



(51) International Patent Classification:

C12N 9/88 (2006.01) C12N 15/82 (2006.01)

(21) International Application Number:

PCT/US2017/040859

(22) International Filing Date:

06 July 2017 (06.07.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/359,819 08 July 2016 (08.07.2016) US

(71) Applicants: **THE UNITED STATES OF AMERICA**, as represented by **THE SECRETARY OF AGRICULTURE** [US/US]; 1400 Independence Ave. S.W., Washington, District of Columbia 20250 (US). **IOWA STATE UNIVERSITY RESEARCH FOUNDATION, INC.** [US/US]; 1805 Collaboration Place, Suite 2100, Ames, Iowa 50010 (US).

(72) Inventors: **HUBER, Steven C.**; ARS/Global Change and Photosynthesis Research, 1201 West Gregory Drive, 190 ERML, Urbana, Illinois 61801 (US). **KIM, Sang Yeol**; ARS/Global Change and Photosynthesis Research, 1201 West Gregory Drive, 190 ERML, Urbana, Illinois 61801 (US). **SPALDING, Martin H.**; Iowa State University, 449 Bessey Hall, 2200 Osborn Drive, Ames, Iowa 50011 (US). **WRIGHT, David A.**; 449 Bessey Hall, 2200 Osborn Drive, Ames, Iowa 50011 (US).

(74) Agent: **GOLDBERG, Joshua B.**; Nath, Goldberg & Meyer, 112 S. West Street, Alexandria, Virginia 22314 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

(54) Title: MUTATED RUBISCO ACTIVASE

Faster Photosynthetic Induction in Rice with truncated RCA

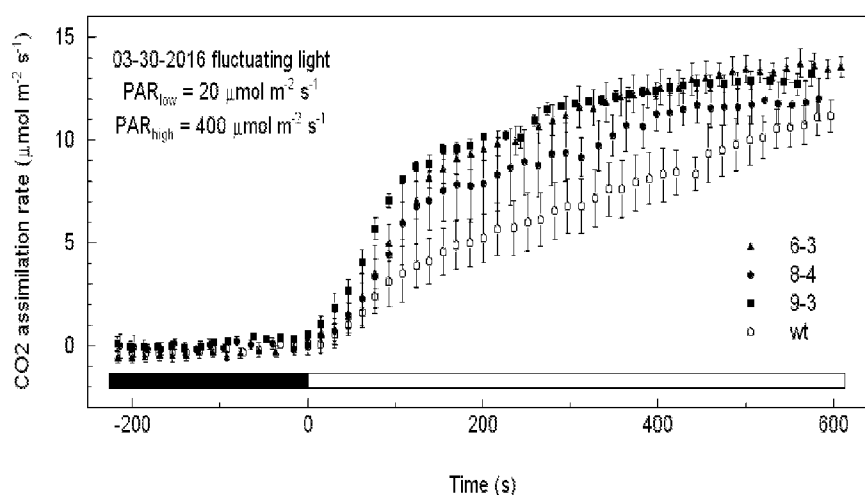


Fig. 20C

(57) Abstract: An altered RuBisCO activase (RCA) α -isoform lacking redox activity at its carboxy terminus and genetically altered plants that encode an altered RCA α -isoform having an altered DNA at its 3' end (or the exon that encodes the redox-sensitive cysteines that are present at the protein's carboxy terminus) so that the encoded RCA α -isoform lacks redox activity at its carboxy terminus are generated. Genetically altered plants that produce the RCA α -isoform lacking redox activity at its carboxy terminus have an increase in at least one agronomic phenotype. Methods of making the genetic alteration in plants and genetically altered plants are described.



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, 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))

MUTATED RUBISCO ACTIVASE

CROSS REFERENCE TO RELATED PATENT APPLICATION

This patent application claims priority to and the benefit of U.S. Patent Application 62/359,819 filed on July 8, 2016, the contents of which are incorporated by reference.

SEQUENCE LISTING

The Sequence Listing submitted via EFS-Web as ASCII compliant text file format (.txt) filed on July 6, 2017, named "SequenceListing_ST25 7-5-17", (created on July 5, 2017, 26 KB), is incorporated herein by reference. This Sequence Listing serves as paper copy of the Sequence Listing required by 37 C.F.R. § 1.821(c) and the Sequence Listing in computer-readable form (CRF) required by 37 C.F.R. § 1.821(e). A statement under 37 C.F.R. § 1.821(f) is not necessary.

FIELD OF THE INVENTION

The invention relates generally to a mutated ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) activase (also referred to as RCA), the genomic sequences that encode the mutated RCA that is truncated, and genetically altered plants that produce the mutated RCA that is truncated.

BACKGROUND OF THE INVENTION

All photosynthetic organisms catalyze the fixation of atmospheric CO₂ by the bifunctional enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO).

RuBisCO catalyzes the first and rate-limiting step in photosynthetic carbon fixation, the transfer of atmospheric CO₂ to ribulose-1,5-bisphosphate. As such, it is the major enzyme that removes CO₂ from the atmosphere, and significant variation in kinetic properties of this enzyme are found among various phylogenetic groups. Well over 1,000 different RuBisCO homologues are available in the public literature (e.g., over 1,000 different RuBisCO homologues are listed in GenBank alone), and the crystal structure of RuBisCO has been solved for several variants of the protein.

RuBisCO contains two competing enzymatic activities: an oxygenase and a carboxylase activity. The oxygenation reaction catalyzed by RuBisCO is a "wasteful" process since it competes with and significantly reduces the net amount of carbon fixed. In addition, its catalytic cycling rate (k_{cat}) at about 3 reactions per second, for the enzymes from higher plants, is relatively slow. To compensate for its low activity, plants deposit large amounts of RuBisCO enzyme in their green tissues. Indeed, RuBisCO accounts for more than 35% of leaf total soluble proteins. Increasing RuBisCO's catalytic efficiency would proportionally increase the rate of photosynthesis and, in turn, increase plant productivity.

For more than 20 years a number of researchers have attempted to improve RuBisCO, using a variety of approaches. See, e.g., Mann, C. C., (1999) Science, 283:314-316, and references cited therein. Indeed, the quest for a better RuBisCO has been called a "Holy Grail" of plant biology. To date, there has been little success in the creation of an improved RuBisCO. Recombination based methods for producing a modified RuBisCO enzyme having increased catalytic efficiency and selectivity for CO₂ are described in U.S. Patent Application No. 09/437,726.

RuBisCO activase (RCA), through ATP hydrolysis and highly specific interaction with RuBisCO causes a conformational change to the RuBisCO protein that allows release of inhibitory sugar phosphates tightly bound at RuBisCO's active sites, thereby activating the enzyme. In addition, a mathematical model indicates that RCA limits non-steady-state photosynthesis in plants receiving fluctuating light (e.g., light flecks formed in a canopy), even at moderate temperatures (Mott K A and Woodrow I E, (2000) J Exp Botany 51:399-406). Thus, there exists a need for improved variants of enzymes involved in carbon fixation, for example RuBisCO and RCA. The present invention meets these and other needs and provides such improvements and opportunities.

SUMMARY OF THE INVENTION

It is an object of this invention to have a mutated ribulose 1,5-bisphosphate carboxylase/oxygenase activase (RCA) α -isoform (also referred to as an "altered" RCA α -isoform) having reduced light sensitive inhibitory activity compared to light sensitive inhibitory activity in a wild-type RCA α -isoform. It is another object of this invention that the mutated RCA α -isoform has reduced amount of redox activity at its carboxy terminus (or no redox activity at its carboxy terminus) compared to the amount of redox activity a wild-type RCA α -isoform has its carboxy terminus. In one embodiment of this invention, the altered RCA α -isoform has less than two redox-sensitive cysteines at its redox active site (one or none cysteines) whereas the wild-type RCA α -isoform has two redox-sensitive cysteines within said redox active site. In another embodiment of this invention, the altered RCA α -isoform has three or more redox-sensitive cysteines at its redox active site whereas the wild-type RCA α -isoform has two redox-sensitive cysteines within said redox active site. It is another object of this invention that the altered RCA α -isoform having less than two redox-sensitive cysteines at its redox active site (one or none cysteines) is a rice altered RCA α -isoform and has at least one carboxy terminal amino acid sequence set forth in SEQ ID NOs: 4, 8, 10, 15, 17, 19, and 24. It is another object of this invention that the altered RCA α -isoform having three or more redox-sensitive cysteines at its redox active site is a rice altered RCA α -isoform and has at least one carboxy terminal amino acid sequence set forth in SEQ ID NOs: 6, 12, 22, 27, and 29. It is further object of this invention to

have an altered (or mutated) plant containing one or more of these altered RCA α -isoforms. It is another object of this invention that these altered RCA α -isoforms have a reduced amount of light sensitive inhibition of RuBisCO activity compared to wild-type RCA's amount of light sensitive inhibition of RuBisCO activity. It is another object of this invention that these mutated RCA α -isoforms have increased activation rate of RuBisCO compared to wild-type RCA's activation rate of RuBisCO.

It is an object of this invention to have an altered plant and part thereof which has an increase in at least one agronomic phenotype compared to the amount of that agronomic phenotype in a wild-type plant grown under similar conditions. It is another object of this invention that the altered plant and part thereof contains an altered RCA α -isoform that lacks a redox active site at the RCA α -isoform's carboxy terminus. It is a further object of this invention that the altered RCA α -isoform causes the increase the agronomic phenotype compared to the agronomic phenotype level in a wild-type plant containing a wild-type RCA α -isoform, and the agronomic phenotype can be increased plant growth, leaf width, seed yield per plant, photosynthetic induction rate, RCA α -isoform activation rate, RuBisCO activation rate, or a combination thereof. It is a further object of this invention to have an altered seed from this altered plant such that the altered seed contains altered plant cells that produce one or more of these altered RCA α -isoforms. It is another object of this invention to have an altered plant cell from this altered plant such that the altered plant cell produces at least one of the altered RCA α -isoforms. It is yet another object of this invention that the altered plant is an altered rice, and that at least one altered RCA α -isoform has a carboxy terminal amino acid sequence that can be SEQ ID NOs: 4, 6, 8, 10, 12, 15, 17, 19, 22, 24, 27, 29, or a combination thereof.

It is an object of this invention to have a method of increasing at least one agronomic phenotype in a transformed plant compared to a wild-type plant's agronomic phenotype by altering a wild-type plant cell's RCA α -isoform's 3' end DNA sequence to produce an altered plant cell containing at least one altered RCA α -isoform containing a DNA alteration at the RCA α -isoform's 3' end, selecting for an altered plant cell that contains the altered RCA α -isoform gene and that produces the altered RCA α -isoform to produce a selected transformed plant cell, such that the altered RCA α -isoform has reduced terminal redox activity compared to the wild-type plant cell's RCA α -isoform's carboxy terminal redox activity; and growing the selected transformed plant cell into the transformed plant that produces at least one altered RCA α -isoform which causes the transformed plant to have an increase in at least one agronomic phenotype compared to the wild-type plant's agronomic phenotype. It is another object of this invention that the agronomic phenotype can be increased plant growth, leaf width, seed yield per plant,

photosynthetic induction rate, RCA α -isoform activation rate, RuBisCO activation rate, or a combination thereof. It is a further object of this invention that the altering step involves transforming the wild-type plant cell with an expression vector containing a promoter operably linked to DNA encoding Cas9 and a promoter operably linked to DNA encoding an appropriate sgRNA having a sequence that targets a genomic sequence prior to or at RCA α -isoform's carboxy terminus redox active site. In an alternative embodiment, the DNA alteration occurs in the RCA α -isoform gene's exon that encodes the redox-sensitive cysteines at the carboxy terminus of the protein.

Another object of this invention is using RNAi to silence the production of RCA α -isoform by transforming a wild-type plant cell with an expression vector containing a promoter operably linked to DNA encodes at least 19 nucleotides that is the reverse complement to the same number of nucleotides of RCA α -isoform within the protein's carboxy terminus. Alternatively, the expression vector could contain a promoter operably linked to DNA that encodes at least 19 nucleotides of the sense coding RCA α -isoform within the protein's carboxy terminus and the complementary anti-sense nucleotides and a linker between the sense and anti-sense sequences. One selected for transformed cells that contain the expression vector and produces the desired polynucleotide and then grows the selected transformed plant cells into a transformed plant.

It is another object of this invention to have a transformed plant and part thereof made by any of the above described methods.

It is an object of this invention to have a method of plant breeding to increase at least one agronomic trait in the plant by crossing a male parent plant with a female parent plant, where one of the parent plants produces an altered RCA α -isoform lacking redox activity at its carboxy terminus, to produce progeny seeds that produce the altered RCA α -isoform; and harvesting the progeny seeds that produce the altered RCA α -isoform from the female parent plant, and that the produced altered RCA α -isoform lacking redox activity at its carboxy terminus which causes the increase in at least one agronomic trait. It is another object of this invention that one of the parent plants contains at least one altered RCA α -isoform gene that encodes the altered RCA α -isoform lacking redox activity at its carboxy terminus and that this altered RCA α -isoform gene is present in the progeny seeds. It is another object of the invention that the agronomic trait can be improved plant growth, leaf width, seed yield per plant, photosynthetic induction rate, RCA α -isoform activation rate, RuBisCO activation rate, or a combination thereof.

It is another object of this invention to inhibit or reduce the light sensitive inhibitory activity of RCA α -isoform, thereby creating a higher activation state for RuBisCO. RCA α -isoform C-terminal extension contains two redox-sensitive cysteines responsible for inactivation

of RCA in the dark or low light. In other plants, this activity may be in a separate RuBisCO alpha gene. In one embodiment, reduction of this redox activity or inhibition of this activity can be caused by disrupting RCA α -isoform or its formation of the activation complex with RCA β -isoform, a decrease in the expression of *RCA* α -isoform gene through the use of co-suppression, antisense, or RNA silencing or interference, or through gene editing to form indels that disrupt the carboxy-terminal region of the *RCA* α -isoform gene.

Constructs and expression cassettes containing nucleotide sequences that can efficiently reduce the expression of a RCA α -isoform's regulatory region are also objects of this invention. Another object of this invention is polynucleotides that encode altered RCA α -isoforms lacking the redox-sensitive cysteines at the protein's carboxy terminus.

In another embodiment, one can reduce RCA α -isoform activity by (a) introducing into a wild-type plant cell at least one expression vector containing a promoter operably linked to a polynucleotide that encodes one or more RCA α -isoform regulatory inhibition sequences or a subsequence thereof, (b) selecting transformed plant cells that express the polynucleotide encoding the inhibition sequence, and (c) growing the selected transformed plant cells into transformed plant where the produced polynucleotide encoding the RCA α -isoform regulatory inhibition sequence modulates (or decreases) the activity of one or more wild-type RCA α -isoform regulatory sequences compared to a corresponding wild-type plant. For example, the expression vector can be introduced by techniques including, but not limited to, electroporation, micro-projectile bombardment, *Agrobacterium*-mediated transfer, and the like.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

FIG. 1 contains the DNA and amino acid sequences of wild-type *O. sativa* cultivar Kitaake RCA α -isoform's carboxy terminus (as encoded in exon 7); SEQ ID NOs: 1 and 2 respectively.

FIG. 2 shows the DNA and amino acid sequences (SEQ ID NO: 3 and 4, respectively) of an altered *O. sativa* cultivar Kitaake RCA α -isoform's carboxy terminus (plant line 2-1) containing 147 bp deletion in intron 6 and exon 7.

FIG. 3 shows the 3' DNA and amino acid carboxy terminal sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 8-1 and 8-2. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 4 bp deletion (SEQ ID NO: 5 and 6), and chromosome 2 has 1 bp insertion (SEQ ID NO: 7 and 8).

FIG. 4 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 8-3. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 2 bp deletion (SEQ ID NO: 9 and 10), and chromosome 2 has 6 bp deletion and 2 bp insertion (SEQ ID NO: 11 and 12).

FIG. 5 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 8-4 and 8-5. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 1 bp insertion (SEQ ID NO: 13 and 8) and chromosome 2 has 1 bp insertion (SEQ ID NO: 14 and 15).

FIG. 6 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 8-6. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 4 bp deletion (SEQ ID NO: 5 and 6) and chromosome 2 has 2 bp deletion and 3 bp insertion (SEQ ID NO: 16 and 17).

FIG. 7 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 8-7 and 8-8. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 4 bp deletion (SEQ ID NO: 5 and 6) and chromosome 2 has 1 bp insertion (SEQ ID NO: 7 and 8).

FIG. 8 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 20-1, 20-2, and 20-3. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 1 bp insertion (SEQ ID NO: 7 and 8) and chromosome 2 has 5 bp deletion and 3 bp insertion (SEQ ID NO: 18 and 19).

FIG. 9 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 1. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 4 bp deletion (SEQ ID NO: 5 and 6) and chromosome 2 has 1 bp insertion (SEQ ID NO: 14 and 15).

FIG. 10 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 6-1 and 6-2. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 1 bp insertion (SEQ ID NO: 14 and 15) and chromosome 2 has 1 bp insertion (SEQ ID NO: 13 and 8).

FIG. 11 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 6-3 and 6-4. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 4 bp deletion (SEQ ID NO: 5 and 6) and chromosome 2 has 1 bp insertion (SEQ ID NO: 7 and 8).

FIG. 12 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 8-10. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 4 bp deletion (SEQ ID NO: 5 and 6) and chromosome 2 has 2 bp deletion (SEQ ID NO: 20 and 19).

FIG. 13 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 9-1. The DNA sequences start with the last nucleotide of intron 6. Chromosomes 1 and 2 have 4 bp deletion (SEQ ID NO: 5 and 6).

FIG. 14 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 9-2. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 1 bp deletion (SEQ ID NO: 21 and 22) and chromosome 2 has 5 bp deletion (SEQ ID NO: 23 and 24).

FIG. 15 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 9-3. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 1 bp deletion (SEQ ID NO: 21 and 22) and chromosome 2 has 1 bp deletion (SEQ ID NO: 25 and 15).

FIG. 16 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 13-1 and 13-2. The DNA sequences start with the last nucleotide of intron 6. Chromosomes 1 and 2 have 1 bp deletion (SEQ ID NO: 21 and 22).

FIG. 17 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 16. The DNA sequences start with the last nucleotide of intron 6. Chromosomes 1 and 2 have 1 bp deletion (SEQ ID NO: 21 and 22).

FIG. 18 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant line 17. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 1 bp deletion (SEQ ID NO: 21 and 22) and chromosome 2 has 9 bp deletion (SEQ ID NO: 26 and 27).

FIG. 19 shows the 3' DNA and amino acid carboxy terminus sequences for the altered *RCA* α -isoform exon 7 in genetically modified *O. sativa* cultivar Kitaake plant lines 21-1 and 21-2. The DNA sequences start with the last nucleotide of intron 6. Chromosome 1 has 16 bp deletion (SEQ ID NO: 28 and 29) and chromosome 2 has 4 bp deletion (SEQ ID NO: 5 and 6).

FIG. 20A shows an immunoblot demonstrating removal of the long, redox-regulated *RCA* α -isoform from the transgenic rice lines 6-3, 8-4 and 9-3 but the presence of *RCA* β -isoform. The staining of the RuBisCO large isoform (RbcL) with coomassie brilliant blue (CBB) in the lower gel shows equal loading of the lanes. FIG. 20B compares the growth of the transgenic rice line 9-3 compared to wild-type (WT) rice during reproductive development. FIG. 20C shows that the transgenic rice lines 6-3, 8-4 and 9-3 have higher CO₂ assimilation rate following an increase in light from low to high irradiance, referred to as photosynthetic induction, compared to the wild-type (WT) rice plant. FIG. 20D shows that the

transgenic rice lines 6-3, 8-4 and 9-3 have increased seed yield per plant compared to wild-type rice plants, while stover amounts are relatively equal.

FIG. 21 shows an alignment of RCA α -isoform carboxy terminal amino acids from nineteen economically important plants, demonstrating the presence of two cysteine amino acids (in bold and underlined) at the carboxy terminus of the RCA α -isoform which are involved in the redox regulation of the protein.

DETAILED DESCRIPTION OF THE INVENTION

RuBisCO activase (also called “RCA”) is needed to activate the primary photosynthetic carbon assimilation enzyme, RuBisCO. In many species, RCA protein is a complex of two related proteins, the alpha- and beta-isoforms, which form a heteromeric complex involved in RuBisCO activation. In rice, the α -isoform is essentially identical in sequence to the shorter β -isoform, except that the carboxy-terminus of the α -isoform is encoded in exons 7a and 7b, while the carboxy-terminus of the β -isoform is encoded only in exon 7a. Exon 7b encodes the two cysteine amino acids that appear in wild-type α -isoform. Not wishing to be bound to any particular hypothesis, it is believed that these two redox-sensitive cysteines can reversibly form a disulfide bond that is responsible for modulating RuBisCO activity in response to varying light intensities. When referring to rice RCA α -isoform gene, exons 7a and 7b is also referred to as “exon 7”. Altered RCA α -isoform lacking one or both cysteine amino acids at the carboxy terminus of the protein results in an increase in altered RCA’s activity, compared to wild-type RCA’s activity. The increase in altered RCA’s activity may result in an increased activation state for RuBisCO, which allows plants to take better advantage of intermittent high light conditions (generally found in the lower leaves in canopies of crop plants) and quicker photosynthetic induction in varying light conditions. This results in the genetically altered plants growing better in alternating low-light/high-light conditions as well as non-fluctuating light conditions.

This invention involves modifying RCA’s α -isoform gene sequence via one or more truncations, deletions, and/or insertions in the exon that encodes RCA’s α -isoform redox-sensitive cysteines, thereby altering the amino acid sequence encoded in that exon (which is at the gene’s 3’ end). By eliminating one or both cysteine amino acids in the RCA’s α -isoform carboxy terminus, one can alter RCA’s activity and thus change the phenotype of the modified plant (*i.e.*, RCA α -isoform activation, RuBisCO activation, increased growth, plant height, leaf width, panicle number, and seed yield per plant). In rice, RCA’s α -isoform redox-sensitive cysteines are encoded in exon 7b.

In addition to improved growth under alternating low-light/high-light conditions, mutated RCA containing with an α -isoform with a carboxy terminal mutation results in plants that have increased growth, height, leaf width, panicle number, RCA α -isoform activation, RuBisCO activation, and seed

yield per plant under constant light during the photoperiod compared to those phenotypes in wild-type plants growing under similar conditions.

The DNA sequence of RuBisCO activase α -isoform in *O. sativa* cultivar Kitaake is identical to the DNA sequence in GenBank Accession XR_001541378 for *O. sativa* Japonica. The DNA and amino acid sequences of RuBisCO activase β -isoform in *O. sativa* cultivar Kitaake is identical to the DNA and amino acid sequences in GenBank Accession XM_015761411 for *O. sativa* Japonica. The DNA and amino acid sequences of the wild-type *O. sativa* cultivar Kitaake exons 7a and 7b are in SEQ ID NO: 1 and 2, respectively; see FIG. 1.

O. sativa cultivar Kitaake has a diploid genome, thus two homologous copies of each chromosome is present in each nucleus. The *RCA* gene is located on the 11th chromosome pair. When the *RCA* gene is altered, the gene on each pair of chromosomes is altered. Thus, as used herein and in the figures, “chromosome 1” and “chromosome 2” refer, arbitrarily, to one chromosome in the pair and to the other chromosome in the pair, respectively; not to the chromosome #1 pair and the chromosome #2 pair. Some plants have the α - and β -isoforms of RCA encoded in separate genes (unlike rice where there is one *RCA* gene that is alternatively spliced to form two transcripts). It is also possible to refer to “chromosome 1” as “transcript 1” and “chromosome 2” as “transcript 2” because they arise from same gene (*RCA* α -isoform) even though the gene on each chromosome can have different mutations/alterations and thus some of the encoded RCA α -isoforms have different amino acid sequences. See FIGs. 3-19 and Table 1, *infra*. Some plants have undergone polyploidy events, so that genes are present in multiple copies, as is the case for soybean and its *RCA* genes. Other plants may have copies of *RCA* genes on more than one chromosome. Regardless of the number of *RCA* genes in a plant's gene, the *RCA* genes can be altered in a similar manner as described in the examples below and produce genetically altered plants with the described phenotype.

The DNA and amino acid sequences of the mutated RCA α -isoform exon 7 in the transformed rice plants discussed herein, as well as other important sequences are in Table 1. See also FIGs. 1-19 for the *RCA* α -isoform carboxy terminal sequences for wild-type rice and transgenic rice lines. Note that the DNA sequences for transgenic rice lines in FIGs. 1 and 3-19 include the last nucleotide of intron 6. FIG. 2 contains a longer portion of intron 6 than the other figures.

Table 1

RCA α -isoform exon 7 in indicated rice plant	Mutation Description	Sequence Numbers (DNA and amino acid)
Wild-type rice (FIG. 1)	not applicable	SEQ ID NOs: 1 and 2
Tr. line 2-1 (FIG. 2)	147 bp del. in chr. 1 and in chr. 2	SEQ ID NOs: 3 and 4
Tr. lines 8-1 & 8-2 (FIG. 3)	4 bp del. in chr. 1	SEQ ID NOs: 5 and 6
	1 bp ins. in chr. 2 (c)*	SEQ ID NOs: 7 and 8

Tr. line 8-3 (FIG. 4)	2 bp del. in chr. 1	SEQ ID NOs: 9 and 10
	6 bp del. & 2 bp ins. in chr. 2 (ag)*	SEQ ID NOs: 11 and 12
Tr. lines 8-4 & 8-5 (FIG. 5)	1 bp ins. in chr. 1 (t)*	SEQ ID NOs: 13 and 8
	1 bp ins. in chr. 2 (a)*	SEQ ID NOs: 14 and 15
Tr. line 8-6 (FIG. 6)	4 bp del. in chr. 1	SEQ ID NOs: 5 and 6
	2 bp del. & 3 bp ins. in chr 2 (cat)*	SEQ ID NOs: 16 and 17
Tr. lines 8-7 & 8-8 (FIG. 7)	4 bp del. in chr. 1	SEQ ID NOs: 5 and 6
	1 bp ins. in chr. 2 (c)*	SEQ ID NOs: 7 and 8
Tr. lines 20-1, 20-2 & 20-3 (FIG. 8)	1 bp ins. in chr. 1 (c)* or (g)*	SEQ ID NOs: 7 and 8
	5 bp del. & 3 bp ins. in chr. 2 (agt)* ^s	SEQ ID NOs: 18 and 19
Tr. line 1 (FIG. 9)	4 bp del. in chr. 1	SEQ ID NOs: 5 and 6
	1 bp ins. in chr. 2 (a)*	SEQ ID NOs: 14 and 15
Tr. lines 6-1 & 6-2 (FIG. 10)	1 bp ins. in chr. 1 (a)*	SEQ ID NOs: 14 and 15
	1 bp ins. in chr. 2 (t)*	SEQ ID NOs: 13 and 8
Tr. lines 6-3 & 6-4 (FIG. 11)	4 bp del. in chr. 1	SEQ ID NOs: 5 and 6
	1 bp ins. in chr. 2 (c)*	SEQ ID NOs: 7 and 8
Tr. line 8-10 (FIG. 12)	4 bp del. in chr. 1	SEQ ID NOs: 5 and 6
	2 bp del. in chr. 2	SEQ ID NOs: 20 and 19
Tr. line 9-1 (FIG. 13)	4 bp del. in chr. 1	SEQ ID NOs: 5 and 6
	4 bp del. in chr. 2	SEQ ID NOs: 5 and 6
Tr. line 9-2 (FIG. 14)	1 bp del. in chr. 1	SEQ ID NOs: 21 and 22
	5 bp del. in chr. 2	SEQ ID NOs: 23 and 24
Tr. line 9-3 (FIG. 15)	1 bp del. in chr. 1	SEQ ID NOs: 21 and 22
	1 bp ins. in chr. 2 (g)*	SEQ ID NOs: 25 and 15
Tr. lines 13-1 & 13-2 (FIG. 16)	1 bp del. in chr. 1	SEQ ID NOs: 21 and 22
	1 bp del. in chr. 2	SEQ ID NOs: 21 and 22
Tr. line 16 (FIG. 17)	1 bp del. in chr. 1	SEQ ID NOs: 21 and 22
	1 bp del. in chr. 2	SEQ ID NOs: 21 and 22
Tr. line 17 (FIG. 18)	1 bp del. in chr. 1	SEQ ID NOs: 21 and 22
	9 bp del. in chr. 2	SEQ ID NOs: 26 and 27
Tr. lines 21-1 & 21-2 (FIG. 19)	16 bp del. in chr. 1	SEQ ID NOs: 28 and 29
	4 bp del. in chr. 2	SEQ ID NOs: 5 and 6

“Tr.” = transgenic; “del.” = deletion; “ins.” = insertion; “chr.” = chromosome (alternatively “transcript”); * = indicates the inserted nucleotide(s); \$ = indicates that the mutation could be considered a 3 bp del. & 1 bp ins.

- 5 In addition, SEQ ID NOs: 37, 38, 39, 40, and 42 lists the carboxy terminal amino acid sequences of five different mutated *RCA* α genes containing mutations in exon 7. These sequences are discussed *infra* in the examples. They are similar to some of the amino acid sequences listed in Table 1, *supra*, except that SEQ ID NOs: 37, 38, 39, 40, and 42 have an initial glycine (G) amino acid which arises from the codon spanning intron 6 and intron 7, whereas the amino acid sequences listed in Table 1 start with the first
- 10 amino acid for which its codon is completely in exon 7.

In the examples below, rice (*Oryza sativa*) cultivar Kitaake is used, but the alteration in RCA α -isoform's gene sequence such that the cysteine amino acids at the carboxy terminus are removed from the altered α -isoform can be made in any plant and have similar phenotypically changes. In fact, as shown in

FIG. 21, the carboxy terminal amino acid sequence of RCA α -isoform is very similar in the following economically important plants (which is not an exhaustive list): *Brassica napus* (rapeseed; SEQ ID NO: 45), *Camelina sativa* (SEQ ID NO: 46), *Arachis duranensis* (SEQ ID NO: 47), *Vigna angularis* (red mung bean; SEQ ID NO: 48), *Cucumis melo* (muskmelon; SEQ ID NO: 49), *Cucumis sativus* (cucumber; SEQ ID NO: 50), *Fragaria vesca* subsp. *vesca* (wild strawberry; SEQ ID NO: 51), *Gossypium hirsutum* (upland cotton; SEQ ID NO: 52), *Theobroma cacao* (cacao; SEQ ID NO: 53), *Morus notabilis* (mulberry; SEQ ID NO: 54), *Prunus mume* (plum; SEQ ID NO: 55), *Beta vulgaris* subsp. *vulgaris* (beet; SEQ ID NO: 56), *Citrus clementina* (clementine orange; SEQ ID NO: 57), *Glycine max* (soybean; SEQ ID NO: 58), *Malus domestica* (apple; SEQ ID NO: 59), *Prunus persica* (peach; SEQ ID NO: 60), *Vitis vinifera* (grape; SEQ ID NO: 61), *Musa acuminata* subsp. *malaccensis* (banana; SEQ ID NO: 62), and *Zea mays* (corn; SEQ ID NO: 63). The two redox-sensitive cysteine amino acids in the carboxy terminus of the α -isoform is underlined in each sequence in FIG. 21. This alignment indicates that the deletion of one or both of these redox-sensitive cysteine amino acids in the RCA α -isoform for these economically important plants, as well as other plants with a similar RCA α -isoform, would result in the modified plants having the same altered phenotype that is described herein. The sequences in FIG. 21 are obtained from public databases which have the full-length amino acid and DNA sequences for RCA α -isoform of those plants, and other plants. One can altered the exon that encodes these redox-sensitive cysteines to generate an altered RCA α -isoform gene that encodes an altered RCA α -isoform with the properties described herein.

Reducing the Regulatory Activity of an RCA α -isoform Polypeptide

Methods are provided to reduce or eliminate the regulatory activity of RuBisCO activase α -isoform in one embodiment by transforming a plant cell with an expression cassette that expresses a polynucleotide that reduces the gene expression, mRNA production, and/or translation of the RCA α -isoform polypeptide. The polynucleotide may reduce the expression of the RCA α -isoform polypeptide directly, by preventing transcription or translation of the RCA α -isoform mRNA, or indirectly, by encoding a polypeptide that reduces the transcription or translation of a RCA α -isoform gene or mRNA. Methods for reducing or eliminating the expression of a gene in a plant are well known in the art and any such method may be used in the present disclosure to reduce the expression of RCA α -isoform polypeptide.

In other embodiments, specific mutations (insertions, deletions, *etc.*) are introduced into the portion of a cell's genome which encodes RCA α -isoform to prevent transcription or translation of a "wild-type" functional α -isoform having "wild-type" activity. These types of mutations (or changes in the DNA encoding RCA α -isoform (exon 7 in rice or the appropriate exon in other plants)) can occur through

gene editing or site directed mutagenesis. In one embodiment, the mutation is introduced in the appropriate exon of *RCA* α -isoform gene which results in the translation of an altered *RCA* α -isoform protein lacking one or both cysteine amino acids at the carboxy terminus that results in improved plant performance and seed yield, as the result of the altered *RCA* having a higher continuous activation state compared to wild-type *RCA*'s activation state.

The expression of a *RCA* α -isoform polypeptide is reduced if the amount of the *RCA* α -isoform polypeptide is less than 100%, 99%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% or 1% of the amount of the same *RCA* α -isoform polypeptide in a control (wild-type) plant. In other embodiments, the amount of the *RCA* α -isoform polypeptide in a modified plant is less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5% or less than 2% of the amount of the same or a related *RCA* α -isoform polypeptide in a control plant. The *RCA* α -isoform polynucleotide expression level and/or amount of polypeptide and/or enzymatic activity may be reduced or altered such that the reduction/alteration is phenotypically sufficient to provide increase in plant biomass, yield seed per plant, and/or RuBisCO's activation status. The level or activity of one or more *RCA* α -isoform polynucleotides, polypeptides or enzymes may be impacted. The expression level of the *RCA* α -isoform polypeptide may be measured directly, for example, by assaying for the quantity of *RCA* α -isoform polypeptide expressed in the plant cell or plant, or indirectly, for example, by measuring the *RCA* activity in the plant cell or plant or by measuring the phenotypic changes in the plant or by measuring RuBisCO levels or activation status. Methods for performing such assays are described elsewhere herein.

In certain embodiments of the invention, the activity of the *RCA* α -isoform polypeptide is reduced or eliminated by transforming a plant cell with an expression cassette containing a polynucleotide encoding a polypeptide that inhibits the activity of a *RCA* α -isoform polypeptide. The activity of a *RCA* α -isoform polypeptide is reduced if the *RCA* α -isoform polypeptide's activity is less than 100%, 99%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% or 1% of the activity of the same *RCA* α -isoform polypeptide in a control plant. In particular embodiments, the *RCA* α -isoform's activity in a modified plant is less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10% or less than 5% of the *RCA* α -isoform's activity in a control plant. The *RCA* α -isoform polypeptide's activity is "eliminated" when it is not detectable by the assay methods described elsewhere herein. Methods of determining the alteration of activity of a *RCA* α -isoform polypeptide are described elsewhere herein.

In other embodiments, *RCA* α -isoform polypeptide's activity may be reduced or eliminated by mutating, disrupting, or excising at least a part of the *RCA* gene encoding the *RCA* α -isoform polypeptide, and, in one embodiment, *RCA* gene's exon that encodes the carboxy terminal redox-sensitive cysteines.

Mutagenized plants that carry mutations in *RCA* α -isoform genes (in one or more chromosomes) also result in reduced production of wild-type *RCA* α -isoform protein and/or reduced activity of the altered *RCA* α -isoform.

Thus, many methods may be used to reduce or eliminate the activity of a wild-type *RCA* α -isoform protein. One or more methods may be used to reduce the activity of a single *RCA* α -isoform polypeptide. One or more methods may be used to reduce the activity of multiple *RCA* α -isoform polypeptides.

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, which are within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Langenheim and Thimann, (1982) *Botany: Plant Biology and Its Relation to Human Affairs*, John Wiley; *Cell Culture and Somatic Cell Genetics of Plants*, vol. 1, Vasil, ed. (1984); Stanier, *et al.*, (1986) *The Microbial World*, 5th ed., Prentice-Hall; Dhringra and Sinclair, (1985) *Basic Plant Pathology Methods*, CRC Press; Maniatis, *et al.*, (1982) *Molecular Cloning: A Laboratory Manual*; *DNA Cloning*, vols. I and II, Glover, ed. (1985); *Oligonucleotide Synthesis*, Gait, ed. (1984); *Nucleic Acid Hybridization*, Hames and Higgins, eds. (1984); and the series *Methods in Enzymology*, Colowick and Kaplan, eds, Academic Press, Inc., San Diego, CA.

Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. The terms defined below are more fully defined by reference to the specification as a whole.

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-Beta Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA). See, e.g., *Diagnostic Molecular Microbiology: Principles and Applications*, Persing, *et al.*, eds., American Society for Microbiology, Washington, DC (1993). The product of amplification is termed an amplicon.

The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refer to those nucleic acids that encode identical or conservatively modified variants of the amino acid sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively modified variation. Every nucleic acid sequence herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; one exception is *Micrococcus rubens*, for which GTG is the methionine codon (Ishizuka, *et al.*, (1993) *J. Gen. Microbiol.* 139:425-32) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid, which encodes a polypeptide of the present invention, is implicit in each described polypeptide sequence and incorporated herein by reference. See Table 2.

Table 2

Amino acid	Nucleic acid codons	Amino acid	Nucleic acid codons
Ala/A	GCT, GCC, GCA, GCG	Leu/L	TTA, TTG, CTT, CTC, CTA, CTG
Arg/R	CGT, CGC, CGA, CGG, AGA, AGG	Lys/K	AAA, AAG
Asn/N	AAT, AAC	Met/M	ATG
Asp/D	GAT, GAC	Phe/F	TTT, TTC
Cys/C	TGT, TGC	Pro/P	CCT, CCC, CCA, CCG
Gln/Q	CAA, CAG	Ser/S	TCT, TCC, TCA, TCG, AGT, AGC
Glu/E	GAA, GAG	Thr/T	ACT, ACC, ACA, ACG
Gly/G	GGT, GGC, GGA, GGG	Trp/W	TGG
His/H	CAT, CAC	Tyr/Y	TAT, TAC
Ile/I	ATT, ATC, ATA	Val/V	GTT, GTC, GTA, GTG
Stop	TAA, TGA, TAG		

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" when the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from 1 to

15 can be so altered. Thus, for example, 1, 2, 3, 4, 5, 7 or 10 alterations can be made. Conservatively modified variants typically provide similar biological activity as the unmodified polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity, or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80% or 90%, preferably 60-90% of the native
 5 protein for its native substrate. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 10 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

15 See also, Creighton, *Proteins*, W.H. Freeman and Co. (1984).

As used herein, "consisting essentially of" means the inclusion of additional sequences to an object polynucleotide where the additional sequences do not selectively hybridize, under stringent hybridization conditions, to the same cDNA as the polynucleotide and where the hybridization conditions include a wash step in 0.1X SSC and 0.1% sodium dodecyl sulfate at 65°C.

20 By "encoding" or "encoded," with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using
 25 the "universal" genetic code. However, variants of the universal code, such as is present in some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma capricolum* (Yamao, *et al.*, (1985) *Proc. Natl. Acad. Sci. USA* 82:2306-9), or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed using these organisms.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known
 30 codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present invention may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledonous plants or dicotyledonous plants as these preferences

have been shown to differ (Murray, *et al.*, (1989) *Nucleic Acids Res.* 17:477-98). Thus, the rice preferred codon for a particular amino acid might be derived from known gene sequences from rice.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell, which comprises a heterologous nucleic acid sequence of the invention, which contains a vector and supports the replication and/or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, plant, amphibian, or mammalian cells. In one embodiment of this invention, host cells are monocotyledonous or dicotyledonous plant cells, including, but not limited to, maize, sorghum, sunflower, soybean, wheat, alfalfa, rice, cotton, canola, lawn grass, barley, millet, and tomato. In another embodiment of this invention, the monocotyledonous host cell is a rice host cell.

The term "hybridization complex" includes reference to a duplex nucleic acid structure formed by two single-stranded nucleic acid sequences selectively hybridized with each other.

The term "introduced" in the context of inserting a nucleic acid into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The terms "isolated" or "isolated nucleic acid" or "isolated protein" refer to material, such as a nucleic acid or a protein, which is substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment. Nucleic acids which are "isolated", as defined herein, are also referred to as "heterologous" nucleic acids. Unless otherwise stated, the term "RuBisCO activase alpha-isoform nucleic acid" and "RCA α -isoform nucleic acid" means a nucleic acid comprising a polynucleotide ("RCA α -isoform polynucleotide") encoding a full length or partial length RCA α -isoform polypeptide with RCA α -isoform activity as defined herein.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known

analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules, which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism.

Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, (1987) *Guide To Molecular Cloning Techniques*, from the series *Methods in Enzymology*, vol. 152, Academic Press, Inc., San Diego, CA; Sambrook, *et al.*, (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., vols. 1-3; and *Current Protocols in Molecular Biology*, Ausubel, *et al.*, eds, Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994 Supplement).

As used herein "operably linked" includes reference to a functional linkage between a first sequence, such as a promoter, and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, cells in or from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. The class of plants which can be used in the methods of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants including species from the genera: *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesis*, *Pelargonium*, *Panieum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Pisum*, *Phaseolus*, *Lolium*, *Oryza*, *Avena*, *Hordeum*, *Secale*, *Allium*, and *Triticum*. *Zea mays*, *Glycine max*, and *Oryza sativa* are among plants that can be improved by this invention.

As used herein, "yield" may include reference to bushels per acre of a grain crop at harvest, as adjusted for grain moisture (15% typically for maize, for example), and/or the volume of biomass generated (for forage crops such as alfalfa, and plant root size for multiple crops). Grain moisture is measured in the grain at harvest. The adjusted test weight of grain is determined to be the weight in pounds per bushel, adjusted for grain moisture level at harvest. Biomass is measured as the weight of harvestable plant material generated.

As used herein, "polynucleotide" includes reference to a deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring nucleotides and/or allow translation into the same amino acid(s) as the naturally occurring nucleotide(s). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including *inter alia*, simple and complex cells.

The terms "polypeptide," "peptide," and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria which comprise genes expressed in plant cells such *Agrobacterium* or *Rhizobium*. Examples are promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, seeds, fibres, xylem vessels, tracheids, or sclerenchyma. Such promoters are referred to as "tissue-preferred." A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" or "regulatable" promoter is a promoter which is under environmental control. Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions or the presence of light. Another type of promoter is a developmentally regulated promoter, for example, a promoter that drives expression during pollen development. Tissue preferred, cell type specific, developmentally regulated, and inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter, which is active under most environmental conditions.

The terms "RuBisCO activase alpha-isoform polypeptide", "RCA α -isoform", "RCA α -isoform protein", and similar variations refer to the same protein/polypeptide. The term is also inclusive of fragments, variants, homologs, alleles or precursors (e.g., preproteins or proproteins) thereof that retain the function of the regulatory/inactivation of RCA in the dark or in low light. In rice, this is the carboxy-terminal extension found only in the alpha-isoform, specifically the two redox-sensitive cysteine residues. Unless otherwise stated, the terms "RuBisCO activase alpha nucleic acid", "RCA α -isoform polynucleotide", "RCA α -isoform gene" and similar variations means a polynucleotide encoding a RCA α -isoform polypeptide.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, organism, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells may express genes/polynucleotides that are not found within the native (non-recombinant or wild-type) form of the cell or express native genes in an otherwise abnormal amount – over-expressed, under-expressed or not expressed at all – compared to the non-recombinant or wild-type cell or organism. In particular, one can alter the genomic DNA of a wild-type plant by molecular biology techniques that are well-known to one of ordinary skill in the art and generate a recombinant plant. The term "recombinant" as used herein does not encompass the alteration of the cell or vector by naturally occurring events (e.g., spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements, which permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

The terms "residue" or "amino acid residue" or "amino acid" are used interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass known analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids. The terms carboxy-terminus, carboxyl-terminal, C-terminal, carboxy-terminus, C-terminus, and similar phrases (with or without the hyphen) refer to the portion of the protein and/or polypeptide distal to the amino terminus. The amino terminus is translated from mRNA first and the carboxy-terminus is translated from mRNA last.

The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 40% sequence identity, preferably 60-90% sequence identity, and most preferably 100% sequence identity (i.e., complementary) with each other.

The terms "stringent conditions" or "stringent hybridization conditions" include reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than 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 can be identified which can be up to 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Optimally, the probe is approximately 500 nucleotides in length, but can vary greatly in length from less than 500 nucleotides to equal to the entire length of the target sequence.

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). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide or Denhardt's. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, (1984) *Anal. Biochem.*, 138:267-84: $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\% \text{ form}) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m , hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity.

For example, if sequences with $\geq 90\%$ identity are sought, the T_m can be decreased 10°C . Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3 or 4°C lower than the thermal melting point (T_m);

5 moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9 or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15 or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T_m , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results

10 in a T_m of less than 45°C (aqueous solution) or 32°C (formamide solution), one can increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes*, part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," Elsevier, New York (1993); and *Current Protocols in Molecular Biology*,

15 chapter 2, Ausubel, *et al.*, eds, Greene Publishing and Wiley-Interscience, New York (1995). Unless otherwise stated, in the present application high stringency is defined as hybridization in 4X SSC, 5X Denhardt's (5 g Ficoll, 5 g polyvinylpyrrolidone, 5 g bovine serum albumin in 500ml of water), 0.1 mg/ml boiled salmon sperm DNA, and 25 mM Na phosphate at 65°C , and a wash in 0.1X SSC, 0.1% SDS at 65°C .

20 As used herein, "transgenic plant" includes reference to a plant, which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. The terms "transgenic", "transformed", "transformation", and "transfection" are similar in meaning to

25 "recombinant". "Transformation", "transgenic", and "transfection" refer to the transfer of a polynucleotide into a host organism or into a cell. Such a transfer of polynucleotides can result in genetically stable inheritance of the polynucleotides or in the polynucleotides remaining extra-chromosomally (not integrated into the chromosome of the cell). Genetically stable inheritance may potentially require the transgenic organism or cell to be subjected for a period of time to one or more

30 conditions which require the transcription of some or all of the transferred polynucleotide in order for the transgenic organism or cell to live and/or grow. Polynucleotides that are transformed into a cell but are not integrated into the host's chromosome remain as an expression vector within the cell. One may need to grow the cell under certain growth or environmental conditions in order for the expression vector to remain in the cell or the cell's progeny. Further, for expression to occur the organism or cell may need to

be kept under certain conditions. Genetically altered organisms or cells containing the recombinant polynucleotide can be referred to as “transgenic” or “transformed” organisms or cells or simply as “transformants”, as well as recombinant organisms or cells or “altered” organism, cell, gene, *etc.*

A genetically altered organism or altered organism is any organism with any changes to its genetic material, whether in the nucleus or cytoplasm (organelle). As such, a genetically altered organism can be a recombinant or transformed organism. A genetically altered organism can also be an organism that was subjected to one or more mutagens or the progeny of an organism that was subjected to one or more mutagens and has mutations in its DNA caused by the one or more mutagens, as compared to the wild-type organism (*i.e.*, organism not subjected to the mutagens). Also, an organism that has been bred to incorporate a mutation into its genetic material is a genetically altered organism.

As used herein, “vector” includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides or polypeptides: (a) “reference sequence,” (b) “comparison window,” (c) “sequence identity,” (d) “percentage of sequence identity,” and (e) “substantial identity.”

As used herein, “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

As used herein, “comparison window” means includes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, and 50, 100 or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of nucleotide and amino acid sequences for comparison are well known in the art. The local homology algorithm (BESTFIT) of Smith and Waterman, (1981) *Adv. Appl. Math* 2:482, may conduct optimal alignment of sequences for comparison; by the homology alignment algorithm (GAP) of Needleman and Wunsch, (1970) *J. Mol. Biol.* 48:443-53; by the search for similarity method (Tfasta and Fasta) of Pearson and Lipman, (1988) *Proc. Natl. Acad. Sci. USA* 85:2444; by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the

PC/Gene program by Intelligenetics, Mountain View, California, GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG® programs (Accelrys, Inc., San Diego, CA).). The CLUSTAL program is well described by Higgins and Sharp, (1988) *Gene* 73:237-44; Higgins and Sharp, (1989) *CABIOS* 5:151-3; Corpet, *et al.*, (1988) *Nucleic Acids Res.* 16:10881-90; Huang, *et al.*, (1992) *Computer Applications in the Biosciences* 8:155-65, and Pearson, *et al.*, (1994) *Meth. Mol. Biol.* 24:307-31. One useful program for optimal global alignment of multiple sequences is PileUp (Feng and Doolittle, (1987) *J. Mol. Evol.*, 25:351-60 which is similar to the method described by Higgins and Sharp, (1989) *CABIOS* 5:151-53). The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, *Current Protocols in Molecular Biology*, Chapter 19, Ausubel *et al.*, eds., Greene Publishing and Wiley-Interscience, New York (1995).

GAP uses the algorithm of Needleman and Wunsch, *supra*, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package are 8 and 2, respectively. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 100. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, and 40, 50 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (see, Henikoff and Henikoff, (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters (Altschul, *et al.*, (1997) *Nucleic Acids Res.* 25:3389-402).

As those of ordinary skill in the art will understand, BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences, which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, (1993) *Comput. Chem.* 17:149-63) and XNU (Claverie and States, (1993) *Comput. Chem.* 17:191-201) low-complexity filters can be employed alone or in combination.

As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences, which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution.

Sequences, which differ by such conservative substitutions, are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, (1988) *Computer Applic. Biol. Sci.* 4:11-17, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched

positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has between 50-100% sequence identity, preferably at least 50% sequence identity, preferably at least 60% sequence identity, preferably at least 70%, more preferably at least 80%, more preferably at least 90%, and most preferably at least 95%, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of between 55-100%, preferably at least 55%, preferably at least 60%, more preferably at least 70%, 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. The degeneracy of the genetic code allows for many amino acids substitutions that lead to variety in the nucleotide sequence that code for the same amino acid, hence it is possible that the DNA sequence could code for the same polypeptide but not hybridize to each other under stringent conditions. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is that the polypeptide, which the first nucleic acid encodes, is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

The terms "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with between 55-100% sequence identity to a reference sequence preferably at least 55% sequence identity, preferably 60% preferably 70%, more preferably 80%, most preferably at least 90% or 95% sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch, *supra*. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. In addition, a peptide can be substantially identical to a second peptide when they differ by a non-conservative change if the epitope that the antibody recognizes is substantially identical. Peptides which are "substantially similar" share sequences as noted above, except that residue positions which are not identical may differ by conservative amino acid changes.

*Reducing the Regulatory Activity of an RCA α -isoform Polypeptide
Polynucleotide-Based Methods:*

In some embodiments of the present invention, a plant is transformed with an expression cassette that is capable of expressing a polynucleotide that inhibits the expression of a RCA α -isoform polypeptide of the invention. The term "expression" as used herein refers to the biosynthesis of a gene product, including the transcription and/or translation of said gene product. For example, for the purposes of the present invention, an expression cassette capable of expressing a polynucleotide that inhibits the expression of at least one RCA α -isoform polypeptide is an expression cassette capable of producing an RNA molecule that inhibits the transcription and/or translation of at least one RCA α -isoform polypeptide of the invention. The "expression" or "production" of a protein or polypeptide from a DNA molecule refers to the transcription and translation of the coding sequence to produce the protein or polypeptide, while the "expression" or "production" of a protein or polypeptide from an RNA molecule refers to the translation of the RNA coding sequence to produce the protein or polypeptide.

Examples of polynucleotides that inhibit the expression of a RCA α -isoform polypeptide are given below.

i. Sense Suppression/Cosuppression

In some embodiments of the invention, inhibition of the expression of a RCA α -isoform polypeptide may be obtained by sense suppression or cosuppression. For cosuppression, an expression cassette is designed to express an RNA molecule corresponding to all or part of exon that encodes the carboxy terminal redox-sensitive cysteines of a messenger RNA encoding a RCA α -isoform polypeptide that in the "sense" orientation. In an alternative embodiment, one would include a poly-T nucleotides between the coding sequences and the poly-A tail so that dsRNA can form by bonding of the poly-T and poly-A nucleotides. Over-expression of the RNA molecule can result in reduced expression of the native gene. Accordingly, multiple plant lines transformed with the cosuppression expression cassette are screened to identify those that show the greatest inhibition of RCA α -isoform polypeptide expression.

The polynucleotide used for cosuppression may correspond to all or part of the coding sequences for RCA α -isoform's exon which encodes the carboxy terminal redox sensitive cysteines, all or part of the 5' and/or 3' untranslated region of a RCA α -isoform polypeptide transcript, or all or part of both the coding sequence and the untranslated regions of a transcript encoding a RCA α -isoform polypeptide. In some embodiments where the polynucleotide comprises all or part of the coding region for the RCA α -isoform polypeptide, the expression cassette is designed to eliminate the start codon of the polynucleotide so that no protein product will be translated.

Cosuppression may be used to inhibit the expression of plant genes to produce plants having undetectable protein levels for the proteins encoded by these genes. See, for example, Broin, *et al.*, (2002)

Plant Cell 14:1417-1432. Cosuppression may also be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Methods for using cosuppression to inhibit the expression of endogenous genes in plants are described in Flavell, *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:3490-3496; Jorgensen, *et al.*, (1996) *Plant Mol. Biol.* 31:957-973; Johansen and Carrington, (2001) *Plant Physiol.* 126:930-938; Broin, *et al.*, (2002) *Plant Cell* 14:1417-1432; Stoutjesdijk, *et al.*, (2002) *Plant Physiol.* 129:1723-1731; Yu, *et al.*, (2003) *Phytochemistry* 63:753-763; and U.S. Patent Nos. 5,034,323, 5,283,184, and 5,942,657. The efficiency of cosuppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the sense sequence and 5' of the polyadenylation signal. See, U.S. Patent Publication No. 2002/0048814. Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous gene, optimally greater than about 65% sequence identity, more optimally greater than about 85% sequence identity, most optimally greater than about 95% sequence identity. See U.S. Patent Nos. 5,283,184 and 5,034,323.

ii. Antisense Suppression

In some embodiments of the invention, inhibition of the expression of the RCA α -isoform polypeptide may be obtained by antisense suppression. For antisense suppression, the expression cassette is designed to express an RNA molecule complementary to all or part of exon which encodes the carboxy terminal redox-sensitive cysteines in a messenger RNA encoding the RCA α -isoform polypeptide. Over-expression of the antisense RNA molecule can result in reduced expression of the native gene.

Accordingly, multiple plant lines transformed with the antisense suppression expression cassette are screened to identify those that show the greatest inhibition RCA α -isoform polypeptide expression.

The polynucleotide for use in antisense suppression may correspond to all or part of the complement of the exon sequence which encodes the the carboxy terminal redox-sensitive cysteines in RCA α -isoform polypeptide, all or part of the complement of the 5' and/or 3' untranslated region of the RCA α -isoform RCA α -isoform transcript, or all or part of the complement of both the coding sequence and the untranslated regions of a transcript encoding the RCA α -isoform polypeptide. In addition, the antisense polynucleotide may be fully complementary (i.e., 100% identical to the complement of the target sequence) or partially complementary (i.e., less than 100% identical to the complement of the target sequence) to the target sequence. Antisense suppression may be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, 300, 400, 450, 500, 550, or greater may be used. Methods for using antisense suppression to inhibit the expression of endogenous genes in plants are described, for example, in Liu, *et al.*, (2002) *Plant Physiol.* 129:1732-1743 and U.S. Patent Nos.

5,759,829 and 5,942,657. Efficiency of antisense suppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the antisense sequence and 5' of the polyadenylation signal. See, U.S. Patent Publication No. 2002/0048814.

iii. *Double-Stranded RNA Interference*

5 In some embodiments of the invention, inhibition of the expression of a RCA α -isoform polypeptide may be obtained by double-stranded RNA (dsRNA) interference. For dsRNA interference, a sense RNA molecule like that described above for cosuppression and an antisense RNA molecule that is fully or partially complementary to the sense RNA molecule are expressed in the same cell, resulting in inhibition of the expression of the corresponding endogenous messenger RNA. In one embodiment, the dsRNA would contain at least 19 contiguous nucleotides present in the RCA α -isoform exon 7b portion of
10 gene (for rice) or at least 19 contiguous in the exon that encodes the carboxy terminal redox-sensitive cysteines for any plant's RCA α -isoform.

Expression of the sense and antisense molecules can be accomplished by designing the expression cassette to comprise both a sense sequence and an antisense sequence. Alternatively, separate
15 expression cassettes may be used for the sense and antisense sequences. Multiple plant lines transformed with the dsRNA interference expression cassette or expression cassettes are then screened to identify plant lines that show the greatest inhibition of RCA α -isoform polypeptide expression. Methods for using dsRNA interference to inhibit the expression of endogenous plant genes are described in Waterhouse, *et al.*, (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964, Liu, *et al.*, (2002) *Plant Physiol.* 129:1732-1743,
20 and WO 99/49029, WO 99/53050, WO 99/61631, and WO 00/49035.

iv. *Hairpin RNA Interference and Intron-Containing Hairpin RNA Interference*

In some embodiments of the invention, inhibition of the expression of a RCA α -isoform polypeptide may be obtained by hairpin RNA (hpRNA) interference or intron-containing hairpin RNA (ihpRNA) interference. These methods are highly efficient at inhibiting the expression of endogenous
25 genes. See, Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38 and the references cited therein.

For hpRNA interference, the expression cassette is designed to express an RNA molecule that hybridizes with itself to form a hairpin structure that comprises a single-stranded loop region and a base-paired stem. The base-paired stem region comprises a sense sequence corresponding to all or part of the endogenous messenger RNA encoding the gene whose expression is to be inhibited, and an antisense
30 sequence that is fully or partially complementary to the sense sequence. Alternatively, the base-paired stem region may correspond to a portion of a promoter sequence controlling expression of the gene to be inhibited. In one embodiment, the dsRNA would contain at least 19 contiguous nucleotides present in the exon that encodes the carboxy terminal redox-sensitive cysteines for any plant's RCA α -isoform. Thus, the base-paired stem region of the molecule generally determines the specificity of the RNA interference.

hpRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and the RNA interference they induce is inherited by subsequent generations of plants. See, for example, Chuang and Meyerowitz, (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk, *et al.*, (2002) *Plant Physiol.* 129:1723-1731; and Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38. Methods for using
 5 hpRNA interference to inhibit or silence the expression of genes are described, for example, in Chuang and Meyerowitz, (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk, *et al.*, (2002) *Plant Physiol.* 129:1723-1731; Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38; Pandolfini *et al.*, *BMC Biotechnology* 3:7, and U.S. Patent Publication No. 2003/0175965. A transient assay for the efficiency of hpRNA constructs to silence gene expression *in vivo* has been described by Panstruga, *et al.*,
 10 (2003) *Mol. Biol. Rep.* 30:135-140.

For ihpRNA, the interfering molecules have the same general structure as for hpRNA, but the RNA molecule additionally comprises an intron that is capable of being spliced in the cell in which the ihpRNA is expressed. The use of an intron minimizes the size of the loop in the hairpin RNA molecule following splicing, and this increases the efficiency of interference. See, for example, Smith, *et al.*, (2000)
 15 *Nature* 407:319-320. In fact, Smith, *et al.*, show 100% suppression of endogenous gene expression using ihpRNA-mediated interference. Methods for using ihpRNA interference to inhibit the expression of endogenous plant genes are described, for example, in Smith, *et al.*, (2000) *Nature* 407:319-320; Wesley, *et al.*, (2001) *Plant J.* 27:581-590; Wang and Waterhouse, (2001) *Curr. Opin. Plant Biol.* 5:146-150; Waterhouse and Helliwell, (2003) *Nat. Rev. Genet.* 4:29-38; Helliwell and Waterhouse, (2003) *Methods*
 20 30:289-295, and U.S. Patent Publication No. 2003/0180945.

The expression cassette for hpRNA interference may also be designed such that the sense sequence and the antisense sequence do not correspond to an endogenous RNA. In this embodiment, the sense and antisense sequence flank a loop sequence that comprises a nucleotide sequence corresponding to all or part of the endogenous messenger RNA of the target gene. Thus, it is the loop region that
 25 determines the specificity of the RNA interference. See, for example, WO 02/00904; Mette, *et al.*, (2000) *EMBO J* 19:5194-5201; Matzke, *et al.*, (2001) *Curr. Opin. Genet. Devel.* 11:221-227; Scheid, *et al.*, (2002) *Proc. Natl. Acad. Sci., USA* 99:13659-13662; Aufsatz, *et al.*, (2002) *Proc. Nat'l. Acad. Sci.* 99(4):16499-16506; Sijen, *et al.*, *Curr. Biol.* (2001) 11:436-440.

v. Amplicon-Mediated Interference

30 Amplicon expression cassettes comprise a plant virus-derived sequence that contains all or part of the target gene but generally not all of the genes of the native virus. The viral sequences present in the transcription product of the expression cassette allow the transcription product to direct its own replication. The transcripts produced by the amplicon may be either sense or antisense relative to the target sequence (i.e., the messenger RNA for the RCA α -isoform polypeptide). In one embodiment, the

transcript targets the *RCA* exon 7b portion of gene (for rice) or, for other types of plants, the exon that encodes the carboxy terminal redox-sensitive cysteines for a plant's *RCA* α -isoform. Methods of using amplicons to inhibit the expression of endogenous plant genes are described, for example, in Angell and Baulcombe, (1997) *EMBO J.* 16:3675-3684, Angell and Baulcombe, (1999) *Plant J.* 20:357-362, and
5 U.S. Patent No. 6,635,805.

vi. *Ribozymes*

In some embodiments, the polynucleotide expressed by the expression cassette of the invention is catalytic RNA or has ribozyme activity specific for the messenger RNA of the *RCA* α -isoform polypeptide. Thus, the polynucleotide causes the degradation of the endogenous messenger RNA,
10 resulting in reduced expression of the *RCA* α -isoform polypeptide. This method is described, for example, in U.S. Patent No. 4,987,071.

vii. *Small Interfering RNA or Micro RNA*

In some embodiments of the invention, inhibition of the expression of *RCA* α -isoform polypeptide may be obtained by RNA interference by expression of a gene encoding a micro RNA (miRNA). miRNAs are regulatory agents consisting of about 22 ribonucleotides. miRNA are highly
15 efficient at inhibiting the expression of endogenous genes. See, for example Javier, *et al.*, (2003) *Nature* 425:257-263.

For miRNA interference, the expression cassette is designed to express an RNA molecule that is modeled on an endogenous miRNA gene. The miRNA gene encodes an RNA that forms a hairpin
20 structure containing a 22-nucleotide sequence that is complementary to another endogenous gene (target sequence). For suppression of *RCA* α -isoform expression, the 22-nucleotide sequence is selected from a *RCA* α -isoform transcript sequence and contains 22 nucleotides of said *RCA* α -isoform sequence in sense orientation and 21 nucleotides of a corresponding antisense sequence that is complementary to the sense sequence. miRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and
25 the RNA interference they induce is inherited by subsequent generations of plants. In one embodiment, the miRNA would target the sequences of the exon that encodes the carboxy terminal redox-sensitive cysteines for any plant's *RCA* α -isoform.

Polypeptide-Based Inhibition of Gene Expression

In one embodiment, the polynucleotide encodes a zinc finger protein that binds to a gene
30 encoding a *RCA* α -isoform polypeptide, resulting in reduced expression of the gene. In particular embodiments, the zinc finger protein binds to a regulatory region of a *RCA* α -isoform gene. In other embodiments, the zinc finger protein binds to a messenger RNA encoding a *RCA* α -isoform polypeptide and prevents its translation. Methods of selecting sites for targeting by zinc finger proteins have been described, for example, in U.S. Patent No. 6,453,242, and methods for using zinc finger proteins to inhibit

the expression of genes in plants are described, for example, in U.S. Patent Publication No. 2003/0037355.

In another embodiment, TALE and dCas9 or dCfp1 could also be used, as well as any DNA binding domain that could bind the promoter. If a repressor protein, such as KRAB or 3xSRDX, is fused
5 to any of these DNA binding domains, then repression is enhanced.

Polypeptide-Based Inhibition of Protein Activity

In some embodiments of the invention, the polynucleotide encodes an antibody that binds to at least one RCA α -isoform polypeptide, and reduces the activity of the RCA α -isoform polypeptide. In another embodiment, the binding of the antibody results in increased turnover of the antibody-RCA α -
10 isoform complex by cellular quality control mechanisms. The expression of antibodies in plant cells and the inhibition of molecular pathways by expression and binding of antibodies to proteins in plant cells are well known in the art. See, for example, Conrad and Sonnewald, (2003) *Nature Biotech.* 21:35-36.

Gene Disruption

In some embodiments of the present invention, the activity of a RCA α -isoform polypeptide may
15 be reduced or eliminated by disrupting the gene encoding the RCA α -isoform polypeptide. The gene sequence encoding the RCA α -isoform polypeptide may be disrupted by any method known in the art. For example, in one embodiment, the gene is disrupted by transposon tagging. In another embodiment, the gene is disrupted by mutagenizing plants using random or targeted mutagenesis, and selecting for plants that have desired traits.

Transposon Tagging

In one embodiment of the invention, transposon tagging is used to reduce or eliminate the RCA α -isoform activity of one or more RCA α -isoform polypeptides. Transposon tagging comprises inserting a transposon within an endogenous RCA α -isoform gene to reduce or eliminate expression of the RCA α -
25 isoform polypeptide. "RuBisCO activase alpha-isoform gene" and RCA α -isoform gene" is intended to mean the gene that encodes a RCA α -isoform polypeptide.

In this embodiment, the expression of one or more RCA α -isoform polypeptides is reduced or eliminated by inserting a transposon within a regulatory region or coding region of the gene encoding the RCA α -isoform polypeptide. A transposon that is within an exon, intron, 5' or 3' untranslated sequence, a promoter, or any other regulatory sequence of a RCA α -isoform gene may be used to reduce or eliminate
30 the expression and/or activity of the encoded RCA α -isoform polypeptide. In one embodiment, the transposon would target exon 7b of the *RCA* gene in rice. For other plants, the transposon would target the exon that encodes the carboxy terminal redox-sensitive cysteines for that plant's RCA α -isoform. In another embodiment, one would have an active or activated DNA transposon in the area of rice's *RCA* gene's exon 7b. The DNA transposon can then move into exon 7b.

Methods for the transposon tagging of specific genes in plants are well known in the art. See, for example, Maes, *et al.*, (1999) *Trends Plant Sci.* 4:90-96; Dharmapuri and Sonti, (1999) *FEMS Microbiol. Lett.* 179:53-59; Meissner, *et al.*, (2000) *Plant J.* 22:265-274; Phogat, *et al.*, (2000) *J. Biosci.* 25:57-63; Walbot, (2000) *Curr. Opin. Plant Biol.* 2:103-107; Gai, *et al.*, (2000) *Nucleic Acids Res.* 28:94-96; 5 Fitzmaurice, *et al.*, (1999) *Genetics* 153:1919-1928). In addition, the TUSC process for selecting Mu insertions in selected genes has been described in Bensen, *et al.*, (1995) *Plant Cell* 7:75-84; Mena, *et al.*, (1996) *Science* 274:1537-1540; and U.S. Patent No. 5,962,764.

Genome Editing and Induced Mutagenesis

In an embodiment, one or more or all of the genetic modifications is a mutation of an endogenous 10 gene which partially or completely inactivates the gene, such as a point mutation, an insertion, or a deletion (or a combination of one or more thereof). Such a genetic modification is usually a mutation introduced to the genome. The point mutation may be a premature stop codon, a splice site mutation, a frame shift mutation or an amino acid substitution mutation that reduces activity of the gene or the encoded polypeptide. The deletion may be of one or more nucleotides within a transcribed exon or 15 promoter of the gene, or extend across or into more than one exon, or extend to deletion of the entire gene. In some embodiments, the deletion is introduced by use of ZF, TALEN or CRISPR technologies. In an alternate embodiment, one or more or all of the genetic modifications is an exogenous polynucleotide encoding an RNA molecule which inhibits expression of the endogenous gene, wherein the exogenous polynucleotide is operably linked to a promoter which is capable of directing expression of the 20 polynucleotide in the plant, or part thereof.

Genome Editing Using Site-Specific Nucleases

Genome editing uses engineered nucleases such as RNA guided DNA endonucleases or nucleases composed of sequence specific DNA binding domains fused to a non-specific DNA cleavage module. These engineered nucleases enable efficient and precise genetic modifications by inducing targeted DNA 25 double stranded breaks that stimulate the cell's endogenous cellular DNA repair mechanisms to repair the induced break. Such mechanisms include, for example, error prone non-homologous end joining (NHEJ) and homology directed repair (HDR).

In the presence of donor plasmid with extended homology arms, HDR can lead to the introduction of single or multiple transgenes to correct or replace existing genes. In the absence of donor 30 plasmid, NHEJ-mediated repair yields small insertion or deletion mutations of the target that cause gene disruption.

Engineered nucleases useful in the methods of the present invention include zinc finger nucleases (ZFNs), transcription activator-like (TAL) effector nucleases (TALEN) and CRISPR/Cas9 type nucleases.

Typically nuclease encoded genes are delivered into cells by plasmid DNA, viral vectors or in vitro transcribed mRNA.

A zinc finger nuclease (ZFN) comprises a DNA-binding domain and a DNA-cleavage domain, wherein the DNA binding domain is comprised of at least one zinc finger and is operatively linked to a DNA-cleavage domain. The zinc finger DNA-binding domain is at the N-terminus of the protein and the DNA-cleavage domain is located at the C-terminus of said protein.

A ZFN must have at least one zinc finger. In one embodiment, a ZFN would have at least three zinc fingers in order to have sufficient specificity to be useful for targeted genetic recombination in a host cell or organism. Typically, a ZFN having more than three zinc fingers would have progressively greater specificity with each additional zinc finger.

The zinc finger domain can be derived from any class or type of zinc finger. In a particular embodiment, the zinc finger domain comprises the Cis2His2 type of zinc finger that is very generally represented, for example, by the zinc finger transcription factors TFIIIA or Sp1. In one embodiment, the zinc finger domain comprises three Cis2His2 type zinc fingers. The DNA recognition and/or the binding specificity of a ZFN can be altered in order to accomplish targeted genetic recombination at any chosen site in cellular DNA. Such modification can be accomplished using known molecular biology and/or chemical synthesis techniques. (See, for example, Bibikova et al., 2002).

The ZFN DNA-cleavage domain is derived from a class of non-specific DNA cleavage domains, for example the DNA-cleavage domain of a Type II restriction enzyme such as FokI (Kim et al., 1996). Other useful endonucleases may include, for example, HhaI, HindIII, Nod, BbvCI, EcoRI, BglI, and AlwI.

A transcription activator-like (TAL) effector nuclease (TALEN) comprises a TAL effector DNA binding domain and an endonuclease domain.

TAL effectors are proteins of plant pathogenic bacteria that are injected by the pathogen into the plant cell, where they travel to the nucleus and function as transcription factors to turn on specific plant genes. The primary amino acid sequence of a TAL effector dictates the nucleotide sequence to which it binds. Thus, target sites can be predicted for TAL effectors, and TAL effectors can be engineered and generated for the purpose of binding to particular nucleotide sequences.

Fused to the TAL effector-encoding nucleic acid sequences are sequences encoding a nuclease or a portion of a nuclease, typically a nonspecific cleavage domain from a type II restriction endonuclease such as FokI (Kim et al., 1996). Other useful endonucleases may include, for example, HhaI, HindIII, Nod, BbvCI, EcoRI, BglI, and AhaI. The fact that some endonucleases (e.g., FokI) only function as dimers can be capitalized upon to enhance the target specificity of the TAL effector. For example, in some cases each FokI monomer can be fused to a TAL effector sequence that recognizes a different DNA target

sequence, and only when the two recognition sites are in close proximity do the inactive monomers come together to create a functional enzyme. By requiring DNA binding to activate the nuclease, a highly site-specific restriction enzyme can be created.

A sequence-specific TALEN can recognize a particular sequence within a preselected target nucleotide sequence present in a cell. Thus, in some embodiments, a target nucleotide sequence can be scanned for nuclease recognition sites, and a particular nuclease can be selected based on the target sequence. In other cases, a TALEN can be engineered to target a particular cellular sequence.

Genome Editing Using Programmable RNA-Guided DNA Endonucleases

Distinct from the site-specific nucleases described above, the clustered regulatory interspaced short palindromic repeats (CRISPR)/Cas system provides an alternative to ZFNs and TALENs for inducing targeted genetic alterations, via RNA-guided DNA cleavage.

CRISPR systems rely on CRISPR RNA (crRNA) and transactivating chimeric RNA (tracrRNA) for sequence-specific cleavage of DNA. Three types of CRISPR/Cas systems exist: in type II systems, Cas9 serves as an RNA-guided DNA endonuclease that cleaves DNA upon crRNA-tracrRNA target recognition. CRISPR RNA base pairs with tracrRNA to form a two-RNA structure that guides the Cas9 endonuclease to complementary DNA sites for cleavage.

The CRISPR system can be portable to plant cells by co-delivery of plasmids expressing the Cas endonuclease and the necessary crRNA components. The Cas endonuclease may be converted into a nickase to provide additional control over the mechanism of DNA repair (Cong, *et al.*, *Science* 339(6121):819-23 2013).

CRISPRs are typically short partially palindromic sequences of 24-40 bp containing inner and terminal inverted repeats of up to 11 bp. Although isolated elements have been detected, they are generally arranged in clusters (up to about 20 or more per genome) of repeated units spaced by unique intervening 20-58 bp sequences. CRISPRs are generally homogenous within a given genome with most of them being identical. However, there are examples of heterogeneity in, for example, the Archaea (Mojica, *et al.*, *Mol. Microbiol.* 36(1):244-6 2000).

As described in the examples below, several indels of the RCA α -isoform gene have been created. These genomic mutations have been shown to reduce accumulation of the alpha-isoform and hence concomitant low light inhibition of RuBisCO activase. The indels are made in the area surrounding and including exon 7, which encodes the carboxy-end extension of the alpha spliced version of the RCA protein in rice.

As described herein, several indel RCA sequences in plants using CRISPRs genome editing have been made that modulate the activity of RCA α -isoform. These genomic changes in and around exon 7 include indels that result in the production of an altered RCA α -isoform protein

having an amino acid sequence at the carboxy end of the protein being one or more of the sequences in SEQ ID NOs: 4, 6, 8, 10, 12, 15, 17, 19, 22, 24, 27, and/or 29. Thus, the invention includes polynucleotide sequences which encode part or all of altered RCA α indel proteins having one or more of these sequences at the carboxy terminus of the protein, including nucleic acid constructs, vectors and the like. In one embodiment, the polynucleotide sequences encode these specific amino acid sequences listed in the above mentioned SEQ ID NOs. The invention also includes modified RCA proteins and their variants. The invention also includes vectors, plant cells, tissues, plant lines, varieties, and hybrids which include the same as well as their use in breeding.

Plant lines 2-1, 8-1, 8-2, 8-3, 8-4, 8-5, 8-6, 8-7, 20-1, 20-2, 20-3, 1, 6-1, 6-2, 6-3, 6-4, 8-10, 9-1, 9-2, 9-3, 13-1, 13-2, 16, 17, 21-1, and/or 21-2 and their progeny which contain the altered RCA α -isoform gene containing an altered exon 7, which encodes an altered RCA α -isoform protein lacking two cysteine amino acids at the protein's carboxy terminus are also included herewith.

TILLING

Yet another embodiment involves the use of TILLING. "TILLING" or "Targeting Induced Local Lesions IN Genomics" refers to a mutagenesis technology useful to generate and/or identify and to eventually isolate mutagenised variants of a particular nucleic acid with modulated expression and/or activity (McCallum, et al., (2000), Plant Physiology 123:439-442; McCallum, et al., (2000) Nature Biotechnology 18:455-457 and Colbert, et al., (2001) Plant Physiology 126:480-484).

TILLING combines high density point mutations with rapid sensitive detection of the mutations. Typically, ethylmethanesulfonate (EMS) is used to mutagenize plant seed. EMS alkylates guanine, which typically leads to mispairing. For example, seeds are soaked in an about 10-20 mM solution of EMS for about 10 to 20 hours; the seeds are washed and then sown. The plants of this generation are known as M1. M1 plants are then self-fertilized. Mutations that are present in cells that form the reproductive tissues are inherited by the next generation (M2). Typically, M2 plants are screened for mutation in the desired gene and/or for specific phenotypes.

TILLING also allows selection of plants carrying mutant variants. These mutant variants may exhibit modified expression, either in strength or in location or in timing (if the mutations affect the promoter, for example). These mutant variants may exhibit higher or lower activity than that exhibited by the gene in its natural form. TILLING combines high-density mutagenesis with high-throughput screening methods. The steps typically followed in TILLING are: (a) EMS mutagenesis (Redei and Koncz, (1992) In Methods in Arabidopsis Research, Koncz, et al., eds. Singapore, World Scientific Publishing Co, pp. 16-82; Feldmann, et al., (1994) In Arabidopsis. Meyerowitz and Somerville, eds, Cold

Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp 137-172; Lightner and Caspar, (1998) In Methods on Molecular Biology 82:91-104; Martinez-Zapater and Salinas, eds, Humana Press, Totowa, N.J.); (b) DNA preparation and pooling of individuals; (c) PCR amplification of a region of interest; (d) denaturation and annealing to allow formation of heteroduplexes; (e) DHPLC, where the presence of a heteroduplex in a pool is detected as an extra peak in the chromatogram; (f) identification of the mutant individual; and (g) sequencing of the mutant PCR product. Methods for TILLING are well known in the art (U.S. Pat. No. 8,071,840).

Other mutagenic methods can also be employed to introduce mutations in a disclosed gene. Methods for introducing genetic mutations into plant genes and selecting plants with desired traits are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as X-rays or gamma rays can be used.

Mutant Plants with Reduced Activity

Additional methods for decreasing or eliminating the expression of endogenous genes in plants are also known in the art and can be similarly applied to the instant invention. These methods include other forms of mutagenesis, such as ethyl methanesulfonate-induced mutagenesis, deletion mutagenesis, and fast neutron deletion mutagenesis used in a reverse genetics sense (with PCR) to identify plant lines in which the endogenous gene has been deleted. For examples of these methods see, Ohshima, *et al.*, (1998) *Virology* 243:472-481; Okubara, *et al.*, (1994) *Genetics* 137:867-874; and Quesada, *et al.*, (2000) *Genetics* 154:421-436.

Mutations that impact gene expression or that interfere with the function of the encoded protein are well known in the art. Insertional mutations in gene exons usually result in null-mutants. Mutations in conserved residues are particularly effective in inhibiting the activity of the encoded protein. Conserved residues of plant RCA α -isoform polypeptides suitable for mutagenesis with the goal to eliminate RCA α -isoform activity have been described. Such mutants can be isolated according to well-known procedures, and mutations in different RCA α -isoform loci can be stacked by genetic crossing. See, for example, Gruis, *et al.*, (2002) *Plant Cell* 14:2863-2882.

In another embodiment of this invention, dominant mutants can be used to trigger RNA silencing due to gene inversion and recombination of a duplicated gene locus. See, for example, Kusaba, *et al.*, (2003) *Plant Cell* 15:1455-1467.

The invention encompasses additional methods for reducing or eliminating the activity of one or more RCA α -isoform polypeptides. Examples of other methods for altering or mutating a genomic nucleotide sequence in a plant are known in the art and include, but are not limited to, the use of

RNA:DNA vectors, RNA:DNA mutational vectors, RNA:DNA repair vectors, mixed-duplex oligonucleotides, self-complementary RNA:DNA oligonucleotides, and recombinogenic oligonucleobases. Such vectors and methods of use are known in the art. See, for example, U.S. Patent Nos. 5,565,350; 5,731,181; 5,756,325; 5,760,012; 5,795,972; and 5,871,984. See also, WO 98/49350, 5 WO 99/07865, WO 99/25821, and Beetham, *et al.*, (1999) *Proc. Natl. Acad. Sci. USA* 96:8774-8778.

Plant Transformation Methods

Numerous methods for introducing foreign genes into plants are known and can be used to insert a RCA α -isoform polynucleotide into a plant host, including biological and physical plant transformation protocols. See, e.g., Miki *et al.*, "Procedure for Introducing Foreign DNA into Plants," in *Methods in Plant Molecular Biology and Biotechnology*, Glick and Thompson, eds., CRC Press, Inc., Boca Raton, 10 pp. 67-88 (1993). The methods chosen vary with the host plant, and include chemical transfection methods such as calcium phosphate, microorganism-mediated gene transfer such as *Agrobacterium* (Horsch *et al.*, *Science* 227:1229-31 (1985)), electroporation, micro-injection, and biolistic bombardment.

Expression cassettes and vectors and *in vitro* culture methods for plant cell or tissue 15 transformation and regeneration of plants are known and available. See, e.g., Gruber *et al.*, "Vectors for Plant Transformation," in *Methods in Plant Molecular Biology and Biotechnology*, *supra*, pp. 89-119.

The isolated polynucleotides or polypeptides may be introduced into the plant by one or more techniques typically used for direct delivery into cells. Such protocols may vary depending on the type of organism, cell, plant or plant cell, i.e. monocot or dicot, targeted for gene modification. Suitable methods 20 of transforming plant cells include microinjection (Crossway, *et al.*, (1986) *Biotechniques* 4:320-334; and U.S. Patent 6,300,543), electroporation (Riggs, *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, direct gene transfer (Paszkowski *et al.*, (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, Sanford, *et al.*, U.S. Patent No. 4,945,050; WO 91/10725; and McCabe, *et al.*, (1988) *Biotechnology* 6:923-926). Also see, Tomes, *et al.*, "Direct DNA Transfer into Intact Plant Cells Via 25 Microprojectile Bombardment". pp. 197-213 in *Plant Cell, Tissue and Organ Culture, Fundamental Methods*, eds. O. L. Gamborg & G.C. Phillips. Springer-Verlag Berlin Heidelberg New York, 1995; U.S. Patent 5,736,369 (meristem); Weissinger, *et al.*, (1988) *Ann. Rev. Genet.* 22:421-477; Sanford, *et al.*, (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou, *et al.*, (1988) *Plant Physiol.* 87:671-674 (soybean); Datta, *et al.*, (1990) *Biotechnology* 8:736-740 (rice); Klein, *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein, *et al.*, (1988) *Biotechnology* 6:559-563 (maize); 30 WO 91/10725 (maize); Klein, *et al.*, (1988) *Plant Physiol.* 91:440-444 (maize); Fromm, *et al.*, (1990) *Biotechnology* 8:833-839; and Gordon-Kamm, *et al.*, (1990) *Plant Cell* 2:603-618 (maize); Hooydaas-Van Slogteren & Hooykaas (1984) *Nature* (London) 311:763-764; Bytebierm, *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet, *et al.*, (1985) *In The Experimental Manipulation of*

Ovule Tissues, ed. G.P. Chapman, *et al.*, pp. 197-209. Longman, NY (pollen); Kaeppler, *et al.*, (1990) *Plant Cell Reports* 9:415-418; and Kaeppler, *et al.*, (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); U.S. Patent No. 5,693,512 (sonication); D'Halluin, *et al.*, (1992) *Plant Cell* 4:1495-1505 (electroporation); Li, *et al.*, (1993) *Plant Cell Reports* 12:250-255; and Christou and Ford, (1995) *Annals of Botany* 75:407-413 (rice); Osjoda, *et al.*, (1996) *Nature Biotech.* 14:745-750; Agrobacterium mediated maize transformation (U.S. Patent 5,981,840); silicon carbide whisker methods (Frame, *et al.*, (1994) *Plant J.* 6:941-948); laser methods (Guo, *et al.*, (1995) *Physiologia Plantarum* 93:19-24); sonication methods (Bao, *et al.*, (1997) *Ultrasound in Medicine & Biology* 23:953-959; Finer and Finer, (2000) *Lett Appl Microbiol.* 30:406-10; Amoah, *et al.*, (2001) *J Exp Bot* 52:1135-42); polyethylene glycol methods (Krens, *et al.*, (1982) *Nature* 296:72-77); protoplasts of monocot and dicot cells can be transformed using electroporation (Fromm, *et al.*, (1985) *Proc. Natl. Acad. Sci. USA* 82:5824-5828) and microinjection (Crossway, *et al.*, (1986) *Mol. Gen. Genet.* 202:179-185).

Agrobacterium-mediated Transformation

The most widely utilized method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium*. *A. tumefaciens* and *A. rhizogenes* are plant pathogenic soil bacteria, which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of plants. See, e.g., Kado, (1991) *Crit. Rev. Plant Sci.* 10:1. Descriptions of the *Agrobacterium* vector systems and methods for *Agrobacterium*-mediated gene transfer are provided in Gruber, *et al.*, *supra*; Miki, *et al.*, *supra*; and Moloney, *et al.*, (1989) *Plant Cell Reports* 8:238.

Similarly, the gene can be inserted into the T-DNA region of a Ti or Ri plasmid derived from *A. tumefaciens* or *A. rhizogenes*, respectively. Thus, expression cassettes can be constructed as above, using these plasmids. Many control sequences are known which when coupled to a heterologous coding sequence and transformed into a host organism show fidelity in gene expression with respect to tissue/organ specificity of the original coding sequence. See, e.g., Benfey and Chua, (1989) *Science* 244:174-81. Particularly suitable control sequences for use in these plasmids are promoters for constitutive leaf-specific expression of the gene in the various target plants. Other useful control sequences include a promoter and terminator from the nopaline synthase gene (NOS). The NOS promoter and terminator are present in the plasmid pARC2, available from the American Type Culture Collection and designated ATCC 67238. If such a system is used, the virulence (*vir*) gene from either the Ti or Ri plasmid must also be present, either along with the T-DNA portion, or via a binary system where the *vir* gene is present on a separate vector. Such systems, vectors for use therein, and methods of transforming plant cells are described in U.S. Patent No. 4,658,082; U.S. Patent Application No. 913,914, filed Oct. 1,

1986, as referenced in U.S. Patent No. 5,262,306; and Simpson, *et al.*, (1986) *Plant Mol. Biol.* 6:403-15 (also referenced in the '306 patent).

Once constructed, these plasmids can be placed into *A. rhizogenes* or *A. tumefaciens* and these vectors used to transform cells of plant species, which are ordinarily susceptible to *Fusarium* or *Alternaria* infection. Several other transgenic plants are also contemplated by the present invention including but not limited to soybean, corn, sorghum, alfalfa, rice, clover, cabbage, banana, coffee, celery, tobacco, cowpea, cotton, melon and pepper. The selection of either *A. tumefaciens* or *A. rhizogenes* will depend on the plant being transformed thereby. In general *A. tumefaciens* is an useful organism for transformation. Most dicotyledonous plants, some gymnosperms, and a few monocotyledonous plants (e.g., certain members of the *Liliales* and *Arales*) are susceptible to infection with *A. tumefaciens*. *A. rhizogenes* also has a wide host range, embracing most dicots and some gymnosperms, which includes members of the *Leguminosae*, *Compositae*, and *Chenopodiaceae*. Monocot plants can now be transformed with some success. European Patent Application No. 604 662 A1 discloses a method for transforming monocots using *Agrobacterium*. European Application No. 672 752 A1 discloses a method for transforming monocots with *Agrobacterium* using the scutellum of immature embryos. Ishida, *et al.*, discuss a method for transforming maize by exposing immature embryos to *A. tumefaciens* (*Nature Biotechnology* 14:745-50 (1996)).

Once transformed, these cells can be used to regenerate transgenic plants. For example, whole plants can be infected with these vectors by wounding the plant and then introducing the vector into the wound site. Any part of the plant can be wounded, including leaves, stems and roots. Alternatively, plant tissue, in the form of an explant, such as cotyledonary tissue or leaf disks, can be inoculated with these vectors, and cultured under conditions, which promote plant regeneration. Roots or shoots transformed by inoculation of plant tissue with *A. rhizogenes* or *A. tumefaciens*, containing the gene coding for the fumonisin degradation enzyme, can be used as a source of plant tissue to regenerate fumonisin-resistant transgenic plants, either via somatic embryogenesis or organogenesis. Examples of such methods for regenerating plant tissue are disclosed in Shahin, (1985) *Theor. Appl. Genet.* 69:235-40; U.S. Patent No. 4,658,082; Simpson, *et al.*, *supra*; and U.S. Patent Application Nos. 913,913 and 913,914, both filed Oct. 1, 1986, as referenced in U.S. Patent No. 5,262,306.

Direct Gene Transfer

Despite the fact that the host range for *Agrobacterium*-mediated transformation is broad, some major cereal crop species and gymnosperms have generally been recalcitrant to this mode of gene transfer, even though some success has recently been achieved in rice (Hiei, *et al.*, (1994) *The Plant Journal* 6:271-82). Several methods of plant transformation, collectively referred to as direct gene transfer, have been developed as an alternative to *Agrobacterium*-mediated transformation.

A generally applicable method of plant transformation is microprojectile-mediated transformation, where DNA is carried on the surface of microprojectiles measuring about 1 to 4 μm . The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds of 300 to 600 m/s which is sufficient to penetrate the plant cell walls and membranes (Sanford, *et al.*, (1987) *Part. Sci. Technol.* 5:27; Sanford, (1988) *Trends Biotech* 6:299; Sanford, (1990) *Physiol. Plant* 79:206; and Klein, *et al.*, (1992) *Biotechnology* 10:268).

Another method for physical delivery of DNA to plants is sonication of target cells as described in Zang, *et al.*, (1991) *BioTechnology* 9:996. Alternatively, liposome or spheroplast fusions have been used to introduce expression vectors into plants. See, e.g., Deshayes, *et al.*, (1985) *EMBO J.* 4:2731; and Christou, *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:3962. Direct uptake of DNA into protoplasts using CaCl_2 precipitation, polyvinyl alcohol, or poly-L-ornithine has also been reported. See, e.g., Hain, *et al.*, (1985) *Mol. Gen. Genet.* 199:161; and Draper, *et al.*, (1982) *Plant Cell Physiol.* 23:451.

Electroporation of protoplasts and whole cells and tissues has also been described. See, e.g., Donn, *et al.*, (1990) *Abstracts of the VIIth Int'l. Congress on Plant Cell and Tissue Culture IAPTC*, A2-38, p. 53; D'Halluin, *et al.*, (1992) *Plant Cell* 4:1495-505; and Spencer, *et al.*, (1994) *Plant Mol. Biol.* 24:51-61.

Genes of interest are reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest change, and as developing nations open up world markets, new crops and technologies will emerge also. In addition, as our understanding of agronomic traits and characteristics such as yield and heterosis increase, the choice of genes for transformation will change accordingly. General categories of genes of interest include, for example, those genes involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. More specific categories of transgenes, for example, include genes encoding important traits for agronomics, insect resistance, disease resistance, herbicide resistance, sterility, grain characteristics, and commercial products. Genes of interest include, generally, those involved in oil, starch, carbohydrate, or nutrient metabolism as well as those affecting kernel size, sucrose loading, and the like.

In certain embodiments the nucleic acid sequences of the present invention can be used in combination ("stacked") with other polynucleotide sequences of interest in order to create plants with a desired phenotype. The combinations generated can include multiple copies of any one or more of the polynucleotides of interest. The polynucleotides of the present invention may be stacked with any gene or combination of genes to produce plants with a variety of desired trait combinations, including but not limited to traits desirable for animal feed such as high oil genes (e.g., U.S. Patent No. 6,232,529); balanced amino acids (e.g., hordothionins (U.S. Patent Nos. 5,990,389; 5,885,801; 5,885,802; and

5,703,049); barley high lysine (Williamson, *et al.*, (1987) *Eur. J. Biochem.* 165:99-106; and WO 98/20122); and high methionine proteins (Pedersen, *et al.*, (1986) *J. Biol. Chem.* 261:6279; Kirihara, *et al.*, (1988) *Gene* 71:359; and Musumura, *et al.*, (1989) *Plant Mol. Biol.* 12:123)); increased digestibility (e.g., modified storage proteins (U.S. Patent Application 10/053,410); and thioredoxins (U.S. Patent Application 10/005,429)). The polynucleotides of the present invention can also be stacked with traits desirable for insect, disease or herbicide resistance (e.g., *Bacillus thuringiensis* toxic proteins (U.S. Patent Nos. 5,366,892; 5,747,450; 5,736,514; 5,723,756; 5,593,881; Geiser, *et al.*, (1986) *Gene* 48:109); lectins (Van Damme, *et al.*, (1994) *Plant Mol. Biol.* 24:825); fumonisin detoxification genes (U.S. Patent No. 5,792,931); avirulence and disease resistance genes (Jones, *et al.*, (1994) *Science* 266:789; Martin, *et al.*, (1993) *Science* 262:1432; Mindrinos, *et al.*, (1994) *Cell* 78:1089); acetolactate synthase (ALS) mutants that lead to herbicide resistance such as the S4 and/or Hra mutations; inhibitors of glutamine synthase such as phosphinothricin or basta (e.g., bar gene); and glyphosate resistance (EPSPS gene)); and traits desirable for processing or process products such as high oil (e.g., U.S. Patent No. 6,232,529); modified oils (e.g., fatty acid desaturase genes (U.S. Patent No. 5,952,544; WO 94/11516)); modified starches (e.g., ADPG pyrophosphorylases (AGPase), starch synthases (SS), starch branching enzymes (SBE) and starch debranching enzymes (SDBE)); and polymers or bioplastics (e.g., U.S. Patent No. 5,602,321; β -ketothiolase, polyhydroxybutyrate synthase, and acetoacetyl-CoA reductase (Schubert, *et al.*, (1988) *J. Bacteriol.* 170:5837-5847) facilitate expression of polyhydroxyalkanoates (PHAs)). One could also combine the polynucleotides of the present invention with polynucleotides affecting agronomic traits such as male sterility (e.g., see U.S. Patent No. 5,583,210), stalk strength, flowering time, or transformation technology traits such as cell cycle regulation or gene targeting (e.g., WO 99/61619; WO 00/17364; WO 99/25821).

In one embodiment, sequences of interest improve plant growth and/or crop yields. For example, sequences of interest include agronomically important genes that result in improved primary or lateral root systems. Such genes include, but are not limited to, nutrient/water transporters and growth inducers. Examples of such genes, include but are not limited to, maize plasma membrane H⁺-ATPase (MHA2) (Frias, *et al.*, (1996) *Plant Cell* 8:1533-44); AKT1, a component of the potassium uptake apparatus in *Arabidopsis*, (Spalding, *et al.*, (1999) *J Gen Physiol* 113:909-18); RML genes which activate cell division cycle in the root apical cells (Cheng, *et al.*, (1995) *Plant Physiol* 108:881); maize glutamine synthetase genes (Sukanya, *et al.*, (1994) *Plant Mol Biol* 26:1935-46) and hemoglobin (Duff, *et al.*, (1997) *J. Biol. Chem* 27:16749-16752, Arredondo-Peter, *et al.*, (1997) *Plant Physiol.* 115:1259-1266; Arredondo-Peter, *et al.*, (1997) *Plant Physiol* 114:493-500 and references cited therein). The sequence of interest may also be useful in expressing antisense nucleotide sequences of genes that that negatively affects root development.

Methods of Use for RCA α -isoform polynucleotide, expression cassettes, and additional polynucleotides

The nucleotides, expression cassettes and methods disclosed herein are useful in modulating RCA activity and thus RuBisCO activity. The increase in RuBisCO activity alters the plant's phenotype, as discussed above and in the examples below.

5 *Nucleic Acids*

The present invention provides, *inter alia*, isolated nucleic acids of RNA, DNA, homologs, paralogs and orthologs and/or chimeras thereof, comprising a RCA α -isoform polynucleotide. This includes naturally occurring as well as synthetic variants and homologs of the sequences.

Sequences homologous, *i.e.*, that share significant sequence identity or similarity, to those
 10 provided herein derived from maize, rice or from other plants of choice, are also an aspect of the invention. Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn (maize), potato, cotton, rape, oilseed rape (including canola), sunflower, alfalfa, clover, sugarcane, and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and
 15 raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, tomatillo, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, Brussels sprouts, and kohlrabi). Other crops, including fruits and vegetables, whose phenotype can be changed and which comprise homologous sequences
 20 include barley; rye; millet; sorghum; currant; avocado; citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries; nuts such as the walnut and peanut; endive; leek; roots such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato; and beans. The homologous sequences may also be derived from woody species, such pine, poplar and eucalyptus, or mint or other labiates. In addition, homologous sequences may be derived from plants that are evolutionarily-related to crop plants,
 25 but which may not have yet been used as crop plants. Examples include deadly nightshade (*Atropa belladonna*), related to tomato; jimson weed (*Datura stramonium*), related to peyote; and teosinte (*Zea* species), related to corn (maize). See FIG. 21 for the homologous carboxy terminal RCA α -isoform amino acid sequences for various economically important plants.

Orthologs and Paralogs

30 Homologous sequences as described above can comprise orthologous or paralogous sequences. Several different methods are known by those of skill in the art for identifying and defining these functionally homologous sequences. Three general methods for defining orthologs and paralogs are described; an ortholog, paralog or homolog may be identified by one or more of the methods described below.

Orthologs and paralogs are evolutionarily related genes that have similar sequence and similar functions. Orthologs are structurally related genes in different species that are derived by a speciation event. Paralogs are structurally related genes within a single species that are derived by a duplication event.

5 Within a single plant species, gene duplication may result in two copies of a particular gene, giving rise to two or more genes with similar sequence and often similar function known as paralogs. A paralog is therefore a similar gene formed by duplication within the same species. Paralogs typically cluster together or in the same clade (a group of similar genes) when a gene family phylogeny is analyzed using programs such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22: 4673-4680; Higgins
10 et al. (1996) *Methods Enzymol.* 266: 383-402). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle (1987) *J. Mol. Evol.* 25: 351-360).

Orthologous genes from different organisms have highly conserved functions, and very often essentially identical functions (Lee et al. (2002) *Genome Res.* 12: 493-502; Remm et al. (2001) *J. Mol. Biol.* 314: 1041-1052). Paralogous genes, which have diverged through gene duplication, may retain
15 similar functions of the encoded proteins. In such cases, paralogs can be used interchangeably with respect to certain embodiments of the instant invention (for example, transgenic expression of a coding sequence).

Variant Nucleotide Sequences in the non-coding regions

The RCA α -isoform nucleotide sequences are used to generate variant nucleotide sequences
20 having the nucleotide sequence of the 5'-untranslated region, 3'-untranslated region, or promoter region that is approximately 70%, 75%, and 80%, 85%, 90% and 95% identical to the original nucleotide sequence. These variants are then associated with natural variation in the germplasm for component traits. The associated variants are used as marker haplotypes to select for the desirable traits.

Variant Amino Acid Sequences of Polypeptides

25 Variant amino acid sequences of the RCA α -isoform polypeptides are generated. In this example, one amino acid is altered. Specifically, the open reading frames are reviewed to determine the appropriate amino acid alteration. The selection of the amino acid to change is made by consulting the protein alignment (with the other orthologs and other gene family members from various species). An amino acid is selected that is deemed not to be under high selection pressure (not highly conserved) and which is
30 rather easily substituted by an amino acid with similar chemical characteristics (i.e., similar functional side-chain). Using a protein alignment, an appropriate amino acid can be changed. Once the targeted amino acid is identified, the procedure outlined herein is followed. Variants having about 70%, 75%, 80%, 85%, 90% and 95% nucleic acid sequence identity are generated using this method. The associated variants are used as marker haplotypes to select for the desirable traits.

The present invention also includes polynucleotides optimized for expression in different organisms. For example, for expression of the polynucleotide in a particular plant, the sequence can be altered to account for specific codon.

Construction of Nucleic Acids

The isolated nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some embodiments, the polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a fungus or bacteria.

The nucleic acids may conveniently comprise sequences in addition to a polynucleotide of the present invention. For example, a multi-cloning site comprising one or more endonuclease restriction sites may be inserted into the nucleic acid to aid in isolation of the polynucleotide. Also, translatable sequences may be inserted to aid in the isolation of the translated polynucleotide of the present invention. For example, a hexa-histidine marker sequence provides a convenient means to purify the proteins of the present invention. The nucleic acid of the present invention - excluding the polynucleotide sequence - is optionally a vector, adapter, or linker for cloning and/or expression of a polynucleotide of the present invention. Additional sequences may be added to such cloning and/or expression sequences to optimize their function in cloning and/or expression, to aid in isolation of the polynucleotide, or to improve the introduction of the polynucleotide into a cell. Typically, the length of a nucleic acid of the present invention less the length of its polynucleotide of the present invention is less than 20 kilobase pairs, often less than 15 kb, and frequently less than 10 kb. Use of cloning vectors, expression vectors, adapters, and linkers is well known in the art. Exemplary nucleic acids include such vectors as: M13, lambda ZAP Express, lambda ZAP II, lambda gt10, lambda gt11, pBK-CMV, pBK-RSV, pBluescript II, lambda DASH II, lambda EMBL 3, lambda EMBL 4, pWE15, SuperCos 1, SurfZap, Uni-ZAP, pBC, pBS+/-, pSG5, pBK, pCR-Script, pET, pSPUTK, p3'SS, pGEM, pSK+/-, pGEX, pSPORTI and II, pOPRSVI CAT, pOPI3 CAT, pXT1, pSG5, pPbac, pMbac, pMC1neo, pOG44, pOG45, pFRT β GAL, pNEO β GAL, pRS403, pRS404, pRS405, pRS406, pRS413, pRS414, pRS415, pRS416, lambda MOSSlox, and lambda MOSElox. Optional vectors for the present invention, include but are not limited to, lambda ZAP II, and pGEX. For a description of various nucleic acids see, e.g., Stratagene Cloning Systems, Catalogs 1995, 1996, 1997 (La Jolla, CA); and, Amersham Life Sciences, Inc, Catalog '97 (Arlington Heights, IL).

Synthetic Methods for Constructing Nucleic Acids

The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang, *et al.*, (1979) *Meth. Enzymol.* 68:90-9; the phosphodiester method of Brown, *et al.*, (1979) *Meth. Enzymol.* 68:109-51; the diethylphosphoramidite method of Beaucage, *et al.*, (1981) *Tetra. Letts.* 22(20):1859-62; the solid phase phosphoramidite triester method described by Beaucage, *et al.*, *supra*, e.g., using an automated

synthesizer, e.g., as described in Needham-VanDevanter, *et al.*, (1984) *Nucleic Acids Res.* 12:6159-68; and, the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis generally produces a single stranded oligonucleotide. This may be converted into double stranded DNA by hybridization with a complementary sequence or by polymerization with a DNA polymerase using the single strand as a template. One of skill will recognize that while chemical synthesis of DNA is limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences.

UTRs and Codon Preference

In general, translational efficiency has been found to be regulated by specific sequence elements in the 5' non-coding or untranslated region (5' UTR) of the RNA. Positive sequence motifs include translational initiation consensus sequences (Kozak, (1987) *Nucleic Acids Res.* 15:8125) and the 5' <G> 7 methyl GpppG RNA cap structure (Drummond, *et al.*, (1985) *Nucleic Acids Res.* 13:7375). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing, *et al.*, (1987) *Cell* 48:691) and AUG sequences or short open reading frames preceded by an appropriate AUG in the 5' UTR (Kozak, *supra*, Rao, *et al.*, (1988) *Mol. and Cell. Biol.* 8:284). Accordingly, the present invention provides 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences.

Further, the polypeptide-encoding segments of the polynucleotides of the present invention can be modified to alter codon usage. Altered codon usage can be employed to alter translational efficiency and/or to optimize the coding sequence for expression in a desired host or to optimize the codon usage in a heterologous sequence for expression in rice. Codon usage in the coding regions of the polynucleotides of the present invention can be analyzed statistically using commercially available software packages such as "Codon Preference" available from the University of Wisconsin Genetics Computer Group. See, Devereaux, *et al.*, (1984) *Nucleic Acids Res.* 12:387-395); or MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.). Thus, the present invention provides a codon usage frequency characteristic of the coding region of at least one of the polynucleotides of the present invention. The number of polynucleotides (3 nucleotides per amino acid) that can be used to determine a codon usage frequency can be any integer from 3 to the number of polynucleotides of the present invention as provided herein. Optionally, the polynucleotides will be full-length sequences. An exemplary number of sequences for statistical analysis can be at least 1, 5, 10, 20, 50 or 100.

Sequence Shuffling

The present invention provides methods for sequence shuffling using polynucleotides of the present invention, and compositions resulting therefrom. Sequence shuffling is described in PCT Publication No. 96/19256. See also, Zhang, *et al.*, (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-9; and Zhao, *et al.*, (1998) *Nature Biotech* 16:258-61. Generally, sequence shuffling provides a means for generating libraries of polynucleotides having a desired characteristic, which can be selected or screened

for. Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides, which comprise sequence regions, which have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property. In some embodiments, the selected characteristic will be an altered K_m and/or K_{cat} over the wild-type protein as provided herein. In other embodiments, a protein or polynucleotide generated from sequence shuffling will have a ligand binding affinity greater than the non-shuffled wild-type polynucleotide. In yet other embodiments, a protein or polynucleotide generated from sequence shuffling will have an altered pH optimum as compared to the non-shuffled wild-type polynucleotide. The increase in such properties can be at least 110%, 120%, 130%, 140% or greater than 150% of the wild-type value.

Recombinant Expression Cassettes

The present disclosure further provides recombinant expression cassettes comprising a nucleic acid of the present disclosure, which is designed for reducing the activity of RCA α . A nucleic acid sequence coding for the desired polynucleotide of the present disclosure, for example a cDNA or a genomic sequence encoding a polypeptide long enough to code for an active protein of the present disclosure, can be used to construct a recombinant expression cassette which can be introduced into the desired host cell. A recombinant expression cassette will typically comprise a polynucleotide of the present disclosure operably linked to transcriptional initiation regulatory sequences which will direct the transcription of the polynucleotide in the intended host cell, such as tissues of a transformed plant.

For example, plant expression vectors may include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site and/or a polyadenylation signal.

Promoters, Terminators, Introns

A plant promoter fragment can be employed which will direct expression of a polynucleotide of the present disclosure in essentially all tissues of a regenerated plant. Such promoters are referred to

herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the 1'- or 2'-promoter derived from T-DNA of *Agrobacterium tumefaciens*, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Pat. No. 5,683,439), the Nos promoter, the RuBisCO promoter, the GRP1-8 promoter, the 35S promoter from cauliflower mosaic virus (CaMV), as described in Odell, et al., (1985) Nature 313:810-2; rice actin (McElroy, et al., (1990) Plant Cell 163-171); ubiquitin (Christensen, et al., (1992) Plant Mol. Biol. 12:619-632 and Christensen, et al., (1992) Plant Mol. Biol. 18:675-89); pEMU (Last, et al., (1991) Theor. Appl. Genet. 81:581-8); MAS (Velten, et al., (1984) EMBO J. 3:2723-30) and maize H3 histone (Lepetit, et al., (1992) Mol. Gen. Genet. 231:276-85 and Atanassova, et al., (1992) Plant Journal 2(3):291-300); ALS promoter, as described in WO 1996/30530 and other transcription initiation regions from various plant genes known to those of skill. For the present disclosure ubiquitin is one useful promoter for expression in monocot plants.

Alternatively, the plant promoter can direct expression of a polynucleotide of the present disclosure in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters may be "inducible" promoters. Environmental conditions that may affect transcription by inducible promoters include pathogen attack, anaerobic conditions or the presence of light. Examples of inducible promoters are the Adh1 promoter, which is inducible by hypoxia or cold stress, the Hsp70 promoter, which is inducible by heat stress and the PPDK promoter, which is inducible by light. Diurnal promoters that are active at different times during the circadian rhythm are also known (U.S. Patent Application Publication No. 2011/0167517).

Examples of promoters under developmental control include promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds or flowers (tissue-specific promoters). The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may become fully or partially constitutive in certain locations.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from a variety of plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or alternatively from any other eukaryotic gene. Examples of such regulatory elements include, but are not limited to, 3' termination and/or polyadenylation regions such as those of the *Agrobacterium tumefaciens* nopaline synthase (nos) gene (Bevan, et al., (1983) Nucleic Acids Res. 12:369-85); the potato proteinase inhibitor

II (PINII) gene (Keil, et al., (1986) Nucleic Acids Res. 14:5641-50 and An, et al., (1989) Plant Cell 1:115-22) and the CaMV 19S gene (Mogen, et al., (1990) Plant Cell 2:1261-72).

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol.

Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold (Buchman and Berg, (1988) Mol. Cell Biol. 8:4395-4405; Callis, et al., (1987) Genes Dev. 1:1183-200). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of maize introns Adh1-S intron 1, 2 and 6, the Bronze-1 intron are known in the art. See generally, The Maize Handbook, Chapter 116, Freeling and Walbot, eds., Springer, New York (1994).

Signal Peptide Sequences

Plant signal sequences, including, but not limited to, signal-peptide encoding DNA/RNA sequences which target proteins to the extracellular matrix of the plant cell (Dratewka-Kos, et al., (1989) J. Biol. Chem. 264:4896-900), such as the Nicotiana plumbaginifolia extension gene (DeLoose, et al., (1991) Gene 99:95-100); signal peptides which target proteins to the vacuole, such as the sweet potato sporamin gene (Matsuka, et al., (1991) Proc. Natl. Acad. Sci. USA 88:834) and the barley lectin gene (Wilkins, et al., (1990) Plant Cell, 2:301-13); signal peptides which cause proteins to be secreted, such as that of PR1b (Lind, et al., (1992) Plant Mol. Biol. 18:47-53) or the barley alpha amylase (BAA) (Rahmatullah, et al., (1989) Plant Mol. Biol. 12:119) or signal peptides which target proteins to the plastids such as that of rapeseed enoyl-Acp reductase (Verwaert, et al., (1994) Plant Mol. Biol. 26:189-202) are useful in the disclosure.

Markers

The vector comprising the sequences from a polynucleotide of the present disclosure will typically comprise a marker gene, which confers a selectable phenotype on plant cells. The selectable marker gene may encode antibiotic resistance, with suitable genes including genes coding for resistance to the antibiotic spectinomycin (e.g., the aada gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance. Also useful are genes coding for resistance to herbicides which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as

phosphinothricin or basta (e.g., the bar gene), or other such genes known in the art. The bar gene encodes resistance to the herbicide basta and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Constructs described herein may comprise a polynucleotide of interest encoding a reporter or marker product. Examples of suitable reporter polynucleotides known in the art can be found in, for example, Jefferson, et al., (1991) in *Plant Molecular Biology Manual*, ed. Gelvin, et al., (Kluwer Academic Publishers), pp. 1-33; DeWet, et al. (1987) *Mol. Cell. Biol.* 7:725-737; Goff, et al., (1990) *EMBO J.* 9:2517-2522; Kain, et al., (1995) *Bio Techniques* 19:650-655 and Chiu, et al., (1996) *Current Biology* 6:325-330. In certain embodiments, the polynucleotide of interest encodes a selectable reporter. These can include polynucleotides that confer antibiotic resistance or resistance to herbicides. Examples of suitable selectable marker polynucleotides include, but are not limited to, genes encoding resistance to chloramphenicol, methotrexate, hygromycin, streptomycin, spectinomycin, bleomycin, sulfonamide, bromoxynil, glyphosate and phosphinothricin.

In some embodiments, the expression cassettes disclosed herein comprise a polynucleotide of interest encoding scorable or screenable markers, where presence of the polynucleotide produces a measurable product. Examples include a β -glucuronidase, or uidA gene (GUS), which encodes an enzyme for which various chromogenic substrates are known (for example, U.S. Pat. Nos. 5,268,463 and 5,599,670); chloramphenicol acetyl transferase and alkaline phosphatase. Other screenable markers include the anthocyanin/flavonoid polynucleotides including, for example, a R-locus polynucleotide, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues, the genes which control biosynthesis of flavonoid pigments, such as the maize C1 and C2, the B gene, the p1 gene and the bronze locus genes, among others. Further examples of suitable markers encoded by polynucleotides of interest include the cyan fluorescent protein (CYP) gene, the yellow fluorescent protein gene, a lux gene, which encodes a luciferase, the presence of which may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry, a green fluorescent protein (GFP) and DsRed2 (Clontech, 2001) where plant cells transformed with the marker gene are red in color, and thus visually selectable. Additional examples include a p-lactamase gene encoding an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin), a xyle gene encoding a catechol dioxygenase that can convert chromogenic catechols, an α -amylase gene and a tyrosinase gene encoding an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone, which in turn condenses to form the easily detectable compound melanin.

The expression cassette can also comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin

phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). Additional selectable markers include phenotypic markers such as β -galactosidase and fluorescent proteins such as green fluorescent protein (GFP) (Su, et al., (2004) Biotechnol Bioeng 85:610-9 and Fetter, et al., (2004) Plant Cell 16:215-28), cyan florescent protein (CYP) (Bolte, et al., (2004) J. Cell Science 117:943-54 and Kato, et al., (2002) Plant Physiol 129:913-42) and yellow florescent protein (PhiYFP.TM. from Evrogen, see, Bolte, et al., (2004) J. Cell Science 117:943-54). For additional selectable markers, see generally, Yarranton, (1992) Curr. Opin. Biotech. 3:506-511; Christopherson, et al., (1992) Proc. Natl. Acad. Sci. USA 89:6314-6318; Yao, et al., (1992) Cell 71:63-72; Reznikoff, (1992) Mol. Microbiol. 6:2419-2422; Barkley, et al., (1980) in The Operon, pp. 177-220; Hu, et al., (1987) Cell 48:555-566; Brown, et al., (1987) Cell 49:603-612; Figge, et al., (1988) Cell 52:713-722; Deuschle, et al., (1989) Proc. Natl. Acad. Aci. USA 86:5400-5404; Fuerst, et al., (1989) Proc. Natl. Acad. Sci. USA 86:2549-2553; Deuschle, et al., (1990) Science 248:480-483; Gossen, (1993) Ph.D. Thesis, University of Heidelberg; Reines, et al., (1993) Proc. Natl. Acad. Sci. USA 90:1917-1921; Labow, et al., (1990) Mol. Cell. Biol. 10:3343-3356; Zambretti, et al., (1992) Proc. Natl. Acad. Sci. USA 89:3952-3956; Bairn, et al., (1991) Proc. Natl. Acad. Sci. USA 88:5072-5076; Wyborski, et al., (1991) Nucleic Acids Res. 19:4647-4653; Hillenand-Wissman, (1989) Topics Mol. Struc. Biol. 10:143-162; Degenkolb, et al., (1991) Antimicrob. Agents Chemother. 35:1591-1595; Kleinschmidt, et al., (1988) Biochemistry 27:1094-1104; Bonin, (1993) Ph.D. Thesis, University of Heidelberg; Gossen, et al., (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Oliva, et al., (1992) Antimicrob. Agents Chemother. 36:913-919; Hlavka, et al., (1985) Handbook of Experimental Pharmacology, Vol. 78 (Springer-Verlag, Berlin); Gill, et al., (1988) Nature 334:721-724. The above list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be used in the compositions and methods disclosed herein.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers, et al., (1987) Meth. Enzymol. 153:253-77. These vectors are plant integrating vectors in that on transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl, et al., (1987) Gene 61:1-11 and Berger, et al., (1989) Proc. Natl. Acad. Sci. USA, 86:8402-6. Another useful vector herein is plasmid pBI101.2 that is available from CLONTECH Laboratories, Inc. (Palo Alto, Calif.).

Expression in Prokaryotes

Prokaryotic cells may be used as hosts for expression. Prokaryotes most frequently are represented by various strains of *E. coli*; however, other microbial strains may also be used. Commonly

used prokaryotic control sequences which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the β -lactamase (penicillinase) and lactose (lac) promoter systems (Chang, *et al.*, (1977) *Nature* 198:1056), the tryptophan (trp) promoter system (Goeddel, *et al.*, (1980) *Nucleic Acids Res.* 8:4057) and the lambda derived P L promoter and N-gene ribosome binding site (Shimatake, *et al.*, (1981) *Nature* 292:128). The inclusion of selection markers in DNA vectors transfected in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol.

The vector is selected to allow introduction of the gene of interest into the appropriate host cell. Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA. Expression systems for expressing a protein of the present invention are available using *Bacillus sp.* and *Salmonella* (Palva, *et al.*, (1983) *Gene* 22:229-35; Mosbach, *et al.*, (1983) *Nature* 302:543-5). The pGEX-4T-1 plasmid vector from Pharmacia is one *E. coli* expression vector for the present invention.

Expression in Eukaryotes

A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, the present invention can be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as expression systems for production of the proteins of the instant invention.

Synthesis of heterologous proteins in yeast is well known. Sherman, *et al.*, (1982) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory is a well-recognized work describing the various methods available to produce the protein in yeast. Two widely utilized yeasts for production of eukaryotic proteins are *Saccharomyces cerevisiae* and *Pichia pastoris*. Vectors, strains, and protocols for expression in *Saccharomyces* and *Pichia* are known in the art and available from commercial suppliers (e.g., Invitrogen). Suitable vectors usually have expression control sequences, such as promoters, including 3-phosphoglycerate kinase or alcohol oxidase, and an origin of replication, termination sequences and the like as desired.

A protein of the present invention, once expressed, can be isolated from yeast by lysing the cells and applying standard protein isolation techniques to the lysates or the pellets. The monitoring of the purification process can be accomplished by using Western blot techniques or radioimmunoassay of other standard immunoassay techniques.

The sequences encoding proteins of the present invention can also be ligated to various expression vectors for use in transfecting cell cultures of, for instance, mammalian, insect, or plant origin. Mammalian cell systems often will be in the form of monolayers of cells although mammalian cell suspensions may also be used. A number of suitable host cell lines capable of expressing intact proteins have been developed in the art, and include the HEK293, BHK21, and CHO cell lines. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter (e.g., the CMV promoter, a HAS *tk* promoter or *pgk* (phosphoglycerate kinase) promoter), an enhancer (Queen, *et al.*, (1986) *Immunol. Rev.* 89:49), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites (e.g., an SV40 large T Ag poly A addition site), and transcriptional terminator sequences. Other animal cells useful for production of proteins of the present invention are available, for instance, from the American Type Culture Collection Catalogue of Cell Lines and Hybridomas (7th ed., 1992).

Appropriate vectors for expressing proteins of the present invention in insect cells are usually derived from the SF9 baculovirus. Suitable insect cell lines include mosquito larvae, silkworm, armyworm, moth, and *Drosophila* cell lines such as a Schneider cell line (see, e.g., Schneider, (1987) *J. Embryol. Exp. Morphol.* 27:353-65).

As with yeast, when higher animal or plant host cells are employed, polyadenylation or transcription terminator sequences are typically incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague *et al.*, *J. Virol.* 45:773-81 (1983)). Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors (Saveria-Campo, "Bovine Papilloma Virus DNA a Eukaryotic Cloning Vector," in *DNA Cloning: A Practical Approach*, vol. II, Glover, ed., IRL Press, Arlington, VA, pp. 213-38 (1985)).

In addition, the RCA α -isoform gene placed in the appropriate plant expression vector can be used to transform plant cells. The polypeptide can then be isolated from plant callus or the transformed cells can be used to regenerate transgenic plants. Such transgenic plants can be harvested, and the appropriate tissues (seed or leaves, for example) can be subjected to large scale protein extraction and purification techniques.

Additional Modifications to Plants

Additional, agronomically important traits such as oil, starch, and protein content can be genetically altered in addition to using traditional breeding methods and are contemplated herein along with the RCA α -isoform modifications. Modifications include increasing content of oleic acid, saturated and unsaturated oils, increasing levels of lysine and sulfur, providing essential amino acids, and also

modification of starch. Hordothionin protein modifications are described in U.S. Patent Nos. 5,703,049, 5,885,801, 5,885,802, and 5,990,389. Another example is lysine and/or sulfur rich seed protein encoded by the soybean 2S albumin described in U.S. Patent No. 5,850,016, and the chymotrypsin inhibitor from barley, described in Williamson, *et al.*, (1987) *Eur. J. Biochem.* 165:99-106.

5 Derivatives of the coding sequences can be made by site-directed mutagenesis to increase the level of preselected amino acids in the encoded polypeptide. For example, the gene encoding the barley high lysine polypeptide (BHL) is derived from barley chymotrypsin inhibitor (see, U.S. Patent Application 08/740,682 and WO 98/20133). Other proteins include methionine-rich plant proteins such as from sunflower seed (Lilley, *et al.*, (1989) *Proceedings of the World Congress on Vegetable Protein*
10 *Utilization in Human Foods and Animal Feedstuffs*, ed. Applewhite (American Oil Chemists Society, Champaign, Illinois), pp. 497-502); corn (Pedersen, *et al.*, (1986) *J. Biol. Chem.* 261:6279; Kirihaara, *et al.*, (1988) *Gene* 71:359); and rice (Musumura, *et al.*, (1989) *Plant Mol. Biol.* 12:123). Other agronomically important genes encode latex, Floury 2, growth factors, seed storage factors, and transcription factors.

15 Insect resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Such genes include, for example, *Bacillus thuringiensis* toxic protein genes (U.S. Patent Nos. 5,366,892; 5,747,450; 5,736,514; 5,723,756; 5,593,881; and Geiser, *et al.*, (1986) *Gene* 48:109); and the like.

20 Genes encoding disease resistance traits include detoxification genes, such as against fumonisin (U.S. Patent No. 5,792,931); avirulence (avr) and disease resistance (R) genes (Jones, *et al.*, (1994) *Science* 266:789; Martin, *et al.*, (1993) *Science* 262:1432; and Mindrinos, *et al.*, (1994) *Cell* 78:1089); and the like.

25 Herbicide resistance traits may include genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance, in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides that act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS-gene mutants encode resistance to the herbicide chlorsulfuron.

30 Sterility genes can also be encoded in an expression cassette and provide an alternative to physical detasseling. Examples of genes used in such ways include male tissue-preferred genes and genes with male sterility phenotypes such as QM, described in U.S. Patent No. 5,583,210. Other genes include kinases and those encoding compounds toxic to either male or female gametophytic development.

The quality of grain is reflected in traits such as levels and types of oils, saturated and unsaturated, quality and quantity of essential amino acids, and levels of cellulose. In corn, modified hordothionin proteins are described in U.S. Patent Nos. 5,703,049, 5,885,801, 5,885,802, and 5,990,389.

Commercial traits can also be encoded on a gene or genes that could increase for example, starch
5 for ethanol production, or provide expression of proteins. Another important commercial use of transformed plants is the production of polymers and bioplastics such as described in U.S. Patent No. 5,602,321. Genes such as β -Ketothiolase, PHBase (polyhydroxybutyrate synthase), and acetoacetyl-CoA reductase (see, Schubert, *et al.*, (1988) *J. Bacteriol.* 170:5837-5847) facilitate expression of polyhydroxyalkanoates (PHAs).

10 Exogenous products include plant enzymes and products as well as those from other sources including prokaryotes and other eukaryotes. Such products include enzymes, cofactors, hormones, and the like. The level of proteins, particularly modified proteins having improved amino acid distribution to improve the nutrient value of the plant, can be increased. This is achieved by the expression of such proteins having enhanced amino acid content.

15 *Hybrid Seed Production.*

Hybrid seed production requires elimination or inactivation of pollen produced by the female parent. Incomplete removal or inactivation of the pollen provides the potential for selfing, raising the risk that inadvertently self-pollinated seed will unintentionally be harvested and packaged with hybrid seed. Once the seed is planted, the selfed plants can be identified and selected; the selfed plants are genetically
20 equivalent to the female inbred line used to produce the hybrid. Typically, the selfed plants are identified and selected based on their decreased vigor relative to the hybrid plants. For example, female selfed plants of e are identified by their less vigorous appearance for vegetative and/or reproductive characteristics, including shorter plant height, small ear size, ear and kernel shape, cob color or other characteristics. Selfed lines also can be identified using molecular marker analyses (see, e.g., Smith and
25 Wych, (1995) *Seed Sci. Technol.* 14:1-8). Using such methods, the homozygosity of the self-pollinated line can be verified by analyzing allelic composition at various loci in the genome.

Because hybrid plants are important and valuable field crops, plant breeders are continually working to develop high-yielding hybrids that are agronomically sound based on stable inbred lines. The availability of such hybrids allows a maximum amount of crop to be produced with the inputs used, while
30 minimizing susceptibility to pests and environmental stresses. To accomplish this goal, the plant breeder must develop superior inbred parental lines for producing hybrids by identifying and selecting genetically unique individuals that occur in a segregating population. The present disclosure contributes to this goal, for example by providing plants that, when crossed, generate male sterile progeny, which can be used as female parental plants for generating hybrid plants.

Use in Breeding Methods

The RCA modulated plants of the disclosure may be used in a plant breeding program. The goal of plant breeding is to combine, in a single variety or hybrid, various desirable traits. For field crops, these traits may include, for example, resistance to diseases and insects, tolerance to heat and drought, tolerance to chilling or freezing, reduced time to crop maturity, greater yield and better agronomic quality. With mechanical harvesting of many crops, uniformity of plant characteristics such as germination and stand establishment, growth rate, maturity and plant and ear height is desirable. Traditional plant breeding is an important tool in developing new and improved commercial crops. This disclosure encompasses methods for producing a plant by crossing a first parent plant with a second parent plant wherein one or both of the parent plants is a transformed plant displaying a phenotype as described herein..

Plant breeding techniques known in the art and used in a plant breeding program include, but are not limited to, recurrent selection, bulk selection, mass selection, backcrossing, pedigree breeding, open pollination breeding, restriction fragment length polymorphism enhanced selection, genetic marker enhanced selection, doubled haploids and transformation. Often combinations of these techniques are used.

The development of hybrids in a plant breeding program requires, in general, the development of homozygous inbred lines, the crossing of these lines and the evaluation of the crosses. There are many analytical methods available to evaluate the result of a cross. The oldest and most traditional method of analysis is the observation of phenotypic traits. Alternatively, the genotype of a plant can be examined.

A genetic trait which has been engineered into a particular plant using gene editing or transformation techniques can be moved into another line using traditional breeding techniques that are well known in the plant breeding arts. For example, a backcrossing approach is commonly used to move a modified plant to an elite inbred line and the resulting progeny would then comprise the modification. Also, if an inbred line was used for the transformation or editing, then those plants could be crossed to a different inbred in order to produce a hybrid plant. As used herein, "crossing" can refer to a simple X by Y cross or the process of backcrossing, depending on the context.

The development of a hybrid in a plant breeding program involves three steps: (1) the selection of plants from various germplasm pools for initial breeding crosses; (2) the selfing of the selected plants from the breeding crosses for several generations to produce a series of inbred lines, which, while different from each other, breed true and are highly homozygous and (3) crossing the selected inbred lines with different inbred lines to produce the hybrids. During the inbreeding process, the vigor of the lines decreases. Vigor is restored when two different inbred lines are crossed to produce the hybrid. An important consequence of the homozygosity and homogeneity of the inbred lines is that the hybrid created by crossing a defined pair of inbreds will always be the same. Once the inbreds that give a superior hybrid

have been identified, the hybrid seed can be reproduced indefinitely as long as the homogeneity of the inbred parents is maintained.

Transgenic plants of the present disclosure may be used to produce, e.g., a single cross hybrid, a three-way hybrid or a double cross hybrid. A single cross hybrid is produced when two inbred lines are crossed to produce the F1 progeny. A double cross hybrid is produced from four inbred lines crossed in pairs (A x B and C x D) and then the two F1 hybrids are crossed again (A x B) times (C x D). A three-way cross hybrid is produced from three inbred lines where two of the inbred lines are crossed (A x B) and then the resulting F1 hybrid is crossed with the third inbred (A x B) x C. Much of the hybrid vigor and uniformity exhibited by F1 hybrids is lost in the next generation (F2). Consequently, seed produced by hybrids is consumed rather than planted.

This invention can be better understood by reference to the following non-limiting examples. It will be appreciated by those skilled in the art that other embodiments of the invention may be practiced without departing from the spirit and the scope of the invention as herein disclosed and claimed.

Example 1. Generation of transgenic rice having altered *RCA* α with altered DNA and amino acid sequence in exon 7

Bacterial strains were grown in LB liquid or on agar plates with appropriate antibiotics for plasmid selection. All cloning was performed using restriction endonucleases and T4 DNA ligase unless indicated otherwise and transformations were performed by electroporation into *E. coli* strain DH10B.

The CRISPR/Cas9 construct used to make targeted mutations in exon 7 of rice *RCA* α -isoform gene. The goal was to modify the following sequence in exon 7: 5'-GGGCAAGGAGCACAGCAAGCAGG-3' (SEQ ID NO: 43). To perform CRISPR/Cas9, the sequence 5'-GGGCAAGGAGCACAGCAAGC-3' (SEQ ID NO: 44; which lacks the 3' AGG PAM sequence) was optimally inserted as oligo pairs into a vector containing a rice U6 promoter, dual BtgZI restriction endonuclease sites for oligo pair insertion, a single guide RNA (sgRNA) sequence, a second rice U6 promoter, dual BsaI restriction endonuclease sites and another sgRNA sequence. The oligo pairs for the BtgZI sites were AEP1650 (5'-TGTTGGGCAAGGAGCACAGCAAGC-3' SEQ ID NO: 30) and AEP1651 (5'-AAACGCTTGCTGTGCTCCTTGCCC-3' SEQ ID NO: 31); and the oligo pairs for the BsaI sites were AEP1652 (5'-GTGTGGGCAAGGAGCACAGCAAGC-3' SEQ ID NO: 32) and AEP1651 (5'-AAACGCTTGCTGTGCTCCTTGCCC-3' SEQ ID NO: 31).

The following three cloning steps were performed. First, pENTR4:sgRNA4 (Bing Yang lab - ISU), containing a dual U6 sgRNA cassette without targeting oligos and a kanamycin selection cassette, was cut open at BtgZI, then the phosphorylated and annealed oligo pair AEP1650 (SEQ ID NO: 30) and AEP1651 (SEQ ID NO: 31) was ligated in to create plasmid pLW1, which now contains a complete single U6 sgRNA. pLW1 was cut at BsaI, then the phosphorylated and annealed oligo pair AEP1651

(SEQ ID NO: 31) and AEP1652 (SEQ ID NO: 32) was ligated in to create pLW4, containing two complete U6 sgRNAs that target SEQ ID NO: 43 of exon 7 of rice RCA. Note that the same site is targeted by both oligo pairs. pBY02:gamma CAS9-ccdB (Bing Yang lab) was joined to pLW4 using Gateway cloning to create pLW6. pLW6 is 16,351 bp and contains the rice RCA exon 7 U6 sgRNA targeting cassette followed by a maize ubiquitin promoter with a rice optimized CAS9 gene and a NOS terminator, a bacterial kanamycin resistance selection cassette and a 2x35S promoter with a hygromycin resistance gene and a 35S terminator for selection of rice transformants. Note that because both pLW4 and pBY02:gamma CAS9-ccdB are kanamycin resistant, the bacterial activity of the 2x35S hygromycin resistance gene cassette was used for selection during Gateway cloning. pLW6 was transformed by electroporation into *Agrobacterium tumefaciens* strain EHA105 for rice transformation.

In pLW6, the U6 promoters expresses the sgRNAs containing the sequence 5'-GGGCAAGGAGCACAGCAAGC-3' (SEQ ID NO: 44) and the ubiquitin promoter expressed the Cas9 RNA, which was translated into Cas9 protein. The Cas9 protein binds the sgRNA then the complex binds the chromosome target sequence 5'-GGGCAAGGAGCACAGCAAGCAGG-3' (SEQ ID NO: 43) because of the presence of the AGG PAM sequence, which in turn triggers the nuclease activity. Nuclease activity causes a chromosome break in the target sequence (SEQ ID NO: 43) at or near a position six nucleotides from the 3' end of the target sequence.

After a chromosome break is detected by the cell, the break is repaired, typically through non-homologous end-joining where DNA at the break point is removed then a repair is made by rejoining the DNA using micro-homology at the remaining ends. This action typically causes small deletions, insertions or base changes at the target site resulting in mutation of existing sequence that may cause frameshifts, additions, deletions or residue changes in the open reading frame of a coding sequence, thereby resulting in alteration of encoded protein as seen in FIGs. 2-19 depicting the sequence of various obtained rice mutants.

Rice cultivar kitaake was used for transformation. All procedures were performed essentially as directed by Cheng, et al. (Rice Transformation by Agrobacterium Infection. *Methods in Biotechnology*, Vol. 3: *Recombination Protein from Plants: Production and Isolation of Clinically Useful Compounds*; pages 1-9 (2009), ed. Cunningham and Porter, Humana Press Inc., Totowa, NJ), except that regeneration was in a 1% CO₂ chamber. Briefly, fresh kitaake seed were dehusked then sterilized in a 95% ethanol bath for 5 minutes followed by a 40% bleach bath for 30 minutes then rinsed in sterile water. Sterile seed were place on callus induction media embryo up for 14 days at 30°C with continuous light. Calli were separated from seed and grown under the same conditions for 2 additional weeks. *A. tumefaciens* strain EHA105 containing pLW6 was grown in liquid callus induction media containing 200 micromolar acetosyringone to an OD600 of 0.1 to 0.2. *A. tumefaciens* strain EHA105 and rice calli were incubated

together for 30 minutes then calli were blotted dry on sterile Whatman paper. Calli and *A. tumefaciens* strain EHA105 were co-cultivated for 3 days at 22°C in light on callus induction media plus 200 µM acetosyringone. Calli were then washed in 50 ml sterile selection media containing 400 µg/ml carbenicillin until the solution was mostly clear, then calli were blotted dry and placed on selection media containing 50 µg/ml hygromycin at 30°C with continuous light for 3 weeks. Calli were moved to fresh selection media with hygromycin and incubated at 30°C with continuous light for 3 more weeks. Resistant calli were picked off the main callus bodies and transferred to selection media with hygromycin then incubated at 30°C with continuous light until they were 2-3 mm in size. 2-3 mm calli were placed on regeneration media with hygromycin and grown at 30°C in continuous light until they turned green (approximately 4 weeks). Shoots from green calli were transferred to fresh regeneration media containing hygromycin in Magenta boxes. When shoots neared the top of the Magenta box, they transferred to rooting media with hygromycin selection and incubated at 25°C with 16 hours of light and 8 hours of dark. Once roots develop, the plants were transferred to soil.

Rice plants were grown either individually in pots or together in large trays using field soil or Berger brand BM custom potting mix containing peatmoss, perlite and bark. Plants were submerged in water 2 to 3 cm over the top of the soil or potting mix and supplemented with Everris water soluble fertilizer diluted to 5 ml/L from a 37 g/L stock and Sprint 330 chelated iron diluted to 5 ml/L in water from a 119 g/L stock. Plants were watered with 1 L of fertilizer and iron chelate twice per week and water was maintained at 2 to 3 cm over the soil or potting mix. Plant growth was at 28°C and approximately 160 µ-Einsteins of light with a 16 hour day length.

Mutation in exon 7 of the rice RCA gene was detected by PCR followed by T7 endonuclease assays and sequence verification for positive events. Mutations are expected to occur at the CRISPR/Cas9 target site by non-homologous end-joining (NHEJ), which is expected to create insertions, deletions or base changes at the target site. These changes are detected by T7 endonuclease using annealed wild type and mutant PCR products, which results base mismatches and subsequent restriction of the PCR product into smaller fragments by T7 endonuclease. Restriction digestion is identified on electrophoresis agarose gels. Mutant identity is confirmed by DNA sequence analysis.

PCR

Primers AEP1722 (5'-CTACTATATCTTGTCTGCATTTTCTC-3' SEQ ID NO: 33) and AEP1701 (5'-CCATGAATGTCACATGTGAATTAG-3' SEQ ID NO: 34) were used to generate an approximate 590 bp PCR product directly from rice leaves as instructed by the Phire PCR kit (Thermo kit #F-130-WH). Briefly, PCR products for the known wild-type rice sample and a transgenic mutated rice samples were mixed then melted and annealed through a temperature range of 100°C, down to 25°C, over a 12 minute period in a PCR machine. 20 µL PCR reactions were performed for each plant along with

positive and negative contamination controls from the kit, positive and negative wild type plant controls and a T7 assay positive control for a mutant and wild type DNA samples. The T7 controls used mutant and wildtype *Chlamydomonas reinhardtii* Arg7 gene sequences from plasmids pDW2638 and pJD67 respectively using primers AEP976 (5'-TGCTCTACTACTGTTTCCTGGCTAC-3' SEQ ID NO: 35) and
 5 AEP1236 (5'-CGCAGGTGTCTGACCGCGACTTTG-3' SEQ ID NO: 36) to yield an approximate 337 bp PCR product. Two μ L of each PCR reaction were run on a 2.5% agarose gel to check PCR product quality for each sample.

T7 assays were performed by mixing 3.5 μ L of wild type and 3.5 μ L of experimental plant PCR product with 13 μ L ddH₂O, then the mixture was heated to 100°C and cooled in 10°C steps for 1 minute
 10 per step down to 20°C. After cooling, 0.2 μ L of T7 endonuclease (NEB#M0302L - 10,000 units per ml) was added to each tube, and these were incubated at 37°C for 2 hours. Note that the T7 control was a mixture of 3.5 μ L mutant and 3.5 μ L wild-type Arg7 PCR product in 13 μ L ddH₂O products, and that 7 μ L of wild-type rice PCR product mixed with 13 μ L ddH₂O was used as a negative control. Controls were treated in the same manner as experimentals. After digestion, 4 μ L of SUDS was added to each reaction,
 15 and the reactions were run on a 2.5% agarose gel. Positive reactions were indicated by the appearance of approximate 100 and a 200 bp band in the T7 positive control, no lower bands in the wild-type rice control and approximate 470 and 120 bp bands in experimentals lanes.

The remaining 11 μ L of each PCR reaction that was determined to be positive by the T7 assay were run on a 2.5% agarose gel, approximate 590 bp bands were excised and cleaned using the IBI
 20 Scientific Gel/PCR product DNA fragment extraction kit (#IB47030) and sequenced using AEP976. Mutations in sequence were identified by typical single peaks in the chromatogram followed by double peaks after the expected mutation site. Initial mutation identity was determined by separating double peaks into consecutive strings of nucleotides that follow the known rice RCA DNA sequence then determining if nucleotides were deleted, inserted or changed at the predicted target site.

FIGs. 2-19 provide the DNA and amino acid sequences of the mutated RCA α gene exon 7 in
 25 chromosomes 1 and 2 for the genetically altered rice plant lines indicated on each figure, and made using the above protocols. For FIGs. 2-19, where applicable, (i) underlined amino acids are from the wild-type exon 7 coding sequence; (ii) the other amino acids arise from the genetic change caused by CRISPR/Cas9; (iii) an asterisk is a stop codon, and (iv) lower case nucleotides are nucleotides that are
 30 inserted into exon 7 (except for the last nucleotide of intron 6). In FIG. 8, it is possible that the indicated inserted cytosine nucleotide could be a guanine nucleotide instead in chromosome 1; and in chromosome 2 the indicated 5 bp deletion with insertion AGT nucleotides could also be viewed as a 3 bp deletion with 1 bp insertion.

Example 2. Phenotypic changes in the transgenic rice plants with altered RCA α DNA & amino acid sequences

1. Plant growth: Wild-type (Kitake) and CRISPR driven transgenics were germinated on the wet paper and transferred to individual pot using custom mix soil (Sungro Horticulture). Fertilizer containing Fe, N, P and K was provided once a week. The plants were grown in HL (continuous high light for 16 hrs) or LL-HL (1 hr high light and 1 hr low light for 16 hrs) with 25°C and 80% humidity. The intensities of high and low light are 40 and 400 ($\mu\text{molm}^{-2}\text{s}^{-1}$), respectively.

2. Western blotting: Total protein was extracted from frozen leaf tissue by grinding in SDS-sample buffer containing 62.5 mM Tris-HCl, pH 8.0, 2% SDS, 1 M urea, 10% glycerol and 0.005% bromophenol blue. Proteins were then resolved on 10% SDS-PAGE gels. Gel-resolved proteins were electrophoretically transferred to low fluorescence PVDF membrane and treated with blocking solution (5% gelatin, Sigma Aldrich) for 1 hr at room temperature. Primary anti-RCA antibody reaction was performed at room temperature for overnight. Immunoblots involving fluorescent secondary antibodies (IRDye 800CW; LI-COR Biosciences, Lincoln, NE, USA) were scanned using a LI-COR Odyssey Infrared Imaging System for visualization. Representative results showing the presence of the RCA α -isoform in wild type rice but not three transgenic lines (transgenic rice lines 6-3, 8-4, and 9-3), are shown in FIG. 20A.

3. Growth phenotype analysis: Six weeks old plants were selected to characterize plant development. Three individual plants of each line (transgenic rice lines 6-3, 8-4, and 9-3 and wild-type rice) were examined to acquire average value of the following characteristics. All plants were grown in a growth chamber with constant light (400 μE) during the 16 hr photoperiod at 25°C. Plant height represents the length from the ground to the top of a plant. The middle of the primary leaf was selected to measure the leaf width. A representative photo of wild type rice and transgenic rice plant line 9-3 during reproductive development is shown in FIG. 20B. Compared to wild-type rice, which were 88.3 ± 1.7 cm tall, transgenic rice lines 6-3 (96 ± 2.2 cm), 8-4 (96 ± 4.1 cm) and 9-3 (98.7 ± 2.6 cm) were considerably taller. The length of individual leaves was not increased, but leaves were wider in the transgenic rice lines than in the wild-type rice. Wild-type rice leaves were 1.07 ± 0.05 cm wide, and the transgenic rice lines' leaves were approximately 20% wider: transgenic rice line 6-3 leaves were 1.30 ± 0.08 wide; transgenic rice line 8-4 leaves were 1.37 ± 0.10 ; and transgenic rice line 9-3 leaves were 1.33 ± 0.05 . Thus, the transgenic rice plants had a greater total leaf area to assimilate atmospheric CO_2 for growth and seed production.

4. Gas exchange: Photosynthetic induction rate was measured with 6 weeks old plants and performed at the end of the dark period. As shown in FIG. 20C, the three transgenic plant lines had faster rates of photosynthetic induction compared to the rate of photosynthetic induction in the wild-type plant

upon transfer of the plants from low to high light, suggesting increased light use efficiency of the transgenic plants compared to the light use efficiency of wild-type plants.

5. Seed yield per plant: For the total seed number, the 14 weeks old plants were selected and counted. At maturity, the three transgenic lines produced more seeds per plant compared to amount of seed produced by wild-type cultivar, indicating that the editing of the *RCA* gene's sequence to remove the di-sulfide bond at the carboxyl terminus of the α -isoform directly impacts seed production (see FIG. 20D). Interestingly, stover yield at maturity was not affected by the editing of the *RCA* gene's sequence.

CRISPR lines deduced amino acid sequences

The wild-type rice *RCA* α -isoform carboxy terminal amino acid sequence in chromosome 1 is GQGAQQAGNLPVPEGCTDPVAKNFDPTARSDDGSCLYTF* (SEQ ID NO: 37; the cysteine amino acids are underlined), and the *RCA* α -isoform carboxy terminal amino acid sequence in chromosome 2 is GSAPSS (SEQ ID NO: 38). SEQ ID NO: 37 is the same as SEQ ID NO: 2, except that SEQ ID NO: 45 lists a glycine (G) as the initial amino acid, and this glycine arises from the codon that is generated during the splicing between exon 6 and exon 7.

Altered rice plant line 6-3 has a 4 bp deletion in *RCA* exon 7 in chromosome 1, which frame shifts to a longer open reading frame (ORF) than the original, non-mutated ORF, and chromosome 2 has a 1 bp deletion and frame shifts to a stop codon (termination) prior to the first cysteine in exon 7. In particular, the altered *RCA* α -isoform carboxy terminal amino acid sequence from the *RCA* gene in chromosome 1 is

GQGAQQVTCLCRKVAPTLLPRTSTQRRGATTAAFTPFKQAGLTLAINYFSFLCFLFVLCIRSRPS HSWA* (SEQ ID NO: 39). SEQ ID NO: 39 is the same as SEQ ID NO: 6 except that SEQ ID NO: 39 lists a glycine (G) as the initial amino acid, and this glycine arises from the codon that is generated during the splicing between exon 6 and exon 7. See also Table 1 *supra*, and the applicable figures for transformed rice plant lines 8-1, 8-2, 8-6, 8-7, 8-8, 1, 6-4, 8-10, 9-1, 9-1, 21-1, and 21-2 which all have a mutation the *RCA* gene exon 7 in chromosomes 1 and/or 2 that encode an *RCA* α -isoform with the same carboxy terminal amino acids.

In altered rice plant line 6-3, the altered *RCA* α -isoform carboxy terminal amino acid sequence in the mutated *RCA* exon 7 gene on chromosome 2 is GQGAQHRSR* (SEQ ID NO: 40). SEQ ID NO: 40 is the same as SEQ ID NO: 8 except that SEQ ID NO: 40 lists a glycine (G) as the initial amino acid, and this glycine arises from the codon that is generated during the splicing between exon 6 and exon 7. See also Table 1 *supra*, and the applicable figures for transformed rice plant lines 8-1, 8-2, 8-4, 8-5, 8-7, 8-8, 20-1, 20-2, 20-3, 6-1, 6-4 which all have a mutation the *RCA* gene exon 7 in chromosomes 1 and/or 2 that encode an *RCA* α -isoform with the same carboxy terminal amino acids.

Altered rice plant line 8-4 has a 1 bp insertion in the mutated *RCA* exon 7 gene on both chromosomes 1 and 2, which results in frame shifts to a stop codon (termination) before the first cysteine encoded within exon 7. The altered *RCA* α -isoform carboxy terminal amino acid sequence encoded by the mutated *RCA* exon 7 gene on chromosome 1 is GQGAQHRS* (SEQ ID NO: 41). The altered *RCA* α -isoform carboxy terminal amino acid sequence for rice plant line 8-4 encoded by the mutated *RCA* exon 7 gene on chromosome 2 is also GQGAQQR* (SEQ ID NO: 41). SEQ ID NO: 41 is the same as SEQ ID NO: 15 except that SEQ ID NO: 41 lists a glycine (G) as the initial amino acid, and this glycine arises from the codon that is generated during the splicing between exon 6 and exon 7. See also Table 1 *supra*, and the applicable figures for transformed rice plant lines 8-5, 6-1, 6-2, 1, and 9-3 which all have a mutation the *RCA* gene exon 7 in chromosomes 1 and/or 2 that encode an *RCA* α -isoform with the same carboxy terminal amino acids.

Altered rice plant line 9-3 has a 1 bp deletion in the mutated *RCA* exon 7 gene on chromosome 1 which results in a frame shift to the longer ORF than the original, non-mutated ORF and a 1 bp insertion in the mutated *RCA* exon 7 gene on chromosome 2 which results in a frame shift to a stop codon before the first cysteine in exon 7. The altered *RCA* α -isoform carboxy terminal amino acid sequence encoded by the mutated *RCA* exon 7 gene on chromosome 1 in altered rice plant line 9-3 is GQGAQQQVTCLCRKVAPTLLPRTSTQRRGATTAAFTPFKQAGLTLAINYFSFLCFLFVLCIRSR PSHSWA* (SEQ ID NO: 42). SEQ ID NO: 42 is the same as SEQ ID NO: 22 except that SEQ ID NO: 42 lists a glycine (G) as the initial amino acid, and this glycine arises from the codon that is generated during the splicing between exon 6 and exon 7. See also Table 1 *supra*, and the applicable figures for transformed rice plant lines 9-2, 13-1, 13-2, 16, and 17 which all have a mutation the *RCA* gene exon 7 in chromosomes 1 and/or 2 that encode an *RCA* α -isoform with the same carboxy terminal amino acids. For transformed rice plant line 9-3, the altered *RCA* α -isoform carboxy terminal amino acid sequence encoded by the mutated *RCA* α gene exon 7 on Chromosome 2 is GQGAQQR* (SEQ ID NO: 41) which is discussed *supra*.

All of the plants lines discussed in this example are independent lines. It is interesting to note that the frame shift to the longer ORF in plant lines 6-3 and 9-3 is not apparent on the Western blot. It is possible that these transformed rice lines do not produce the altered *RCA* α -isoform but these transformed plants still have an increase or improvement in the desired agronomic phenotypes and result in increased activity in RuBisCO.

Example 3. Altering *RCA* α sequence in other plants

As shown in FIG. 21, the genomes of *Brassica napus*, *Camelina sativa*, *Arachis duranensis*, *Vigna angularis*, *Cucumis melo*, *Cucumis sativus*, *Fragaria vesca* subsp. *vesca*, *Gossypium hirsutum*, *Theobroma cacao*, *Morus notabilis*, *Prunus mume*, *Beta vulgaris* subsp. *vulgaris*, *Citrus clementina*,

Glycine max, *Malus domestica*, *Prunus persica*, *Vitis vinifera*, *Musa acuminata* subsp. *malaccensis*, and *Zea mays* are all predicted to have RCA α -isoforms with very similar carboxy terminal amino acid sequences and two cysteine amino acids involved in redox regulation (*i.e.*, redox-sensitive cysteines). One can alter RCA α -isoform genomic sequences in these plants, and other plants, to remove one or both cysteines involved in the redox regulation of RuBisCO's activity, thereby increasing RuBisCO's activity state and generating altered plants with the desired agronomic phenotypes described herein for the altered rice. Using the protocols described in Example 1 and the published sequences of RCA α -isoform, one can create primers that would generate an sgRNA that targets the carboxy terminal sequences of the plant's RCA α -isoform. One can transform wild-type plant cells with an expression vector containing U6 promoters operably linked to a polynucleotide encoding sgRNA for the desired target sequence, a promoter operably linked to a polynucleotide encoding Cas9, and a selection marker. The transformed plant cells are selected for those cells that produce the selection marker and/or Cas9 and the desired sgRNA sequence. The selected transformed plant cells. One can isolate the DNA from the transformed plant cells to determine the alteration in RCA α -isoform carboxy terminal sequence. The transformed plant cells can be induced to grow into transformed plants and assess for an increase in at least one agronomic phenotype described herein (*e.g.*, RCA α -isoform activation, RuBisCO activation, plant growth, leaf width, seed yield per plant, and photosynthetic induction rate) when grown under constant light during the photoperiod compared to the level of the agronomic phenotype in the wild-type plant. These transformed plants will have an altered RCA α -isoform with an altered carboxy terminal amino acid sequence, a loss of redox activity, and increased activation of RuBisCO.

Example 4. Silencing RCA α -isoform via RNAi

Using bioinformatics tools for rice, RCA α -isoform mRNA and RCA β -isoform mRNA are compared and found to be nearly identical, except for the absence of a 85 bp segment from the α -isoform mRNA resulting from alternative splicing. Presence of the 85 bp region in β -isoform mRNA ensures that exon 7a is encoded, and the resulting protein is the β -isoform. The absence of the 85 bp region in α -isoform mRNA ensures that exon 7b is encoded, and the resulting protein is the α -isoform. Because of this arrangement, 18 bp sequences flanking the 85 bp region in the β -isoform mRNA is used for RNAi. This strategy reduces the chance of targeting the β -isoform mRNA because the sequence is less than 20 bp on either side of the 85 bp region in question, and it represents 36 bp of consecutive base homology in the α -isoform mRNA. Thus, 5'-GACTGGTTCCTTCTACGGGCAAGGAGCACAGCAAGC-3' (SEQ ID NO: 64) is the target sequence for which dsRNA is to be generated.

This 36 bp sequence is placed in an expression vector such that two copies, in opposite orientation (sense and anti-sense sequence), with a spacer sequence placed between the sense and anti-sense sequence so that a hairpin loop can be founded and the sense sequence and anti-sense sequence can

bind, forming dsRNA. The expression vector contains a left border sequence, CaMV35S promoter driving expression of the hygB resistance gene with a 35S terminator, a rice UBI promoter, EcoRI restriction endonuclease site, the sense target sequence (SEQ ID NO: 64), BamHI restriction endonuclease site, a generic spacer sequence with a chloramphenicol resistance gene for *E. coli* selection, XhoI restriction endonuclease site, the anti-sense sequence of the target sequence (anti-sense of SEQ ID NO: 64), XbaI restriction endonuclease site, a NOS terminator and right border sequence all cloned into an *Agrobacterium* T-DNA vector. One such example of the expression vector is pHellsgate 8.0 plasmid.

The completed *Agrobacterium* T-DNA vector is electroporated into *Agrobacterium tumefaciens* strain EHA105 by electroporation and selected on LB media supplemented with 100 µg/ml of streptomycin for T-DNA vector selection. *Agrobacterium* colonies are grown at 30°C then isolated and grown for vector verification via restriction digestion and running on a gel.

Kitaake rice callus is generated on callus induction media then infected with the EHA105 *Agrobacterium tumefaciens* strain transformed with the vector. Infected callus are grown on co-cultivation media then selected on selection media supplemented with hygromycin B at 50 µg/ml. Once callus are selected, regeneration media supplemented with 50 µg/ml hygromycin B is used to produce rice shoots followed by rooting on root generation media again supplemented with 50 µg/ml hygromycin B. Specific protocol details are described in Cheng, *et al.* 2009 and also described above in Example 1.

Potential transgenic plants are screened for the presence of the expression vector containing the RNAi sequence by Southern blot hybridization. After transformation is verified, expression of the RNAi transgene is confirmed RT-PCR, and then the RCA α-isoform gene expression is analyzed by RT-PCR. RCA α-isoform gene silencing or reduction of gene expression is also verified by examining the genetically altered plants for morphological changes described above (broader leaves, larger size, increased seed yield/plant) and by examining RCA α-isoform protein expression using antibodies in a Western blot assay. The genetically altered rice are distinguished from wild-type rice via Western blot because the genetically altered rice produce RCA β-isoform only whereas the wild-type rice produce RCA α-isoform and RCA β-isoform.

All publications, patents, and patent applications mentioned herein are indicative of the level of ordinary skill in the art to which this invention pertains. All publications, patents, and patent applications mentioned herein are herein incorporated by reference to the same extent as if each individual document was specifically and individually indicated by reference.

The invention has been described with reference to various embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention as described in the appended claims.

CLAIMS

We, the inventors, claim as follows:

5 Claim 1. An altered ribulose 1,5-bisphosphate carboxylase/oxygenase activase (RCA) α -isoform having reduced light sensitive inhibitory activity compared to light sensitive inhibitory activity in a wild-type RCA α -isoform.

10 Claim 2. The altered RCA α -isoform of Claim 1, wherein said wild-type RCA α -isoform has a redox active site in its carboxy terminus and two redox-sensitive cysteines within said redox active site, and wherein said altered RCA α -isoform has less than two redox-sensitive cysteines at its redox active site.

15 Claim 3. The altered RCA α -isoform of Claim 2, wherein said altered RCA α -isoform has no redox-sensitive cysteine at its redox active site.

20 Claim 4. The altered RCA α -isoform of Claim 3, wherein said altered RCA α -isoform is in rice and has a carboxy terminal amino acid sequence comprising at least one sequence selected from the group consisting SEQ ID NOs: 4, 8, 10, 15, 17, 19, and 24.

25 Claim 5. The altered RCA α -isoform of Claim 1, wherein said wild-type RCA α -isoform has a redox active site in its carboxy terminus and two redox-sensitive cysteines within said redox active site, and wherein said altered RCA α -isoform has at least three cysteines at its redox active site.

30 Claim 6. The altered RCA α -isoform of Claim 5, wherein said altered RCA α -isoform is in rice and has a carboxy terminal amino acid sequence comprising at least one sequence selected from the group consisting SEQ ID NOs: 6, 12, 22, 27, and 29.

35 Claim 7. An altered plant, and its progeny, comprising said altered RCA α -isoform of Claim 1.

40 Claim 8. An altered plant and part thereof, and its progeny, having an increase in at least one agronomic phenotype compared to said at least one agronomic phenotype of a wild-type plant grown under similar conditions, said altered plant and part thereof, and its progeny, comprising an altered RCA α -isoform lacking a redox active site at said RCA α -isoform's carboxy terminus, wherein said altered RCA α -isoform causes said increase said agronomic phenotype, and wherein said agronomic phenotype is

selected from the group consisting of plant growth, leaf width, seed yield per plant, and photosynthetic induction rate.

5 Claim 9. An altered seed from said altered plant of Claim 8, wherein said altered seed contains altered plant cells that produce said altered RCA α -isoform.

Claim 10. An altered plant cell from said altered plant of Claim 8, wherein said altered plant cell produces said altered RCA α -isoform.

10 Claim 11. The altered plant and its progeny of Claim 8, wherein said altered plant is rice, and wherein said altered RCA α -isoform has a carboxy terminal amino acid sequence selected from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 15, 17, 19, 22, 24, 27, 29, and a combination thereof.

15 Claim 12. A method of increasing at least one agronomic phenotype in a transformed plant compared to a wild-type plant's agronomic phenotype, said method comprising

- (i) altering a wild-type plant cell's RCA α -isoform's 3' end DNA sequence to produce an altered plant cell comprising at least one altered RCA α -isoform comprising a DNA alteration at said RCA α -isoform's 3' end;
- (ii) selecting for said altered plant cell that contains said altered RCA α -isoform and produces an
20 altered RCA α -isoform to produce a selected transformed plant cell, wherein said altered RCA α -isoform has reduced carboxy terminal redox activity compared to said wild-type plant cell's RCA α -isoform's carboxy terminal redox activity; and
- (iii) growing said selected transformed plant cell into said transformed plant, wherein said
25 transformed plant produces said altered RCA α -isoform and said altered RCA α -isoform causes said transformed plant to have an increase in said at least one agronomic phenotype compared to said wild-type plant's agronomic phenotype.

30 Claim 13. The method of Claim 12, wherein said at least one agronomic phenotype is selected from the group consisting of plant growth, leaf width, seed yield per plant, and photosynthetic induction rate.

Claim 14. A transformed plant and part thereof and its progeny made by the method of Claim 12.

Claim 15. A transformed plant and part thereof of Claim 14, wherein said transformed plant is transformed rice and said altered RCA α -isoform has a carboxy terminal amino acid sequence selected

from the group consisting of SEQ ID NOs: 4, 6, 8, 10, 12, 15, 17, 19, 22, 24, 27, 29, and a combination thereof.

Claim 16. A method of plant breeding to increase at least one agronomic trait comprising:

- 5 (i) crossing a male parent plant with a female parent plant, wherein one of said parent plants produces an altered RCA α -isoform lacking redox activity at its carboxy terminus, to produce progeny seed that produce said altered RCA α -isoform; and
- (ii) harvesting said progeny seed from said female parent plant, wherein said produced altered RCA α -isoform lacking redox activity in said carboxy terminus causes said increase in
- 10 said at least one agronomic trait.

Claim 17. The method of Claim 16, wherein said one of said parent plants comprises at least one altered RCA α -isoform gene that encodes said altered RCA α -isoform lacking redox activity at said carboxy terminus.

Claim 18. The method of Claim 16, wherein said agronomic trait is selected from the group consisting of plant growth, leaf width, seed yield per plant, and photosynthetic induction rate.

Claim 19. A method of reducing a genetically altered plant's RCA redox activity compared to a wild-type plant's RCA redox activity, said method comprising

- 20 (i) transforming a wild-type plant cell with an expression vector to produce a transformed plant cell, wherein said expression vector comprising a polynucleotide encoding a dsRNA having a sequence that encodes at least nineteen contiguous nucleotides of said wild-type plant's RCA carboxy terminus which has redox activity;
- 25 (ii) selecting said transformed plant cell that produces said dsRNA to provide a selected transform plant cell; and
- (iii) induces said selected transform plant cell to grow into a genetically altered plant, wherein said produced dsRNA reduces the amount of RCA redox activity in said genetically altered plant compared to the amount of RCA redox activity in said wild-type plant.

Wild-type rice RCA α exon 7

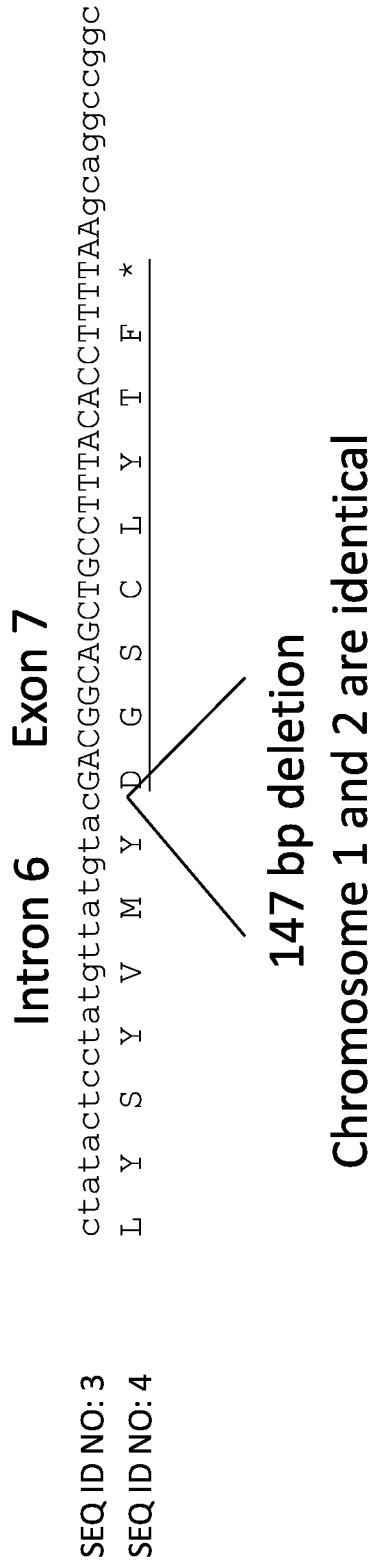
FIG. 1

Intron 6		Exon 7
SEQ ID NO: 1	g	GGCAAGGAGCACAGCAAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCT
SEQ ID NO: 2		<u>Q G A Q Q A Q A G N L P V P E G C T D P</u>
SEQ ID NO: 1		GTTGCCAAGAACTTCGACCCCAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA
SEQ ID NO: 2		<u>V A K N F D D P T A R S D D G S C L Y T F *</u>

Underlined amino acids are the original exon 7 coding sequence
Asterisk is a stop codon

Rice plant line 2-1 bi-allelic deletion

FIG. 2



Underlined amino acids are the remnants of the exon 7 coding sequence
Uppercase nucleotides are the remnants of the exon 7
Lower case nucleotides are the remnants of intron 6 sequence
Asterisk is a stop codon

FIG. 3

Chromosome 1 – 4 bp deletion		
Intron 6	Exon 7	
SEQ ID NO: 5	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAAGAACTTCGACCCCAA	
SEQ ID NO: 6	<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T Q	
SEQ ID NO: 5	CGCGGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAAGCAGGCCGGCTTGACGGCTTGCTATTAAATTAT	
SEQ ID NO: 6	R R G A T A A A F T P F K Q A G L T L A I N Y	
SEQ ID NO: 5	TTCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA	
SEQ ID NO: 6	F S F L C F L F V L C I R S R P S H S W A *	
Chromosome 2 – 1 bp insertion (c)		
Intron 6	Exon 7	
SEQ ID NO: 7	gGGCAAGGAGCACAGCAcAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAG	
SEQ ID NO: 8	<u>Q G A Q</u> H S R *	
SEQ ID NO: 7	AACTTCGACCCCAACGGCGGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA	

FIG. 4

Rice plant line 8-3 bi-allelic mutations

Chromosome 1 – 2 bp deletion

Intron 6	Exon 7
SEQ ID NO: 9	gGGCAAAGGAGCACACAGCACAGGTAACCTGCCCTGTGCCGGAAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGACCC
SEQ ID NO: 10	<u>Q G A Q H R</u> *
SEQ ID NO: 9	AACGGCGAGGAGCGACGACGGCAGCTGCCTTTACACCTTTTAA

Chromosome 2 – 6 bp deletion and 2 bp insertion (ag)

Intron 6	Exon 7
SEQ ID NO: 11	gGGCAAAGGAGCACAGagGGTAACCTGCCCTGTGCCGGAAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGACCCAA
SEQ ID NO: 12	<u>Q G A Q R V T C L C R K V A P T L L P R T S T Q</u>
SEQ ID NO: 11	CGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCCTTTTAAGCAGGCGCGGCTTGACGCTTGCTATTAAATTATTC
SEQ ID NO: 12	R R G A T T A A A F T P F K Q A G L T L A I N Y F
SEQ ID NO: 11	TCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 12	S F L C F L F V L C I R S R P S H S W A *

Rice plant lines 8-4 and 8-5 bi-allelic mutations

FIG. 5

Chromosome 1 – 1 bp insertion (t)

Intron 6	Exon 7
SEQ ID NO: 13	gGGCAAGGAGCACAGCAtAGCAGGTAAcCTGCCCTGTGCCGGAAgGTtGCACCCGACCCCTGTTGCCAAAGAACTT
SEQ ID NO: 8	<u>Q G A Q</u> H S R *
SEQ ID NO: 13	CGACCCAAcGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

Chromosome 2 – 1 bp insertion (a)

Intron 6	Exon 7
SEQ ID NO: 14	gGGCAAGGAGCACAGCAaAGCAGGTAAcCTGCCCTGTGCCGGAAgGTtGCACCCGACCCCTGTTGCCAAAGAACTT
SEQ ID NO: 15	<u>Q G A Q Q</u> S R *
SEQ ID NO: 14	CGACCCAAcGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

FIG. 6

Rice plant lines 8-6 bi-allelic mutations

Intron 6	Exon 7	Chromosome 1 – 4 bp deletion
SEQ ID NO: 5	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGACCCAA	
SEQ ID NO: 6		<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T Q
SEQ ID NO: 5	CGCGAGGAGCGACGCGCAGCTGCCCTTTACACCTTTAAAGCAGGCCGGCTTGACGGCTTGCTATTAAATTATTTTC	
SEQ ID NO: 6		R R G A T T A A A F T P F K Q A G L T L A I N Y F
SEQ ID NO: 5	TCCTTTTATGTTTCTCTTTGTTCTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA	
SEQ ID NO: 6		S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 2 bp deletion and 3 bp insertion (cat)

Intron 6	Exon 7
SEQ ID NO: 16	gGGCAAGGAGCACAGCACatCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCC
SEQ ID NO: 17	
SEQ ID NO: 16	CAAGAACTTCGACCCCAACGGCGAGGAGCGACGACGGCAGCTGCCTTTACACCTTTTAA

Rice plant lines 8-7 and 8-8 bi-allelic mutations **FIG. 7**

Intron 6	Exon 7	Chromosome 1 – 4 bp deletion
SEQ ID NO: 5		gGGCAAGGAGCACAGCAGGTAACCTGCCTGTGCCGGAAGTTGCACCGACCCCTGTGCCAAGAACTTCGACCCAA
SEQ ID NO: 6		<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T Q
SEQ ID NO: 5		CGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCCGGCTTGACGCTTGCTATTAAATTATTC
SEQ ID NO: 6		R R G A T T A A A F T P F K Q A G L T L A I N Y F
SEQ ID NO: 5		TCCTTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCAATGA
SEQ ID NO: 6		S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 1 bp insertion (c)

Intron 6	Exon 7
SEQ ID NO: 7	gGGCAAGGAGCACAGCAcAGCAGGTAACCTGCCTGTGCCGGAAGTTGCACCGA
SEQ ID NO: 8	
SEQ ID NO: 7	CCCTGTTGCCAAGAACTTCGACCCAACGGCGAGGAGCGACGCGCAGCTGCCTTTACACCTTTTAA

FIG. 8

Rice plant lines 20-1, 20-2 and 20-3 bi-allelic mutations

Chromosome 1 – 1 bp insertion (c) *

Intron 6	Exon 7
SEQ ID NO: 7	gGGCAAGGAGCACAGCAcAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGA
SEQ ID NO: 8	<u>Q G A Q</u> H S R *
SEQ ID NO: 7	CCCAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

Chromosome 2 – 5 bp deletion with a 3 bp insertion (agt) \$

Intron 6	Exon 7
SEQ ID NO: 18	dGGCAAGGAGCACAGAGagtAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGA
SEQ ID NO: 19	<u>Q G A Q</u> S R *
SEQ ID NO: 18	CCCAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

- * May be a g insertion instead of a c
- \$ May be a 3 bp deletion with a 1 bp insertion

FIG. 9

Rice plant line 1 bi-allelic mutations

Chromosome 1 – 4 bp deletion

Intron 6	Exon 7
SEQ ID NO: 5	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 6	<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T
SEQ ID NO: 5	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCCGGCTTGACGCTTGCTATTAAAT
SEQ ID NO: 6	Q R G A T T A A A F T P F K Q A G L T L A I N
SEQ ID NO: 5	TATTTCTCCTTTTATGTTTTCCTTTGTTCTGTATCAGATCGCGCCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 6	Y F S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 1 bp insertion (a)

Intron 6	Exon 7
SEQ ID NO: 14	gGGCAAGGAGCACAGCAaAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGA
SEQ ID NO: 15	<u>Q G A Q Q</u> S R *
SEQ ID NO: 14	CCCAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

FIG. 10

Rice plant line 6-1 and 6-2 bi-allelic mutations

Chromosome 1 – 1 bp insertion (a)

Intron 6	Exon 7
SEQ ID NO: 14	gGGCAAGGAGCACAGCAaAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGA
SEQ ID NO: 15	<u>Q G A Q Q</u> S R *
SEQ ID NO: 14	CCCAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

Chromosome 2 – 1 bp insertion (t)

Intron 6	Exon 7
SEQ ID NO: 13	gGGCAAGGAGCACAGCA ^t AGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAA
SEQ ID NO: 8	<u>Q G A Q H S R</u> *
SEQ ID NO: 13	CTTCGACCCCAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

FIG. 11

Rice plant line 6-3 and 6-4 bi-allelic mutations

Chromosome 1 – 4 bp deletion

Intron 6	Exon 7
SEQ ID NO: 5	gGGCAAGGAGCACAGCAGGTAAcCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGACCCCAACGG
SEQ ID NO: 6	<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T Q R
SEQ ID NO: 5	CGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCCGGCTTGACGCTTGCTATTAAATTATTCTCCTTT
SEQ ID NO: 6	R G A T A A A F T P F K Q A G L T L A I N Y F S F
SEQ ID NO: 5	TTATGTTTTCTCTTTGTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 6	L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 1 bp insertion (c)

Intron 6	Exon 7
SEQ ID NO: 7	gGGCAAGGAGCACAGCAcAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGACCC
SEQ ID NO: 8	<u>Q G A Q</u> H S R *
SEQ ID NO: 7	AACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAA

FIG. 12

Rice plant line 8-10 bi-allelic mutations

Chromosome 1 – 4 bp deletion

Intron 6	Exon 7
SEQ ID NO: 5	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 6	<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T
SEQ ID NO: 5	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTACACCTTTTAAGCAGGCCGGCTTGACGCTTGCTATTAAAT
SEQ ID NO: 6	Q R R G A T T A A A F T P F K Q A G L T L A I N
SEQ ID NO: 5	TATTTCCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCCAAGCCATAGCTGGGGCATGA
SEQ ID NO: 6	Y F S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 2 bp deletion

Intron 6	Exon 7
SEQ ID NO: 20	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGACCCAAC
SEQ ID NO: 19	<u>Q G A Q Q</u> S R *
SEQ ID NO: 20	GGCGAGGAGCGACGACGGCAGCTGCCCTTACACCTTTTAA

Rice plant line 9-1 bi-allelic mutations

FIG. 13

Chromosome 1 – 4 bp deletion

Intron 6	Exon 7
SEQ ID NO: 5	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGTTGCACCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 6	<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T
SEQ ID NO: 5	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCCGGCTTGACGCTTGCTATTAAT
SEQ ID NO: 6	Q R G A T T A A A F T P F K Q A G L T L A I N
SEQ ID NO: 5	TATTTCTCCTTTTATGTTTCTCTTGTGTATCAGATCGCGCCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 6	Y F S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 4 bp deletion

Intron 6	Exon 7
SEQ ID NO: 5	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGTTGCACCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 6	<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T
SEQ ID NO: 5	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCCGGCTTGACGCTTGCTATTAAT
SEQ ID NO: 6	Q R R G A T T A A A F T P F K Q A G L T L A I N
SEQ ID NO: 5	TATTTCTCCTTTTATGTTTCTCTTGTGTATCAGATCGCGCCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 6	Y F S F L C F L F V L C I R S R P S H S W A *

Rice plant line 9-2 bi-allelic mutations

FIG. 14

Chromosome 1 – 1 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 21	gGGCAAGGAGCACAGCAGGTAACCTGCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 22	<u>Q G A Q Q</u> Q V T C L C R K V A P T L L P R T S T
SEQ ID NO: 21	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCGCGCTTGACGCTTGCTATTAAATTAT
SEQ ID NO: 22	Q R R G A T T A A A F T P F K Q A G L T L A I N Y
SEQ ID NO: 21	TTCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 22	F S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 5 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 23	gGGCAAGGAGCAAGCAGGTAACCTGCCTGTGCCGGAAGGTTGCACCGACCCCTGTTGCCAAGAACTTCGACCC
SEQ ID NO: 24	<u>Q G A S R</u> *
SEQ ID NO: 23	AACGGCGAGGAGCGACGACGGCAGCTGCCTTTACACCTTTTAA

Rice plant line 9-3 bi-allelic mutations

FIG. 15

Chromosome 1 – 1 bp deletion

Intron 6	Exon 7
SEQ ID NO: 21	gGGCAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 22	<u>Q G A Q Q</u> Q V T C L C R K V A P T L L P R T S T
SEQ ID NO: 21	CAACGGCGAGGAGCGACGCGCAGCTGCCCTTTACACCTTTTAAGCAGGCCGGCTTGACGGCTTGCTATTAAATTAT
SEQ ID NO: 22	Q R R G A T T A A A F T P F K Q A G L T L A I N Y
SEQ ID NO: 21	TTCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 22	F S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 1 bp insertion (g)

Intron 6	Exon 7
SEQ ID NO: 25	gGGCAAGGAGCACAGCAGCagAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAA
SEQ ID NO: 15	<u>Q G A Q Q</u> S R *
SEQ ID NO: 25	CTTCGACCCCAACGGCGAGGAGCGACGACGGCAGCTGCCCTTACACCTTTTAA

Rice plant lines 13-1 and 13-2 bi-allelic mutations **FIG. 16**

Chromosome 1 – 1 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 21	gGGCAAGGAGCACAGCAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 22	<u>Q G A Q Q</u> Q V T C L C R K V A P T L L P R T S T
SEQ ID NO: 21	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTACACCTTTAAAGCAGCGCGGCTTGACGCTTGCTATTAAATTAT
SEQ ID NO: 22	Q R G A T T A A A F T P F K Q A G L T L A I N Y
SEQ ID NO: 21	TTCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 22	F S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 1 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 21	gGGCAAGGAGCACAGCAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 22	<u>Q G A Q Q</u> Q V T C L C R K V A P T L L P R T S T
SEQ ID NO: 21	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTACACCTTTAAAGCAGCGCGGCTTGACGCTTGCTATTAAATTAT
SEQ ID NO: 22	Q R R G A T T A A A F T P F K Q A G L T L A I N Y
SEQ ID NO: 21	TTCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 22	F S F L C F L F V L C I R S R P S H S W A *

Rice plant line 16 bi-allelic mutations

FIG. 17

Chromosome 1 – 1 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 21	gGGCAAGGAGCACAGCAGCAGGTAACCTGCCTGTGCCGGAAGTTGCACCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 22	<u>Q G A Q Q</u> Q V T C L C R K V A P T L L P R T S T
SEQ ID NO: 21	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCGCGGCTTGACGCTTGCTATTAAATTAT
SEQ ID NO: 22	Q R R G A T T A A A F T P F K Q A G L T L A I N Y
SEQ ID NO: 21	TTCTCCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCAATGA
SEQ ID NO: 22	F S F L C F L F V L C I R S R P S H S W A *

Chromosome 2 – 1 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 21	gGGCAAGGAGCACAGCAGCAGGTAACCTGCCCTGTGCCGGAAGTTGCACCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 22	<u>Q G A Q Q</u> Q V T C L C R K V A P T L L P R T S T
SEQ ID NO: 21	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCGCGGCTTGACGCTTGCTATTAAATTAT
SEQ ID NO: 22	Q R R G A T T A A A F T P F K Q A G L T L A I N Y
SEQ ID NO: 21	TTCTCCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCAATGA
SEQ ID NO: 22	F S F L C F L F V L C I R S R P S H S W A *

Rice plant line 17 bi-allelic mutations

Chromosome 1 – 1 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 21	gGGCAAGGAGCACAGCAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTGCCAAGAACTTCGACC
SEQ ID NO: 22	<u>Q G A Q Q</u> Q Q V T C L C R K V A P T L L P R T S T
SEQ ID NO: 21	CAACGGCGAGGAGCGACGACGCGCAGCTGCCCTTTACACCTTTTAAGCAGGCCGGCTTGACGCTTGCTATTAAATTAT
SEQ ID NO: 22	Q R R G A T A A A F T P F K Q A G L T L A I N Y
SEQ ID NO: 21	TTCTCCCTTTTATGTTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 22	F S F L C F L F V L C I R S R P S H S W A *
Chromosome 2 – 9 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 26	gGGCAAGGAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTGCCCCAAG
SEQ ID NO: 27	<u>Q G A G N L P V P E G C T D P V A K</u>
SEQ ID NO: 26	AACTTCGACCCAAACGGCGAGGAGCGACGACGCGCAGCTGCCCTTTACACCTTTTAA
SEQ ID NO: 27	<u>N F D P T A R S D D G S C L Y T F *</u>

FIG. 19

Rice plant lines 21-1 and 21-2 bi-allelic mutations

Chromosome 1 – 16 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 28	gGGCAGGTAACCTGCCTGTGCCGGAAGTTGCACCGACCCCTGTTGCCAAGAACTTCGACCCCAACGG
SEQ ID NO: 29	<u>Q</u> V T C L C R K V A P T L L P R T S T Q R
SEQ ID NO: 28	CGAGGAGCGACGACGCTGCCTTTACACCTTTAAGCAGGCGCGCTTGACGCTTGCTATTAAAT
SEQ ID NO: 29	R G A T T A A A F T P F K Q A G L T L A I N
SEQ ID NO: 28	TATTTCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 29	Y F S F L C F L F V L C I R S R P S H S W A *
Chromosome 2 – 4 bp deletion	
Intron 6	Exon 7
SEQ ID NO: 5	gGGCAAAGGAGCACAGCAGGTAACCTGCCCTGTGCCGGAAGGTTGCACCCGACCCCTGTTGCCAAGAACTTCGACC
SEQ ID NO: 6	<u>Q G A Q Q</u> V T C L C R K V A P T L L P R T S T
SEQ ID NO: 5	CAACGGCGAGGAGCGACGACGGCAGCTGCCCTTTACACCTTTTAAGCAGGCGCGCTTGACGCTTGCTATTAAAT
SEQ ID NO: 6	Q R R G A T T A A A F T P F K Q A G L T L A I N
SEQ ID NO: 5	TATTTCTCCTTTTATGTTTCTCTTTGTTCTGTGTATCAGATCGCGCCCAAGCCATAGCTGGGCATGA
SEQ ID NO: 6	Y F S F L C F L F V L C I R S R P S H S W A *

Confirmation at the RCA protein level

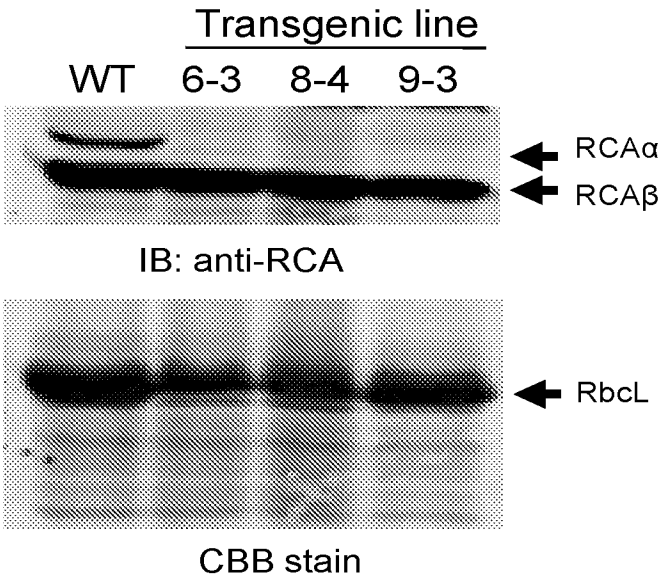


FIG. 20A

Robust plant growth

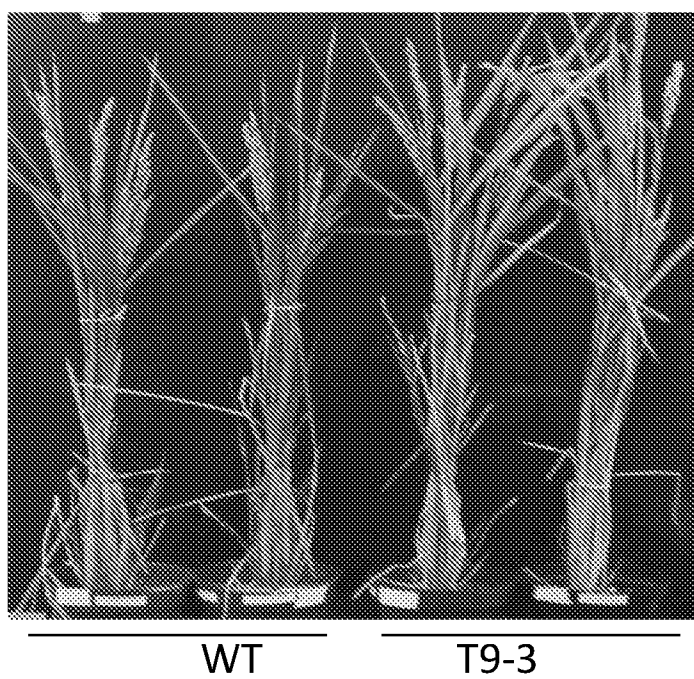


Fig. 20B

Faster Photosynthetic Induction in Rice with truncated RCA

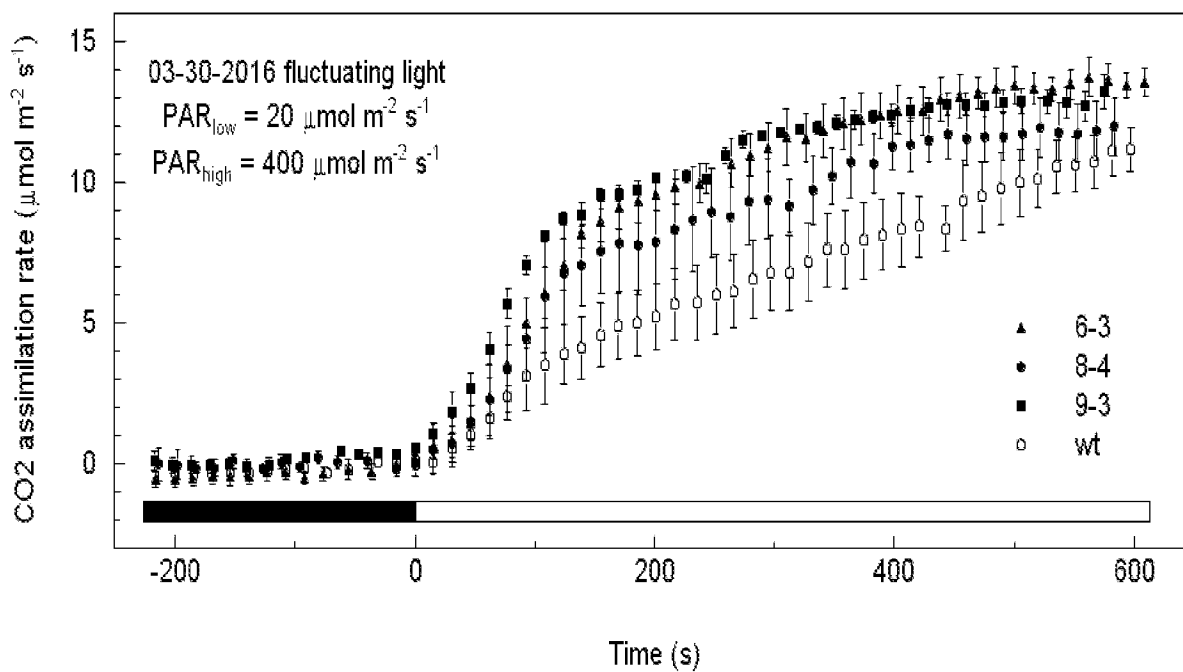


Fig. 20C

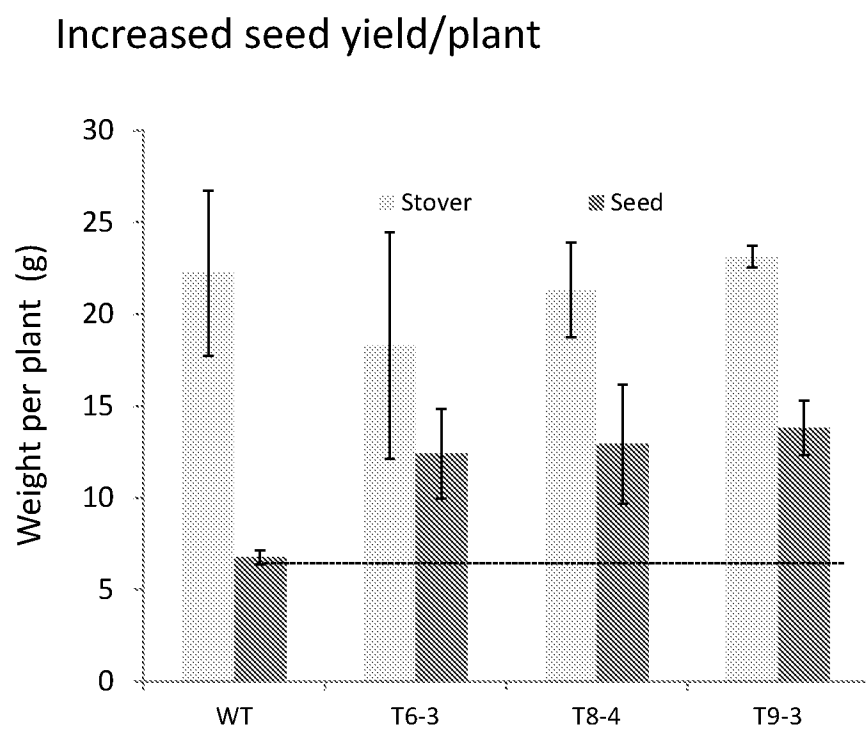


FIG. 20D

FIG. 21

<i>Brassica napus</i>	QYLNEAALGDANADAIGRGTFYGKAAQQVNI PVPEGCTDPQADNFDPTARSDDGTCVYNF	(SEQ ID NO: 45)
<i>Camelina sativa</i>	TYLSQAALGDANADAIDRGTFYGKAAQQVNI PVPEGCTDPAAQNFDPTARSDDGTCVYNF	(SEQ ID NO: 46)
<i>Arachis duranensis</i>	KYLNEAALGDANQDAIKSGSFYGKAAQQIGVPVPEGCTDPNASNFDPTARSDDGTCLYTF	(SEQ ID NO: 47)
<i>Vigna angularis</i>	KYLNEAALGDANEDSIKRGTFYGKAAQQINVPVPEGCTDPNAENFDPTARSDDGTCLYTD	(SEQ ID NO: 48)
<i>Cucumis melo</i>	KYLSEAAALGDANEDAIKRGTFYGKAAQQVSLPIPEGCTDPNADNFDPTARSDDGSCNYVF	(SEQ ID NO: 49)
<i>Cucumis sativus</i>	KYLSEAAALGDANEDAIKSGTFYGKAAQQVSLPIPEGCTDPNADNFDPTARSDDGSCNYVL	(SEQ ID NO: 50)
<i>Fragaria vesca</i>	QYLSEAAALGDANADAIDRGTFYGKAAQQVRVPVPEGCTDPTANNFDPTARSDDGTCLYTF	(SEQ ID NO: 51)
subsp. <i>vesca</i>		
<i>Gossypium hirsutum</i>	KYLSEAAALGEANEDSINRGTFYGKAAQQVGVPVPEGCTDPNADNFDPTARSDDGTCYQF	(SEQ ID NO: 52)
<i>Theobroma cacao</i>	KYLSEAAALGDANEDSIKRGTFYGKAAQQIGVPVPEGCTDPGADNFDPTARSDDGSCYTK-	(SEQ ID NO: 53)
<i>Morus notabilis</i>	KYLSEAAALGDANQDSIKRGTFYGKAAQQINIPVPEGCTDPLAANFDPTARSDNGSCQY--	(SEQ ID NO: 54)
<i>Prunus mume</i>	KYLSDAALGDANQDAIKRGEFYGKAAQQIKVPVPEGCTDPSAANFDPTARSDNGSCQYE-	(SEQ ID NO: 55)
<i>Beta vulgaris</i>	KYLNEAALGDANKDSIDSGAFFGQAAQQGNIPIPEGCTDPNAKNFDPTARSDDGSCLYTF	(SEQ ID NO: 56)
subsp. <i>vulgaris</i>		
<i>Citrus clementina</i>	KYLSEAAALGDANDDAIKNGSFYGKAAQQVKVPVPEGCTDPAAANYDPTARSDDGSCNYQF	(SEQ ID NO: 57)
<i>Glycine max</i>	KYLNEAALGDANDDAIKTGNFYGQAAQQVHVVPVPEGCTDPTAENYDPTARSDDGSCYKF	(SEQ ID NO: 58)
<i>Malus domestica</i>	KYLKEAALGDANDDAIKSGSFYGKAAQQVNLVPVPEGCTDPSAANYDPTARSDDGSCLYQF	(SEQ ID NO: 59)
<i>Prunus persica</i>	KYLKEAALGDANDDAIKSGNFYGKAAQLVNLVPVPEGCTDPSAANYDPTARSDNGSCLYQF	(SEQ ID NO: 60)
<i>Vitis vinifera</i>	KYLSEAAALGDANEDAIKSGSFYGKAAQQVNLVPVPEGCTDPSANNFDPTARSDDGSCLYQI	(SEQ ID NO: 61)
<i>Musa acuminata</i>	KYLSEAAALGDANEDAINTGQFYGKAAQQVGVPVPEGCTDPIATNFDPTARSDDGSCLYPV	(SEQ ID NO: 62)
subsp. <i>malaccensis</i>		
<i>Zea mays</i>	KYLSEAAALGDANDD---DDLYGKAAQHVRVPVPEGCTDPKAGNFDPTARSDDGSCVYN-	(SEQ ID NO: 63)

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/040859**A. CLASSIFICATION OF SUBJECT MATTER****C12N 9/88(2006.01)i, C12N 15/82(2006.01)i**

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)

C12N 9/88; C12N 15/82; A01H 5/00; C12N 15/87

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords:ribulose 1,5-bisphosphate carboxylase/oxygenase activase, light, cysteine, redox active site

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHANG, NING et al., 'Light modulation of Rubisco in Arabidopsis requires a capacity for redox regulation of the larger Rubisco activase isoform' Proceedings of the National Academy of Sciences, 5 March 2002, Vol. 99, No. 5, pp. 3330-3334 See abstract; page 3330, right column, 2nd paragraph; page 3331, left column, 1st paragraph; page 3333, right column, 1st paragraph; page 3334, left column, 3rd paragraph-right column, 1st paragraph; figure 3.	1-3,7-10
Y A		5,12-14,16-19 4,6,11,15
Y	WANG, DAFU et al., 'Increased sensitivity of oxidized large isoform of ribulose-1, 5-bisphosphate carboxylase/oxygenase (Rubisco) activase to ADP inhibition is due to an interaction between its carboxyl extension and nucleotide-binding pocket', Journal of Biological Chemistry, 1 September 2006, Vol. 281, No. 35, pp. 25241-25249 See abstract; page 25244, right column, 3rd paragraph; page 25245, right column, 1st paragraph.	5
Y	US 2004-0241860 A1 (KASTER, JR., LARRY V. et al.) 02 December 2004 See claims 14 and 19.	12-14,16-19



Further documents are listed in the continuation of Box C.



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

"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

18 October 2017 (18.10.2017)

Date of mailing of the international search report

18 October 2017 (18.10.2017)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

KIM, Sun Hee

Telephone No. +82-42-481-5405



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/040859

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MORENO, JOAQUIN et al., 'Redox modulation of Rubisco conformation and activity through its cysteine residues', Journal of Experimental Botany, Epub. 22 January 2008, Vol. 59, No. 7, pp. 1605-1614 See the whole document.	1-19
A	US 2014-0182015 A1 (BROEKAERT, WILLEM et al.) 26 June 2014 See claims 1-2; paragraphs [0011], [0027]-[0032] and [0562]-[0565].	1-19
A	NCBI, GenBank accession no. ABG22615.1 (05 May 2011) See the whole document.	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2017/040859

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2004-0241860 A1	02/12/2004	AT 429807 T	15/05/2009
		AU 4686199 A	05/01/2000
		CA 2335142 A1	23/12/1999
		CA 2335142 C	10/04/2012
		DK 1098965 T3	03/08/2009
		EP 1098965 A1	16/05/2001
		EP 1098965 A4	23/04/2003
		EP 1098965 B1	29/04/2009
		EP 1098965 B8	19/08/2009
		EP 2083075 A1	29/07/2009
		ES 2325101 T3	25/08/2009
		PT 1098965 E	15/07/2009
		US 6750377 B1	15/06/2004
		WO 99-66034 A1	23/12/1999
US 2014-0182015 A1	26/06/2014	AR 064512 A1	08/04/2009
		BR PI0721100 A2	11/02/2014
		CA 2671401 A1	26/06/2008
		CN 101583720 A	18/11/2009
		CN 102676575 A	19/09/2012
		CN 102676575 B	30/09/2015
		EP 2069505 A2	17/06/2009
		EP 2395091 A2	14/12/2011
		EP 2395091 A3	01/08/2012
		EP 2395092 A2	14/12/2011
		EP 2395092 A3	01/08/2012
		EP 2395093 A2	14/12/2011
		EP 2395093 A3	01/08/2012
		EP 2405011 A2	11/01/2012
		EP 2405011 A3	25/07/2012
		MX 2009005524 A	08/06/2009
		US 2010-0077502 A1	25/03/2010
		US 8569575 B2	29/10/2013
		US 9428761 B2	30/08/2016
		WO 2008-074891 A2	26/06/2008
		WO 2008-074891 A3	04/09/2008