSYWNSSQSYVLTKGWSRYVKEKQLDAG
DVVLFHRH RADGGRFFIGWRRRGDSSSSSDSYRNLQSNSSLQYYPHAGAQAVENQRGNSK
TLRLFGVNM ECQIDSDWSEPSTPDGFTTCPTNHDQFPIYPEH FPPPYYMDVSFTGDVHQT
SSQQG

CDS SEQ ID NO: 109
ATGTCAATAAATCAATATTCAAGCGAGTTCTACTACCATTCTCTCATGTGGCAACAACAG
CAGCAACACCACCATCAAAACGAAGTCGTGGAGGAAAAAGAAGCTCTTTTCGAGAAACCC
TTAACCCCAAGTGACGTCGGAAAACTAAACCGCCTAGTCATCCCTAAACAGCACGCCGAG
AGATACTTCCCTCTCGCCGCCGCCGCGGTAGACGCCGTGGAGAAGGGATTACTCCTCTGC
TTCGAGGACGAGGAAGGTAAGCCATGGAGATTCAGATACTCTTATTGGAATAGTAGCCAG
AGTTACGTCTTGACCAAAGGATGGAGCAGATATGTTAAAGAGAAGCAACTTGACGCCGGC
GACGTTGTTCTCTTTCATCGCCACCGTGCTGACGGTGGAAGATTCTTCATTGGCTGGAGA
AGACGCGGCGACTCTTCCTCCTCCTCCGACTCTTATCGCAATCTTCAATCTAATTCCTCG
CTCCAATATTATCCTCATGCAGGGGCTCAAGCGGTGGAGAACCAGAGAGGTAACTCCAAG
ACATTGAGACTTTTTGGAGTGAACATGGAGTGCCAGATAGACTCAGACTGGTCCGAGCCA
TCCACACCTGACGGTTTTACCACATGTCCAACCAATCACGACCAGTTTCCTATCTACCCT
GAACACTTTCCTCCTCCGTACTACATGGACGTAAGTTTCACAGGAGATGTGCACCAGACG
AGTAGCCAACAAGGATAG

Bra000434
Cover 96% identity 47%

SEQ ID NO: 110
MMTNLSLAREGEEEEEEAGAKKPTEEVEREHM FDKVVTPSDVGKLNRLVI PKQHAERYFP
LDSSTNEKGLI LNFEDLTGKSWRFRYSYWNSSQSYVMTKGWSRFVKDKKLDAGDIVSFLR
CVGDTGRDSRLFI DWRRRPKVPDYTTSTSHFPAGAMFPRFYSFQTATTSTSYNPYNHQQP
RHHHSGYCYPQI PREFGYGYVVRSVDQRAVVADPLVI ESVPVM MHGGARVNQAAVGTAGK
RLRLFGVDMECGESGGTNSTEEESSSSGGSLPRGGASPSSSMFQLRLGNSSEDDH LFKKG
KSSLPFN LDQ

SEQ ID NO: 111
ATGATGACAAATTTGTCTCTTGCAAGAGAAGGAGAAGAAGAAGAAGAAGAGGCAGGAGCA
AAGAAGCCCACAGAAGAAGTGGAGAGAGAGCACATGTTCGACAAAGTGGTGACTCCAAGT
GACGTCGGGAAACTAAACCGACTCGTGATCCCAAAGCAACACGCGGAGAGATACTTCCCT
TTAGATTCATCCACAAACGAGAAGGGTTTGATTCTAAACTTCGAAGATCTCACGGGAAAG
TCATGGAGGTTCCGTTACTCTTACTGGAACAGCAGTCAGAGCTATGTCATGACTAAAGGT
TGGAGCCGTTTCGTTAAAGACAAGAAGCTAGACGCTGGAGATATTGTCTCTTTCCTGAGA
TGTGTCGGAGACACAGGAAGGGACAGCCGCTTGTTTATCGATTGGAGGAGACGACCTAAA

GTCCCTGACTACACGACATCGACTTCTCACTTTCCTGCCGGAGCTATGTTCCCTAGGTTT
TACAGTTTTCAGACAGCAACTACTTCCACAAGTTACAATCCCTATAATCATCAGCAGCCA
CGTCATCATCACAGTGGTTACTGTTATCCTCAAATCCCGAGAGAATTTGGATATGGGTAT
GTCGTTAGGTCAGTAGATCAGAGGGCGGTGGTGGCTGATCCGTTAGTGATCGAATCTGTG
CCGGTGATGATGCACGGAGGAGCTCGAGTGAACCAGGCGGCTGTTGGAACGGCCGGGAAA
AGGCTGAGGCTTTTTGGAGTCGATATGGAATGTGGCGAGAGTGGAGGAACAAACAGTACG
GAGGAAGAATCTTCATCTTCCGGTGGGAGTTTGCCACGTGGCGGTGCTTCTCCGTCTTCC
TCTATGTTTCAGCTGAGGCTTGGAAACAGCAGTGAAGATGATCACTTATTTAAGAAAGGA
AAGTCTTCATTGCCTTTTAATTTGGATCAATAA

Bra040478
Cover 96% identity 48%

SEQ ID NO: 112
MMTNLSLAREGEAQVKKPI EEVEREHMFDKVVTPSDVGKLNRLVIPKQHAERYFPLDSSS
NEKGLLLNFEDLTGKSWRFRYSYWNSSQSYVMTKGWSRFVKDKKLDAGDIVSFQRCVGDS
RLFI DWRRRPKVPDYPTSTAHFAAGAMFPRFYSFPTATTSTCYDLYNHQPPRHHHIGYGY
PQI PREFGYGYFVRSVDQRAVVADPLVI ESVPVMMRGGARVSQEVVGTAGKRLRLFGVDM
EEESSSSGGSLPRAGGGGASSSSSLFQLRLGSSCEDDH FSKKGKSSLPFDLDQ

SEQ ID NO: 113
ATGATGACCAACTTGTCTCTTGCAAGGGAAGGAGAAGCACAAGTAAAGAAGCCCATAGAA
GAAGTTGAGAGAGAGCACATGTTCGACAAAGTGGTGACTCCAAGCGACGTAGGGAAACTA
AACAGACTCGTGATCCCAAAGCAACACGCAGAGAGATACTTCCCTCTAGATTCATCCTCA
AACGAGAAAGGTTTGCTTCTAAACTTTGAAGATCTAACAGGAAAGTCATGGAGGTTCCGT
TACTCTTACTGGAACAGTAGCCAGAGCTATGTCATGACTAAAGGTTGGAGTCGTTTCGTT
AAAGACAAGAAGCTTGACGCCGGAGATATTGTCTCTTTCCAGAGATGTGTCGGAGACAGC
CGCTTGTTTATCGATTGGAGGAGACGACCTAAAGTCCCTGACTATCCGACATCGACTGCT
CACTTTGCTGCAGGAGCTATGTTCCCTAGGTTTTACAGTTTTCCGACAGCAACTACTTCG
ACATGTTACGATCTGTACAATCATCAGCCGCCACGTCATCATCACATTGGTTACGGTTAT
CCACAGATTCCGAGAGAATTTGGATACGGGTATTTCGTTAGGTCAGTGGACCAGAGAGCG
GTGGTGGCTGATCCGTTGGTGATCGAATCTGTGCCGGTGATGATGCGCGGAGGAGCTCGA
GTTAGTCAGGAGGTTGTTGGAACGGCCGGGAAGAGGCTGAGGCTTTTTGGAGTCGATATG
GAGGAAGAATCTTCATCTTCCGGTGGGAGTTTGCCGCGTGCCGGAGGTGGCGGTGCTTCT
TCATCTTCCTCTTTGTTTCAGCTGAGACTTGGGAGCAGCTGTGAAGATGATCACTTCTCT
AAGAAAGGAAAGTCTTCATTGCCTTTTGATTTGGATCAATAA

Bra004501
Cover 74% identity 45%

SEQ ID NO: 114
M M MTNLSLSREGEEEEEEEQEEAKKPM EEVEREHM FDKVVTPSDVGKLNRLVI PKQYAER
YFPLDSSTNEKGLLLNFEDLAGKSWRFRYSYWNSSQSYVMTKGWSRFVKDKKLDAGDIVS
FQRCVGDSGRDSRLFIDWRRRPKVPDHPTSIAHFAAGSM FPRFYSFPTATSYNLYNYQQP
RHHHHSGYNYPQI PREFGYGYLVDQRAVVADPLVIESVPVMM HGGAQVSQAVVGTAGKRL
RLFGVDMEEESSSSGGSLPRGDASPSSSLFQLRLGSSSEDDH FSKKGKSSLPFDLDQ

SEQ ID NO: 133
ATGATGATGACAAACTTGTCTCTTTCAAGAGAAGGAGAAGAGGAGGAAGAAGAAGAACAA
GAAGAGGCCAAGAAGCCCATGGAAGAAGTAGAGAGAGAGCACATGTTCGACAAAGTGGTG
ACTCCAAGCGATGTTGGTAAACTAAACCGGCTCGTGATCCCAAAGCAATACGCAGAGAGA

TACTTCCCTTTAGATTCATCCACAAACGAGAAAGGTTTGCTTCTAAACTTCGAAGATCTC
GCAGGAAAGTCATGGAGGTTCCGTTACTCTTACTGGAACAGTAGTCAGAGCTATGTCATG
ACTAAAGGTTGGAGCCGTTTCGTTAAAGACAAAAAGCTAGACGCCGGAGATATTGTCTCT
TTCCAGAGATGTGTCGGAGATTCAGGAAGAGACAGCCGCTTGTTTATTGATTGGAGGAGA
AGACCTAAAGTTCCTGACCATCCGACATCGATTGCTCACTTTGCTGCCGGATCTATGTTT
CCTAGGTTTTACAGTTTTCCGACAGCAACTAGTTACAATCTTTACAACTATCAGCAGCCA
CGTCATCATCATCACAGTGGTTATAATTATCCTCAAATTCCGAGAGAATTTGGATACGGG
TACTTGGTGGATCAAAGAGCCGTGGTGGCTGATCCGTTGGTGATTGAATCTGTGCCGGTG
ATGATGCACGGAGGAGCTCAAGTTAGTCAGGCGGTTGTTGGAACGGCCGGGAAGAGGCTG
AGGCTTTTTGGAGTCGATATGGAGGAAGAATCTTCATCTTCCGGTGGGAGTTTGCCACGT
GGTGACGCTTCTCCGTCTTCCTCTTTGTTTCAGCTGAGACTTGGAAGCAGCAGTGAAGAT
GATCACTTCTCTAAGAAAGGAAAGTCCTCATTGCCTTTTGATTTGGATCAATAA

Bra003482

Cover 79% identity 44%

SEQ ID NO: 115

MNQEEENPVEKASSM EREHM FEKVVTPSDVGKLNRLVIPKQHAERYFPLDNNSDSSKGLL
LNFEDRTGNSWRFRYSYWNSSQSYVMTKGWSRFVKDKKLDAGDIVSFQRDPGNKDKLFI D
WRRRPKI PDHHHQFAGAMFPRFYSFSHPQNLYHRYQQDLGIGYYVSSMERNDPTAVI ESV
PLI MQRRAAHVAAI PSSRGEKRLRLFGVDM ECGGGGGSVNSTEEESSSSGGGGGVSMASV
GSLLQLRLVSSDDESLVAMEAASVDEDHHLFTKKGKSSLSFDLDRK

SEQ ID NO: 116

ATGAATCAAGAAGAAGAGAATCCTGTGGAAAAAGCCTCTTCAATGGAGAGAGAGCACATG
TTTGAAAAAGTAGTAACACCAAGCGACGTAGGCAAACTAAACCGACTCGTGATCCCAAAG
CAACACGCGGAGAGATACTTCCCTTTAGACAACAATTCTGACAGCAGCAAAGGTTTGCTT
CTAAACTTCGAAGACCGAACAGGAAACTCATGGAGATTCCGTTACTCTTACTGGAACAGT
AGCCAGAGTTATGTCATGACAAAAGGTTGGAGCCGCTTCGTCAAAGACAAGAAGCTTGAT
GCTGGCGACATCGTTTCTTTTCAGAGAGATCCTGGTAATAAAGACAAGCTTTTCATTGAT
TGGAGGAGACGACCAAAGATTCCAGATCATCATCATCAATTCGCTGGAGCTATGTTCCCT
AGGTTTTACTCTTTCTCTCATCCTCAGAACCTTTATCATCGATATCAACAAGATCTTGGA
ATTGGGTATTATGTGAGTTCAATGGAGAGAAATGATCCAACGGCTGTAATTGAATCTGTG
CCGTTGATAATGCAAAGGAGAGCAGCACACGTGGCTGCTATACCTTCATCAAGAGGAGAG
AAGAGGTTAAGGCTGTTTGGAGTGGACATGGAGTGCGGCGGCGGCGGAGGAAGTGTGAAT
AGCACGGAGGAAGAGTCGTCGTCTTCCGGTGGTGGCGGCGGCGTTTCTATGGCTAGTGTT
GGTTCTCTTCTCCAATTGAGGCTAGTGAGCAGTGATGATGAGTCTTTGGTAGCAATGGAA
GCTG CAAGTGTCGATGA GGATCATCACTTGTTTACAAAGAAAG GAAAGTCTTCTTTGTCT
TTCGATTTGGATAGAAAATGA

Bra007646

Cover 74% identity 45%

SEQ ID NO: 117

MNQEN KKPLEEASTSMEREN MFDKVVTPSDVGKLNRLVIPKQHAERYFPLDNSSTNNKGL
LLDFEDRTGSSWRFRYSYWNSSQSYVMTKGWSRFVKDKKLDAGDIVSFQRDPCNKDKLYI
DWRRRPKI PDHHQFAGAMFPRFYSFPHPQM PTSFESSHNLYHHRFQRDLGIGYYPTAVIE
SVPVI MQRREAQVAN MASSRGEKRLRLFGVDVECGGGGGSVNSTEEESSSSGGSMSRGG
VSMAGVGSLLQLRLVSSDDESLVAMEGATVDEDHH LFTTKKGKSSLSFDLDI

CDS SEQ ID NO: 118
ATGAATCAAGAAAACAAGAAGCCTTTGGAAGAAGCTTCGACTTCAATGGAGAGAGAGAAC
ATGTTCGACAAAGTAGTAACACCAAGCGACGTAGGGAAACTAAACCGACTCGTGATCCCA
AAGCAACACGCAGAGAGATACTTCCCTTTAGACAACTCCTCAACAAACAACAAAGGGTTG
CTTCTAGACTTCGAAGACCGTACAGGAAGCTCATGGAGATTCCGTTACTCTTACTGGAAC
AGTAGCCAAAGTTATGTCATGACAAAAGGTTGGAGCCGTTTTGTCAAAGACAAGAAGCTT
GATGCTGGTGACATCGTGTCTTTTCAAAGAGATCCCTGTAATAAAGACAAGCTTTACATA
GATTGGAGGAGACGACCAAAGATTCCAGATCATCATCAGTTCGCCGGAGCTATGTTCCCT
AGGTTTTACTCTTTCCCTCACCCTCAGATGCCGACAAGTTTTGAAAGTAGTCACAACCTT
TATCATCATCGGTTTCAACGAGATCTTGGAATTGGGTATTATCCAACGGCTGTGATTGAA
TCTGTGCCGGTGATAATGCAAAGGAGAGAAGCACAAGTGGCTAATATGGCTTCATCAAGA
GGAGAGAAGAGGTTAAGGCTGTTTGGAGTGGACGTGGAGTGCGGCGGCGGAGGAGGAGGA
AGTGTGAATAGCACGGAGGAAGAGTCGTCGTCTTCCGGTGGTAGTATGTCACGTGGCGGC
GTTTCTATGGCTGGTGTTGGTTCTCTCCTTCAGTTGAGGTTAGTGAGCAGTGATGATGAG
TCTTTAGTAGCGATGGAAGGTGCTACTGTCGATGAGGATCATCACTTGTTTACAACTAAG
AAAGGAAAGTCTTCTTTGTCTTTCGATTTGGATATATGA

Bra014415
Cover 48% identity 60%
SEQ ID NO: 119
MERKSNDLERSENI DSQNKKMNLEEERPVQEASSMEREHM FDKVVTPSDVGKLNRLVI PK
QHAERYFPLDNNSSDNN KGLLLNFEDRIGI LWSFRYSYWNSSQSYVMTKGWSRFVKDKKL
DAGDIVSFHRGSCNKDKLFI DWKRRPKI PDHQVVGAM FPRFYSYPYPQIQASYERHNLYH
RYQRDIGIGYYVRSM ERYDPTAVIESVPVI MQRRAHVATMASSRGEKRLRLFGVDM ECVR
GGRGGGGSVNSTEEESSTSGGSISRGGVSMAGVGSPLQLRLVSSDGDDQSLVARGAARVD
EDHH LFTKKGKSSLSFDLDK

CDS SEQ ID NO: 120
ATGGAGAGGAAGTCCAATGATCTTGAGAGATCTGAGAATATTGATTCTCAAAACAAGAAG
ATGAATCTAGAAGAAGAGAGGCCTGTACAAGAAGCTTCTTCGATGGAGAGAGAGCACATG
TTCGACAAAGTAGTAACACCAAGCGACGTTGGGAAACTAAACCGGCTGGTGATCCCAAAG
CAACACGCAGAGCGATACTTCCCTTTAGACAATAATTCCTCAGACAACAACAAAGGTTTG
CTTCTAAACTTCGAAGATCGAATAGGAATCTTATGGAGTTTCCGTTACTCCTACTGGAAC
AGTAGCCAAAGTTATGTAATGACTAAAGGCTGGAGCCGTTTCGTCAAAGACAAGAAGCTT
GATGCTGGCGACATAGTTTCTTTTCATAGAGGTTCTTGTAATAAAGACAAGCTTTTCATT
GATTGGAAGAGACGACCAAAGATTCCTGATCACCAAGTCGTCGGAGCTATGTTCCCTAGG
TTTTACTCTTACCCTTATCCTCAGATACAGGCTAGTTATGAACGTCACAACCTTTATCAT
CGATATCAACGAGATATAGGAATTGGGTATTATGTGAGGTCAATGGAGAGATATGATCCA
ACGGCTGTAATTGAATCTGTGCCGGTGATAATGCAAAGGAGAGCACATGTGGCTACTATG
GCTTCATCAAGAGGAGAGAAGAGGTTAAGGCTTTTTGGAGTGGATATGGAGTGCGTCAGA
GGCGGCCGAGGAGGAGGAGGAAGTGTGAATAGCACGGAGGAAGAGTCTTCGACTTCCGGT
GGTAGTATCTCACGTGGCGGCGTTTCTATGGCTGGTGTTGGCTCTCCACTCCAGTTGAGG
TTAGTGAGCAGTGACGGTGATGATCAGTCTCTAGTAGCTAGGGGAGCTGCTAGGGTTGAT
GAGGATCATCACTTGTTTACAAAGAAAGGAAAGTCTTCTTTGTCTTTCGATTTGGATAAA
TGA

Bra038346
Cover 51% identity 57%
SEQ ID NO: 121
MVFSCI DESSSTSESFSPATATATATATKFSAPPLPPLRLNRMRSGGSNVVLDSKNGVDI

DSRKLSSSKYKGVVPQPNGRWGAQIYVKHQRVWLGTFCDEEEAAHSYDIAARKFRGRDAV
VNFKTFLASEDDNGELCFLEAHSKAEIVDM LRKHTYADELAQSNKRSGANTNTNTTQSHT
VSRTREVLFEKVVTPSDVGKLNRLVI PKQHAEKYFPLPSLSVTKGVLI N FEDVTGKVWRF
RYSYWNSSQSYVLTKGWSRFVKEKNLRAGDVVTFERSTGSDRQLYI DWKI RSGPSKNPVQ
VVVRLFGVDI FNVTSAKPSNVVDACGGKRSRDVDM FALRCSKKHAII NAL

CDS SEQ ID NO: 122
ATGGTATTCAGTTGCATAGACGAGAGCTCTTCCACTTCAGAATCTTTTTCACCCGCAACC
GCAACCGCAACCGCAACCGCCACAAAGTTCTCTGCTCCTCCGCTTCCACCGTTACGCCTC
AACCGGATGAGAAGCGGTGGAAGCAACGTCGTGTTGGATTCAAAGAATGGCGTAGATATT
GATTCACGGAAGCTATCGTCGTCAAAGTACAAAGGCGTGGTTCCTCAGCCCAACGGAAGA
TGGGGAGCTCAGATTTACGTGAAGCACCAGCGAGTTTGGCTGGGCACTTTCTGCGATGAA
GAGGAAGCTGCTCACTCCTACGACATAGCCGCCCGTAAATTCCGTGGCCGTGACGCCGTT
GTCAACTTCAAAACCTTCCTCGCCTCAGAGGACGACAACGGCGAGTTATGTTTCCTTGAA
GCTCACTCCAAGGCCGAGATCGTCGACATGTTGAGGAAACACACTTACGCTGACGAGCTT
GCGCAGAGCAATAAACGCAGCGGAGCGAATACGAATACGAATACGACTCAAAGCCACACC
GTTTCGAGAACACGTGAAGTGCTTTTCGAGAAGGTTGTCACGCCTAGCGACGTTGGTAAG
CTAAACCGCCTCGTGATACCTAAACAGCACGCGGAGAAATATTTTCCGTTACCGTCACTG
TCGGTGACTAAAGGCGTTCTGATCAACTTCGAAGACGTGACGGGTAAGGTGTGGCGGTTC
CGTTACTCATACTGGAACAGTAGTCAAAGTTACGTGTTGACCAAGGGATGGAGTCGGTTC
GTTAAGGAGAAGAATCTCCGAGCCGGTGATGTCGTTACTTTCGAGAGATCGACCGGTTCA
GACCGGCAGCTTTATATTGATTGGAAAATCCGGTCTGGTCCGAGCAAAAACCCTGTTCAG
GTTGTG GTTAGG CTTTTCG GAGTTG ACATCTTCAACGTG ACAAGCG CGAAG CCGA G CAAC
GTTGTAGACGCGTGCGGTGGAAAGAGATCTCGGGATGTTGATATGTTTGCGCTACGGTGT
TCCAAAAAAACACGCTATAATCAATGCTTTGTGA

Zea mays
GRMZM2G053008
Cover 74% identity 47%

SEQ ID NO: 123
MAASPSSPLTAPPEPVTPPSPWTITDGAISGTLPAAEAFAVHYPGYPSSPARAARTLGGL
PGLAKVRSSDPGARLELRFRPEDPYCHPAFGQSRASTGLLLRLSKRKGAAAPCAHVVARV
RTAYYFEGMADFQHVVPVHAAQTRKRKHSDSQNDNENFGSDKTGHDEADGDVM MLVPPLF
SVKDRPTKIALVPSSNAISKTMH RGVVQERWEMNVGPTLALPFNTQVVPEKINWEDHI RK
NSVEWGWQMAVCKLFDERPVWPRQSLYERFLDDNVHVSQNQFKRLLFRAGYYFSTGPFGK
FWI RRGYDPRKDSESQIYQRI DFRM PPELRYLLRLKNSESRKWADMCKLETMPSQSFIYL
QLYELKDDFIQAEI RKPSYQSVCSRSTGWFSKPM IKTLRLQVSI RLLSLLHNEEAKNLLR
NAHELI ERSKKQEALSRSELSI EYNDADQVSAAHTGTEDQVGPNNSDSEDVDDEEEEEEL
EGYDSPPMADDI HEFTLGDSYAFGEGFSNGYLEEVLRSLPLQEDGQKKLCDAPI NADASD

CDS SEQ ID NO: 124
ATGGCCGCCTCGCCCTCTTCACCCTTGACAGCGCCGCCAGAGCCGGTGACCCCGCCGTCC
CCATGGACCATCACAGACGGAGCCATCTCTGGCACGCTCCCAGCAGCCGAGGCCTTCGCA
GTGCACTACCCGGGCTACCCCTCCTCTCCCGCCCGCGCCGCCCGCACCCTCGGCGGTCTC
CCCGGCCTCGCCAAGGTCCGGAGTTCCGATCCCGGCGCCCGCCTCGAGCTCCGCTTCCGC
CCCGAGGACCCCTACTGCCATCCAGCCTTTGGCCAGTCCCGCGCCTCCACTGGCCTTCTG
CTGCGCCTCTCCAAGCGCAAAGGAGCTGCGGCACCTTGTGCCCATGTGGTCGCTCGTGTC
CGGACTGCTTACTACTTCGAAGGTATGGCAGATTTTCAACATGTTGTTCCAGTGCATGCT
GCACAAACAAGAAAAAGAAAACACTCAGATTCTCAAAATGATAATGAGAATTTTGGTAGT
GATAAGACAGGACATGATGAAGCAGATGGAGATGTCATGATGTTGGTACCCCCTCTCTTT
TCAGTG AAG GATAGG CCAACAAAG ATAGCG CTTGTACCATCGTCCAATG CCATATCTAAA

ACCATGCACAGGGGAGTTGTACAAGAACGGTGGGAGATGAATGTTGGACCAACTCTGGCG
CTTCCGTTCAACACTCAAGTTGTCCCGGAGAAGATTAATTGGGAAGACCACATTAGAAAG
AATTCTGTAGAATGGGGTTGGCAAATGGCTGTTTGCAAATTGTTTGATGAGCGCCCTGTG
TGGCCAAGGCAATCACTTTATGAGCGGTTCCTTGATGATAATGTGCATGTCTCTCAAAAC
CAATTCAAAAGGCTTCTGTTTAGAGCTGGATACTACTTCTCTACTGGACCCTTTGGAAAA
TTTTGGATCAGAAGAGGATATGACCCTCGTAAAGACTCTGAGTCACAAATATATCAGAGA
ATTGATTTTCGCATGCCTCCCGAGCTACGATATCTTCTAAGGCTGAAGAATTCTGAGTCT
CGAAAGTGGGCAGATATGTGCAAGCTTGAAACAATGCCATCACAGAGTTTCATCTACCTG
CAATTATATGAACTGAAGGATGATTTTATTCAAGCAGAAATTCGAAAACCTTCTTATCAA
TCAGTTTGTTCACGTTCTACAGGATGGTTTTCTAAGCCAATGATCAAAACCCTGAGGTTG
CAAGTGAGCATAAGGCTCCTCTCTTTATTGCATAATGAAGAGGCTAAAAACTTGTTGAGG
AATGCCCATGAGCTTATTGAAAGGTCCAAGAAGCAGGAAGCCCTTTCGAGATCTGAGCTG
TCAATAGAATATAATGATGCTGATCAAGTTTCTGCCGCACATACTGGAACTGAGGATCAA
GTCGGCCCTAACAACTCTGATAGTGAAGATGTGGATGATGAAGAAGAGGAAGAGGAATTG
GAGGGTTATGATTCTCCACCTATGGCAGATGATATTCATGAGTTCACCTTAGGTGATTCC
TATGCATTTGGTGAAGGCTTCTCGAATGGATACCTCGAAGAAGTACTGCGCAGCTTGCCA
TTGCAGGAAGACGGCCAAAAGAAATTATGTGATGCTCCTATCAACGCTGATGCAAGTGAT
GGAGAGTTTGAAATTTACGAACAGCCCAGTGATGATGAAGATTCTGATGGCTAG

GRMZM2G102059_T01
Cover 47% identity 62%
SEQ ID NO: 125
M EFASSSSRFSREEDEEEEQEEEEEEEEASPREI PFMTAAATADTGAAASSSSPSAAASS
GPAAAPRSSDGAGASGSGGGGSDDVQVI EKEHM FDKVVTPSDVGKLNRLVIPKQHAEKYF
PLDAAAN EKGQLLSFEDRAGKLWRFRYSYWNSSQSYVMTKGWSRFVKEKRLDAGDTVSFC
RGAGDTARDRLFI DWKRRADSRDPH RM PRLPLPMAPVASPYGPWGGGGGGAGGFFM PPA
PPATLYEHHRFRQALDFRNI NAAAAPARQLLFFGSAGM PPRASMPQQQQPPPPPHPPLHS
IM LVQPSPAPPTASVPM LLDSVPLVNSPTAASKRVRLFGVNLDNPQPGTSAESSQDANAL
SLRTPGWQRPGPLRFFESPQRGAESSAASSPSSSSSSKREAHSSLDLDL

CDS SEQ ID NO: 126
ATGGAGTTCGCGAGCTCTTCGAGTAGGTTTTCCAGGGAGGAGGACGAGGAGGAAGAGCAG
GAGGAAGAGGAGGAGGAGGAGGAGGCGTCTCCGCGCGAGATCCCCTTCATGACAGCGGCA
GCGACGGCCGACACCGGAGCCGCCGCCTCCTCGTCCTCGCCTTCCGCGGCGGCCTCATCG
GGTCCTGCTGCTGCCCCCCGCTCGAGCGACGGCGCCGGGGCGTCCGGGAGCGGCGGCGGC
GGGAGCGACGACGTGCAGGTGATCGAGAAGGAGCACATGTTCGACAAGGTGGTGACGCCC
AGCGACGTGGGGAAGCTCAACCGGCTGGTGATCCCGAAGCAGCACGCGGAGAAGTACTTC
CCGCTGGACGCGGCGGCCAACGAGAAGGGCCAGCTGCTCAGCTTCGAGGACCGCGCCGGT
AAGCTCTGGCGCTTCCGCTACTCCTACTGGAACAGCAGCCAGAGCTACGTCATGACCAAG
GGCTGGAGCCGCTTCGTCAAGGAGAAGCGCCTCGACGCCGGCGACACCGTCTCCTTCTGC
CGCGGCGCCGGCGACACCGCGCGGGACCGCCTCTTCATCGACTGGAAGCGCCGCGCCGAC
TCCCGCGACCCGCACCGCATGCCGCGCCTCCCGCTCCCCATGGCGCCCGTCGCGTCGCCC
TACGGCCCCTGGGGCGGCGGCGGCGGCGGCGGCGCGGGCGGTTTCTTCATGCCGCCCGCG
CCGCCCGCCACACTCTACGAGCACCACCGCTTCCGCCAGGCCCTCGACTTCCGCAACATC
AACGCCGCGGCCGCGCCGGCCAGGCAGCTCCTCTTCTTCGGCTCAGCCGGCATGCCCCCG
CGCGCGTCCATGCCGCAGCAGCAGCAGCCGCCTCCGCCCCCGCACCCGCCTCTGCACAGC
ATTATGTTGGTGCAACCCAGCCCCGCGCCGCCCACGGCCAGCGTGCCCATGCTTCTCGAC
TCGGTACCGCTCGTCAACAGCCCAACGGCAGCGTCGAAGCGCGTCCGCCTGTTTGGGGTC
AACCTCGACAACCCGCAACCAGGCACAAGTGCGGAGTCAAGCCAAGATGCCAACGCATTG
TCGCTGAGGACACCGGGATGGCAAAGGCCGGGGCCGTTGAGGTTCTTCGAATCGCCTCAA
CGCGGCGCCGAGTCATCTGCAGCCTCCTCGCCGTCGTCATCGTCGTCCTCCAAGAGAGAA

GCGCACTCGTCCTTGGATCTCGATCTGTGA

GRMZM2G098443_T01
Cover 47% identity 63%
SEQ ID NO: 127
M EFTTPPPATRSGGGEERAAAEHNQHHQQQHATVEKEHM FDKVVTPSDVGKLNRLVIPKQ
HAEKYFPLDAAANEKGLLLSFEDRTGKPWRFRYSYWNSSQSYVMTKGWSRFVKEKRLDAG
DTVSFGRGISEAARDRLFI DWRCRPDPPVVHHQYHH RLPLPSAVVPYAPWAAHAHHHHYP
ADGHTEPVTPCLCATLVATEM RASSSQLSLTRSNLSRPPQPRIARVDGAQPRPSSSPRQP
QSLWCRSCQPQPRRTADVP

CDS SEQ ID NO: 128
ATGGAGTTCACCACTCCCCCGCCCGCGACCCGGTCGGGCGGCGGAGAGGAGAGGGCGGCT
GCTGAGCACAACCAGCACCACCAGCAGCAGCATGCGACGGTGGAGAAGGAGCACATGTTC
GACAAGGTGGTGACGCCGAGCGACGTCGGGAAGCTGAACCGGCTGGTGATCCCGAAGCAG
CACGCGGAGAAGTACTTCCCGCTGGACGCGGCGGCGAACGAGAAGGGCCTCCTGCTCAGC
TTCGAGGACCGCACGGGGAAGCCCTGGCGCTTCCGCTACTCCTACTGGAACAGTAGCCAG
AGCTACGTGATGACCAAGGGCTGGAGCCGCTTCGTCAAGGAGAAGCGCCTCGACGCCGGG
GACACAGTCTCCTTCGGCCGCGGCATCAGCGAGGCGGCGCGCGACAGGCTTTTCATCGAC
TGGCGGTGCCGACCCGACCCGCCCGTCGTGCACCACCAGTACCACCACCGCCTCCCTCTC
CCCTCCGCCGTCGTCCCCTACGCGCCGTGGGCGGCGCACGCGCACCACCACCACTACCCA
GCAGATGGGCACACGGAACCAGTAACACCTTGCCTGTGCGCCACACTCGTTGCCACTGAA
ATGAGAGCATCATCTTCGCAACTGTCACTCACACGCTCCAACCTCTCCAGGCCGCCACAA
CCTAGAATAGCCAGAGTCGATGGCGCCCAGCCACGGCCGTCGTCGTCACCACGCCAGCCA
CAGTCGTTGTGGTGCCGGTCGTGCCAACCGCAACCACGGCGAACGGCCGACGTTCCTTGA

GRMZM2G082227_T01
Cover 45% identity 64%

SEQ ID NO: 129
M EFTAPPPATRSGGGEERAAAEHHQQQQQATVEKEHM FDKVVTPSDVGKLNRLVIPKQHA
ERYFPLDAAANDKGLLLSFEDRAGKPWRFRYSYWNSSQSYVMTKGWSRFVKEKRLDAGDT
VSFGRGVGEAARGRLFI DWRRRPDPPVVHHQYHH HRLPLPSAVVPYAPWAAAAHAHHHHY
PAAGVGAARTTTTTTTTVLH HLPPSPSPLYLDTRRRHVGYDAYGAGTRQLLFYRPHQQPS
TTVM LDSVPVRLPPTPGQHAEPPPPAVASSASKRVRLFGVNLDCAAAAGSEEENVGGWRT
SAPPTQQASSSSSYSSGKARCSLNLDL

CDS SEQ ID NO: 130
ATGGAGTTCACCGCTCCCCCGCCCGCGACCCGGTCGGGCGGCGGCGAGGAGAGGGCGGCT
GCTGAGCACCACCAGCAGCAGCAGCAGGCGACGGTGGAGAAGGAGCACATGTTCGACAAG
GTGGTGACGCCGAGCGACGTCGGGAAGCTGAACCGGCTGGTGATCCCGAAGCAGCACGCG
GAGAGGTACTTCCCGCTGGACGCGGCGGCGAACGACAAGGGCCTGCTGCTCAGCTTCGAG
GACCGCGCGGGGAAGCCCTGGCGCTTCCGCTACTCCTACTGGAACAGCAGCCAGAGCTAC
GTGATGACCAAGGGCTGGAGCCGCTTCGTCAAGGAGAAGCGCCTCGACGCCGGGGACACC
GTCTCCTTCGGCCGCGGCGTCGGCGAGGCGGCGCGCGGCAGGCTCTTCATCGACTGGCGG
CGCCGACCCGACCCGCCCGTCGTGCACCACCAGTACCACCACCACCGCCTCCCTCTCCCC
TCCGCCGTCGTCCCCTACGCGCCGTGGGCGGCGGCGGCGCACGCGCACCACCACCACTAC
CCAGCAGCTGGGGTCGGTGCCGCCAGGACGACGACGACGACGACGACGACGGTGCTCCAC
CACCTGCCGCCCTCGCCCTCCCCGCTCTACCTTGACACCCGCCGCCGCCACGTCGGCTAC
GACGCCTACGGGGCCGGCACCAGGCAACTTCTCTTCTACAGGCCGCACCAGCAGCCCTCC
ACGACGGTGATGCTGGACTCCGTGCCGGTACGGTTACCGCCAACGCCAGGGCAGCACGCC
GAGCCGCCGCCCCCCGCCGTGGCGTCGTCAGCCTCGAAGCGGGTGCGCCTGTTCGGGGTG

AACCTCGACTGCGCCGCCGCCGCCGGCTCAGAGGAGGAGAACGTCGGCGGGTGGAGGACT
AGTGCGCCGCCGACGCAGCAGGCGTCCTCCTCCTCATCCTACTCTTCCGGGAAAGCGAGG
TGCTCCTTGAACCTTGACTTGTGA

GRMZM2G024948_T01
Cover 46% identity 63%
SEQ ID NO: 131
M DQFAASG RFSREEEADEEQEDASNSMREISFMP PAAASSSSAAASASASASTSASACAS
GSSSAPFRSASASGDAAGASGSGGPADADAEAEAVEKEHMFDKVVTPSDVGKLNRLVI PK
QYAEKYFPLDAAANEKGLLLSFEDSAGKHWRFRYSYWNSSQSYVMTKGWSRFVKEKRLVA
GDTVSFSRAAAEDARHRLFI DWKRRVDTRGPLRFSGLALPM PLPSSHYGGPHHYSPWGFG
GGGGGGGGFFMPPSPPATLYEH RLRQGLDFRSMTTTYPAPTVGRQLLFFGSARM PPHHAP
PPQPRPFSLPLH HYTVQPSAAGVTAASRPVLLDSVPVI ESPTTAAKRVRLFGVNLDNN PD
GGGEASHQGDALSLQMPGWQQRTPTLRLLELPRHGGESSAASSPSSSSSSKREARSALDL
DL

CDS SEQ ID NO: 132
ATGGACCAGTTCGCCGCGAGCGGGAGGTTCTCTAGAGAGGAGGAGGCGGACGAGGAGCAG
GAGGATGCGTCCAATTCCATGCGCGAGATCTCCTTCATGCCGCCGGCTGCGGCCTCGTCA
TCTTCGGCGGCTGCTTCCGCGTCCGCGTCCGCCTCCACCAGCGCATCCGCGTGTGCATCG
GGAAGCAGCAGCGCCCCCTTCCGCTCCGCCTCCGCGTCGGGGGATGCCGCCGGAGCGTCG
GGGAGCGGCGGCCCAGCGGACGCGGACGCGGAGGCGGAGGCGGTGGAGAAGGAGCACATG
TTCGACAAGGTGGTCACGCCGAGCGACGTGGGGAAGCTCAACCGGCTGGTGATCCCGAAG
CAGTACGCGGAGAAGTACTTCCCGCTGGACGCGGCGGCCAACGAGAAGGGCCTCCTCCTC
AGCTTCGAGGACAGCGCCGGCAAGCACTGGCGCTTCCGCTACTCCTACTGGAACAGCAGC
CAGAGCTACGTCATGACCAAGGGCTGGAGCCGCTTCGTCAAGGAGAAGCGCCTCGTCGCC
GGGGACACCGTCTCCTTCTCCCGCGCCGCCGCCGAGGACGCGCGCCACCGCCTCTTCATC
GACTGGAAGCGCCGGGTCGACACCCGCGGCCCGCTTCGTTTCTCCGGCCTCGCGCTGCCG
ATGCCGCTGCCGTCGTCGCACTACGGCGGGCCCCACCACTACAGCCCGTGGGGCTTCGGC
GGCGGCGGCGGCGGCGGCGGCGGATTCTTCATGCCGCCCTCGCCGCCCGCCACGCTCTAC
GAGCACCGCCTCAGACAGGGCCTCGACTTCCGCAGCATGACGACGACCTACCCCGCGCCG
ACCGTGGGGAGGCAGCTCCTGTTTTTCGGCTCGGCCAGGATGCCTCCTCATCACGCGCCG
CCGCCCCAGCCGCGCCCGTTCTCGCTGCCGCTGCATCACTACACGGTGCAACCGAGCGCC
GCCGGCGTCACCGCCGCGTCACGGCCGGTCCTTCTTGACTCGGTGCCGGTCATCGAGAGC
CCGACGACCGCCGCGAAGCGCGTGCGGCTGTTCGGCGTCAACCTGGACAACAACCCAGAT
GGCGGCGGCGAGGCTAGCCATCAGGGCGATGCATTGTCATTGCAGATGCCCGGGTGGCAG
CAAAGGACTCCAACTCTAAGGCTACTAGAATTGCCTCGCCATGGCGGGGAGTCCTCCGCG
GCGTCGTCTCCGTCGTCGTCGTCTTCCTCCAAGAGGGAGGCGCGTTCAGCTTTGGATCTC
GATCTGTGA

GRMZM2G328742_T01
Cover 55% identity 64%
SEQ ID NO: 134
MATN HLSQGQHQHPQAWPWGVAMYTNLHYH HQQH HHYEKEH LFEKPLTPSDVGKLNRLVI
PKQHAERYFPLSSSGAGDKGLI LCFEDDDDDEAAAANKPWRFRYSYWTSSQSYVLTKGWS
RYVKEKQLDAGDVVRFQRM RGFGM PDRLFISHSRRGETTATAATTVPPAAAAVRVVVAPA
QSAGADHQQQQQPSPWSPMCYSTSGSYSYPTSSPANSQHAYHRHSADH DHSNNMQHAGES
QSDRDNRSCSAASAPPPPSRRLRLFGVN LDCGPGPEPETPTAMYGYMHQSPYAYNNWGSP
YQHDEEI

CDS 135

ATGGCCACGAACCATCTCTCCCAAGGGCAGCACCAGCACCCGCAGGCCTGGCCCTGGGGC
GTGGCCATGTACACCAACCTACACTACCACCACCAGCAGCACCACCACTACGAGAAGGAG
CACCTGTTCGAGAAGCCGCTGACGCCGAGCGACGTGGGCAAGCTCAACAGGCTGGTGATC
CCCAAGCAGCACGCCGAGAGGTACTTCCCTCTCAGCAGCAGCGGCGCCGGCGACAAAGGC
CTCATCCTGTGCTTCGAGGACGACGACGACGACGAGGCTGCCGCCGCCAACAAGCCGTGG
CGGTTCCGCTACTCGTACTGGACCAGCAGCCAGAGCTACGTGCTCACCAAGGGCTGGAGC
CGCTACGTCAAGGAGAAGCAGCTTGACGCCGGCGACGTCGTGCGCTTCCAGAGGATGCGT
GGTTTCGGCATGCCCGACCGCCTGTTCATCAGCCACAGCCGCCGCGGCGAGACTACTGCT
ACTGCTGCAACAACAGTGCCCCCCGCTGCTGCTGCCGTGCGCGTAGTAGTGGCACCTGCA
CAGAGCGCTGGCGCAGACCACCAGCAGCAGCAGCAGCCGTCGCCTTGGAGCCCAATGTGC
TACAGCACATCAGGCTCGTACTCGTACCCCACCAGCAGCCCAGCCAATTCCCAGCATGCC
TACCACCGCCACTCAGCTGACCATGACCACAGCAACAACATGCAACATGCAGGAGAATCT
CAGTCCGACAGAGACAACAGGAGCTGCAGTGCAGCTTCGGCACCGCCGCCACCGTCGCGG
CGGCTCCGGCTGTTCGGCGTAAACCTCGACTGCGGCCCGGGGCCGGAGCCGGAGACACCA
ACGGCGATGTACGGCTACATGCACCAAAGCCCCTACGCTTACAACAACTGGGGCAGTCCA
TACCAGCATGACGAGGAGATTTAA

GRMZM2G142999_T01
Cover 44% identity 64%

SEQ ID NO: 136
MEFTPAHAHARVVEDSERPRGGVAWVEKEHM FEKVVTPSDVGKLNRLVIPKQHAERYFPA
LDASSAAAAAAAAAAGGGKGLVLSFEDRAGKAWRFRYSYWNSSQSYVMTKGWSRFVKEKR
LGAGDTVLFARGAGGARGRFFI DFRRRRQDLAFLQPTLASAQRLLPLPSVPICPWQDYGA
SAPAPNRHVLFLRPQVPAAVVLKSVPVHVAASAVEATMSKRVRLFGVNLDCPPDAEDSAT
VPRGRAASTTLLQLPSPSSSTSSSTAGKDVCCLDLGL

CDS SEQ ID NO: 137
ATGGAGTTCACGCCCGCGCATGCGCATGCCCGTGTCGTTGAGGATTCCGAGAGGCCTCGC
GGCGGCGTGGCCTGGGTGGAGAAGGAGCACATGTTCGAGAAGGTGGTCACCCCGAGCGAC
GTGGGGAAGCTCAATCGCCTGGTCATCCCAAAGCAGCACGCGGAGCGCTACTTCCCCGCG
CTGGACGCCTCGTCCGCCGCGGCGGCGGCGGCGGCAGCAGCCGCGGGAGGCGGGAAGGGG
CTGGTGCTCAGCTTCGAGGACCGGGCGGGGAAGGCGTGGCGCTTCCGCTACTCGTACTGG
AACAGCAGCCAGAGCTACGTGATGACCAAAGGTTGGAGCCGCTTCGTGAAGGAGAAGCGC
CTCGGTGCCGGGGACACAGTCTTGTTCGCGCGCGGCGCGGGCGGCGCGCGGCCGCTTC
TTCATCGATTTCCGCCGCCGTCGCCAGGATCTCGCGTTCCTGCAGCCGACGCTGGCGTCT
GCGCAGCGACTCCTGCCGCTGCCGTCGGTGCCCATCTGCCCGTGGCAGGACTACGGCGCC
TCGGCTCCGGCGCCCAACCGGCACGTGCTGTTCCTGCGGCCGCAGGTGCCGGCCGCCGTA
GTGCTCAAGTCGGTCCCCGTGCACGTTGCTGCATCCGCGGTGGAGGCGACCATGTCGAAG
CGCGTCCGCCTGTTCGGGGTGAACCTCGACTGCCCGCCGGACGCCGAAGACAGCGCCACA
GTCCCCCGGGGCCGGGCGGCGTCGACGACGCTTCTGCAACTGCCCTCGCCATCGTCGTCA
ACATCCTCCTCGACGGCAGGGAAGGACGTGTGCTGTTTGGATCTTGGACTGTGA

GRMZM2G125095_T01
Cover 85% identity 40%
SEQ ID NO: 138
M EFRPAHARVFEDSERPRGGVAWLEKEHMFEKVVTPSDVGKLNRLVI PKQHAERYFPALD
ASAAAASASASAGGGKAGLVLSFEDRAGKAWRFRYSYWNSSQSYVMTKGWSRFVKEKRLG
AGDTVLFARGAGATRGRFFI DFRRRRH ELAFLQPPLASAQRLLPLPSVPICPWQGYGASA
PAPSRHVLFLRPQVPAAVVLTSVPVRVAASAVEEATRSKRVRLFGVNLDCPPDAEDGATA

TRTPSTLLQLPSPSSSTSSSTGGKDVRSLDLGL

CDS SEQ ID NO: 139
ATGGAGTTCAGGCCCGCGCATGCCCGTGTCTTCGAGGATTCCGAGAGGCCTCGCGGCGGC
GTGGCGTGGCTGGAGAAGGAGCACATGTTCGAGAAAGTGGTCACCCCGAGCGACGTGGGG
AAGCTCAATCGCCTGGTCATCCCGAAGCAGCACGCCGAGCGCTACTTCCCCGCGCTGGAC
GCCTCGGCCGCCGCGGCGTCGGCATCGGCGTCGGCGGGCGGCGGGAAGGCGGGGCTGGTG
CTCAGCTTCGAGGACCGGGCGGGGAAGGCGTGGCGCTTCCGCTACTCGTACTGGAACAGC
AGCCAGAGCTACGTGATGACCAAGGGATGGAGCCGCTTCGTGAAAGAGAAGCGCCTCGGT
GCCGGGGACACGGTATTGTTCGCGCGCGGCGCGGGCGCCACGCGCGGCCGCTTCTTCATC
GATTTCCGCCGCCGCCGCCACGAGCTCGCGTTCCTGCAGCCGCCGCTGGCGTCTGCGCAG
CGCCTCCTGCCGCTCCCGTCGGTGCCCATCTGCCCGTGGCAGGGCTACGGCGCCTCCGCT
CCGGCGCCAAGCCGGCACGTGCTGTTCCTGCGGCCGCAGGTGCCGGCCGCCGTAGTGCTC
ACGTCGGTGCCCGTGCGCGTCGCCGCATCCGCGGTGGAGGAGGCGACGAGGTCGAAGCGC
GTCCGCCTGTTCGGGGTGAACCTCGACTGCCCGCCGGACGCCGAAGACGGTGCCACAGCC
ACCCGGACGCCGTCGACGCTTCTGCAGCTGCCCTCGCCATCGTCGTCAACATCCTCCTCC
ACGGGAGGCAAGGATGTGCGTTCTTTGGATCTTGGACTTTGA

Tricum aeseirum

TRAES3BF098300010CFD_tl
Cover: 42% ident 60%

SEQ ID NO: 140
MGVEI LSSMVEHSFQYSSGVSTATTESGTAGTPPRPLSLPVAIADESVTSRSASSRFKGVVPQPNGRWGAQIYERH
ARVWLGTFPDQDSAARAYDVASLRYRGRDVAFNFPCAAVEGELAFLAAHSKAEIVDMLRKQTYADELRQGLRRG
RGMGARAQPTPSWAREPLFEKAVTPSDVGKLNRLVVPKQHAEKH FPLKRTPETPTTTGKGVLLNFEDGEGKVWR
FRYSYWNSSQSYVLTKGWSRFVREKGLGAGDSILFSCSLYEQEKQFFIDCKKNTSMNGGKSASPLPVGVTTKGEQV
RVVRLFGVDISGVKRGRAATATAEQGLQELFKRQCVAPGQHSPALGAFAL

CDS SEQ ID NO: 141
ATGGGGGTGGAAATCCTGAGCTCCATGGTGGAGCACTCCTTCCAGTACTCTTCCGGCGTG
TCCACGGCCACGACGGAGTCAGGCACCGCCGGAACACCGCCGAGGCCTTTGAGCCTACCT
GTCGCCATCGCCGACGAGTCCGTGACCTCGCGGTCGGCGTCGTCTCGGTTCAAGGGCGTG
GTGCCGCAGCCAAACGGGCGATGGGGCGCCCAGATCTACGAGCGCCACGCTCGCGTCTGG
CTCGGCACGTTCCCAGACCAGGACTCGGCGGCGCGCGCCTACGACGTAGCCTCGCTCAGG
TACCGCGGCCGCGACGTCGCCTTCAACTTCCCGTGCGCGGCCGTGGAGGGGGAGCTCGCC
TTCCTGGCGGCGCACTCCAAGGCTGAGATAGTGGACATGCTCCGGAAGCAGACCTACGCC
GATGAACTCCGCCAGGGCCTGCGGCGCGGCCGTGGCATGGGGGCGCGCGCGCAGCCGACG
CCGTCGTGGGCGCGGGAGCCCCTTTTCGAGAAGGCCGTGACCCCTAGCGATGTCGGCAAG
CTCAATCGCCTCGTAGTGCCGAAGCAGCACGCCGAGAAGCACTTCCCCCTGAAGCGCACG
CCGGAGACGCCGACCACCACCGGCAAGGGCGTGCTGCTCAACTTCGAGGACGGCGAGGGG
AAGGTGTGGAGGTTCCGGTACTCGTACTGGAACAGCAGCCAGAGCTACGTGCTCACCAAA
GGCTGGAGCCGCTTCGTCCGGGAGAAGGGCCTAGGTGCCGGCGACTCCATCCTATTCTCG
TGCTCGCTGTACGAACAGGAGAAGCAGTTCTTCATCGACTGCAAGAAGAACACTAGCATG
AACGGAGGCAAATCGGCGTCGCCGCTGCCAGTGGGGGTGACTACCAAAGGAGAACAAGTT
CGCGTCGTTAGGCTATTCGGTGTCGACATCTCGGGAGTGAAGAGGGGGCGAGCGGCGACG
GCAACGGCGGAGCAAGGCCTGCAGGAGTTGTTCAAGAGGCAATGCGTGGCACCCGGCCAG
CACTCTCCTGCCCTAGGTGCCTTCGCCTTATAG

TRAES3BF062700040CFD_tl
Cover 47% ident 55%

SEQ ID NO: 142
MASGKPTNHGMEDDNDM EYSSAESGAEDAAEPSSSPVLAPPRAAPSSRFKGVVPQPNGRW
GAQIYEKHSRVWLGTFPDEDAAVRAYDVAALRFRGPDAVI NHQRPTAAEEAGSSSSRSEL
DPELGFLADHSKAEIVDMLRKHTYDDELRQGLRRGRGRAQPTPAWARELLFEKAVTPSDV
GKLNRLVVPKQQAEKHFPPTTAAATGSNGKGVLLNFEDGEGKVWRFRYSYWNSSQSYVLT
KGWSRFVKETGLRAGDTVAFYRSAYGNDTEDQLFIDYKKMNKNDDAADAAISDENETGHV
AVKLFGVDIAGGGMAGSSGG

CDS SEQ ID NO: 143
ATGGCATCTGGCAAGCCGACAAACCACGGGATGGAGGACGACAACGACATGGAGTACTCC
TCCGCGGAATCGGGGGCCGAGGACGCGGCGGAGCCGTCGTCGTCGCCGGTGCTGGCGCCG
CCCCGGGCGGCTCCATCGTCGCGGTTCAAGGGCGTCGTGCCGCAGCCCAACGGGCGGTGG
GGAGCGCAGATCTACGAGAAGCACTCGCGGGTGTGGCTCGGAACGTTCCCCGACGAGGAC
GCCGCCGTGCGCGCCTACGACGTGGCCGCGCTCCGCTTCCGCGGCCCGGACGCCGTCATC
AACCACCAGCGACCGACGGCCGCGGAGGAGGCCGGCTCGTCGTCGTCCAGGAGCGAGCTG
GATCCAGAGCTCGGCTTCCTTGCCGACCACTCCAAGGCCGAGATCGTCGACATGCTCCGG
AAGCACACCTACGACGACGAGCTCCGTCAGGGCCTGCGCCGCGGCCGCGGGCGCGCGCAG
CCGACGCCGGCGTGGGCACGAGAGCTCCTCTTCGAGAAGGCCGTGACCCCGAGCGACGTC
GGCAAGCTCAACCGCCTCGTGGTGCCGAAGCAGCAGGCCGAGAAGCACTTCCCTCCGACC
ACTGCGGCGGCCACCGGCAGCAACGGCAAGGGCGTGCTGCTCAACTTCGAGGACGGCGAA
GGGAAGGTGTGGCGCTTCCGGTACTCGTACTGGAACAGCAGCCAGAGCTACGTGCTCACC
AAGGGCTGGAGCCGCTTCGTCAAGGAGACGGGCCTCCGCGCCGGCGACACCGTGGCGTTC
TACCGGTCGGCGTACGGGAATGACACGGAGGATCAGCTCTTCATCGACTACAAGAAGATG
AACAAGAATGACGATGCTGCGGACGCGGCGATTTCCGATGAGAATGAGACAGGCCATGTC
GCCGTCAAGCTCTTCGGCGTTGACATTGCCGGTGGAGGGATGGCGGGATCATCAGGTGGC
TGA

TRAES3BF062600010CFD_tl
Cover 43% ident 58%

SEQ ID NO: 144
MASGKPTNHGMEDDNDM EYSSAESGAEDAAEPSSSPVLAPPRAAPSSRFKGVVPQPNGRW
GAQIYEKHSRVWLGTFPDEDAAARAYDVAALRFRGPDAVI NHQRPTAAEEAGSSSSRSEL
DPELGFLADHSKAEIVDMLRKHTYDDELRQGLRRGRGRAQPTPAWARELLFEKAVTPSDV
GKLNRLVVPKQQAEKHFPPTTAAATGSNGKGVLLNFEDGEGKVWRFRYSYWNSSQSYVLT
KGWSRFVKETGLRAGDTVAFYRSAYGNDTEDQLFIDYKKMNKNDDAADAAISDENETGHV
AVKLFGVDIAGGGMAGSSGG

CDS SEQ ID NO: 145
ATGGCATCTGGCAAGCCGACAAACCACGGGATGGAGGACGACAACGACATGGAGTACTCC
TCCGCGGAATCGGGGGCCGAGGACGCGGCGGAGCCGTCGTCGTCGCCGGTGCTGGCGCCG
CCCCGGGCGGCTCCATCGTCGCGGTTCAAGGGCGTCGTGCCGCAGCCCAACGGGCGGTGG
GGAGCGCAGATCTACGAGAAGCACTCGCGGGTGTGGCTCGGAACGTTCCCCGACGAGGAC
GCCGCCGCGCGCGCCTACGACGTGGCCGCGCTCCGCTTCCGCGGCCCGGACGCCGTCATC

AACCACCAGCGACCGACGGCCGCGGAGGAGGCCGGCTCGTCGTCGTCCAGGAGCGAGCTG
GATCCAGAGCTCGGCTTCCTCGCCGACCACTCCAAGGCCGAGATCGTCGACATGCTCCGG
AAGCACACCTACGACGACGAGCTCCGTCAGGGCCTGCGCCGCGGCCGCGGGCGCGCGCAG
CCGACGCCGGCGTGGGCACGAGAGCTCCTCTTCGAGAAGGCCGTGACCCCGAGCGACGTC
GGCAAGCTCAACCGCCTCGTGGTGCCGAAGCAGCAGGCCGAGAAGCACTTCCCTCCGACC
ACTGCGGCGGCCACCGGCAGCAACGGCAAGGGCGTGCTGCTCAACTTCGAGGACGGCGAA
GGGAAGGTGTGGCGCTTCCGGTACTCGTACTGGAACAGCAGCCAGAGCTACGTGCTCACC
AAGGGCTGGAGCCGCTTCGTCAAGGAGACGGGCCTCCGCGCCGGCGACACCGTGGCGTTC
TACCGGTCGGCGTACGGGAATGACACGGAGGATCAGCTCTTCATCGACTACAAGAAGATG
AACAAGAATGACGATGCTGCGGACGCGGCGATTTCCGATGAGAATGAGACAGGCCATGTC
GCCGTCAAGCTCTTCGGCGTTGACATTGCCGGTGGAGGGATGGCGGGATCATCAGGTGGC
TGA

CLAIMS:

1. A plant wherein said plant does not produce a functional NGAL2 polypeptide or does not produce functional NGAL2 and NGAL3 polypeptides.
2. A plant according to claim 1 wherein the expression of a nucleic acid sequence encoding a NGAL2 polypeptide or the activity of a NGAL2 polypeptide is reduced or abolished.
3. A plant according to claim 1 or 2 wherein the expression of a nucleic acid sequence encoding a NGAL3 polypeptide or the activity of a NGAL3 polypeptide is reduced or abolished.
4. A plant according to a preceding claim wherein the NGAL2 polypeptide comprises a sequence comprising or consisting of SEQ ID NO: 3, a functional variant or homologue thereof.
5. A plant according to a preceding claim wherein the *SOD7* nucleic acid sequence encoding a NGAL2 polypeptide comprises a nucleic acid sequence comprising or consisting of SEQ ID NO: 1 or 2, a functional variant or homologue thereof.
6. A plant according to claim 5 wherein the functional variant or homologue comprises a sequence comprising or consisting of SEQ ID No 6 or 7.
7. A plant according to a preceding claim wherein the NGAL3 polypeptide comprises a sequence comprising or consisting of SEQ ID NO: 5, a functional variant or homologue thereof.
8. A plant according to a preceding claim wherein the *NGAL3* nucleic acid sequence encoding a NGAL3 polypeptide comprises a nucleic acid sequence comprising or consisting of SEQ ID NO: 4, a functional variant or homologue thereof.
9. A plant according to claim 8 wherein the functional variant or homologue thereof comprises a sequence comprising or consisting of SEQ ID NO: 6 or 7.
10. A plant according to a preceding claim wherein the endogenous *SOD7* nucleic acid sequence or its promoter carries a functional mutation.
11. A plant according to any of claims 1-9 wherein said plant comprises an RNA interference construct that reduces the expression of *SOD7,* a functional variant or homolog thereof.
12. A plant according to a preceding claim wherein the endogenous *NGAL3* nucleic acid sequence or its promoter carries a functional mutation.

13. A plant according to any of claims 2 to 11 wherein said plant comprises an RNA interference construct that reduces the expression of *NGAL3,* a functional variant or homolog thereof.
14. A plant according to a preceding claim wherein said plant is a crop plant.
15. A plant according to a preceding claim wherein said plant is a monocotyledonous plant or dicotyledonous plant.
16. A plant according to a preceding claim wherein said plant is selected from maize, rice, wheat, oilseed rape/canola, sorghum, soybean, sunflower, alfalfa, potato, tomato, tobacco, grape, barley, pea, bean, field bean, lettuce, cotton, sugar cane, sugar beet, broccoli or other vegetable brassicas or poplar.
17. A method for altering a plant phenotype comprising reducing or abolishing the expression of a nucleic acid sequence encoding a NGAL2 polypeptide or reducing or abolishing the activity of a NGAL2 or reducing or abolishing the expression of a nucleic acid sequences encoding NGAL2 and NGAL3 polypeptides or reducing or abolishing the activity of a NGAL2 and NGAL3 polypeptide relative to a control plant.
18. A method for making a plant with an altered phenotype comprising reducing or abolishing the expression of a nucleic acid sequence encoding a NGAL2 polypeptide or reducing or abolishing the activity of a NGAL2 or reducing or abolishing the expression of a nucleic acid sequences encoding NGAL2 and NGAL3 polypeptides or reducing or abolishing the activity of a NGAL2 and NGAL3 polypeptide relative to a control plant..
19. A method according to claim 17 or 18 comprising reducing or abolishing the expression of a nucleic acid sequence encoding a NGAL2 polypeptide and reducing or abolishing expression of a nucleic acid sequence encoding a NGAL3 polypeptide.
20. A method according to claim 17 or 18 comprising reducing or abolishing activity of a NGAL2 polypeptide and reducing or abolishing activity of a NGAL3 polypeptide.
21. A method according to claim 17 or 18 comprising reducing or abolishing expression of a nucleic acid sequence encoding a NGAL2 polypeptide and reducing or abolishing activity of a NGAL3 polypeptide.
22. A method according to claim 17 or 18 comprising reducing or abolishing expression of a nucleic acid sequence encoding a NGAL3 polypeptide and reduced or abolished activity of a NGAL2 polypeptide.

23. A method according to any of claims 17 to 22 wherein the NGAL2 polypeptide comprises a sequence comprising or consisting of SEQ D NO: 1 or 2, a functional variant or homologue thereof.
24. A method according to any of claims 17 to 22 wherein the *SOD7* nucleic acid sequence encoding a NGAL2 polypeptide comprises a nucleic acid sequence comprising or consisting of SEQ ID NO: 1 or 2, a functional variant or homologue thereof.
25. A method according to claim 24 wherein the functional variant or homologue comprises a nucleic acid sequence as shown in SEQ Id NO:49-145.
26. A method according to any of claims 17-25 wherein the NGAL3 polypeptide comprises a sequence comprising or consisting of SEQ ID NO: 5, a functional variant or homologue thereof.
27. A method according to claim 17 to 26 wherein the *NGAL3* nucleic acid sequence encoding a NGAL3 polypeptide comprises a nucleic acid sequence comprising or consisting of SEQ ID NO: 4, a functional variant or homologue thereof.
28. A method according to claim 27 wherein the functional variant or homologue comprises a nucleic acid sequence comprising or consisting of SEQ ID NOs:49-145.
29. A method according to any of claims 17 to 28 wherein said method comprises introducing a functional mutation in a nucleic acid sequence encoding a NGAL2 and/or NGAL3 protein or peptide in a plant or in its corresponding promoter.
30. A method according to claim 29 wherein said mutation is introduced using T-DNA insertion, chemical mutagenesis or genome editing.
31. A method according to claim 30 comprising using TILLING.
32. A method according to any of claims 17 to 28 comprising silencing of the *SOD7* and/or *NGAL3* nucleic acid sequence.
33. A method according to claim 32 comprising introducing a RNAi, shRNA, snRNA, dsRNA, siRNA, miRNA, ta-siRNA or co-suppression molecule which targets the *SOD7* or *NGAL3* nucleic acid sequence gene into a plant.
34. A method according to any of claims 17 to 33 wherein said phenotype is characterised by increased seed size relative to a control plant.
35. A plant obtained or obtainable by the method of any of claims 17 to 34.
36. An isolated nucleic acid comprising a sequence comprising or consisting of SEQ ID NO: 1 or 2 or a functional variant or homologue thereof.
37. A vector comprising an isolated nucleic acid according to claim 36.

38. A silencing nucleic acid construct targeting sequence comprising or consisting of SEQ ID NO: 1, 2 or 3 or a functional variant, part or homologue thereof.

Figure 1

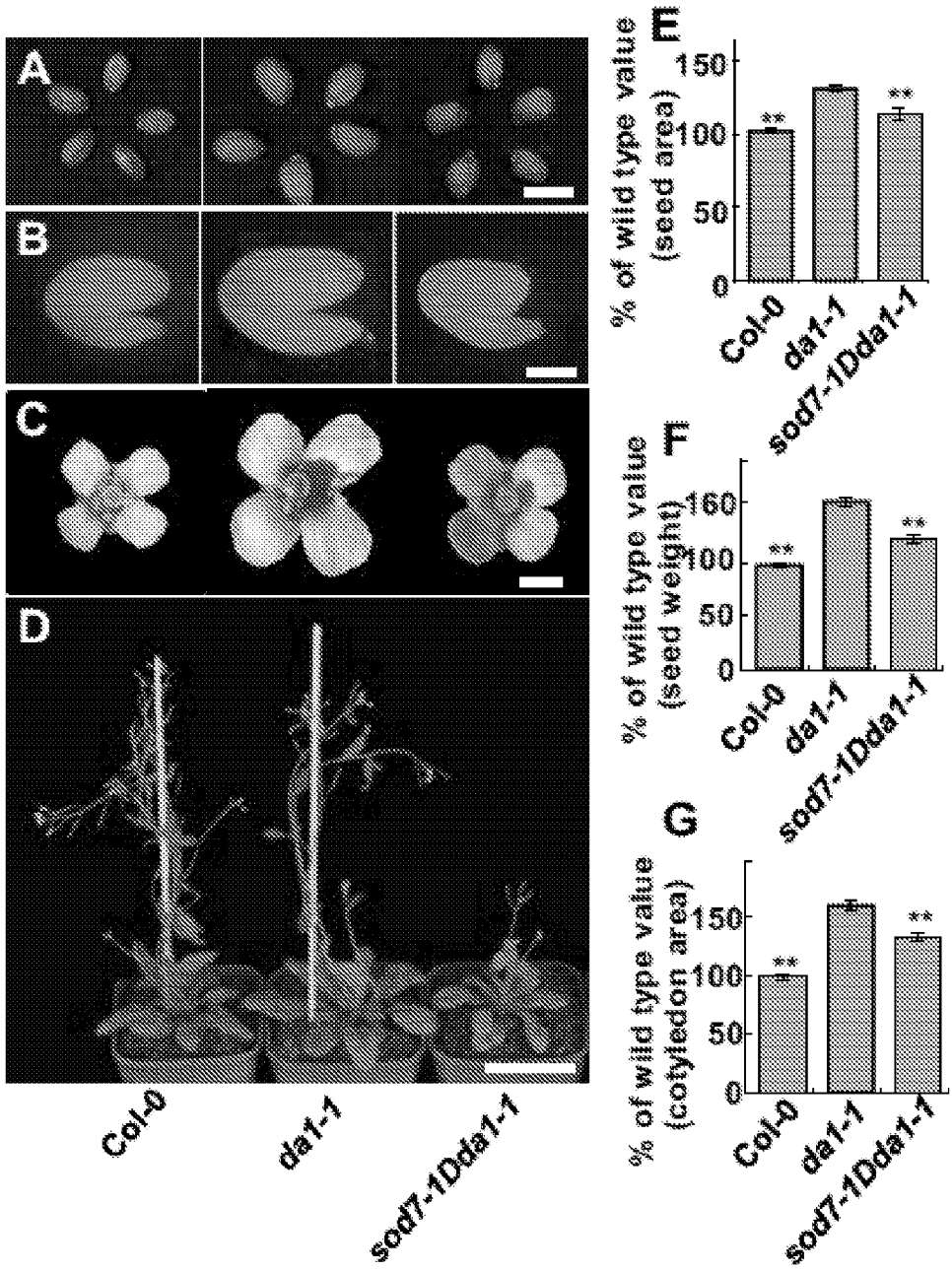

A
B
C
D
Col-0
da1-1
sod7-1Dda1-1
E
% of wild type value (seed area)
150
100
50
0
**
Col-0
da1-1
sod7-1Dda1-1
F
% of wild type value (seed weight)
160
100
50
0
**
Col-0
da1-1
sod7-1Dda1-1
G
% of wild type value (cotyledon area)
150
100
50
0
**
Col-0
da1-1
sod7-1Dda1-1

Figure 2

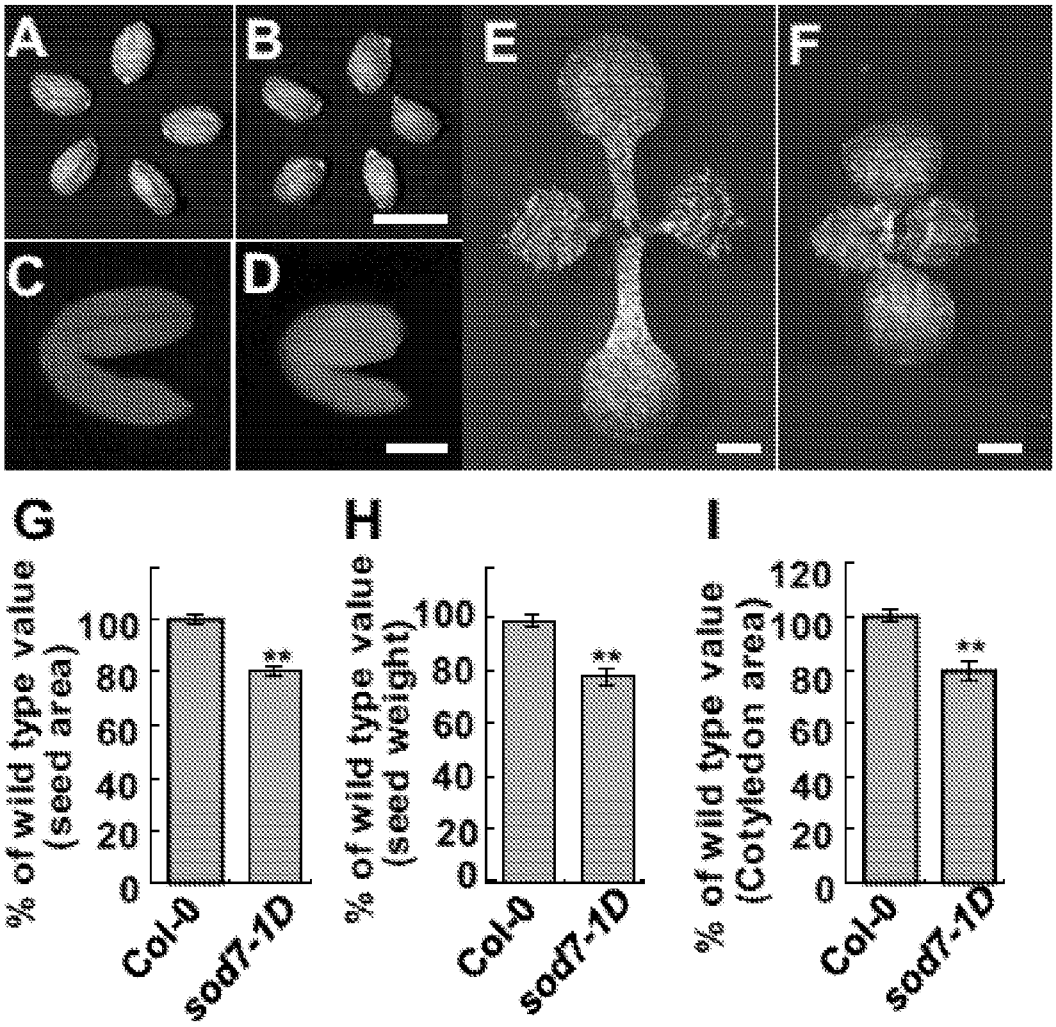

A
B
E
F
C
D
G
% of wild type value (seed area)
100
80
60
40
20
0
Col-0
sod7-1D
**
H
% of wild type value (seed weight)
100
80
60
40
20
0
Col-0
sod7-1D
**
I
% of wild type value (Cotyledon area)
120
100
80
60
40
20
0
Col-0
sod7-1D
**

Figure 3

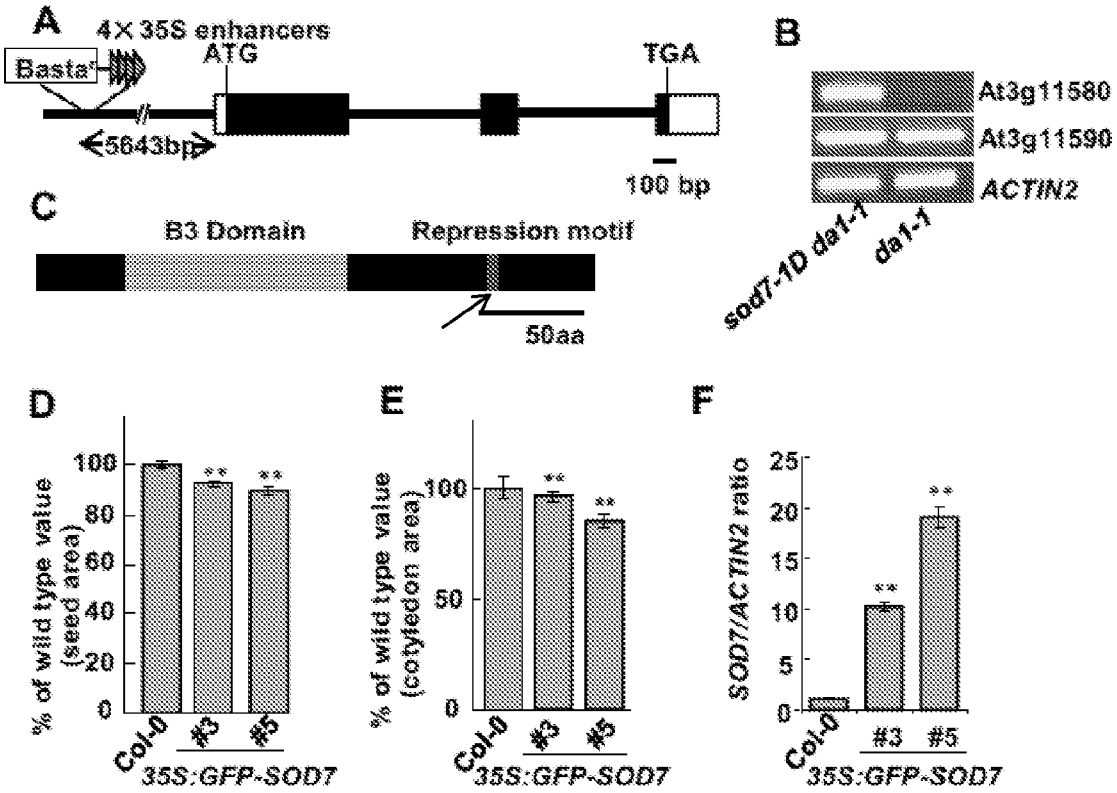

A
4×35S enhancers
Basta
ATG
TGA
5643bp
100 bp
B
At3g11580
At3g11590
ACTIN2
sod7-1D da1-1
da1-1
C
B3 Domain
Repression motif
50aa
D
% of wild type value (seed area)
Col-0
#3
#5
35S:GFP-SOD7
E
% of wild type value (cotyledon area)
Col-0
#3
#5
35S:GFP-SOD7
F
SOD7/ACTIN2 ratio
Col-0
#3
#5
35S:GFP-SOD7

Figure 4

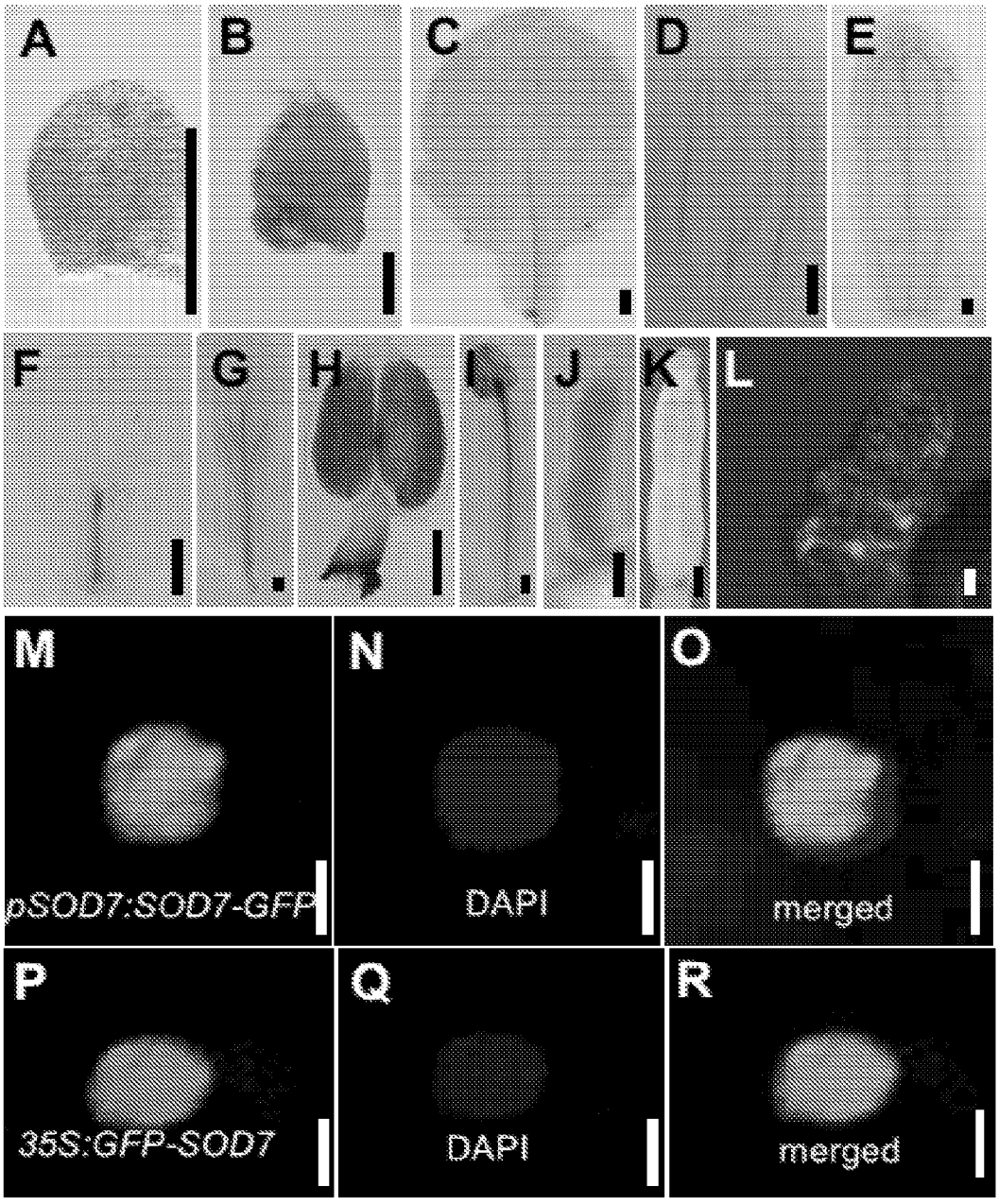

A
B
C
D
E
F
G
H
I
J
K
L
M
pSOD7:SOD7-GFP
N
DAPI
O
merged
P
35S:GFP-SOD7
Q
DAPI
R
merged

Figure 5

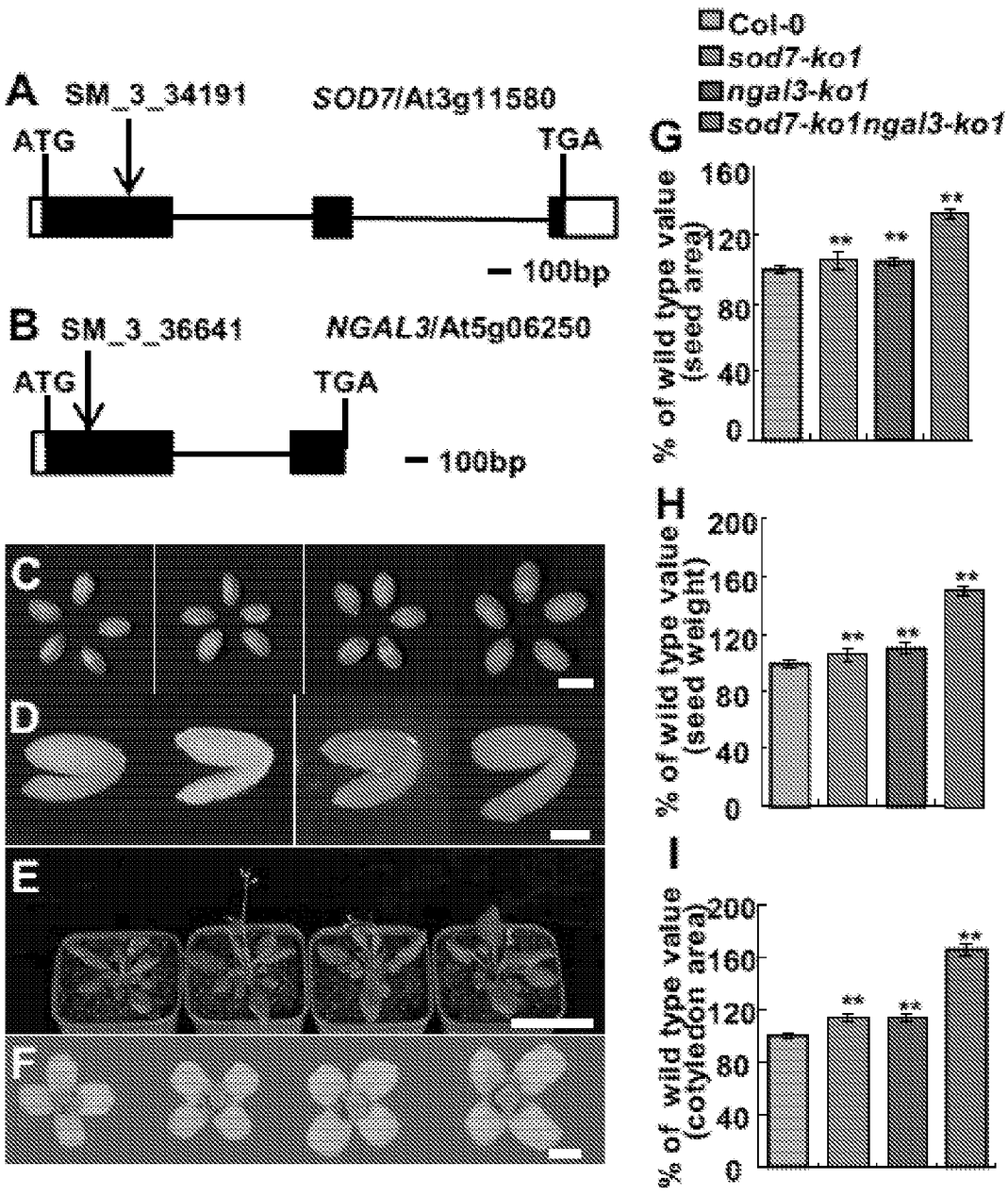

A SM_3_34191
SOD7/At3g11580
ATG
TGA
100bp
B SM_3_36641
NGAL3/At5g06250
ATG
TGA
100bp
C
D
E
F
G
Col-0
sod7-ko1
ngal3-ko1
sod7-ko1ngal3-ko1
% of wild type value (seed area)
H
% of wild type value (seed weight)
I
% of wild type value (cotyledon area)
0
40
80
120
160
200
**

Figure 6

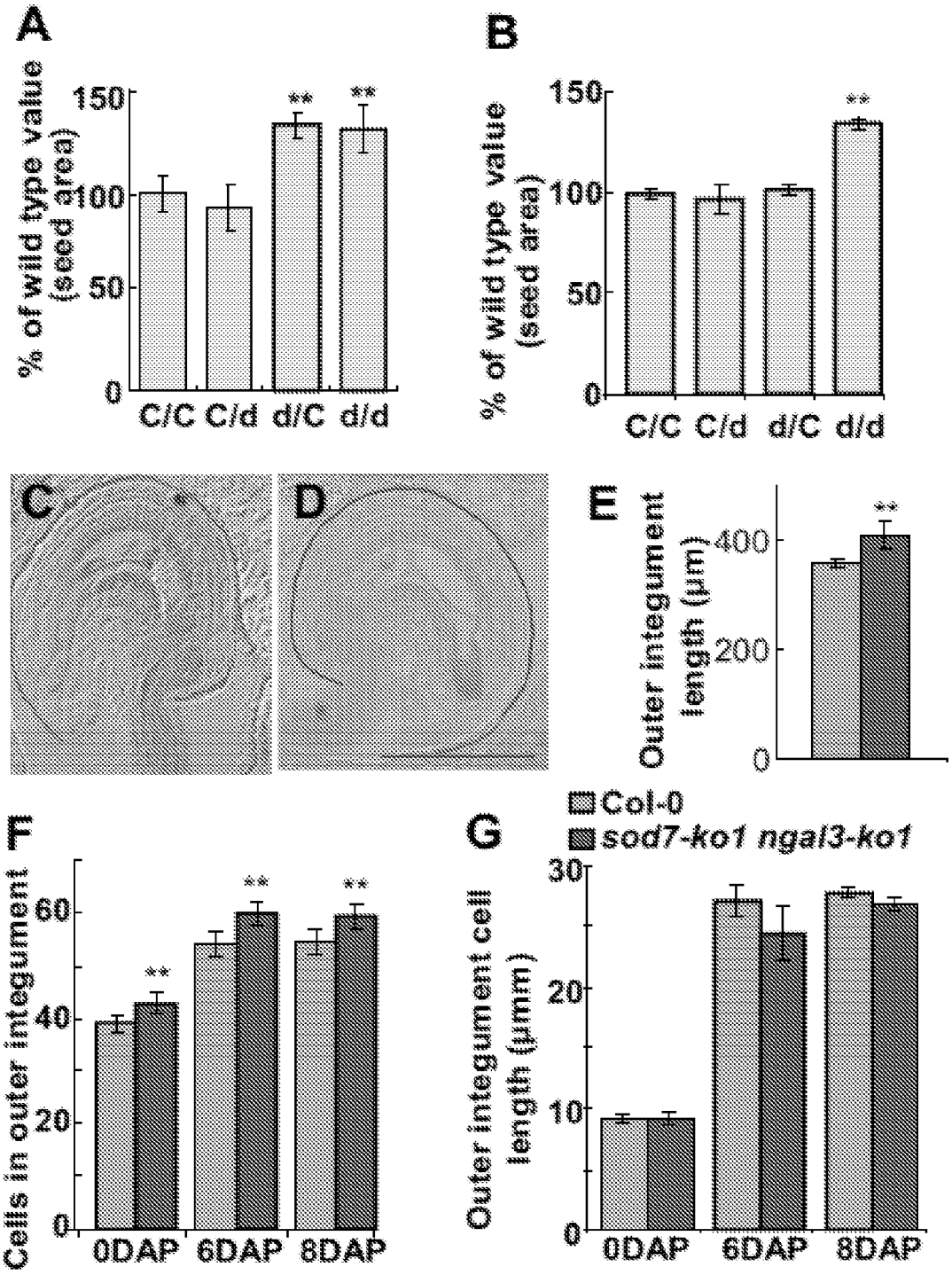

A
% of wild type value (seed area)
150
100
50
0
**
**
C/C C/d d/C d/d
B
% of wild type value (seed area)
150
100
50
0
**
C/C C/d d/C d/d
C
D
E
Outer integument length (µm)
400
200
0
**
F
Cells in outer integument
60
40
20
0
**
**
0DAP 6DAP 8DAP
G
Col-0
sod7-ko1 nga13-ko1
Outer integument cell length (µmm)
30
20
10
0
0DAP 6DAP 8DAP

Figure 7

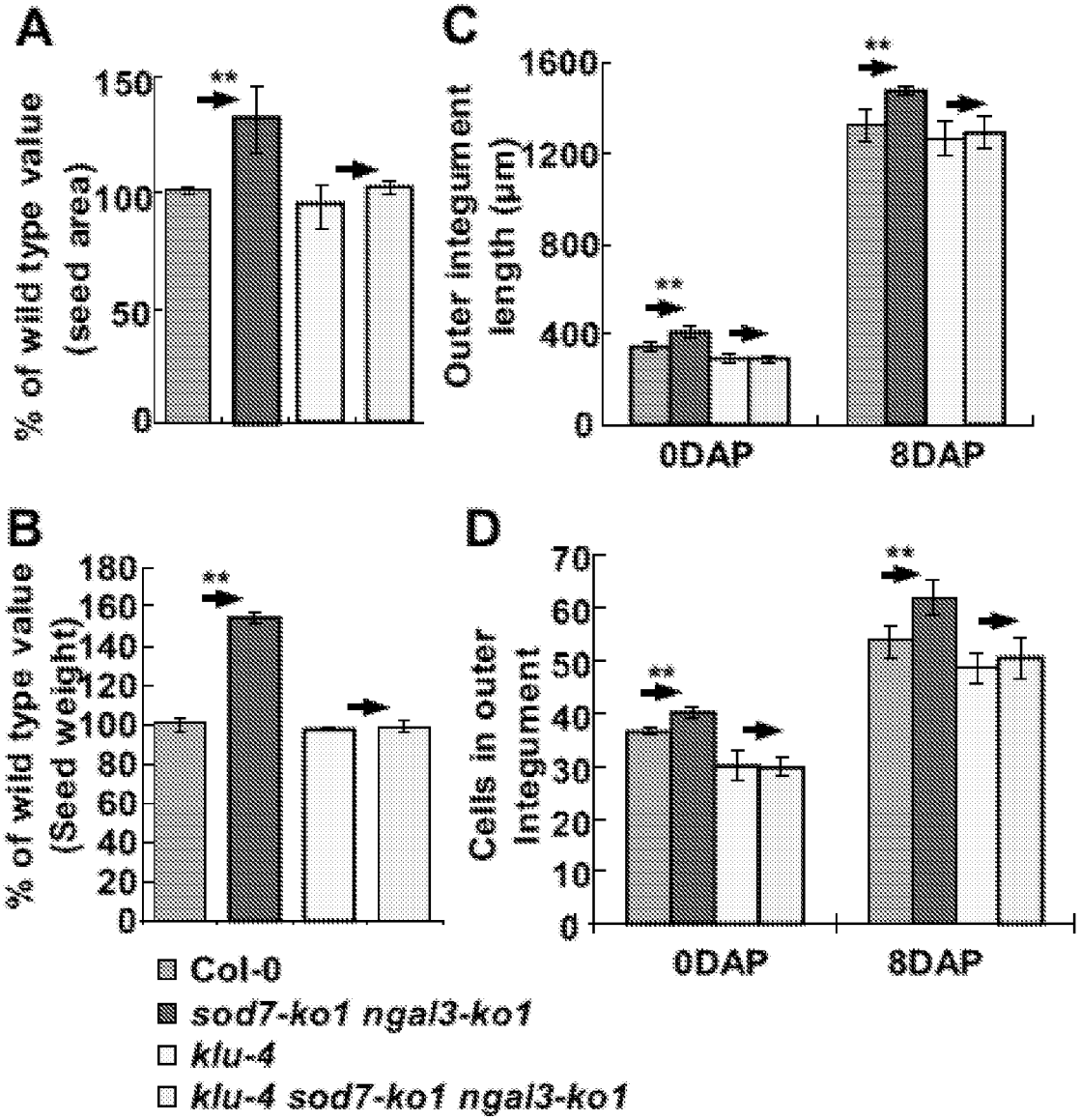

A
% of wild type value (seed area)
150
100
50
0
**
C
Outer integument length (µm)
1600
1200
800
400
0
**
**
0DAP
8DAP
B
% of wild type value (Seed weight)
180
160
140
120
100
80
60
40
20
0
**
D
Cells in outer Integument
70
60
50
40
30
20
10
0
**
**
0DAP
8DAP
Col-0
sod7-ko1 ngal3-ko1
klu-4
klu-4 sod7-ko1 ngal3-ko1

Figure 8

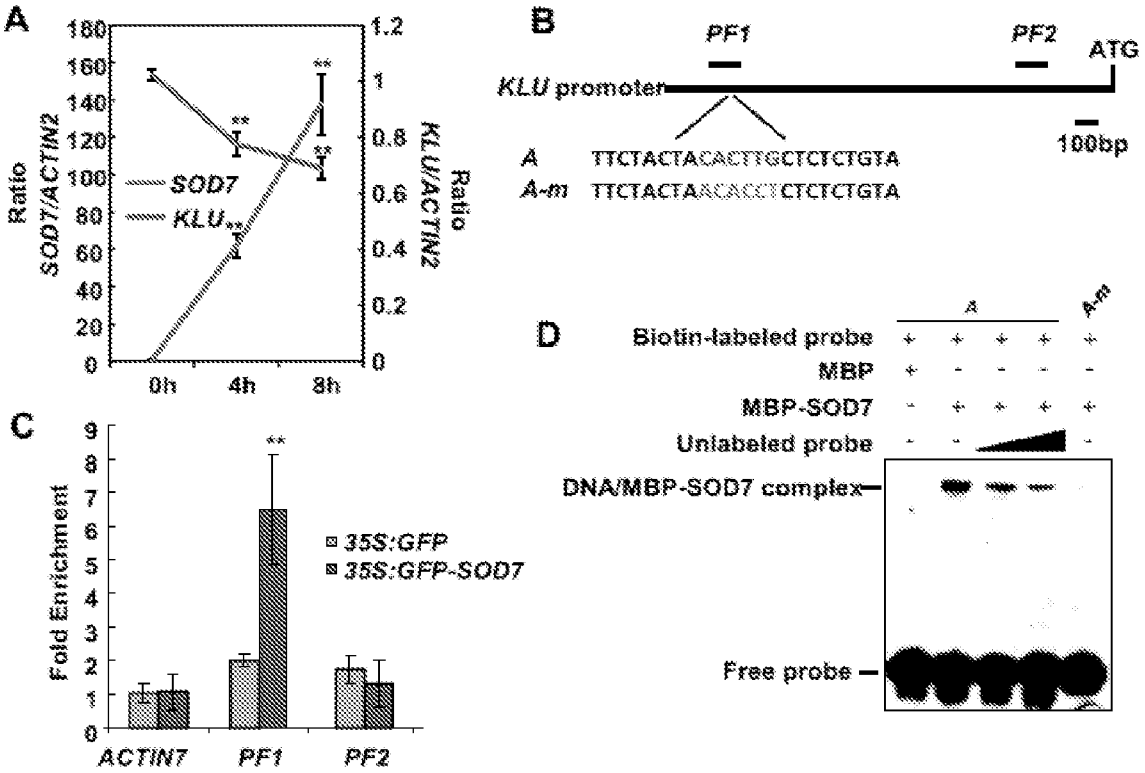

A
Ratio SOD7/ACTIN2
Ratio KLU/ACTIN2
SOD7
KLU
0h
4h
8h
B
PF1
PF2
ATG
KLU promoter
100bp
A TTCTACTACACTTGCTCTCTGTA
A-m TTCTACTAACACCTCTCTCTGTA
C
Fold Enrichment
35S:GFP
35S:GFP-SOD7
ACTIN7
PF1
PF2
D
Biotin-labeled probe
MBP
MBP-SOD7
Unlabeled probe
DNA/MBP-SOD7 complex
Free probe

Figure 9

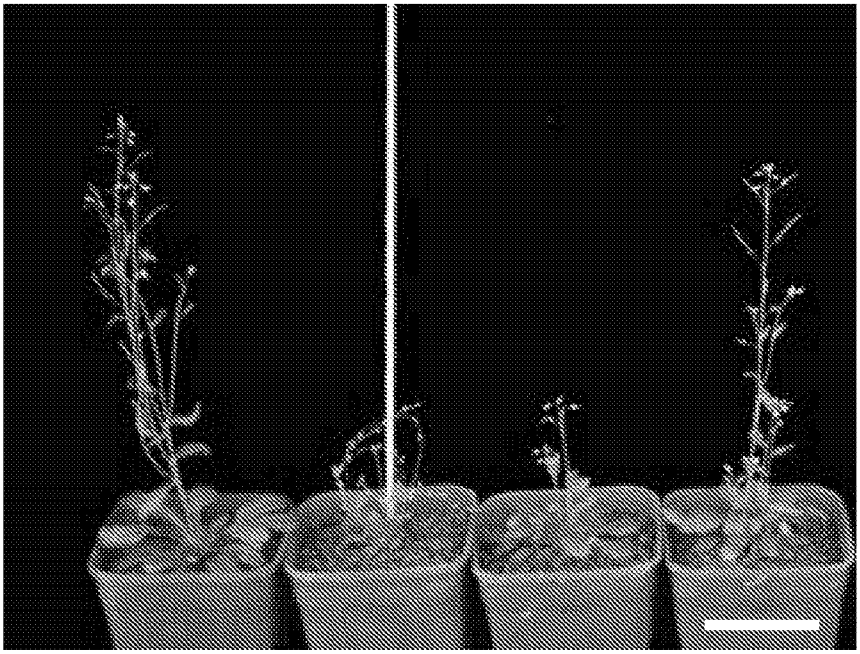


Figure 10

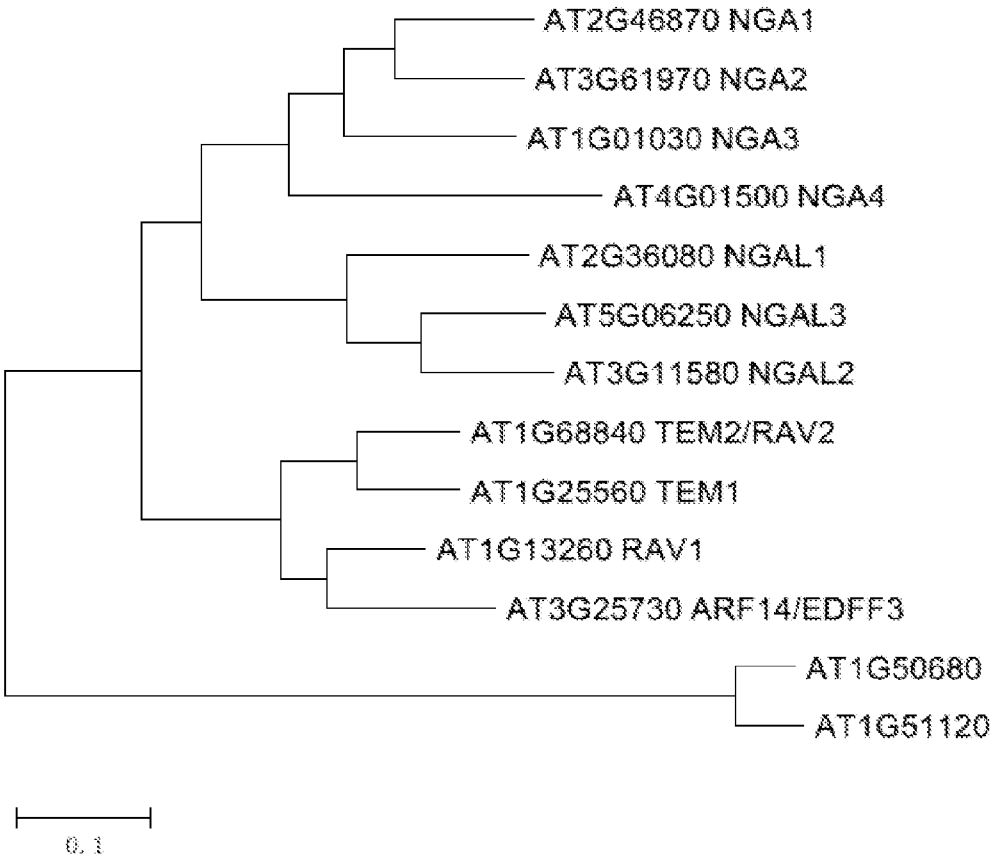

AT2G46870 NGA1
AT3G61970 NGA2
AT1G01030 NGA3
AT4G01500 NGA4
AT2G36080 NGAL1
AT5G06250 NGAL3
AT3G11580 NGAL2
AT1G68840 TEM2/RAV2
AT1G25560 TEM1
AT1G13260 RAV1
AT3G25730 ARF14/EDFF3
AT1G50680
AT1G51120
0.1

Figure 11

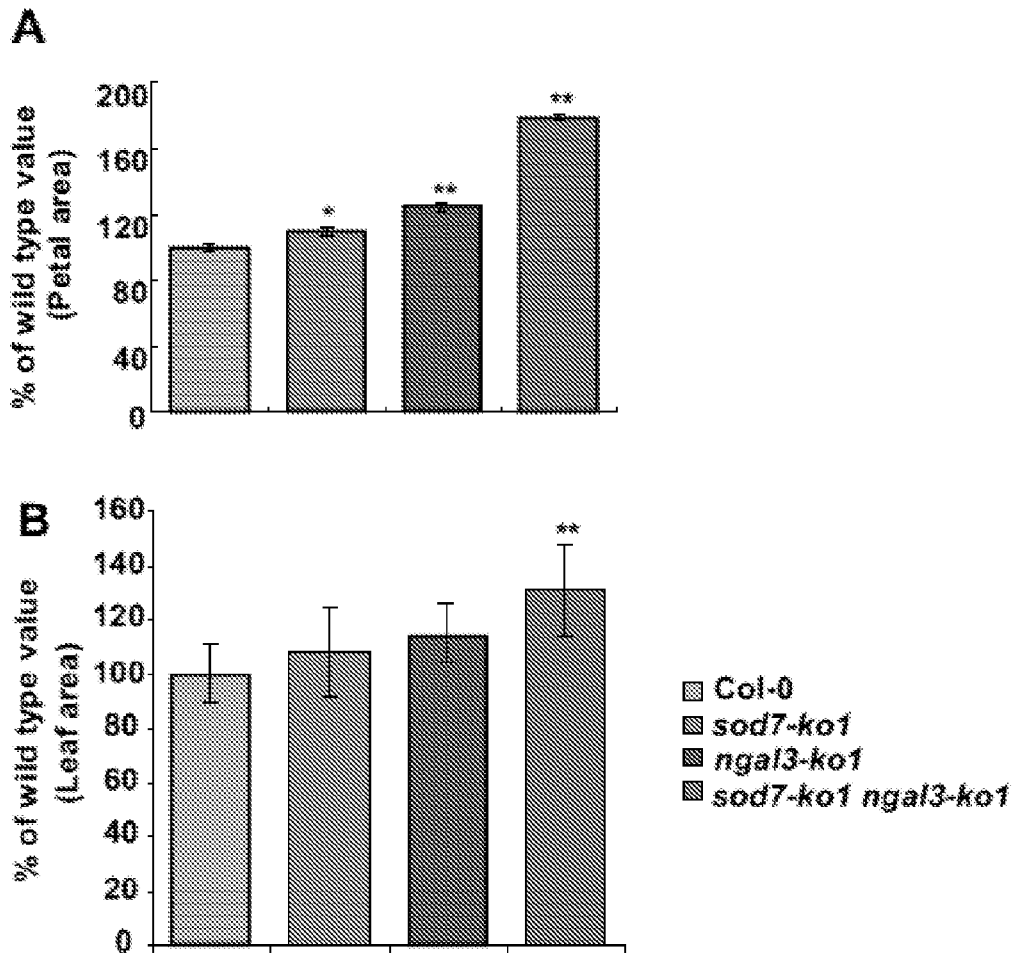
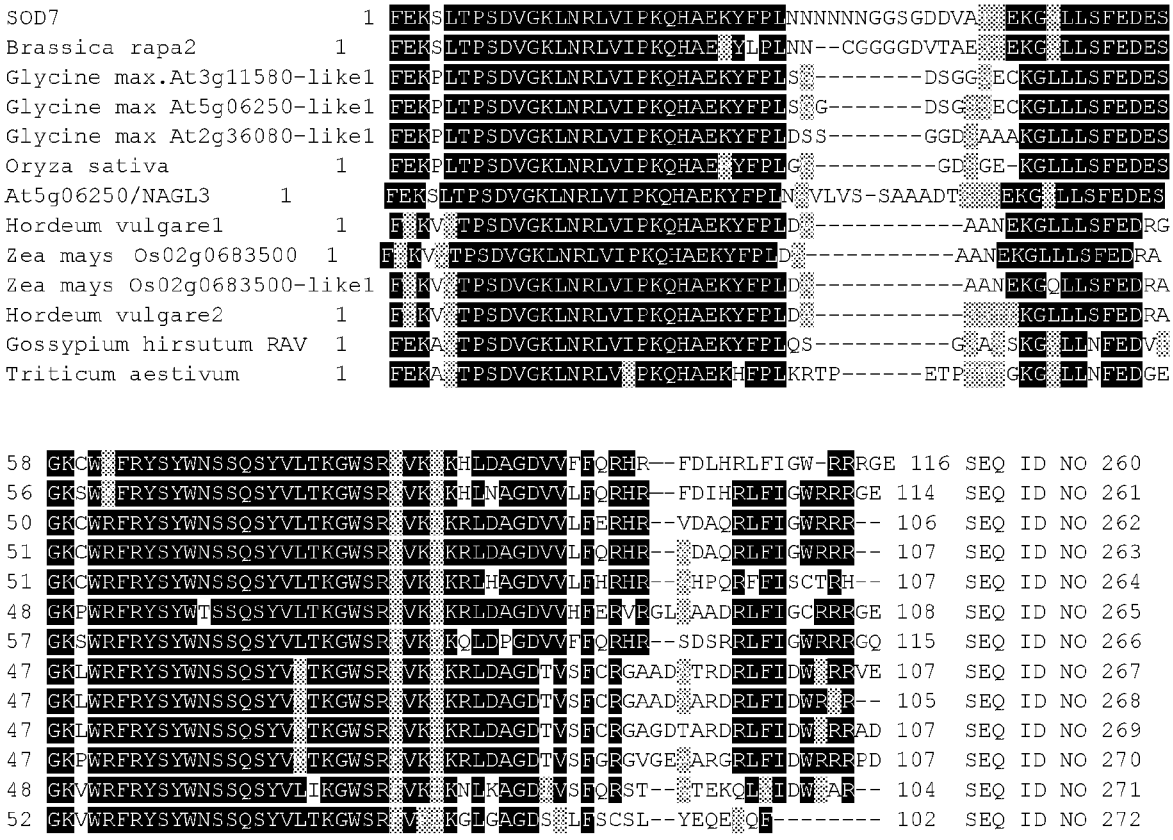

A
% of wild type value (Petal area)
200
160
120
80
40
0
*
**
**
B
% of wild type value (Leaf area)
160
140
120
100
80
60
40
20
0
**
Col-0
sod7-ko1
ngal3-ko1
sod7-ko1 ngal3-ko1

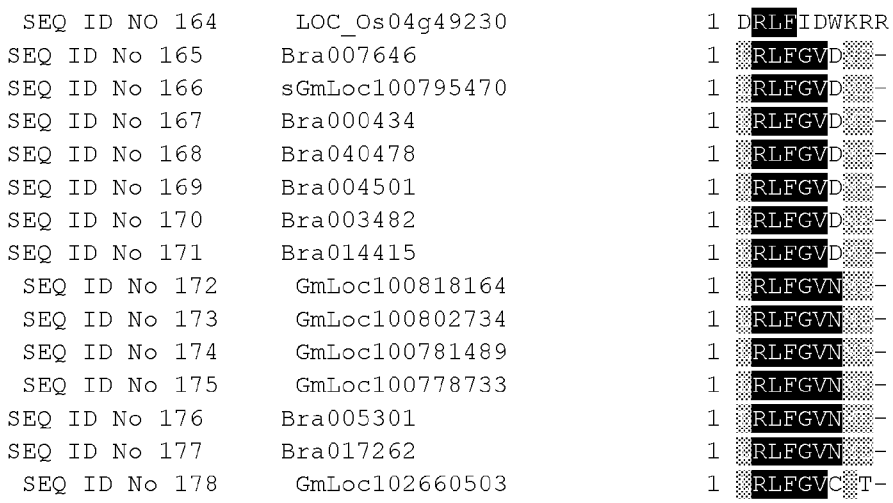
Figure 12

A


SOD7 1 FEKSLTPSDVGKLNRLVIPKQHAEKYFPLNNNNNNGGSGDDVA EKG LLSFEDES
Brassica rapa2 1 FEKSLTPSDVGKLNRLVIPKQHAE YLPLNN--CGGGGDVTAE EKG LLSFEDES
Glycine max.At3g11580-like1 FEKPLTPSDVGKLNRLVIPKQHAEKYFPLS--------DSGG ECKGLLLSFEDES
Glycine max At5g06250-like1 FEKPLTPSDVGKLNRLVIPKQHAEKYFPLS G-------DSG ECKGLLLSFEDES
Glycine max At2g36080-like1 FEKPLTPSDVGKLNRLVIPKQHAEKYFPLDSS-------GGD AAAKGLLLSFEDES
Oryza sativa 1 FEKPLTPSDVGKLNRLVIPKQHAE YFPLG---------GD GE-KGLLLSFEDES
At5g06250/NAGL3 1 FEKSLTPSDVGKLNRLVIPKQHAEKYFPLN VLVS-SAAADT EKG LLSFEDES
Hordeum vulgare1 1 F KV TPSDVGKLNRLVIPKQHAEKYFPLD----------AANEKGLLLSFEDRG
Zea mays Os02g0683500 1 F KV TPSDVGKLNRLVIPKQHAEKYFPLD----------AANEKGLLLSFEDRA
Zea mays Os02g0683500-like1 F KV TPSDVGKLNRLVIPKQHAEKYFPLD----------AANEKGQLLSFEDRA
Hordeum vulgare2 1 F KV TPSDVGKLNRLVIPKQHAEKYFPLD---------- KGLLLSFEDRA
Gossypium hirsutum RAV 1 FEKA TPSDVGKLNRLVIPKQHAEKYFPLQS---------G A SKG LLNFEDV
Triticum aestivum 1 FEKA TPSDVGKLNRLV PKQHAEK FPLKRTP------ETP GKG LLNFEDGE

58 GKCW FRYSYWNSSQSYVLTKGWSR VK KHLDAGDVVF FQRHR--FDLHRLFIGW-RRRGE 116 SEQ ID NO 260
56 GKSW FRYSYWNSSQSYVLTKGWSR VK KHLNAGDVVL FQRHR--FDIHRLFIGWRRRGE 114 SEQ ID NO 261
50 GKCWRFRYSYWNSSQSYVLTKGWSR VK KRLDAGDVVL FERHR--VDAQRLFIGWRRR-- 106 SEQ ID NO 262
51 GKCWRFRYSYWNSSQSYVLTKGWSR VK KRLDAGDVVL FQRHR--DAQRLFIGWRRR-- 107 SEQ ID NO 263
51 GKCWRFRYSYWNSSQSYVLTKGWSR VK KRLHAGDVVL FHRHR--HPQRFFISCTRH-- 107 SEQ ID NO 264
48 GKPWRFRYSYWTSSQSYVLTKGWSR VK KRLDAGDVVH FERVRGL AADRLFIGCRRRGE 108 SEQ ID NO 265
57 GKSWRFRYSYWNSSQSYVLTKGWSR VK KQLDPGDVVF FQRHR--SDSRRLFIGWRRRGQ 115 SEQ ID NO 266
47 GKLWRFRYSYWNSSQSYV TKGWSR VK KRLDAGDTVS FCRGAAD TRDRLFIDW RRVE 107 SEQ ID NO 267
47 GKLWRFRYSYWNSSQSYV TKGWSR VK KRLDAGDTVS FCRGAAD ARDRLFIDWP R-- 105 SEQ ID NO 268
47 GKLWRFRYSYWNSSQSYV TKGWSR VK KRLDAGDTVS FCRGAGDTARDRLFIDW RRAD 107 SEQ ID NO 269
47 GKPWRFRYSYWNSSQSYV TKGWSR VK KRLDAGDTVS FGRGVGE ARGRLFIDWRRRPD 107 SEQ ID NO 270
48 GKVWRFRYSYWNSSQSYVL KGWSR VK KNLKAGD VS FQRST--TEKQL IDW AR-- 104 SEQ ID NO 271
52 GKVWRFRYSYWNSSQSYVLTKGWSR V KGLGAGDS LFSCSL--YEQE QE-------- 102 SEQ ID NO 272

B


SEQ ID NO 164 LOC_Os04g49230 1 DRLFIDWKRR
SEQ ID No 165 Bra007646 1 RLFGVD -
SEQ ID No 166 sGmLoc100795470 1 RLFGVD -
SEQ ID No 167 Bra000434 1 RLFGVD -
SEQ ID No 168 Bra040478 1 RLFGVD -
SEQ ID No 169 Bra004501 1 RLFGVD -
SEQ ID No 170 Bra003482 1 RLFGVD -
SEQ ID No 171 Bra014415 1 RLFGVD -
SEQ ID No 172 GmLoc100818164 1 RLFGVN -
SEQ ID No 173 GmLoc100802734 1 RLFGVN -
SEQ ID No 174 GmLoc100781489 1 RLFGVN -
SEQ ID No 175 GmLoc100778733 1 RLFGVN -
SEQ ID No 176 Bra005301 1 RLFGVN -
SEQ ID No 177 Bra017262 1 RLFGVN -
SEQ ID No 178 GmLoc102660503 1 RLFGVC T-

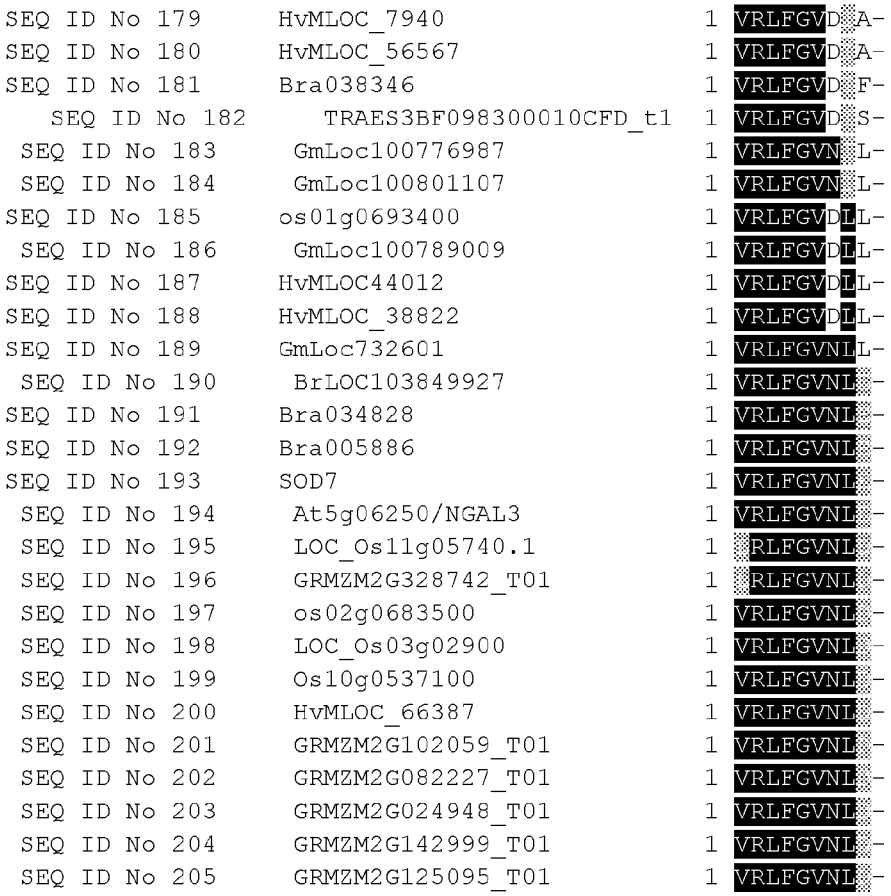
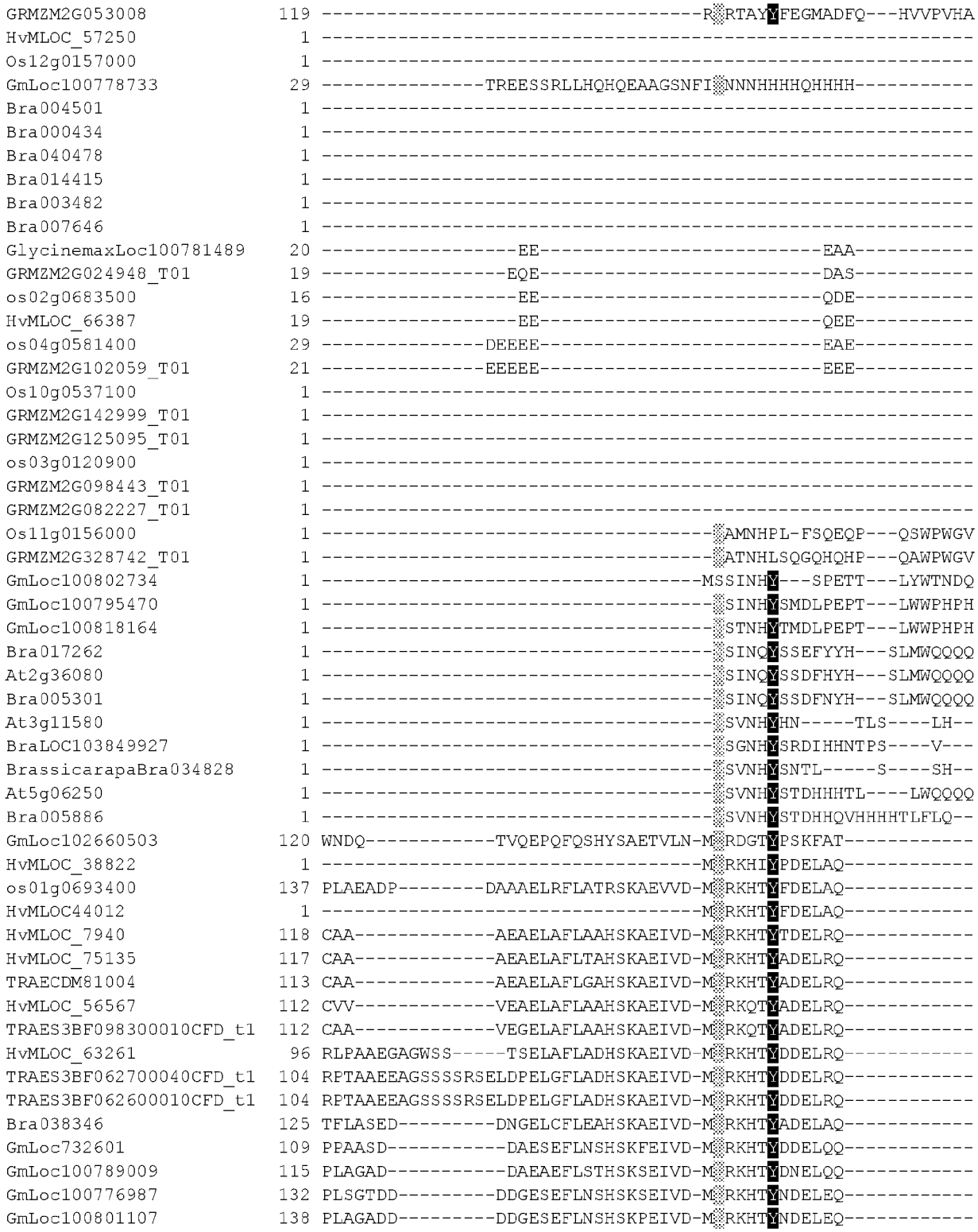

SEQ ID No 179 HvMLOC_7940 1 VRLFGVD A-
SEQ ID No 180 HvMLOC_56567 1 VRLFGVD A-
SEQ ID No 181 Bra038346 1 VRLFGVD F-
SEQ ID No 182 TRAES3BF098300010CFD_t1 1 VRLFGVD S-
SEQ ID No 183 GmLoc100776987 1 VRLFGVN L-
SEQ ID No 184 GmLoc100801107 1 VRLFGVN L-
SEQ ID No 185 os01g0693400 1 VRLFGVDLL-
SEQ ID No 186 GmLoc100789009 1 VRLFGVDLL-
SEQ ID No 187 HvMLOC44012 1 VRLFGVDLL-
SEQ ID No 188 HvMLOC_38822 1 VRLFGVDLL-
SEQ ID No 189 GmLoc732601 1 VRLFGVNLL-
SEQ ID No 190 BrLOC103849927 1 VRLFGVNL -
SEQ ID No 191 Bra034828 1 VRLFGVNL -
SEQ ID No 192 Bra005886 1 VRLFGVNL -
SEQ ID No 193 SOD7 1 VRLFGVNL -
SEQ ID No 194 At5g06250/NGAL3 1 VRLFGVNL -
SEQ ID No 195 LOC_Os11g05740.1 1 RLFGVNL -
SEQ ID No 196 GRMZM2G328742_T01 1 RLFGVNL -
SEQ ID No 197 os02g0683500 1 VRLFGVNL -
SEQ ID No 198 LOC_Os03g02900 1 VRLFGVNL -
SEQ ID No 199 Os10g0537100 1 VRLFGVNL -
SEQ ID No 200 HvMLOC_66387 1 VRLFGVNL -
SEQ ID No 201 GRMZM2G102059_T01 1 VRLFGVNL -
SEQ ID No 202 GRMZM2G082227_T01 1 VRLFGVNL -
SEQ ID No 203 GRMZM2G024948_T01 1 VRLFGVNL -
SEQ ID No 204 GRMZM2G142999_T01 1 VRLFGVNL -
SEQ ID No 205 GRMZM2G125095_T01 1 VRLFGVNL -

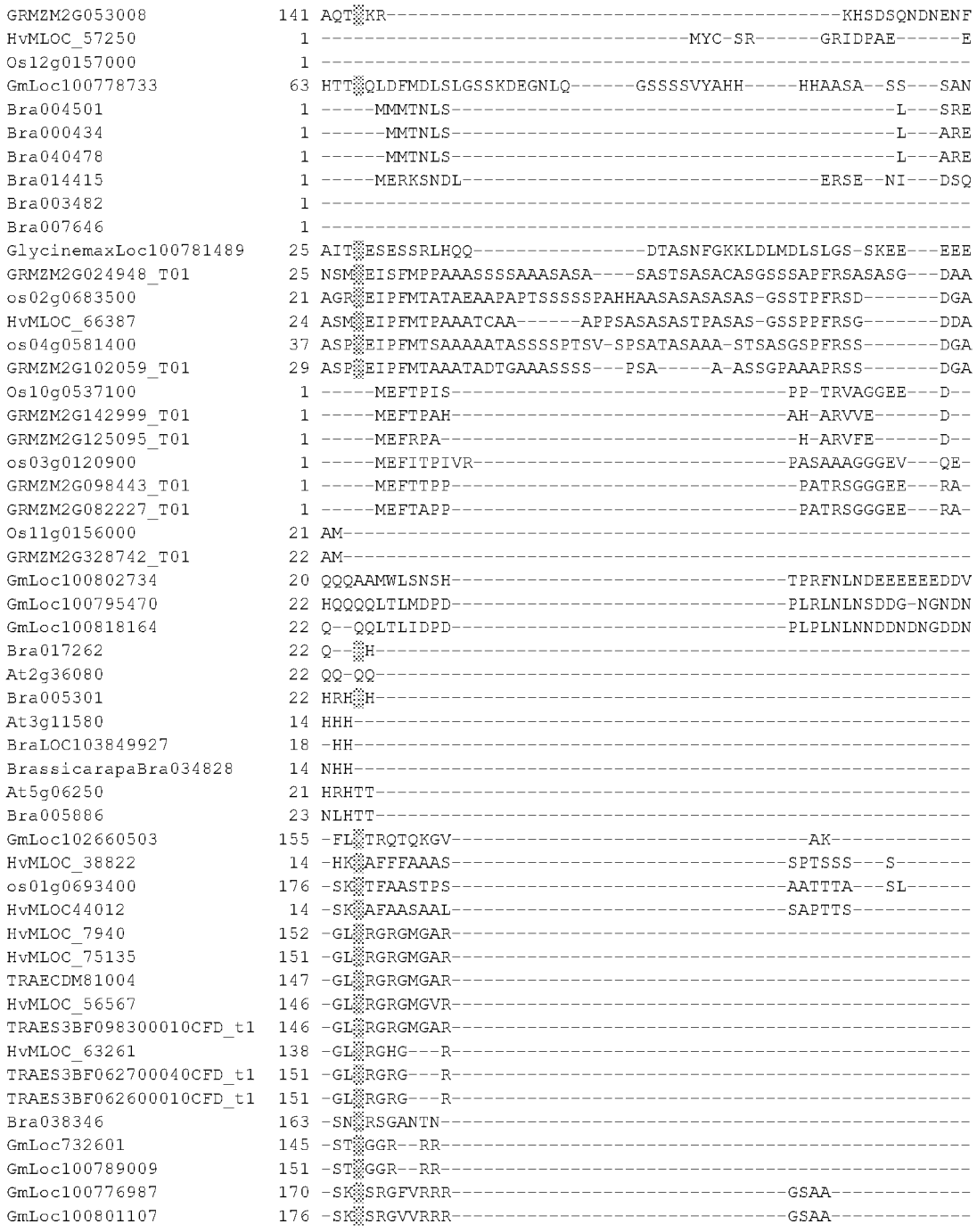
Figure 13

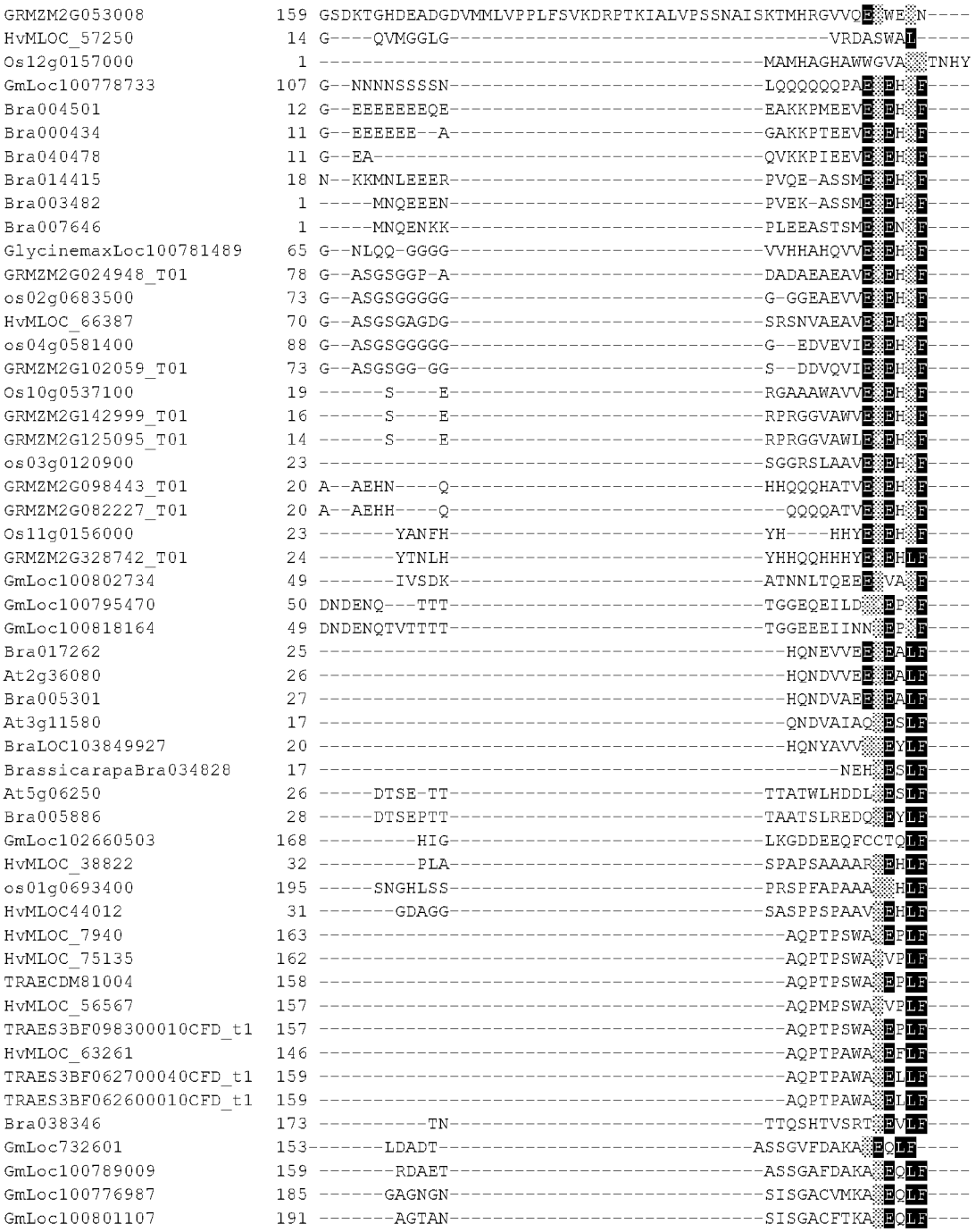
```
GRMZM2G053008          SEQ ID NO 207  1 ----------------------------MAAS----------------PSSPLTAPPEPV
HvMLOC_57250           SEQ ID NO 208  1 ------------------------------------------------------------
Os12g0157000           SEQ ID NO 209  1 ------------------------------------------------------------
GmLoc100778733         SEQ ID NO 210  1 ------------------------------------------------------------
Bra004501              SEQ ID NO 211  1 ------------------------------------------------------------
Bra000434              SEQ ID NO 212  1 ------------------------------------------------------------
Bra040478              SEQ ID NO 213  1 ------------------------------------------------------------
Bra014415              SEQ ID NO 214  1 ------------------------------------------------------------
Bra003482              SEQ ID NO 215  1 ------------------------------------------------------------
Bra007646              SEQ ID NO 216  1 ------------------------------------------------------------
GlycinemaxLoc100781489 SEQ ID NO 217  1 ------------------------------------------------------------
GRMZM2G024948_T01      SEQ ID NO 218  1 ------------------------------------------------------------
os02g0683500           SEQ ID NO 219  1 ------------------------------------------------------------
HvMLOC_66387           SEQ ID NO 220  1 ------------------------------------------------------------
os04g0581400           SEQ ID NO 221  1 ------------------------------------------------------------
GRMZM2G102059_T01      SEQ ID NO 222  1 ------------------------------------------------------------
Os10g0537100           SEQ ID NO 223  1 ------------------------------------------------------------
GRMZM2G142999_T01      SEQ ID NO 224  1 ------------------------------------------------------------
GRMZM2G125095_T01      SEQ ID NO 225  1 ------------------------------------------------------------
os03g0120900           SEQ ID NO 226  1 ------------------------------------------------------------
GRMZM2G098443_T01      SEQ ID NO 227  1 ------------------------------------------------------------
GRMZM2G082227_T01      SEQ ID NO 228  1 ------------------------------------------------------------
Os11g0156000           SEQ ID NO 229  1 ------------------------------------------------------------
GRMZM2G328742_T01      SEQ ID NO 230  1 ------------------------------------------------------------
GmLoc100802734         SEQ ID NO 231  1 ------------------------------------------------------------
GmLoc100795470         SEQ ID NO 232  1 ------------------------------------------------------------
GmLoc100818164         SEQ ID NO 233  1 ------------------------------------------------------------
Bra017262              SEQ ID NO 234  1 ------------------------------------------------------------
At2g36080              SEQ ID NO 235  1 ------------------------------------------------------------
Bra005301              SEQ ID NO 236  1 ------------------------------------------------------------
At3g11580              SEQ ID NO 237  1 ------------------------------------------------------------
BraLOC103849927        SEQ ID NO 238  1 ------------------------------------------------------------
BrassicarapaBra034828  SEQ ID NO 239  1 ------------------------------------------------------------
At5g06250              SEQ ID NO 240  1 ------------------------------------------------------------
Bra005886              SEQ ID NO 241  1 ------------------------------------------------------------
GmLoc102660503         SEQ ID NO 242  1 -------MIGVEKVTICMRIEVNT------------------------EKGRR------A
HvMLOC_38822           SEQ ID NO 243  1 ------------------------------------------------------------
os01g0693400           SEQ ID NO 244  1 ----------MDSS-SCLVDDTNSGGSS---------TDKL---------RALAAAAAET
HvMLOC44012            SEQ ID NO 245  1 ------------------------------------------------------------
HvMLOC_7940            SEQ ID NO 246  1 --------MGVEIL-SSTGEHSS------QYSSGAASTATT-------ESGVGGRPPTAP
HvMLOC_75135           SEQ ID NO 247  1 --------MGMEIL-SSTVEHCS------QYSSSA-STATT-------ESGAAGRSTTAL
TRAECDM81004           SEQ ID NO 248  1 --------MGVEIL-SSMVEDSS------QYSSGA-STATT-------ESGTTGRALTAL
HvMLOC_56567           SEQ ID NO 249  1 --------MGVEIL-SSMVEHSF------QYSSGA-SSATA-------ESGAVGTPPRHL
TRAES3BF098300010CFD_t1 SEQ ID NO 250 1 --------MGVEIL-SSMVEHSF------QYSSGV-STATT-------ESGTAGTPPRPL
HvMLOC_63261           SEQ ID NO 251  1 MASSKPTNP--EVD-ND-----M------ECSS-------P-------ESGAEDAV-ESS
TRAES3BF062700040CFD_t1 SEQ ID NO 252 1 MASGKPTNHGMEDD-ND-----M------EYSS-------A-------ESGAEDAA-EPS
TRAES3BF062600010CFD_t1 SEQ ID NO 253 1 MASGKPTNHGMEDD-ND-----M------EYSS-------A-------ESGAEDAA-EPS
Bra038346              SEQ ID NO 254  1 -----------MVF-SCIDESSST---SESFSPAT-ATAT-------ATATKFSAPPLPP
GmLoc732601            SEQ ID NO 255  1 ----------MDGG-CVTDETT-TSSDS--------------------LSV------PP-
GmLoc100789009         SEQ ID NO 256  1 ----------MDGG-SVTDETT-TTSNS--------------------LSVPANLSPPP-
GmLoc100776987         SEQ ID NO 257  1 ----------MDAI-SCLDESTTTESLSIS---------QAKPSSTIMSSEKASPSPPPP
GmLoc100801107         SEQ ID NO 258  1 ----------MDAI-SCMDESTTTESLSISLSPTS-SSEKAKPSSMITSSEKVSLSPPPS
```

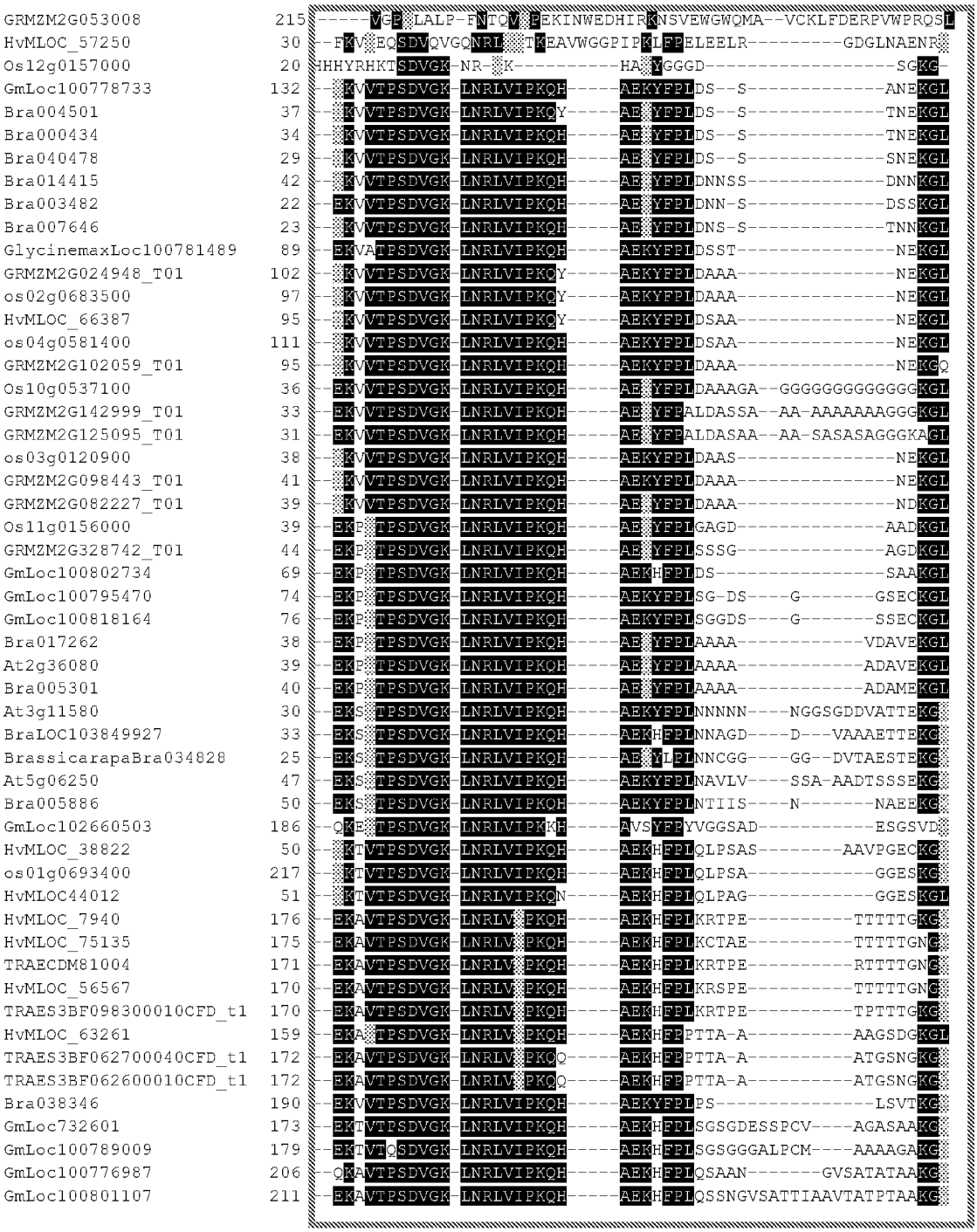
```
GRMZM2G053008                17 TPPSPWTITDGAISGTLPAAEAFAVHYPGYPSSP---ARAARTLGGLPGLAKVRSSDPGA
HvMLOC_57250                  1 ------------------------------------------------------------
Os12g0157000                  1 ------------------------------------------------------------
GmLoc100778733                1 ------------------------------------------------------------
Bra004501                     1 ------------------------------------------------------------
Bra000434                     1 ------------------------------------------------------------
Bra040478                     1 ------------------------------------------------------------
Bra014415                     1 ------------------------------------------------------------
Bra003482                     1 ------------------------------------------------------------
Bra007646                     1 ------------------------------------------------------------
GlycinemaxLoc100781489        1 ------------------------------------------------------------
GRMZM2G024948_T01             1 ------------------------------------------------------------
os02g0683500                  1 ------------------------------------------------------------
HvMLOC_66387                  1 ------------------------------------------------------------
os04g0581400                  1 ------------------------------------------------------------
GRMZM2G102059_T01             1 ------------------------------------------------------------
Os10g0537100                  1 ------------------------------------------------------------
GRMZM2G142999_T01             1 ------------------------------------------------------------
GRMZM2G125095_T01             1 ------------------------------------------------------------
os03g0120900                  1 ------------------------------------------------------------
GRMZM2G098443_T01             1 ------------------------------------------------------------
GRMZM2G082227_T01             1 ------------------------------------------------------------
Os11g0156000                  1 ------------------------------------------------------------
GRMZM2G328742_T01             1 ------------------------------------------------------------
GmLoc100802734                1 ------------------------------------------------------------
GmLoc100795470                1 ------------------------------------------------------------
GmLoc100818164                1 ------------------------------------------------------------
Bra017262                     1 ------------------------------------------------------------
At2g36080                     1 ------------------------------------------------------------
Bra005301                     1 ------------------------------------------------------------
At3g11580                     1 ------------------------------------------------------------
BraLOC103849927               1 ------------------------------------------------------------
BrassicarapaBra034828         1 ------------------------------------------------------------
At5g06250                     1 ------------------------------------------------------------
Bra005886                     1 ------------------------------------------------------------
GmLoc102660503               24 LMDC-WQIS-G-------------VHESSDC-SEIK-----FAFDAVVKRAR---HEENN
HvMLOC_38822                  1 ------------------------------------------------------------
os01g0693400                 32 APLE-RMGS-GA----------SAVVDAAEPGAEADSGSGGRVCGGGGGGAG---GAGGK
HvMLOC44012                   1 ------------------------------------------------------------
HvMLOC_7940                  39 SLP---------------------VSIAD---ES--------------ATSR---SASAQ
HvMLOC_75135                 38 SLP---------------------VAITD---ES--------------VTSR---SASAQ
TRAECDM81004                 38 SLP---------------------VAIAD---ES--------------VTS-------AQ
HvMLOC_56567                 38 SLP---------------------VAIAD---ES--------------LTS--------R
TRAES3BF098300010CFD_t1      38 SLP---------------------VAIAD---ES--------------VTS--------R
HvMLOC_63261                 32 ------------------------------------------------------SPVA
TRAES3BF062700040CFD_t1      34 SSP-----------------------------------------------VL---APPRA
TRAES3BF062600010CFD_t1      34 SSP-----------------------------------------------VL---APPRA
Bra038346                    38 LRLN-RMRS-GG----------SNVVLDSKNG-----------------VD---IDSRK
GmLoc732601                  22 --PS-RVGS-VA----------SAVVDPDGCCVS----------------GE---AESRK
GmLoc100789009               28 --LS-LVGS-GA----------TAVVYPDGCCVS----------------GE---AESRK
GmLoc100776987               41 NRLC-RVGS-GA----------SAVVDSDGGGGG--------------GSTE---VESRK
GmLoc100801107               49 NRLC-RVGS-GA----------SAVVDPDGGGSG----------------AE---VESRK
```

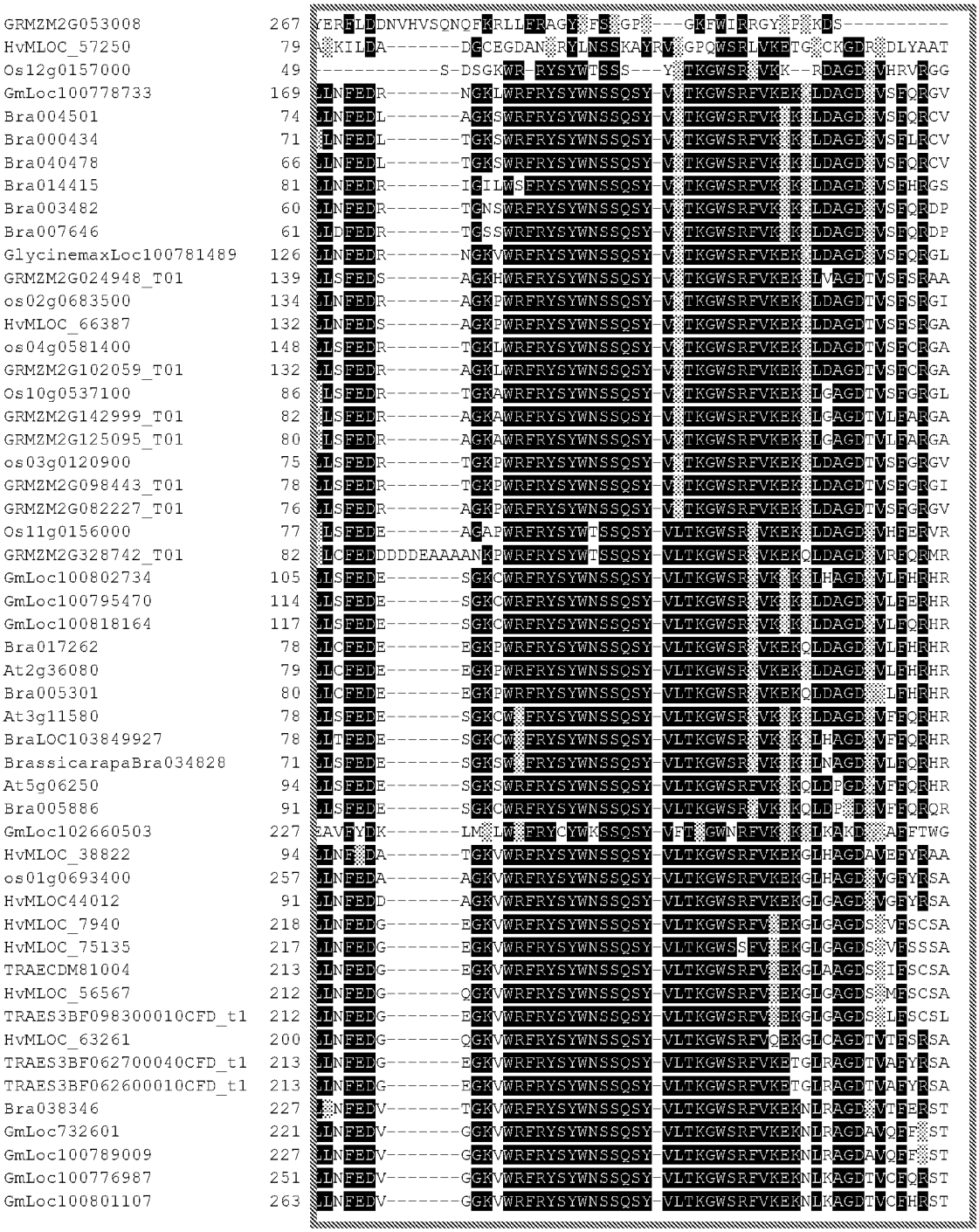
```
GRMZM2G053008              74 RLELRFRPEDPYCHPAFGQSRAST---GLLLRLSKRKGAAAPCAHVVA------------
HvMLOC_57250                1 ------------------------------------------------------------
Os12g0157000                1 ------------------------------------------------------------
GmLoc100778733              1 ---------------------MELMQ-EVKG-YSDGREEEEEEEEAAEEII----------
Bra004501                   1 ------------------------------------------------------------
Bra000434                   1 ------------------------------------------------------------
Bra040478                   1 ------------------------------------------------------------
Bra014415                   1 ------------------------------------------------------------
Bra003482                   1 ------------------------------------------------------------
Bra007646                   1 ------------------------------------------------------------
GlycinemaxLoc100781489      1 ---------------------MELMQ-QVKGNYSDSREEEE--------------------
GRMZM2G024948_T01           1 ---------------------MDQFA--ASGRFSREEEADE--------------------
os02g0683500                1 ---------------------MEFTT---SSRFSKE--EED--------------------
HvMLOC_66387                1 ---------------------MEFTA--TSSRFSKGEEEVE--------------------
os04g0581400                1 ---------------------MEFAT--TSSRFSKEEEEEEEGEQEMEQEQ----------
GRMZM2G102059_T01           1 ---------------------MEFAS--SSSRFSREEDEEEE--------Q----------
Os10g0537100                1 ------------------------------------------------------------
GRMZM2G142999_T01           1 ------------------------------------------------------------
GRMZM2G125095_T01           1 ------------------------------------------------------------
os03g0120900                1 ------------------------------------------------------------
GRMZM2G098443_T01           1 ------------------------------------------------------------
GRMZM2G082227_T01           1 ------------------------------------------------------------
Os11g0156000                1 ------------------------------------------------------------
GRMZM2G328742_T01           1 ------------------------------------------------------------
GmLoc100802734              1 ------------------------------------------------------------
GmLoc100795470              1 ------------------------------------------------------------
GmLoc100818164              1 ------------------------------------------------------------
Bra017262                   1 ------------------------------------------------------------
At2g36080                   1 ------------------------------------------------------------
Bra005301                   1 ------------------------------------------------------------
At3g11580                   1 ------------------------------------------------------------
BraLOC103849927             1 ------------------------------------------------------------
BrassicarapaBra034828       1 ------------------------------------------------------------
At5g06250                   1 ------------------------------------------------------------
Bra005886                   1 ------------------------------------------------------------
GmLoc102660503             60 AAAQKFKGVVSQQNGNWGAQIYAHQQRIWLGTFKSEREAAMAYDSASIKLRSGECHRNFP
HvMLOC_38822                1 ------------------------------------------------------------
os01g0693400               77 LPSSKFKGVVPQPNGRWGAQIYERHQRVWLGTFAGEDDAARAYDVAAQRFRGRDAVTNFR
HvMLOC44012                 1 ------------------------------------------------------------
HvMLOC_7940                58 STSSRFKGVVPQPNGRWGAQIYERHARVWLGTFPDEDSAARAYDVAALRYRGREAATNFP
HvMLOC_75135               57 PASSRFKGVVPQPNGRWGSQIYERHARVWLGTFPDQDSAARAYDVASLRYRGRDAATNFP
TRAECDM81004               53 SAPSRFKGVVPQPNGRWGSQIYERHARVWLGTFPDQDLAARAYDVAALRYRGRDAATNFP
HvMLOC_56567               52 SASSRFKGVVPQPNGRWGAQIYERHARVWLGTFPDQDSAARAYDVASLRYRGGDAAFNFP
TRAES3BF098300010CFD_t1    52 SASSRFKGVVPQPNGRWGAQIYERHARVWLGTFPDQDSAARAYDVASLRYRGRDVAFNFP
HvMLOC_63261               36 APSSRFKGVVPQPNGRWGAQIYEKHSRVWLGTFGDEEAAACAYDVAALRFRGRDAVTNHQ
TRAES3BF062700040CFD_t1    44 APSSRFKGVVPQPNGRWGAQIYEKHSRVWLGTFPDEDAAVRAYDVAALRFRGPDAVINHQ
TRAES3BF062600010CFD_t1    44 APSSRFKGVVPQPNGRWGAQIYEKHSRVWLGTFPDEDAAARAYDVAALRFRGPDAVINHQ
Bra038346                  65 LSSSKYKGVVPQPNGRWGAQIYVKHQRVWLGTFCDEEEAAHSYDIAARKFRGRDAVVNFK
GmLoc732601                49 LPSSKYKGVVPQPNGRWGAQIYEKHQRVWLGTFNEEDEAARAYDIAALRFRGPDAVTNFK
GmLoc100789009             55 LPSSKYKGVVPQPNGRWGAQIYEKHQRVWLGTFNEEDEAARAYDIAAHRFRGRDAVTNFK
GmLoc100776987             72 LPSSKYKGVVPQPNGRWGSQIYEKHQRVWLGTFNEEDEAARAYDVAVQRFRGKDAVTNFK
GmLoc100801107             78 LPSSKYKGVVPQPNGRWGAQIYEKHQRVWLGTFNEEDEAARAYDIAAQRFRGKDAVTNFK
```

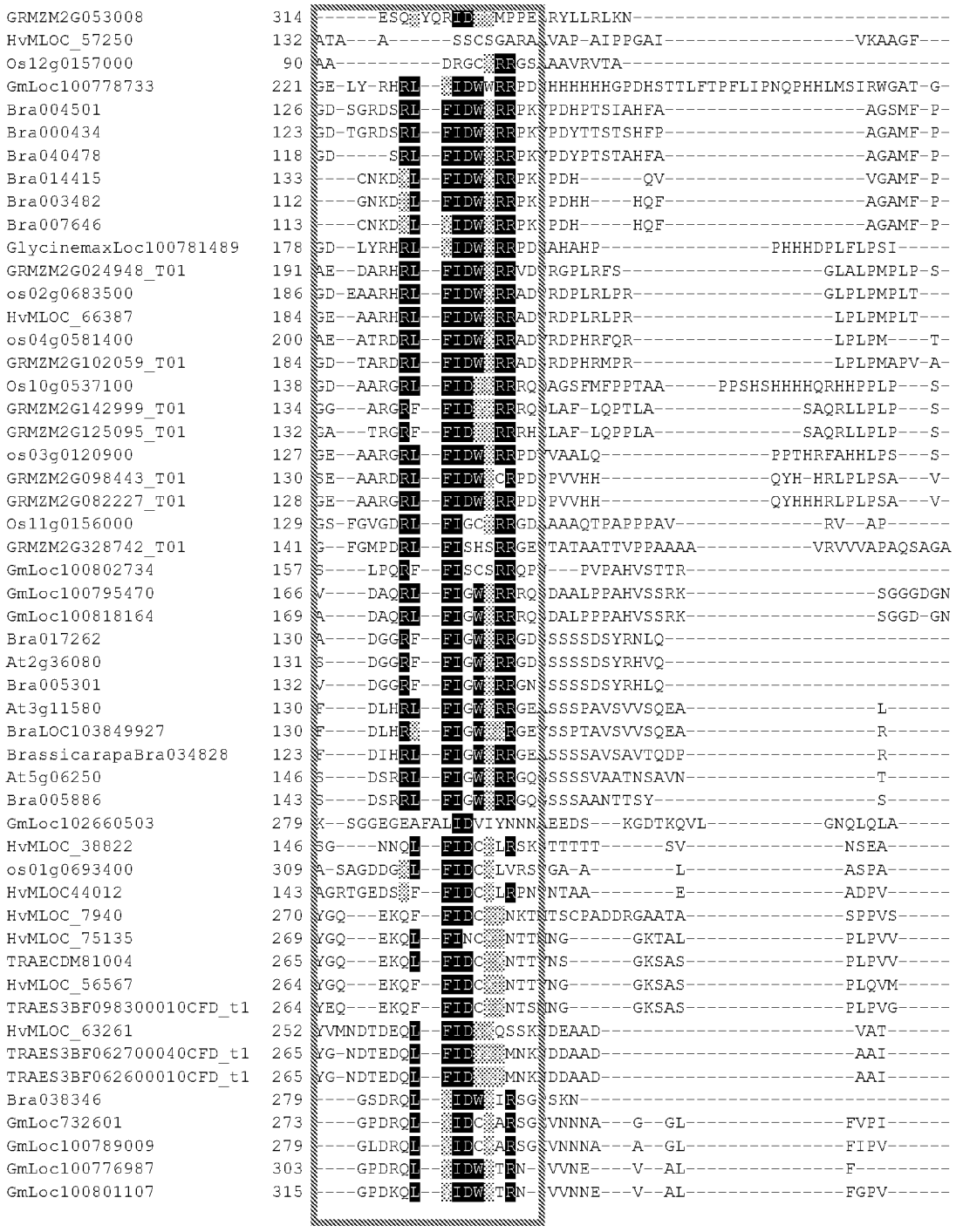

GRMZM2G053008 119 ----------R RTAYYFEGMADFQ---HVVPVHA
HvMLOC_57250 1 ----------
Os12g0157000 1 ----------
GmLoc100778733 29 ----------TREESSRLLHQHQEAAGSNFI NNNHHHHQHHHH----------
Bra004501 1 ----------
Bra000434 1 ----------
Bra040478 1 ----------
Bra014415 1 ----------
Bra003482 1 ----------
Bra007646 1 ----------
GlycinemaxLoc100781489 20 ----------EE----------EAA----------
GRMZM2G024948_T01 19 ----------EQE----------DAS----------
os02g0683500 16 ----------EE----------QDE----------
HvMLOC_66387 19 ----------EE----------QEE----------
os04g0581400 29 ----------DEEEE----------EAE----------
GRMZM2G102059_T01 21 ----------EEEEE----------EEE----------
Os10g0537100 1 ----------
GRMZM2G142999_T01 1 ----------
GRMZM2G125095_T01 1 ----------
os03g0120900 1 ----------
GRMZM2G098443_T01 1 ----------
GRMZM2G082227_T01 1 ----------
Os11g0156000 1 ----------AMNHPL-FSQEQP---QSWPWGV
GRMZM2G328742_T01 1 ----------ATNHLSQGQHQHP---QAWPWGV
GmLoc100802734 1 ----------MSSINHY---SPETT---LYWTNDQ
GmLoc100795470 1 ----------SINHYSMDLPEPT---LWWPHPH
GmLoc100818164 1 ----------STNHYTMDLPEPT---LWWPHPH
Bra017262 1 ----------SINQYSSEFYYH---SLMWQQQQ
At2g36080 1 ----------SINQYSSDFHYH---SLMWQQQQ
Bra005301 1 ----------SINQYSSDFNYH---SLMWQQQQ
At3g11580 1 ----------SVNHYHN-----TLS----LH--
BraLOC103849927 1 ----------SGNHYSRDIHHNTPS----V---
BrassicarapaBra034828 1 ----------SVNHYSNTL-----S----SH--
At5g06250 1 ----------SVNHYSTDHHHTL----LWQQQQ
Bra005886 1 ----------SVNHYSTDHHQVHHHHTLFLQ--
GmLoc102660503 120 WNDQ----------TVQEPQFQSHYSAETVLN-M RDGTYPSKFAT----------
HvMLOC_38822 1 ----------M RKHIYPDELAQ----------
os01g0693400 137 PLAEADP--------DAAAELRFLATRSKAEVVD-M RKHTYFDELAQ----------
HvMLOC44012 1 ----------M RKHTYFDELAQ----------
HvMLOC_7940 118 CAA----------AEAELAFLAAHSKAEIVD-M RKHTYTDELRQ----------
HvMLOC_75135 117 CAA----------AEAELAFLTAHSKAEIVD-M RKHTYADELRQ----------
TRAECDM81004 113 CAA----------AEAELAFLGAHSKAEIVD-M RKHTYADELRQ----------
HvMLOC_56567 112 CVV----------VEAELAFLAAHSKAEIVD-M RKQTYADELRQ----------
TRAES3BF098300010CFD_t1 112 CAA----------VEGELAFLAAHSKAEIVD-M RKQTYADELRQ----------
HvMLOC_63261 96 RLPAAEGAGWSS-----TSELAFLADHSKAEIVD-M RKHTYDDELRQ----------
TRAES3BF062700040CFD_t1 104 RPTAAEEAGSSSSRSELDPELGFLADHSKAEIVD-M RKHTYDDELRQ----------
TRAES3BF062600010CFD_t1 104 RPTAAEEAGSSSSRSELDPELGFLADHSKAEIVD-M RKHTYDDELRQ----------
Bra038346 125 TFLASED---------DNGELCFLEAHSKAEIVD-M RKHTYADELAQ----------
GmLoc732601 109 PPAASD-----------DAESEFLNSHSKFEIVD-M RKHTYDDELQQ----------
GmLoc100789009 115 PLAGAD-----------DAEAEFLSTHSKSEIVD-M RKHTYDNELQQ----------
GmLoc100776987 132 PLSGTDD---------DDGESEFLNSHSKSEIVD-M RKHTYNDELEQ----------
GmLoc100801107 138 PLAGADD---------DDGESEFLNSHSKPEIVD-M RKHTYNDELEQ----------


GRMZM2G053008 141 AQT KR KHSDSQNDNENF
HvMLOC_57250 1 MYC-SR GRIDPAE E
Os12g0157000 1
GmLoc100778733 63 HTT QLDFMDLSLGSSKDEGNLQ GSSSSVYAHH HHAASA SS SAN
Bra004501 1 MMMTNLS L SRE
Bra000434 1 MMTNLS L ARE
Bra040478 1 MMTNLS L ARE
Bra014415 1 MERKSNDL ERSE NI DSQ
Bra003482 1
Bra007646 1
GlycinemaxLoc100781489 25 AIT ESESSRLHQQ DTASNFGKKLDLMDLSLGS SKEE EEE
GRMZM2G024948_T01 25 NSM EISFMPPAAASSSSAAASASA SASTSASACASGSSSAPFRSASASG DAA
os02g0683500 21 AGR EIPFMTATAEAAPAPTSSSSSPAHHAASASASASAS GSSTPFRSD DGA
HvMLOC_66387 24 ASM EIPFMTPAAATCAA APPSASASASTPASAS GSSPPFRSG DDA
os04g0581400 37 ASP EIPFMTSAAAAATASSSSPTSV SPSATASAAA STSASGSPFRSS DGA
GRMZM2G102059_T01 29 ASP EIPFMTAAATADTGAAASSSS PSA A ASSGPAAAPRSS DGA
Os10g0537100 1 MEFTPIS PP TRVAGGEE D
GRMZM2G142999_T01 1 MEFTPAH AH ARVVE D
GRMZM2G125095_T01 1 MEFRPA H ARVFE D
os03g0120900 1 MEFITPIVR PASAAAGGGEV QE
GRMZM2G098443_T01 1 MEFTTPP PATRSGGGEE RA
GRMZM2G082227_T01 1 MEFTAPP PATRSGGGEE RA
Os11g0156000 21 AM
GRMZM2G328742_T01 22 AM
GmLoc100802734 20 QQQAAMWLSNSH TPRFNLNDEEEEEEDDV
GmLoc100795470 22 HQQQQLTLMDPD PLRLNLNSDDG NGNDN
GmLoc100818164 22 Q QQLTLIDPD PLPLNLNNDDNDNGDDN
Bra017262 22 Q H
At2g36080 22 QQ QQ
Bra005301 22 HRH H
At3g11580 14 HHH
BraLOC103849927 18 HH
BrassicarapaBra034828 14 NHH
At5g06250 21 HRHTT
Bra005886 23 NLHTT
GmLoc102660503 155 FL TRQTQKGV AK
HvMLOC_38822 14 HK AFFFAAAS SPTSSS S
os01g0693400 176 SK TFAASTPS AATTTA SL
HvMLOC44012 14 SK AFAASAAL SAPTTS
HvMLOC_7940 152 GL RGRGMGAR
HvMLOC_75135 151 GL RGRGMGAR
TRAECDM81004 147 GL RGRGMGAR
HvMLOC_56567 146 GL RGRGMGVR
TRAES3BF098300010CFD_t1 146 GL RGRGMGAR
HvMLOC_63261 138 GL RGHG R
TRAES3BF062700040CFD_t1 151 GL RGRG R
TRAES3BF062600010CFD_t1 151 GL RGRG R
Bra038346 163 SN RSGANTN
GmLoc732601 145 ST GGR RR
GmLoc100789009 151 ST GGR RR
GmLoc100776987 170 SK SRGFVRRR GSAA
GmLoc100801107 176 SK SRGVVRRR GSAA


GRMZM2G053008 159 GSDKTGHDEADGDVMMLVPPLFSVKDRPTKIALVPSSNAISKTMHRGVVQEIWEIN----
HvMLOC_57250 14 G----QVMGGLG----------VRDASWAL-----
Os12g0157000 1 ----------MAMHAGHAWWGVAITNHY
GmLoc100778733 107 G--NNNNSSSSN----------LQQQQQQPAEIEHIF----
Bra004501 12 G--EEEEEEEQE----------EAKKPMEEVEIEHIF----
Bra000434 11 G--EEEEEE--A----------GAKKPTEEVEIEHIF----
Bra040478 11 G--EA----------QVKKPIEEVEIEHIF----
Bra014415 18 N--KKMNLEEER----------PVQE-ASSMEIEHIF----
Bra003482 1 -----MNQEEEN----------PVEK-ASSMEIEHIF----
Bra007646 1 -----MNQENKK----------PLEEASTSMEIENIF----
GlycinemaxLoc100781489 65 G--NLQQ-GGGG----------VVHHAHQVVEIEHIF----
GRMZM2G024948_T01 78 G--ASGSGGP-A----------DADAEAEAVEIEHIF----
os02g0683500 73 G--ASGSGGGGG----------G-GGEAEVVEIEHIF----
HvMLOC_66387 70 G--ASGSGAGDG----------SRSNVAEAVEIEHIF----
os04g0581400 88 G--ASGSGGGGG----------G--EDVEVIEIEHIF----
GRMZM2G102059_T01 73 G--ASGSGG-GG----------S--DDVQVIEIEHIF----
Os10g0537100 19 ------S----E----------RGAAAWAVVEIEHIF----
GRMZM2G142999_T01 16 ------S----E----------RPRGGVAWVEIEHIF----
GRMZM2G125095_T01 14 ------S----E----------RPRGGVAWIEIEHIF----
os03g0120900 23 ----------SGGRSLAAVEIEHIF----
GRMZM2G098443_T01 20 A--AEHN----Q----------HHQQQHATVEIEHIF----
GRMZM2G082227_T01 20 A--AEHH----Q----------QQQQATVEIEHIF----
Os11g0156000 23 -------YANFH----------YH----HHYEIEHIF----
GRMZM2G328742_T01 24 -------YTNLH----------YHHQQHHHYEIEHLF----
GmLoc100802734 49 -------IVSDK----------ATNNLTQEEEIVAIF----
GmLoc100795470 50 DNDENQ---TTT----------TGGEQEILDIEPIF----
GmLoc100818164 49 DNDENQTVTTTT----------TGGEEEIINNIEPIF----
Bra017262 25 ----------HQNEVVEEIEALF----
At2g36080 26 ----------HQNDVVEEIEALF----
Bra005301 27 ----------HQNDVAEEIEALF----
At3g11580 17 ----------QNDVAIAQIESLF----
BraLOC103849927 20 ----------HQNYAVVIEYLF----
BrassicarapaBra034828 17 ----------NEHIESLF----
At5g06250 26 -----DTSE-TT----------TTATWLHDDLIESLF----
Bra005886 28 -----DTSEPTT----------TAATSLREDQIEYLF----
GmLoc102660503 168 ---------HIG----------LKGDDEEQFCCTQLF----
HvMLOC_38822 32 ---------PLA----------SPAPSAAAARIEHLF----
os01g0693400 195 -----SNGHLSS----------PRSPFAPAAAIHLF----
HvMLOC44012 31 -------GDAGG----------SASPPSPAAVIEHLF----
HvMLOC_7940 163 ----------AQPTPSWAIEPLF----
HvMLOC_75135 162 ----------AQPTPSWAIVPLF----
TRAECDM81004 158 ----------AQPTPSWAIEPLF----
HvMLOC_56567 157 ----------AQPMPSWAIVPLF----
TRAES3BF098300010CFD_t1 157 ----------AQPTPSWAIEPLF----
HvMLOC_63261 146 ----------AQPTPAWAIEFLF----
TRAES3BF062700040CFD_t1 159 ----------AQPTPAWAIELLF----
TRAES3BF062600010CFD_t1 159 ----------AQPTPAWAIELLF----
Bra038346 173 ----------TN----------TTQSHTVSRTIEVLF----
GmLoc732601 153-------LDADT----------ASSGVFDAKAIEQLF----
GmLoc100789009 159 -------RDAET----------ASSGAFDAKAIEQLF----
GmLoc100776987 185 ------GAGNGN----------SISGACVMKAIEQLF----
GmLoc100801107 191 -------AGTAN----------SISGACFTKAIEQLF----


GRMZM2G053008 215
HvMLOC_57250 30
Os12g0157000 20
GmLoc100778733 132
Bra004501 37
Bra000434 34
Bra040478 29
Bra014415 42
Bra003482 22
Bra007646 23
GlycinemaxLoc100781489 89
GRMZM2G024948_T01 102
os02g0683500 97
HvMLOC_66387 95
os04g0581400 111
GRMZM2G102059_T01 95
Os10g0537100 36
GRMZM2G142999_T01 33
GRMZM2G125095_T01 31
os03g0120900 38
GRMZM2G098443_T01 41
GRMZM2G082227_T01 39
Os11g0156000 39
GRMZM2G328742_T01 44
GmLoc100802734 69
GmLoc100795470 74
GmLoc100818164 76
Bra017262 38
At2g36080 39
Bra005301 40
At3g11580 30
BraLOC103849927 33
BrassicarapaBra034828 25
At5g06250 47
Bra005886 50
GmLoc102660503 186
HvMLOC_38822 50
os01g0693400 217
HvMLOC44012 51
HvMLOC_7940 176
HvMLOC_75135 175
TRAECDM81004 171
HvMLOC_56567 170
TRAES3BF098300010CFD_t1 170
HvMLOC_63261 159
TRAES3BF062700040CFD_t1 172
TRAES3BF062600010CFD_t1 172
Bra038346 190
GmLoc732601 173
GmLoc100789009 179
GmLoc100776987 206
GmLoc100801107 211

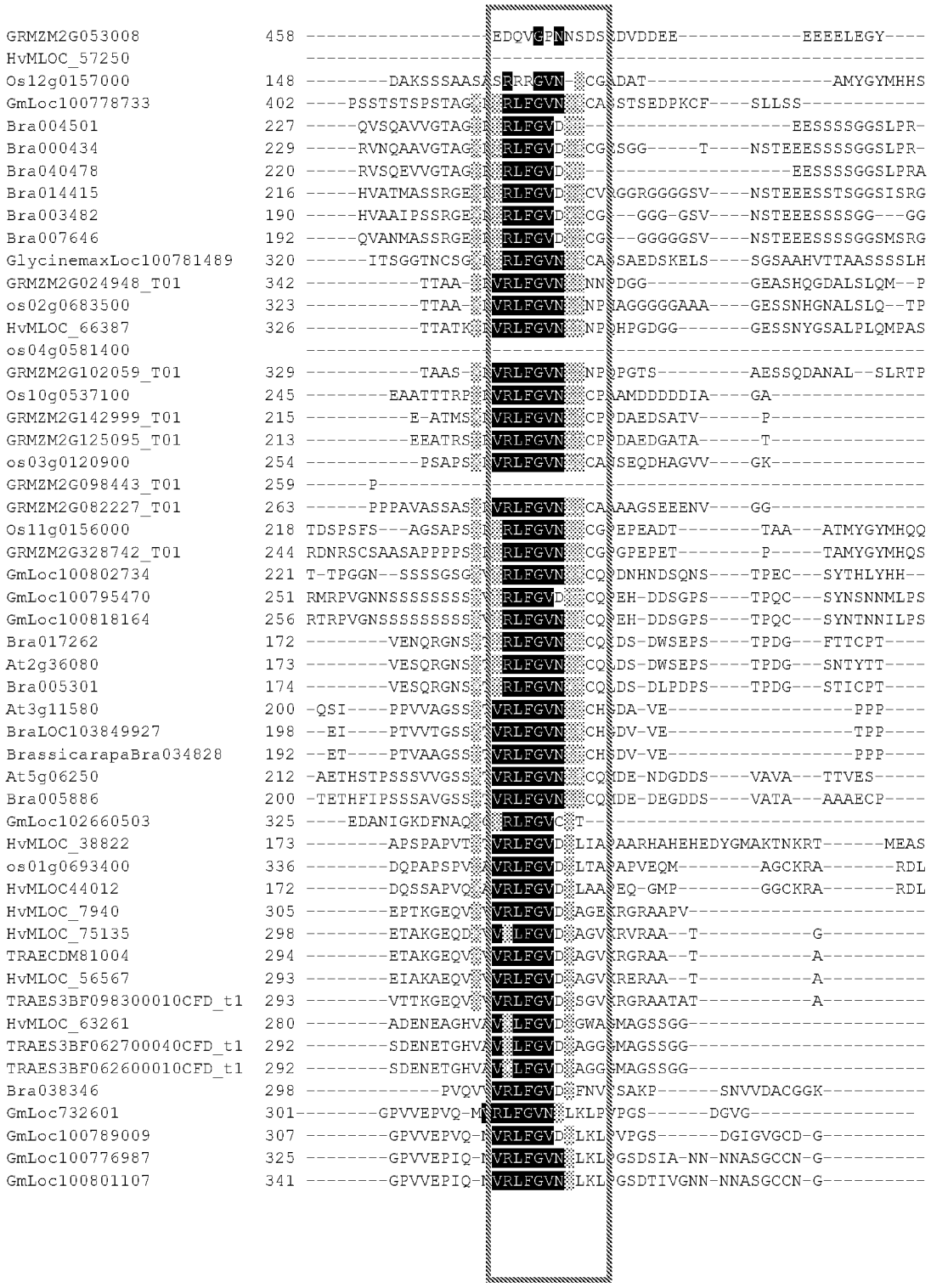

GRMZM2G053008 267
HvMLOC_57250 79
Os12g0157000 49
GmLoc100778733 169
Bra004501 74
Bra000434 71
Bra040478 66
Bra014415 81
Bra003482 60
Bra007646 61
GlycinemaxLoc100781489 126
GRMZM2G024948_T01 139
os02g0683500 134
HvMLOC_66387 132
os04g0581400 148
GRMZM2G102059_T01 132
Os10g0537100 86
GRMZM2G142999_T01 82
GRMZM2G125095_T01 80
os03g0120900 75
GRMZM2G098443_T01 78
GRMZM2G082227_T01 76
Os11g0156000 77
GRMZM2G328742_T01 82
GmLoc100802734 105
GmLoc100795470 114
GmLoc100818164 117
Bra017262 78
At2g36080 79
Bra005301 80
At3g11580 78
BraLOC103849927 78
BrassicarapaBra034828 71
At5g06250 94
Bra005886 91
GmLoc102660503 227
HvMLOC_38822 94
os01g0693400 257
HvMLOC44012 91
HvMLOC_7940 218
HvMLOC_75135 217
TRAECDM81004 213
HvMLOC_56567 212
TRAES3BF098300010CFD_t1 212
HvMLOC_63261 200
TRAES3BF062700040CFD_t1 213
TRAES3BF062600010CFD_t1 213
Bra038346 227
GmLoc732601 221
GmLoc100789009 227
GmLoc100776987 251
GmLoc100801107 263

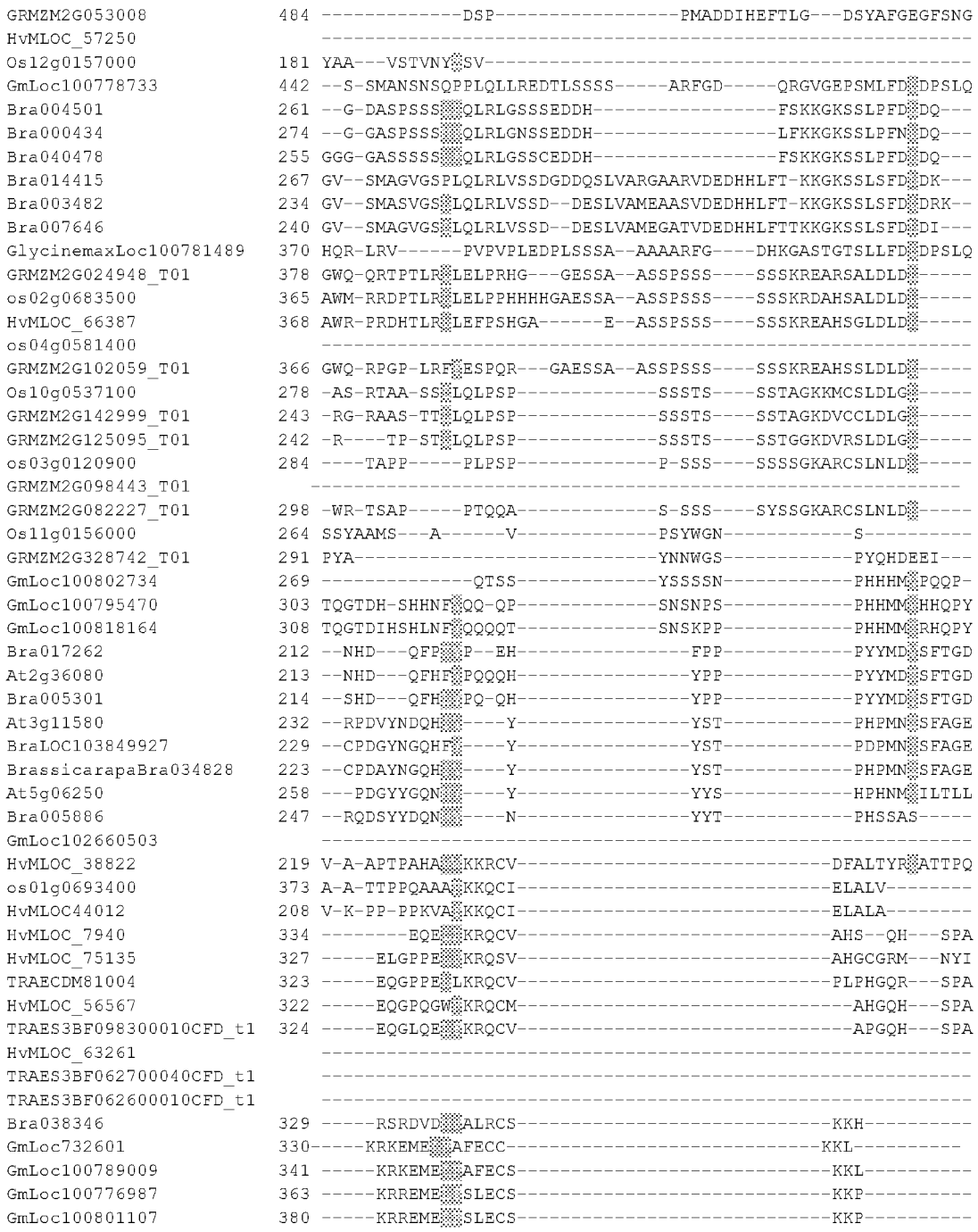

GRMZM2G053008 314 -----ESQ YQR ID MPPE RYLLRLKN
HvMLOC_57250 132 ATA---A------SSCSGARA VAP-AIPPGAI------------------VKAAGF---
Os12g0157000 90 AA----------DRGC RR GS AAVRVTA-----------------------------
GmLoc100778733 221 GE-LY-RHRL--IDW RR PD HHHHHHGPDHSTTLFTPFLIPNQPHHLMSIRWGAT-G-
Bra004501 126 GD-SGRDSRL--FIDW RR PK PDHPTSIAHFA------------------AGSMF-P-
Bra000434 123 GD-TGRDSRL--FIDW RR PK PDYTTSTSHFP-----------------AGAMF-P-
Bra040478 118 GD-----SRL--FIDW RR PK PDYPTSTAHFA-----------------AGAMF-P-
Bra014415 133 ----CNKD L--FIDW RR PK PDH------QV-----------------VGAMF-P-
Bra003482 112 ----GNKD L--FIDW RR PK PDHH----HQF-----------------AGAMF-P-
Bra007646 113 ----CNKD L--IDW RR PK PDH-----HQF-----------------AGAMF-P-
GlycinemaxLoc100781489 178 GD--LYRHRL--IDW RR PD AHAHP----------------PHHHDPLFLPSI-----
GRMZM2G024948_T01 191 AE--DARHRL--FIDW RR VD RGPLRFS-------------------GLALPMPLP-S-
os02g0683500 186 GD-EAARHRL--FIDW RR AD RDPLRLPR-----------------GLPLPMPLT---
HvMLOC_66387 184 GE--AARHRL--FIDW RR AD RDPLRLPR-----------------LPLPMPLT---
os04g0581400 200 AE--ATRDRL--FIDW RR AD RDPHRFQR-----------------LPLPM----T-
GRMZM2G102059_T01 184 GD--TARDRL--FIDW RR AD RDPHRMPR-----------------LPLPMAPV-A-
Os10g0537100 138 GD--AARGRL--FID RR RQ AGSFMFPPTAA-----PPSHSHHHHQRHHPPLP---S-
GRMZM2G142999_T01 134 GG---ARGRF--FID RR RQ LAF-LQPTLA--------------SAQRLLPLP---S-
GRMZM2G125095_T01 132 GA---TRGRF--FID RR RH LAF-LQPPLA--------------SAQRLLPLP---S-
os03g0120900 127 GE--AARGRL--FIDW RR PD VAALQ----------------PPTHRFAHHLPS---S-
GRMZM2G098443_T01 130 SE--AARDRL--FIDW CR PD PVVHH----------------QYH-HRLPLPSA---V-
GRMZM2G082227_T01 128 GE--AARGRL--FIDW RR PD PVVHH----------------QYHHHRLPLPSA---V-
Os11g0156000 129 GS-FGVGDRL--FIGC RR GD AAAQTPAPPPAV--------------RV--AP------
GRMZM2G328742_T01 141 G--FGMPDRL--FISHS RR GE TATAATTVPPAAAA-----------VRVVVAPAQSAGA
GmLoc100802734 157 S----LPQRF--FISCS RR QP ---PVPAHVSTTR-----------------------
GmLoc100795470 166 V----DAQRL--FIGW RR RQ DAALPPAHVSSRK-------------------SGGGDGN
GmLoc100818164 169 A----DAQRL--FIGW RR RQ DALPPPAHVSSRK-------------------SGGD-GN
Bra017262 130 A----DGGRF--FIGW RR GD SSSSDSYRNLQ--------------------------
At2g36080 131 S----DGGRF--FIGW RR GD SSSSDSYRHVQ--------------------------
Bra005301 132 V----DGGRF--FIGW RR GN SSSSDSYRHLQ--------------------------
At3g11580 130 F----DLHRL--FIGW RR GE SSSPAVSVVSQEA------------------L------
BraLOC103849927 130 F----DLHR--FIGW RGE SSPTAVSVVSQEA------------------R------
BrassicarapaBra034828 123 F----DIHRL--FIGW RR GE SSSSAVSAVTQDP------------------R------
At5g06250 146 S----DSRRL--FIGW RR GQ SSSSVAATNSAVN-----------------T------
Bra005886 143 S----DSRRL--FIGW RR GQ SSSAANTTSY---------------------S------
GmLoc102660503 279 K--SGGEGEAFALIDVIYNNN EEDS---KGDTKQVL-----------GNQLQLA-----
HvMLOC_38822 146 SG----NNQL--FIDC LR SK TTTTT------SV---------------NSEA------
os01g0693400 309 A-SAGDDG L--FIDC LVRS GA-A--------L---------------ASPA------
HvMLOC44012 143 AGRTGEDS F--FIDC LR PN NTAA--------E---------------ADPV------
HvMLOC_7940 270 YGQ---EKQF--FIDC NK TN TSCPADDRGAATA---------------SPPVS-----
HvMLOC_75135 269 YGQ---EKQL--FINC NTT NG------GKTAL---------------PLPVV-----
TRAECDM81004 265 YGQ---EKQL--FIDC NTT NS------GKSAS---------------PLPVV-----
HvMLOC_56567 264 YGQ---EKQF--FIDC NTT NG------GKSAS---------------PLQVM-----
TRAES3BF098300010CFD_t1 264 YEQ---EKQF--FIDC NTS NG------GKSAS---------------PLPVG-----
HvMLOC_63261 252 YVMNDTDEQL--FID QSSK DEAAD-----------------------VAT------
TRAES3BF062700040CFD_t1 265 YG-NDTEDQL--FID MNK DDAAD-----------------------AAI------
TRAES3BF062600010CFD_t1 265 YG-NDTEDQL--FID MNK DDAAD-----------------------AAI------
Bra038346 279 ----GSDRQL--IDW R SG SKN-----------------------------------
GmLoc732601 273 ----GPDRQL--IDC A SG VNNNA---G--GL---------------FVPI------
GmLoc100789009 279 ----GLDRQL--IDC A SG VNNNA---A--GL---------------FIPV------
GmLoc100776987 303 ----GPDRQL--IDW TRN- VVNE----V--AL---------------F----------
GmLoc100801107 315 ----GPDKQL--IDW TRN- VVNNE---V--AL---------------FGPV------

```
GRMZM2G053008            338 ------SESRKWADMCKLETMPSQSFIYLQLYELKDDFIQAEIRKPSYQSVCSRSTGWFS
HvMLOC_57250                 ------------------------------------------------------------
Os12g0157000             109 --------NGGWSMCYSTSG---SSYDT--S-----------------------------
GmLoc100778733           275 RLYSLPSPTPPRHHEH---------LNYNNNA-----------MYH-------------
Bra004501                162 RFYSFPTAT---SYNL---------YNYQQP-----------------------------
Bra000434                159 RFYSFQTATTSTSYNP---------YNHQQP-----------------------------
Bra040478                150 RFYSFPTATTSTCYDL---------YNHQPP-----------------------------
Bra014415                160 RFYSYPYPQIQASYER-------------------------------------------
Bra003482                141 RFYSFSHPQN-------------------------------------------------
Bra007646                141 RFYSFPHPQMPTSFES-------------------------------------------
GlycinemaxLoc100781489   213 RLYSLPPTMPPRYHHDHHFH---HHLNYNNLF-----------T---------------
GRMZM2G024948_T01        226 SHYGGPHHYSPWGFGGGGG----------------------------------GGGGFFM
os02g0683500             222 ------SHYAPWGIGGG-------------------------------------GGFFV
HvMLOC_66387             218 ------SHYSPWGLGAG-----------------------------------ARGFFM
os04g0581400             232 ------SPYGPWGGGA-GA---------------------------------SSCRPRR
GRMZM2G102059_T01        219 ------SPYGPWGGGG-GG---------------------------------GAGGFFM
Os10g0537100             185 ------VPLCPWRDYTTAYG---GGYGY-------------------------------
GRMZM2G142999_T01        170 ------VPICPWQDYG-----------------------------------------
GRMZM2G125095_T01        168 ------VPICPWQGYG-----------------------------------------
os03g0120900             163 ------IPFAPWAHHH---G---H-------G------------AAAAA-AAAAGARFLL
GRMZM2G098443_T01        165 ------VPYAPWAA--HAHH---HHYPADGHT------------EPVTPCLCATLVATEM
GRMZM2G082227_T01        164 ------VPYAPWAAAAAHAHH---HHYPAAGVG------------AARTTTTTTTTVLHHL
Os11g0156000             164 -AAQNAGEQQPWSPMCYSTS---GGGSY-------------------------------
GRMZM2G328742_T01        186 -DHQQQQQPSPWSPMCYSTS---GSYSY-------------------------------
GmLoc100802734           183 -----------SSASFYSAH---PP--Y-------------------------------
GmLoc100795470           202 SNK-----NEGWTRGFYSAH---HP--Y-------------------------------
GmLoc100818164           204 SSKNEGDVGVGWTRGFYPAH---HP--Y-------------------------------
Bra017262                157 ------------------------------------------------------------
At2g36080                158 ------------------------------------------------------------
Bra005301                159 ------------------------------------------------------------
At3g11580                160 -----VNTTAYWSGLT-------TP--Y-------------------------------
BraLOC103849927          160 -----VNTTAYWSGLT-------TP--Y-------------------------------
BrassicarapaBra034828    153 -----ANTTAYWNGLT-------TP--Y-------------------------------
At5g06250                176 -----SS------MGA-------LS--Y-------------------------------
Bra005886                170 -----SS------MT--------------------------------------------
GmLoc102660503           318 ------------------------------------------------------------
HvMLOC_38822             173 ------------------------------------------------------------
os01g0693400             336 ------------------------------------------------------------
HvMLOC44012              172 ------------------------------------------------------------
HvMLOC_7940              305 ------------------------------------------------------------
HvMLOC_75135             298 ------------------------------------------------------------
TRAECDM81004             294 ------------------------------------------------------------
HvMLOC_56567             293 ------------------------------------------------------------
TRAES3BF098300010CFD_t1  293 ------------------------------------------------------------
HvMLOC_63261             280 ------------------------------------------------------------
TRAES3BF062700040CFD_t1  292 ------------------------------------------------------------
TRAES3BF062600010CFD_t1  292 ------------------------------------------------------------
Bra038346                298 ------------------------------------------------------------
GmLoc732601              301-------------------------------------------------------------
GmLoc100789009           307 ------------------------------------------------------------
GmLoc100776987           325 ------------------------------------------------------------
GmLoc100801107           341 ------------------------------------------------------------
```

```
GRMZM2G053008            392 KPMIKTL-RLQVSIR--------------------------LLSLLHNE----------
HvMLOC_57250                 ------------------------------------------------------------
Os12g0157000             127 ----------------------------------------ANSYAYHRSV----------
GmLoc100778733           301 ------P-FHHHGAGSGINATTHHYNNYHEMSSTTT--SGSAGSVFYHRS-TPPISMPLA
Bra004501                181 --------RHHHHSG----------YNYPQIPRE-------FGYGYLV------------
Bra000434                181 --------RH-HHSG----------YCYPQIPRE-------FGYGYVVRS-V--------
Bra040478                172 --------RH-HHIG----------YGYPQIPRE-------FGYGYFVRS-V--------
Bra014415                176 ------------H------------NLYHRYQRD-------IGIGYYVRS-M--------
Bra003482                151 --------------------------LYHRYQQD-------LGIGYYVSS-M--------
Bra007646                157 ------------SHN----------LYHHRFQRD-------LGIGYY--------------
GlycinemaxLoc100781489   243 --------FQQHQYQQLGAATTTHHNNYGY-------QNSGSGSLYYLRSSMSMGG----
GRMZM2G024948_T01        252 PPSPPATLYEH-RLR---Q-----GLDFRSMTTTYPAPTVGRQLLFFGSARMPPHHAPPP
os02g0683500             238 QPSPPATLYEH-RLR---Q-----GLDFRAFNPA-A--AMGRQVLLFGSAR-IPPQAP--
HvMLOC_66387             235 PPSPPATLYEH-RLR---Q-----GFDFRGMNPSYP--TMGRQVILFGSAARMPPHGPAP
os04g0581400             251 PPRSTSI----TAFA---R-----AST----SATS------TPLCRRGSSS----SSAPQ
GRMZM2G102059_T01        238 PPAPPATLYEHHRFR---Q-----ALDFRNINAAA---APARQLLFFGSAGMPPRASMPQ
Os10g0537100             204 --------------------------GYG----GGSTPASSRHVLFL--------------
GRMZM2G142999_T01        180 ----------------------------------ASAPAPNRHVLFL--------------
GRMZM2G125095_T01        178 ----------------------------------ASAPAPSRHVLFL--------------
os03g0120900             191 PPSST-PIYDHHRRH-------AHAVGYDAYA-----AATSRQVLFY-------------
GRMZM2G098443_T01        202 RASSS-QLSLTRSNLS--RPPQPRIARVDGAQPRPSSSPRQPQSLWC-------------
GRMZM2G082227_T01        203 PPSPS-PLYLDTRRR----------HVGYDAY------GAGTRQLLFY-------------
Os11g0156000             188 ----------PT-----------SPANSY-----------------AY-------------
GRMZM2G328742_T01        210 ----------PT-----------SSPANSQH----------------AYH------------
GmLoc100802734           195 ----------PA-----------HH---------------------F--------------
GmLoc100795470           220 ----------PT-----------HH---------------------LHH------------
GmLoc100818164           227 ----------PT-----------HH-----------------------------------
Bra017262                157 ------------------------------------------------------------
At2g36080                158 ------------------------------------------------------------
Bra005301                159 ------------------------------------------------------------
At3g11580                174 ----------RQ-----------VH---------------------AST------------
BraLOC103849927          174 ----------RQ-----------VH---------------------AST------------
BrassicarapaBra034828    167 ----------RQ-----------VH---------------------AST------------
At5g06250                184 ----------HQ-----------IH---------------------ATS------------
Bra005886                174 ----------------------AP---------------------PYS------------
GmLoc102660503           318 ------------------------------------------------------------
HvMLOC_38822             173 ------------------------------------------------------------
os01g0693400             336 ------------------------------------------------------------
HvMLOC44012              172 ------------------------------------------------------------
HvMLOC_7940              305 ------------------------------------------------------------
HvMLOC_75135             298 ------------------------------------------------------------
TRAECDM81004             294 ------------------------------------------------------------
HvMLOC_56567             293 ------------------------------------------------------------
TRAES3BF098300010CFD_t1  293 ------------------------------------------------------------
HvMLOC_63261             280 ------------------------------------------------------------
TRAES3BF062700040CFD_t1  292 ------------------------------------------------------------
TRAES3BF062600010CFD_t1  292 ------------------------------------------------------------
Bra038346                298 ------------------------------------------------------------
GmLoc732601              301------------------------------------------------------------
GmLoc100789009           307 ------------------------------------------------------------
GmLoc100776987           325 ------------------------------------------------------------
GmLoc100801107           341 ------------------------------------------------------------
```

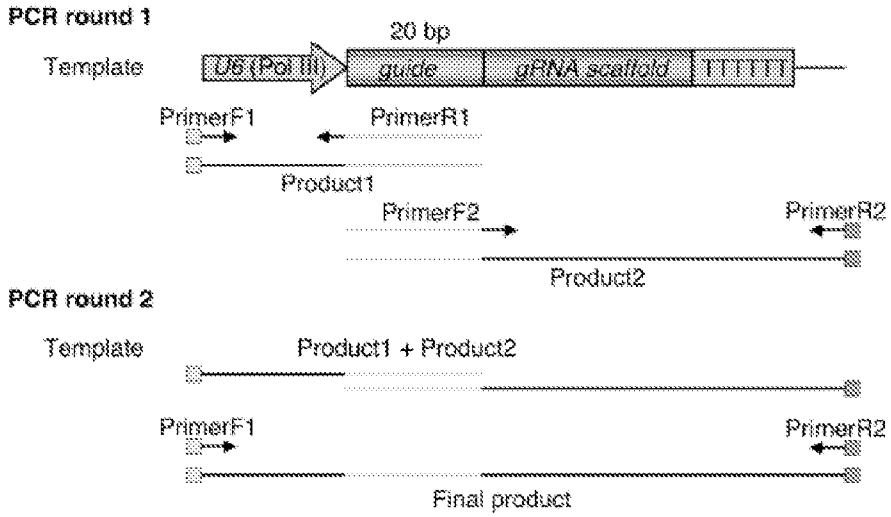
```
GRMZM2G053008            414 -----EAKNLLRNAHELIER--SKKQEALSRSELSIEYNDA--DQVSAAHTGT-------
HvMLOC_57250                 ------------------------------------------------------------
Os12g0157000             137 ----------------------------------------DDHSDHAGSRA-----
GmLoc100778733           351 DHQTLNTRQ------QQQQQQQQEGAGNVSLSPMIIDSVPVAHHLHHQQHHGGKSSG---
Bra004501                204 ------------------------DQRAVVADPLVIESVPVMMHGG-A------------
Bra000434                206 ------------------------DQRAVVADPLVIESVPVMMHGG-A------------
Bra040478                197 ------------------------DQRAVVADPLVIESVPVMMRGG-A------------
Bra014415                196 ------------------------ERYD---PTAVIESVPVIMQRR-A------------
Bra003482                169 ------------------------ERND---PTAVIESVPLIMQRRAA-------------
Bra007646                175 -------------------------------PTAVIESVPVIMQRREA-------------
GlycinemaxLoc100781489   284 -------------------GDQNLQGRGSNIVPMIIDSVPVNVAHHNNNRHGNGG-----
GRMZM2G024948_T01        303 QP-----RPFSLPLHHYTVQP-SAAGVTAASRPVLLDSVPVIESP---------------
os02g0683500             283 -L---LARAPSPLHHHYTLQP-SGDGVRAAGSPVVLDSVPVIESP---------------
HvMLOC_66387             284 LL---VPRPPPPLHFTVQQQGSDAGGSVTAGSPVVLDSVPVIESP---------------
os04g0581400             285 GRGFISTRPCHRRRRHLRLLT-NSTLRCTTRAP-----------------------------
GRMZM2G102059_T01        287 QQ--QPPPPPHPPLHSIMLVQ-PSPAPPTASVPMLLDSVPLVNSP---------------
Os10g0537100             221 -------RPQV--------------------PAAVVLKSVPVHVAATSAVQ----------
GRMZM2G142999_T01        193 -------RPQV--------------------PAAVVLKSVPVHVAASAV------------
GRMZM2G125095_T01        191 -------RPQV--------------------PAAVVLTSVPVRVAASAV------------
os03g0120900             225 -------RPLPPQQ--------------QHHPAVVLESVPVRMTAGH---A--EP-----
GRMZM2G098443_T01        246 -------RSC---------------------------------QPQPRRTA--DV-----
GRMZM2G082227_T01        234 -------RPH---Q--------------QPSTTVMLDSVPVRLPPTPGQHA--EP-----
Os11g0156000             198 -------RRAADHDH-----------------------------GDMHHADES---PRD
GRMZM2G328742_T01        223 -------RHSADHDH-----------------------------SNNMQHAGES---QSD
GmLoc100802734           200 ---------PFPYQ-----------------------------PHSLHAPGGGSQGQNE
GmLoc100795470           227 -------HQPSPYQ-----------------------------QQHDCLHAGRGSQGQNQ
GmLoc100818164           231 -------HHPSPYHH----------------------------QQDDSLHAVRGSQGQNQ
Bra017262                157 ----------SNSSL----------------------------QY--YPHAG--AQA---
At2g36080                158 ----------SNASL----------------------------QY--YPHAG--AQA---
Bra005301                159 ----------SNASL----------------------------QY--YPHAG--VQA---
At3g11580                181 --------TYPN-IH----------------------------QE--YSHYG--AVVDHA
BraLOC103849927          181 -------SSYPN-IH----------------------------QE--YSHYG--AVA---
BrassicarapaBra034828    174 -------SSYPNNIH----------------------------QE--YSHYG--PVA---
At5g06250                191 -------NYSNPPSH----------------------------SE--YSHYG--AAVATA
Bra005886                179 -------NYSNRPAH----------------------------SE--YSHYG--AAVATA
GmLoc102660503           318 ------------------------------------------GSEE-----GED------
HvMLOC_38822             173 ------------------------------------------------------------
os01g0693400             336 ------------------------------------------------------------
HvMLOC44012              172 ------------------------------------------------------------
HvMLOC_7940              305 ------------------------------------------------------------
HvMLOC_75135             298 ------------------------------------------------------------
TRAECDM81004             294 ------------------------------------------------------------
HvMLOC_56567             293 ------------------------------------------------------------
TRAES3BF098300010CFD_t1  293 ------------------------------------------------------------
HvMLOC_63261             280 ------------------------------------------------------------
TRAES3BF062700040CFD_t1  292 ------------------------------------------------------------
TRAES3BF062600010CFD_t1  292 ------------------------------------------------------------
Bra038346                298 ------------------------------------------------------------
GmLoc732601              301------------------------------------------------------------
GmLoc100789009           307 ------------------------------------------------------------
GmLoc100776987           325 ------------------------------------------------------------
GmLoc100801107           341 ------------------------------------------------------------
```


GRMZM2G053008 458 EDQVGPNNSDS DVDDEE EEEELEGY
HvMLOC_57250
Os12g0157000 148 DAKSSSAASASRRRGVN CGA DAT AMYGYMHHS
GmLoc100778733 402 PSSTSTSPSTAG RLFGVN CA STSEDPKCF SLLSS
Bra004501 227 QVSQAVVGTAG RLFGVD EESSSSGGSLPR
Bra000434 229 RVNQAAVGTAG RLFGVD CG SGG T NSTEEESSSSGGSLPR
Bra040478 220 RVSQEVVGTAG RLFGVD EESSSSGGSLPRA
Bra014415 216 HVATMASSRGE RLFGVD CV GGRGGGGSV NSTEEESSTSGGSISRG
Bra003482 190 HVAAIPSSRGE RLFGVD CG GGG GSV NSTEEESSSSGG GG
Bra007646 192 QVANMASSRGE RLFGVD CG GGGGGSV NSTEEESSSSGGSMSRG
GlycinemaxLoc100781489 320 ITSGGTNCSG RLFGVN CA SAEDSKELS SGSAAHVTTAASSSSLH
GRMZM2G024948_T01 342 TTAA VRLFGVN NN DGG GEASHQGDALSLQM P
os02g0683500 323 TTAA VRLFGVN NP AGGGGGAAA GESSNHGNALSLQ TP
HvMLOC_66387 326 TTATK VRLFGVN NP HPGDGG GESSNYGSALPLQMPAS
os04g0581400
GRMZM2G102059_T01 329 TAAS VRLFGVN NP PGTS AESSQDANAL SLRTP
Os10g0537100 245 EAATTTRP VRLFGVN CP AMDDDDDIA GA
GRMZM2G142999_T01 215 E ATMS VRLFGVN CP DAEDSATV P
GRMZM2G125095_T01 213 EEATRS VRLFGVN CP DAEDGATA T
os03g0120900 254 PSAPS VRLFGVN CA SEQDHAGVV GK
GRMZM2G098443_T01 259 P
GRMZM2G082227_T01 263 PPPAVASSAS VRLFGVN CA AAGSEEENV GG
Os11g0156000 218 TDSPSFS AGSAPS RLFGVN CG EPEADT TAA ATMYGYMHQQ
GRMZM2G328742_T01 244 RDNRSCSAASAPPPPS RLFGVN CG GPEPET P TAMYGYMHQS
GmLoc100802734 221 T TPGGN SSSSGSG RLFGVN CQ DNHNDSQNS TPEC SYTHLYHH
GmLoc100795470 251 RMRPVGNNSSSSSSSS RLFGVD CQ EH DDSGPS TPQC SYNSNNMLPS
GmLoc100818164 256 RTRPVGNSSSSSSSSS RLFGVN CQ EH DDSGPS TPQC SYNTNNILPS
Bra017262 172 VENQRGNS RLFGVN CQ DS DWSEPS TPDG FTTCPT
At2g36080 173 VESQRGNS RLFGVN CQ DS DWSEPS TPDG SNTYTT
Bra005301 174 VESQRGNS RLFGVN CQ DS DLPDPS TPDG STICPT
At3g11580 200 QSI PPVVAGSS VRLFGVN CH DA VE PPP
BraLOC103849927 198 EI PTVVTGSS VRLFGVN CH DV VE TPP
BrassicarapaBra034828 192 ET PTVAAGSS VRLFGVN CH DV VE PPP
At5g06250 212 AETHSTPSSSVVGSS VRLFGVN CQ DE NDGDDS VAVA TTVES
Bra005886 200 TETHFIPSSSAVGSS VRLFGVN CQ DE DEGDDS VATA AAAECP
GmLoc102660503 325 EDANIGKDFNAQ RLFGVC T
HvMLOC_38822 173 APSPAPVT VRLFGVD LIA AARHAHEHEDYGMAKTNKRT MEAS
os01g0693400 336 DQPAPSPV VRLFGVD LTA APVEQM AGCKRA RDL
HvMLOC44012 172 DQSSAPVQ VRLFGVD LAA EQ GMP GGCKRA RDL
HvMLOC_7940 305 EPTKGEQV VRLFGVD AGE RGRAAPV
HvMLOC_75135 298 ETAKGEQD VRLFGVD AGV RVRAA T G
TRAECDM81004 294 ETAKGEQV VRLFGVD AGV RGRAA T A
HvMLOC_56567 293 EIAKAEQV VRLFGVD AGV RERAA T A
TRAES3BF098300010CFD_t1 293 VTTKGEQV VRLFGVD SGV RGRAATAT A
HvMLOC_63261 280 ADENEAGHVA VRLFGVD GWA MAGSSGG
TRAES3BF062700040CFD_t1 292 SDENETGHVA VRLFGVD AGG MAGSSGG
TRAES3BF062600010CFD_t1 292 SDENETGHVA VRLFGVD AGG MAGSSGG
Bra038346 298 PVQV VRLFGVD FNV SAKP SNVVDACGGK
GmLoc732601 301 GPVVEPVQ M RLFGVN LKLP PGS DGVG
GmLoc100789009 307 GPVVEPVQ N VRLFGVD LKL VPGS DGIGVGCD G
GmLoc100776987 325 GPVVEPIQ N VRLFGVN LKL GSDSIA NN NNASGCCN G
GmLoc100801107 341 GPVVEPIQ N VRLFGVN LKL GSDTIVGNN NNASGCCN G


GRMZM2G053008 484 ----------DSP----------PMADDIHEFTLG---DSYAFGEGFSNG
HvMLOC_57250
Os12g0157000 181 YAA---VSTVNY▒SV----------
GmLoc100778733 442 --S-SMANSNSQPPLQLLREDTLSSSS-----ARFGD-----QRGVGEPSMLFD▒DPSLQ
Bra004501 261 --G-DASPSSS▒QLRLGSSSEDDH----------FSKKGKSSLPFD▒DQ---
Bra000434 274 --G-GASPSSS▒QLRLGNSSEDDH----------LFKKGKSSLPFN▒DQ---
Bra040478 255 GGG-GASSSSS▒QLRLGSSCEDDH----------FSKKGKSSLPFD▒DQ---
Bra014415 267 GV--SMAGVGSPLQLRLVSSDGDDQSLVARGAARVDEDHHLFT-KKGKSSLSFD▒DK---
Bra003482 234 GV--SMASVGS▒LQLRLVSSD--DESLVAMEAASVDEDHHLFT-KKGKSSLSFD▒DRK--
Bra007646 240 GV--SMAGVGS▒LQLRLVSSD--DESLVAMEGATVDEDHHLFTTKKGKSSLSFD▒DI---
GlycinemaxLoc100781489 370 HQR-LRV------PVPVPLEDPLSSSA--AAAARFG----DHKGASTGTSLLFD▒DPSLQ
GRMZM2G024948_T01 378 GWQ-QRTPTLR▒LELPRHG---GESSA--ASSPSSS----SSSKREARSALDLD▒-----
os02g0683500 365 AWM-RRDPTLR▒LELPPHHHHGAESSA--ASSPSSS----SSSKRDAHSALDLD▒-----
HvMLOC_66387 368 AWR-PRDHTLR▒LEFPSHGA------E--ASSPSSS----SSSKREAHSGLDLD▒-----
os04g0581400
GRMZM2G102059_T01 366 GWQ-RPGP-LRF▒ESPQR---GAESSA--ASSPSSS----SSSKREAHSSLDLD▒-----
Os10g0537100 278 -AS-RTAA-SS▒LQLPSP----------SSSTS----SSTAGKKMCSLDLG▒-----
GRMZM2G142999_T01 243 -RG-RAAS-TT▒LQLPSP----------SSSTS----SSTAGKDVCCLDLG▒-----
GRMZM2G125095_T01 242 -R----TP-ST▒LQLPSP----------SSSTS----SSTGGKDVRSLDLG▒-----
os03g0120900 284 ----TAPP-----PLPSP----------P-SSS----SSSSGKARCSLNLD▒-----
GRMZM2G098443_T01
GRMZM2G082227_T01 298 -WR-TSAP-----PTQQA----------S-SSS----SYSSGKARCSLNLD▒-----
Os11g0156000 264 SSYAAMS---A------V----------PSYWGN----------S----------
GRMZM2G328742_T01 291 PYA----------YNNWGS----------PYQHDEEI---
GmLoc100802734 269 ----------QTSS----------YSSSSN----------PHHHM▒PQQP-
GmLoc100795470 303 TQGTDH-SHHNF▒QQ-QP----------SNSNPS----------PHHMM▒HHQPY
GmLoc100818164 308 TQGTDIHSHLNF▒QQQQT----------SNSKPP----------PHHMM▒RHQPY
Bra017262 212 --NHD---QFP▒P--EH----------FPP----------PYYMD▒SFTGD
At2g36080 213 --NHD---QFHF▒PQQQH----------YPP----------PYYMD▒SFTGD
Bra005301 214 --SHD---QFH▒PQ-QH----------YPP----------PYYMD▒SFTGD
At3g11580 232 --RPDVYNDQH▒----Y----------YST----------PHPMN▒SFAGE
BraLOC103849927 229 --CPDGYNGQHF▒----Y----------YST----------PDPMN▒SFAGE
BrassicarapaBra034828 223 --CPDAYNGQH▒----Y----------YST----------PHPMN▒SFAGE
At5g06250 258 ---PDGYYGQN▒----Y----------YYS----------HPHNM▒ILTLL
Bra005886 247 --RQDSYYDQN▒----N----------YYT----------PHSSAS-----
GmLoc102660503
HvMLOC_38822 219 V-A-APTPAHA▒KKRCV----------DFALTYR▒ATTPQ
os01g0693400 373 A-A-TTPPQAAA▒KKQCI----------ELALV---------
HvMLOC44012 208 V-K-PP-PPKVA▒KKQCI----------ELALA---------
HvMLOC_7940 334 ---------EQE▒KRQCV----------AHS--QH---SPA
HvMLOC_75135 327 -----ELGPPE▒KRQSV----------AHGCGRM---NYI
TRAECDM81004 323 -----EQGPPE▒LKRQCV----------PLPHGQR---SPA
HvMLOC_56567 322 -----EQGPQGW▒KRQCM----------AHGQH---SPA
TRAES3BF098300010CFD_t1 324 -----EQGLQE▒KRQCV----------APGQH---SPA
HvMLOC_63261
TRAES3BF062700040CFD_t1
TRAES3BF062600010CFD_t1
Bra038346 329 -----RSRDVD▒ALRCS----------KKH----------
GmLoc732601 330-----KRKEME▒AFECC----------KKL----------
GmLoc100789009 341 -----KRKEME▒AFECS----------KKL----------
GmLoc100776987 363 -----KRREME▒SLECS----------KKP----------
GmLoc100801107 380 -----KRREME▒SLECS----------KKP----------

```
GRMZM2G053008            511 YLEEVLRSLPLQEDGQKKL-CDAPINADASD
HvMLOC_57250                 -------------------------------
Os12g0157000                 -------------------------------
GmLoc100778733           489 YRQ----------------------------
Bra004501                    -------------------------------
Bra000434                    -------------------------------
Bra040478                    -------------------------------
Bra014415                    -------------------------------
Bra003482                    -------------------------------
Bra007646                    -------------------------------
GlycinemaxLoc100781489   417 YHRH---------------------------
GRMZM2G024948_T01            -------------------------------
os02g0683500                 -------------------------------
HvMLOC_66387                 -------------------------------
os04g0581400                 -------------------------------
GRMZM2G102059_T01            -------------------------------
Os10g0537100                 -------------------------------
GRMZM2G142999_T01            -------------------------------
GRMZM2G125095_T01            -------------------------------
os03g0120900                 -------------------------------
GRMZM2G098443_T01            -------------------------------
GRMZM2G082227_T01            -------------------------------
Os11g0156000                 -------------------------------
GRMZM2G328742_T01            -------------------------------
GmLoc100802734               -------------------------------
GmLoc100795470           336 YY-----------------------------
GmLoc100818164           343 YY-----------------------------
Bra017262                237 VHQTSS-Q-----QG-----------------
At2g36080                240 MNRTS--------------------------
Bra005301                240 VHQTRS-P-----QG-----------------
At3g11580                258 ALEQVGDG-----RG-----------------
BraLOC103849927          255 AMEQVGDG-----RR-----------------
BrassicarapaBra034828    249 AMEQVGDG-----RG-----------------
At5g06250                283 -------------------------------
Bra005886                    -------------------------------
GmLoc102660503               -------------------------------
HvMLOC_38822             248 CPRSRDQL-----EGVQAAGSTFAL------
os01g0693400                 -------------------------------
HvMLOC44012                  -------------------------------
HvMLOC_7940              352 ------------------L-GAFVL------
HvMLOC_75135             350 CYSIG------------TI-GPLMLN-----
TRAECDM81004             346 ------------------L-GAFVL------
HvMLOC_56567             343 ------------------L-GDFAL------
TRAES3BF098300010CFD_t1  345 ------------------L-GAFAL------
HvMLOC_63261                 -------------------------------
TRAES3BF062700040CFD_t1      -------------------------------
TRAES3BF062600010CFD_t1      -------------------------------
Bra038346                345 ----A------------II-NAL--------
GmLoc732601              346 ----K------------VI-GAL--------
GmLoc100789009           357 ----K------------VI-GAL--------
GmLoc100776987           379 ----K------------II-GAL--------
GmLoc100801107           396 ----K------------II-GAL--------
```

Figure 14



gRNA sequence (SEQ ID NO: 146)

gacggccagtgccaagcttCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTGCCCTTGGATCA
TGAACCAACGGCCTGGCTGTATTTGGTGGTTGTGTAGGGAGATGGGGAGAAGAAAAGCCCGATT
CTCTTCGCTGTGATGGGCTGGATGCATGCGGGGGAGCGGGAGGCCCAAGTACGTGCACGGTGAG
CGGCCCACAGGGCGAGTGTGAGCGCGAGAGGCGGGAGGAACAGTTTAGTACCACATTGCCCAG
CTAACTCGAACGCGACCAACTTATAAACCCGCGCGCTGTCGCTTGTGT**GGGAAGGAAGAGAC**
AGATTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGCTTTTTTTGTCCCTTCGAAGGGCAATTCTGCAGATATCCATCAC*ACT*
GGCGGCCGCTCGAGGTCGaagcttgcatgcctgcagg

os11g01560000
A-R1
GGACTGGGGTTGCTCCTGGG**ACACAAGCGACAGCGCGCGGG** (SEQ ID NO: 147)
A-F2
CCCAGGAGCAACCCCAGTCC**GTTTTAGAGCTAGAAATAGCA** (SEQ ID NO: 148)
B-R1
TGCTATTTCTAGCTCTAAAAC**ACACAAGCGACAGCGCGCGGG** (SEQ ID NO: 149)
B-F2
GCCCCTGACGCCCAGTGACG**GTTTTAGAGCTAGAAATAGCA** (SEQ ID NO: 150)

Os12g0157000
C-R1
GGGGGTGCCCCTGGGCGAGA**ACACAAGCGACAGCGCGCGGG** (SEQ ID NO: 152)
C-F2
TCTCGCCCAGGGGCACCCCC**GTTTTAGAGCTAGAAATAGCA** (SEQ ID NO: 153)
D-R1
CTCGTAGTGGTGGTGGTAGT**ACACAAGCGACAGCGCGCGGG** (SEQ ID NO: 154)
D-F2
ACTACCACCACCACTACGAG**GTTTTAGAGCTAGAAATAGCA** (SEQ ID NO: 155)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/050245

A. CLASSIFICATION OF SUBJECT MATTER
INV. A01H5/00 A01H5/10 C07K14/415 C12N15/82
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A01H C07K C12N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal , BIOSIS, Sequence Search , EMBASE, MEDLINE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2004/031349 A2 (MENDEL BIOTECHNOLOGY INC [US] ; JIANG CAI-ZHONG [US] ; HEARD JACQUELINE) 15 April 2004 (2004-04-15) Claims 1, 3-9 ; pages 32, 34, 41, 62, 110, 115-117 , 247 ; Tables 4-6; Seq. ID Nos 395 and 396 ---	1-6, 10, 11, 14-18, 23,24, 29-33 , 35-38
X	US 2005/193443 Al (DALE ROCK CHRISTOPHER [US] ET AL) 1 September 2005 (2005-09-01)	36-38
A	claims 1, 10, 11, 13, 22, 25, 29, 30, 36, 37; pargraphs [0020] , [0122] and [0140] ---	1-35
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" documentwhich may throw doubts on priority claim(s) orwhich is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 April 2016	26/04/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Kurz , Birgit

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2016/050245

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13fer1 (a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13fer1 (a)).

      [ ] on paper or in the form of an image file (Rule 13fer1 (b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/050245

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SWAMINATHAN K ET AL: "The plant B3 superfamily", TRENDS IN PLANT SCIENCE, ELSEVIER SCIENCE, OXFORD, GB, vol. 13, no. 12, 1 December 2008 (2008-12-01), pages 647-655, XP025694257, ISSN: 1360-1385, DOI: 10.1016/J.TPLANTS.2008.09.006 [retrieved on 2008-11-26] cited in the application abstract; pages 647, 649, 650; Figures 1 and 2	1-38
A	M. IKEDA ET AL: "A Novel Group of Transcriptional Repressors in Arabidopsis", PLANT AND CELL PHYSIOLOGY, vol. 50, no. 5, 26 March 2009 (2009-03-26), pages 970-975, XP055263546, UK ISSN: 0032-0781, DOI: 10.1093/pcp/pcp048 cited in the application abstract; pages 971-974; Table 1	1-38
A	LI YUNHAI ET AL: "Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana", GENES AND DEVELOPMENT, COLD SPRING HARBOR LABORATORY PRESS, PLAINVIEW, NY, US, vol. 22, no. 10, 1 May 2008 (2008-05-01), pages 1331-1336, XP002512707, ISSN: 0890-9369, DOI: 10.1101/GAD.463608 cited in the application abstract; page 1332	1-38
X, P	YUEYING ZHANG ET AL: "Transcription Factors S0D7/NGAL2 and DPA4/NGAL3 Act Redundantly to Regulate Seed Size by Directly Repressing KLU Expression in Arabidopsis thaliana", THE PLANT CELL, vol. 27, no. 3, 1 March 2015 (2015-03-01), pages 620-632, XP055263586, US ISSN: 1040-4651, DOI: 10.1105/tpc.114.135368 abstract; pages 621-624, 628, 629; Figures 2, 3 and 5	1-38

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2016/050245

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004031349 A2	15-04-2004	AT 539158 T	15-01-2012
		AU 2003285856 A1	23-04-2004
		BR 0314389 A	12-07-2005
		DK 1546336 T3	10-04-2012
		EP 1546336 A2	29-06-2005
		EP 2270166 A2	05-01-2011
		EP 2270167 A2	05-01-2011
		EP 2272962 A2	12-01-2011
		ES 2380017 T3	07-05-2012
		PT 1546336 E	09-04-2012
		US 2007033671 A1	08-02-2007
		US 2007101454 A1	03-05-2007
		WO 2004031349 A2	15-04-2004
US 2005193443 A1	01-09-2005	US 2005193443 A1	01-09-2005
		US 2009083877 A1	26-03-2009