



(51) International Patent Classification:

C07K 14/415 (2006.01) *C12N 15/82* (2006.01)
C12N 15/29 (2006.01) *A01H 5/00* (2006.01)
C12N 15/113 (2010.01)

(21) International Application Number:

PCT/CN2014/083982

(22) International Filing Date:

8 August 2014 (08.08.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

201310343713.3 8 August 2013 (08.08.2013) CN

(71) Applicant: INSTITUTE OF GENETICS AND DEVELOPMENTAL BIOLOGY, CHINESE ACADEMY OF SCIENCES [CN/CN]; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN).

(72) Inventors: **CHU, Chengcai**; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN). **LIU, Linchuan**; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN). **TONG, Hongning**; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN). **HU, Bin**; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN). **LIANG, Chengzhen**; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN). **CHE, Ronghui**; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN). **XU, Fan**; No. 1 West Beichen Road, Chaoyang District, Beijing 100101 (CN).

(74) Agent: NTD PATENT & TRADEMARK AGENCY LIMITED; 10th Floor, Block A, Investment Plaza, 27 Jinrongdajie, Xicheng District, Beijing 100033 (CN).

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

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

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: BG1 COMPOSITIONS AND METHODS TO INCREASE AGRONOMIC PERFORMANCE OF PLANTS

(57) Abstract: Methods and compositions that affect yield and other agronomic characteristics in plants are disclosed. Methods of transgenic modulation and marker-assisted breeding methods improve grain size and grain yield in rice are also disclosed. Increased expression of BG1 results in increased grain size and yield.



CROSS REFERENCE

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

FIELD

BACKGROUND

SUMMARY

1

(BG1) controls the elongation of both grains and internodes. In an embodiment, transgenic plants overexpressing BG1 had increased seed size and panicle length, whereas knockdown of BG1 by RNAi led to smaller seed size. BG1 can be induced by auxin and *Bg1-D* mutant showed increased sensibility to IAA and enhanced polar auxin transport, which altered endogenous IAA distribution. Use of BG1 in molecular breeding of rice and other plants are disclosed.

Methods of improving an agronomic characteristic of a plant, the method includes modulating the expression of (i) a polynucleotide encoding an amino acid sequence comprising SEQ ID NO: 1 or an amino acid sequence that is at least 95% identical to one of SEQ ID NO: 1 (ii) a polynucleotide that hybridizes under stringent hybridization conditions to a fragment of polynucleotide comprising SEQ ID NO: 2, wherein the fragment comprises at least 100 contiguous nucleotides of SEQ ID NO: 2 (iii) a polynucleotide that encodes an amino acid sequence that is at least 90% identical to SEQ ID NO: 1, (iv) a polynucleotide encoding a polypeptide comprising one or more deletions or insertions or substitutions of amino acids compared to SEQ ID NO: 1.

In an embodiment, the expression of the polynucleotide encoding a polypeptide having at least 95% identity to SEQ ID NO: 1 is increased by transforming the plant with a recombinant polynucleotide operably linked to a heterologous promoter. In an embodiment, the expression of an endogenous polynucleotide encoding a polypeptide having at least 95% identity to SEQ ID NO: 1 is increased by upregulating a regulatory element operably associated with the endogenous polynucleotide. In an embodiment, the expression of the polynucleotide is increased by expressing the polynucleotide under a regulatory element of SEQ ID NO: 3.

In an embodiment, the agronomic characteristic is selected from the group consisting of (i) an increase in grain size, (ii) an increase in grain weight, (iii) an increase in panicle length, (iv) an increase in grain yield (v) an increase in grain filling rate, (vi) an increase in biomass. The increase in agronomic characteristic is measured with respect to a control plant that does not exhibit elevated levels of BG1 (or a variant or an ortholog/homolog thereof). In an embodiment, the agronomic performance is an increase in plant biomass. In an embodiment, the grain weight is increased in relation to a control plant not having an increased expression of the polynucleotide.

In an embodiment, the plant is a monocot. In an embodiment, the plant is rice or maize. In an embodiment, the plant is a dicot. In an embodiment, the plant is soybean.

In an embodiment, methods of improving yield of a plant include increasing the expression of a polynucleotide that encodes a polypeptide comprising the amino acid

sequence selected from the group consisting of SEQ ID NOS: 1, 4-27 or an allelic variant thereof.

In an embodiment, methods of improving rice grain yield include the expression of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a variant thereof.

In an embodiment, methods of marker assisted selection of a plant or identifying a native trait associated with increased yield, include:

a. performing marker-assisted selection of plants that have one or more variations in genomic regions encoding a protein comprising SEQ ID NO: 1 or a variant thereof or a regulatory sequence thereof; and

b. identifying the plant that exhibits higher yield.

In an embodiment, methods of identifying one or more alleles in a population of rice plants that are associated with increased grain yield includes:

a. evaluating in a population of rice plants for one or more allelic variations in (i) a genomic region, the genomic region encoding a polypeptide or (ii) the regulatory region controlling the expression of the polypeptide, wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 1 or a sequence that is 95% identical to SEQ ID NO: 1;

b. obtaining phenotypic values of increased yield for the one or more rice plants in the population;

c. associating the allelic variations in the genomic region with the phenotype; and

d. identifying the one or more alleles that are associated with increased yield.

In an embodiment, an isolated polynucleotide includes for example, (i) encoding an amino acid sequence comprising one of SEQ ID NOS: 1, 4-27 or an amino acid sequence that is at least 95% identical to one of SEQ ID NOS: 1, 4-27 (ii) hybridizing under stringent hybridization conditions to a fragment of polynucleotide selected from the group consisting of SEQ ID NO: 2, wherein the fragment comprises at least 100 contiguous nucleotides of SEQ ID NO: 2 (iii) that encodes an amino acid sequence that is at least 90% identical to SEQ ID NO: 1, (iv) a polynucleotide encoding a polypeptide comprising one or more deletions or insertions or substitution of amino acids compared to SEQ ID NO: 1, wherein the polynucleotide encodes a polypeptide involved in the regulation of grain width, weight or yield.

In an embodiment, a recombinant expression cassette includes the polynucleotide that is operably linked to a regulatory element, wherein the expression

cassette is functional in a plant cell. In an embodiment, a host cell includes the expression cassette. In an embodiment, a transgenic plant includes the recombinant expression cassette.

In an embodiment, a transgenic plant part includes a plant regulatory element that operably regulates the expression of a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a variant or an ortholog thereof, wherein the regulatory element is heterologous to the polynucleotide.

In an embodiment, the polynucleotide that regulates the expression comprises a sequence selected from the group consisting of SEQ ID NOS: 7-8 or a functional promoter fragment thereof.

In an embodiment, methods of marker assisted selection of a maize plant for improved yield, includes:

a. detecting one or more genetic variations in a genomic region comprising a polynucleotide encoding a protein selected from the group consisting of SEQ ID NOS: 12-14 or a variant thereof or in a regulatory sequence controlling the expression of the polynucleotide thereof, in a plurality of maize plants; and

b. associating the one or more variations with increased yield and thereby selecting the maize plant that has the one or more variations associated with higher yield.

In an embodiment, methods of identifying one or more alleles associated with increased yield in a population of maize plants, the method comprising:

a. evaluating in a population of maize plants one or more genetic variations in (i) a genomic region encoding a polypeptide or (ii) a regulatory region controlling the expression of the polypeptide, wherein the polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NOS: 12-14 or a sequence that is 95% identical to SEQ ID NOS: 12-14;

b. obtaining yield data for one or more maize plants in the population;

c. associating the one or more genetic variations in the genomic region encoding the polypeptide or in the regulatory region controlling the expression of the polypeptide with yield, thereby identifying one or more alleles associated with increased yield.

In an embodiment, the one or more genetic variations is in the coding region of the polynucleotide. In an embodiment, the regulatory region is a promoter element. In an embodiment, the yield is grain yield

A method of improving grain width, weight or yield of a plant includes the steps of modulating the expression of (i) a polynucleotide encoding an amino acid sequence

comprising one of SEQ ID NOS: 1, 4-27 or an amino acid sequence that is at least 95% identical to one of SEQ ID NOS: 1, 4-27 (ii) a polynucleotide that hybridizes under stringent hybridization conditions to a fragment of a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NOS: 1, 4-27, wherein the
5 fragment comprises at least 100 contiguous nucleotides of a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NOS: 1, 4-27 (iii) a polynucleotide that encodes an amino acid sequence that is at least 90% identical to SEQ ID NO: 1, (iv) a polynucleotide encoding a polypeptide comprising one or more deletions or insertions or substitution of amino acids compared to SEQ ID NO: 1,
10 wherein the polynucleotide encodes a polypeptide involved in the regulation of grain width, weight or yield.

In an embodiment, the polynucleotide that comprises a fragment of SEQ ID NO: 2, is sufficient to up-regulate the endogenous expression of the polynucleotide that encodes a polypeptide. In an embodiment, the modulation of the expression is
15 achieved through RNA interference.

In an embodiment, the modulation of the expression is achieved through mutagenesis. In an embodiment, the modulation of the expression is achieved through microRNA mediated gene silencing. In an embodiment, the modulation of the expression is achieved through promoter-mediated gene suppression. In an
20 embodiment, the modulation of the expression is achieved through targeted mutagenesis of an endogenous regulatory element.

In an embodiment, the grain weight is increased in relation to a control plant not expressing the polynucleotide.

In an embodiment, the modulation of expression is performed in a monocot. In
25 an embodiment, the plant is rice. In an embodiment, the plant is a dicot.

A method of improving the length to width aspect of rice grain includes the steps of modulating the expression of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or an allelic variant thereof.

A method of improving rice grain yield, the method includes modulating the
30 expression of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a variant thereof.

A method of marker assisted selection of a plant for improved grain size or yield, the method includes the steps of: performing marker-assisted selection of plants that have one or more variations in genomic regions encoding a protein comprising SEQ ID
35 NO: 1 or a variant thereof or a regulatory sequence thereof; and identifying the plant that produces grains with improved grain quality and/or higher yield.

A method of identifying alleles in rice plants or rice germplasm that are associated with improved grain quality and/or increased yield, the method includes the steps of obtaining a population of rice plants, wherein one or more plants exhibit improved grain quality and/or increased yield; evaluating allelic variations with respect to the polynucleotide sequence encoding a protein comprising a polypeptide comprising: SEQ ID NO: 1 or in the genomic region that regulates the expression of the polynucleotide encoding the protein; obtaining phenotypic values of improved grain quality and/or increased yield for a plurality of rice plants in the population; associating the allelic variations in the genomic region associated with the polynucleotide with the phenotype; and identifying the alleles that are associated with improved grain quality and/or increased yield.

In an embodiment, the polynucleotides described herein encode a polypeptide involved in the regulation of grain width, weight or yield.

In an embodiment, a host cell includes the recombinant polynucleotides disclosed herein. In an embodiment, a transgenic plant includes the recombinant expression cassette disclosed herein.

An isolated polynucleotide that operably regulates the expression of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a variant thereof. In an embodiment, the polynucleotide that regulates the expression comprises a sequence of SEQ ID NO: 3 or a functional promoter fragment thereof.

Table 1 Sequence Description

The sequence descriptions and Sequence Listing attached hereto, and incorporated herein by reference, comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical J.* 219(2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NO:	Polynucleotide /polypeptide	Identity	Description
SEQ ID NO: 1	polypeptide	OsBG1 protein	Oryza sativa
SEQ ID NO: 2	polynucleotide	OsBG1- cDNA	Oryza sativa

SEQ ID NO: 3	polynucleotide	OsBG1 promoter	Oryza sativa
SEQ ID NO: 4	polypeptide	AT1G54200	Arabidopsis thaliana
SEQ ID NO: 5	polypeptide	AT3G13980	Arabidopsis thaliana
SEQ ID NO: 6	polypeptide	AT5G12050	Arabidopsis thaliana
SEQ ID NO: 7	polypeptide	AT3G42800	Arabidopsis thaliana
SEQ ID NO: 8	polypeptide	Bra037972	Brassica rapa
SEQ ID NO: 1	polypeptide	Bra030898	Brassica rapa
SEQ ID NO: 10	polypeptide	Os03g0175800	Oryza sativa
SEQ ID NO: 11	polypeptide	Os10g0397600	Oryza sativa
SEQ ID NO: 12	polypeptide	GRMZM2G007134	Zea mays
SEQ ID NO: 13	polypeptide	GRMZM2G438606	Zea mays
SEQ ID NO: 14	polypeptide	GRMZM2G178852	Zea mays
SEQ ID NO: 15	polypeptide	Sb01g023080	Sorghum bicolor
SEQ ID NO: 16	polypeptide	Sb01g045450	Sorghum bicolor
SEQ ID NO: 17	polypeptide	Bradi1g72920	Brachypodium distachyon
SEQ ID NO: 18	polypeptide	Medtr5g049030	Medicago truncatula
SEQ ID NO: 19	polypeptide	Medtr3g139450	Medicago truncatula
SEQ ID NO: 20	polypeptide	POPTR_0003s06210	Populus trichocarpa
SEQ ID NO: 21	polypeptide	POPTR_0001s17050	Populus trichocarpa
SEQ ID NO: 22	polypeptide	Glyma04g05780	Glycine max
SEQ ID NO: 23	polypeptide	Glyma07g04110	Glycine max
SEQ ID NO: 24	polypeptide	Glyma16g00820	Glycine max
SEQ ID NO: 25	polypeptide	Si036603m	Setaria italica
SEQ ID NO: 26	polypeptide	Si036624m	Setaria italica
SEQ ID NO: 27	polypeptide	GSVIVG0101678000 1	Vitis vinifera
SEQ ID NO: 28	oligo	SF1	Site-Finding PCR primer
SEQ ID NO: 29	oligo	SFP1	Site-Finding PCR primer
SEQ ID NO: 30	oligo	SFP2	Site-Finding PCR primer
SEQ ID NO: 31	oligo	GSP1	Site-Finding PCR primer
SEQ ID NO: 32	oligo	GSP2	Site-Finding PCR primer
SEQ ID NO: 33	oligo	GeneRacer 5' primer	RACE primer
SEQ ID NO: 34	oligo	GeneRacer RNA Oligo	RACE primer
SEQ ID NO: 35	oligo	Sequence	RACE primer
SEQ ID NO: 36	oligo	BG1-1R	RACE primer
SEQ ID NO: 37	oligo	BG1-2R	RACE primer
SEQ ID NO: 38	oligo	BG1-3R	RACE primer
SEQ ID NO: 39	oligo	BG1-F	Real-time PCR primer
SEQ ID NO: 40	oligo	BG1-R	Real-time PCR primer
SEQ ID NO: 41	oligo	Ubiquitin-F	Real-time PCR primer
SEQ ID NO: 42	oligo	Ubiquitin-R	Real-time PCR primer
SEQ ID NO: 43	oligo	Actin-F	Real-time PCR primer
SEQ ID NO: 44	oligo	Actin-R	Real-time PCR primer
SEQ ID NO: 45	oligo	CDKA1-F	Real-time PCR primer
SEQ ID NO: 46	oligo	CDKA1-R	Real-time PCR primer
SEQ ID NO: 47	oligo	E2F2-F	Real-time PCR primer
SEQ ID NO: 48	oligo	E2F2-R	Real-time PCR primer
SEQ ID NO: 49	oligo	H1-F	Real-time PCR primer

SEQ ID NO: 50	oligo	H1-R	Real-time PCR primer
SEQ ID NO: 51	oligo	CYCD1;1-F	Real-time PCR primer
SEQ ID NO: 52	oligo	CYCD1;1-R	Real-time PCR primer
SEQ ID NO: 53	oligo	CYCT1-F	Real-time PCR primer
SEQ ID NO: 54	oligo	CYCT1-R	Real-time PCR primer
SEQ ID NO: 55	oligo	OsEXPA5-F	Real-time PCR primer
SEQ ID NO: 56	oligo	OsEXPA5-R	Real-time PCR primer
SEQ ID NO: 57	oligo	OsEXPA10-F	Real-time PCR primer
SEQ ID NO: 58	oligo	OsEXPA10-R	Real-time PCR primer
SEQ ID NO: 59	oligo	OsEXPB3-F	Real-time PCR primer
SEQ ID NO: 60	oligo	OsEXPB3-R	Real-time PCR primer
SEQ ID NO: 61	oligo	OsEXPB4-F	Real-time PCR primer
SEQ ID NO: 62	oligo	OsEXPB4-R	Real-time PCR primer
SEQ ID NO: 63	oligo	OsEXPB7-F	Real-time PCR primer
SEQ ID NO: 64	oligo	OsEXPB7-R	Real-time PCR primer
SEQ ID NO: 65	oligo	BG1OX-F	Vector Primer
SEQ ID NO: 66	oligo	BG1OX-R	Vector Primer
SEQ ID NO: 67	oligo	BG1RNAi-F	Vector Primer
SEQ ID NO: 68	oligo	BG1RNAi-R	Vector Primer
SEQ ID NO: 69	oligo	BG1GUS-F	Vector Primer
SEQ ID NO: 70	oligo	BG1GUS-R	Vector Primer
SEQ ID NO: 71	oligo	BG1EGFP-F	Vector Primer
SEQ ID NO: 72	oligo	BG1EGFP-R	Vector Primer
SEQ ID NO: 73	oligo	BG1DNAOX-F	Vector Primer
SEQ ID NO: 74	oligo	BG1DNAOX-R	Vector Primer
SEQ ID NO: 75	oligo	BG1 RNAi fragment	dsRNA

In another aspect, the present disclosure relates to a recombinant expression cassette comprising a nucleic acid as described. Additionally, the present disclosure relates to a vector containing the recombinant expression cassette. Further, the vector containing the recombinant expression cassette can facilitate the transcription and translation of the nucleic acid in a host cell. The present disclosure also relates to the host cells able to express the polynucleotide of the present disclosure. A number of host cells could be used, such as but not limited to, microbial, mammalian, plant or insect.

In yet another embodiment, the present disclosure is directed to a transgenic plant or plant cells, containing the nucleic acids of the present disclosure. Preferred plants containing the polynucleotides of the present disclosure include but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, tomato and millet. In another embodiment, the transgenic plant is a maize plant or plant cells. Another embodiment is the transgenic seeds from the transgenic nitrate uptake-associated polypeptide of the disclosure operably linked to a promoter that

drives expression in the plant. The plants of the disclosure can have improved grain quality as compared to a control plant.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the phenotype of the *Bg1-D* mutant. (A). 10-day-old seedlings grown on 1/2 MS medium. Scale bar, 5 cm. (B and C). Growth phenotype of 2-month-old (B) and 4-month-old (C) plants grown in a paddy field. Scale bar, 10 cm. (D and E). The young panicles (D) and harvested panicles (E) showed an increased length for *Bg1-D*. Scale bar, 5 cm. (F). The mature seed showed an increased size for *Bg1-D*. Scale bar, 10 mm.

Fig. 2 shows the histological analyses of spikelet hulls before heading. (A). Spikelets of WT and *Bg1-D* plants. Scale bar, 2 mm. (B). Cross-sections of spikelet hulls. Dashed line indicates the position of cross-section. Scale bar, 0.5 mm. (C). Magnified view of the cross-section boxed in B. Scale bar, 0.1 mm. (D-F). Total area (D), cell number (E) and cell area (F) in the outer parenchyma layer of the spikelet hulls formed by WT and *Bg1-D* plants. All data are given as means $\pm$ S.D. (n=15). A Students t-test was used to generate the P values. **P<0.01.

Fig. 3 shows identification of *BG1* gene. (A). Schematic map of the T-DNA flanking region of the *Bg1-D* mutant. A candidate for the ORF responsible for the phenotype (Os03g0175800, which encodes a novel protein with unknown function, colored blue), referred to as *BG1*, is located 0.8-kb downstream of the T-DNA insertion. The rice Actin1 promoter is shown by the green arrow, LB and RB are left and right borders of the T-DNA. (B). The expression level of *BG1* mRNA in the wild-type and *Bg1-D* plants. Data were obtained from three independent replicates. Values are showed as means $\pm$ S.D.

Fig. 4 shows the phylogenetic analysis of *BG1*. MEGA4 Neighbor-Joining tree was inferred from the amino acid sequences of the *BG1* (Os03g0175800) homologs among green plants. Bootstrap values are based on 1000 replications and are indicated in their respective nodes. The scale bar indicates genetic distance based on branch length. SEQ ID NOS: 1, 4-27 represent the sequences shown in the phylogeny of *BG1*.

Fig. 5 shows the expression pattern of *BG1*. (A). Expression of *BG1* in various organs analyzed by quantitative RT-PCR analysis. Root, culm, leaf and sheath were harvested in 2-month-old wild-type (Nipponpare) plants. Different length of young panicles (YP) were used for analysis, number indicates the length in centimeters. Data were obtained from three independent replicates. (B). GUS staining of *PROBG1::GUS*

transgenic line tissues. (a) Microscopic observation of a cross-section of young culm. (b) young spikelet. (c) Developing hull. (d) Cross-section of the hull. (e) mature grain.

Fig. 6 shows increased grain size and weight in transgenic rice plants overexpressing BG1. (A). Gross morphology of wild-type (WT), OE-1, OE-2 and OE-3 transgenic plants. Scale bar, 20 cm. (B). Comparison of panicle length of the wild-type and transgenic plants is shown. Scale bar, 5 cm. (C). The grain morphologies of wild-type and transgenic plants are shown. Scale bar, 5 mm. (D) Relative expression level of BG1 mRNA in wild-type and transgenic plants. Data were obtained from three independent replicates. (E-F). Comparison of the grain length (E), grain width (F), and 1000-grain weight (G) of wild-type and transgenic plants are shown. (A-E). OE-1, OE-2, OE-3 indicate three representative transgenic lines of BG1 overexpression. (E-G) Values are showed as means $\pm$ S.D. (n=20 in E and F, n=5 in G). **P<0.01 (t-test).

Fig. 7 shows phenotypes BG1-RNAi plants (Ri). (A and B). Overall morphology of wild-type (WT) and BG1-RNAi plants are shown. 60-day-old (A), Scale bar, 10 cm. 100-day-old (B). Scale bar, 10 cm. (C) Comparison of panicle length of the wild-type and BG1-RNAi plants. Scale bar, 5 cm. (D). The grain morphologies of wild-type and BG1-RNAi plants are shown. Scale bar, 5 mm. (E) Relative expression level of BG1 mRNA in wild-type and BG1-RNAi plants. Data were obtained from three independent replicates. (F). Comparison of 1000-grain weight of the wild-type and BG1-RNAi plants. Values are showed as means $\pm$ S.D. (n=5). *P<0.05, **P<0.01 (t-test). (A-F). Ri-1, Ri-2 indicate two representative transgenic lines of BG1-RNAi plants.

Fig. 8 shows that BG1 localizes primarily to the plasma membrane. (A). Confocal micrographs of rice root showed plasma membrane targeting of BG1. Rice transformed with pCambia 35S::eGFP (upper panel) and pCambia 35S::BG1-eGFP (lower panel) plasmids are shown. Bar, 50 μ M. DIC indicates the differential interference contrast phase, eGFP indicates the enhanced green fluorescence of proteins, and merged indicates the merging of eGFP and DIC. (B). Immunoblot analysis of BG1 in the protein extracts of seedlings. (a). Microsomal pellet (P) and soluble (S) fraction were isolated in 7-day-old young seedlings. (b). Fractions of two-phase membrane partitioning, plasma membrane (P) and other membrane (O). PEPC (phosphoenolpyruvate carboxylase) was used as cytosolic control, LHCII and aquaporin PIP2;1 were used as other membrane and plasma membrane control.

Fig. 9 shows that BG1 is an auxin response gene. Relative expression level of BG1 under different phytohormone treatments. (B) Time course of BG1 response to 10 μ M IAA treatment. (C) Cycloheximide leads to increased expression level of BG1. (D)

BG1 response to different auxin compounds. (A-D) Data were obtained from three independent replicates. Bars represent means $\pm$ SD (n=3) (E) BG1 gene promote auxin-induced growth in the elongating leaf sheath.

Fig. 10 shows that Bg1-D displayed enhanced polar auxin transport (PAT).

Comparison of PAT between wild-type, Bg1-D and BG1-RNAi (Ri2) plants in dark-grown coleoptiles. Bars represent means $\pm$ SD (n=5). Asterisks indicate the significance of differences determined by Student's t-test: ** $0.001 \leq P < 0.01$.

Fig. 11 shows the expression level of genes involved in cell cycle and cell expansion in WT and Bg1-D. (A) Genes involved in cell cycle G1 to S phase were upregulated in Bg1-D. (B) Expansin gene involved in cell expansion were upregulated in Bg1-D. Bars represent means $\pm$ SD (n=3).

Fig. 12 shows increased grain yield in transgenic rice plants overexpressing BG1 by its native promoter. (A). Gross morphology of wild-type (WT), OE-5 and OE-7 transgenic plants. Scale bar, 20 cm. (B). The grain morphologies of wild-type and transgenic plants. Scale bar, 5 mm. (C). Comparison of the total panicles per plant harvested in the wild-type and transgenic plants. Scale bar, 5 cm. (D and E). Comparison of the 1000-grain weight (D) and yield per plant of wild-type and transgenic plants (E). (A-E). OE-5, OE-7 indicate two representative transgenic lines of BG1 overexpression driven by BG1 native promoter. (D-E) Values are showed as means $\pm$ S.D. (n=10 in D and n=25 in G). * $0.01 \leq P < 0.05$, ** $0.001 \leq P < 0.01$ (student's t-test).

Fig. 13 shows accelerated growth of Bg1-D seedlings. (A) 7-day-old seedlings of wild-type and Bg1-D plants. Scale bar, 2 cm. (B) Comparison of plant height between wild-type and Bg1-D plants after seed germination in a week. (C) Comparison of length of leaf sheath between wild-type and Bg1-D plants in 7-day-old seedlings. Data are given as mean $\pm$ S.E. (n=20).

Fig. 14 shows increased cell length in the spikelet hull of Bg1-D. (A) Close-up view of longitudinal sections of spikelet hull. Scale bar, 50 μ m. (B) Comparison of cell length of palea and lemma cell layer between wild-type and Bg1-D plants. Data are given as mean $\pm$ SE (n=15). A Student's t-test was used to generate the P values. ** $0.001 \leq P < 0.01$.

Fig. 15 demonstrates that Bg1-D showed increased grain filling rate. (A) Comparison of fresh weight of endosperm in WT and Bg1-D after fertilization. (B) Comparison of dry weight of endosperm in WT and Bg1-D after fertilization. Data are given as mean $\pm$ SE (n=40).

Fig. 16 shows phylogenetic and amino acid sequences analysis of BG1 in rice and Arabidopsis. (A) Phylogenetic relationships of BG1. (B) Sequence alignment of BG1 (SEQ ID NOS: 4, 5, 7, 10 or 1, and 11 in that order). Conserved amino acids are highlighted in shades of black and gray: white letters with black background (100% identity), white letters with gray background (80% identity), and black letters with gray background (60% identity). See also Global Identity (Table 2) and Global Similarity (Table 3) disclosed herein.

Fig. 17 shows that overexpression of OsBG1 leads to increased organ size in Arabidopsis. (A) Elongated hypocotyls of 5-day-old seedlings overexpressing OsBG1 (B-D) OsBG1 promotes increased leaf size (B), stem length (C) and seed size (D). 3# and 50# indicate two representative transgenic lines of OsBG1 overexpression in Arabidopsis driven by 35S promoter.

Fig. 18 shows that Activation of BG1 led to increased polar auxin transport (PAT) (A) Phenotypes of 7-day-old seedlings treated with 0.1 μ M NAA. (B) Comparison of the length of root between WT and Bg1-D plants. Data are given as mean $\pm$ SE (n=20). A Student's t-test was used to generate the P values. ** 0.001 $\leq$ P < 0.01.

Fig. 19 demonstrates that activation of BG1 leads to increased biomass. The biomass was calculated by the dry weight of mature organs. OE-1, OE-2 and OE-3 indicated three independent transgenic lines. Ri-1 and Ri-2 indicated two representative transgenic plants of BG1-RNAi. Bars represent means $\pm$ SD (n=20).

DETAILED DESCRIPTION

Grain size is a desirable component of grain yield for example, in rice and has been under selection since cereals were first domesticated.

A method of producing a seed, the method comprising: (a) crossing a first plant with a second plant, wherein at least one of the first plant and the second plant comprises a recombinant DNA construct, wherein the recombinant DNA construct comprises a polynucleotide operably linked to at least one heterologous regulatory element, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V or the Clustal W method of alignment, using the respective default parameters, when compared to SEQ ID NOS: 1, 4-27; and (b) selecting a seed of the crossing of step (a), wherein the seed comprises the recombinant DNA construct. A plant grown from the seed may exhibit at least one trait selected from the group consisting of: increased abiotic stress tolerance,

increased yield, increased biomass, and altered root architecture, when compared to a control plant not comprising the recombinant DNA construct. The polypeptide may be over-expressed in at least one tissue of the plant, or during at least one condition of abiotic stress, or both. The plant may be selected from the group consisting of: maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, sugar cane and switchgrass.

A method of producing a plant that exhibits an increase in at least one trait selected from the group consisting of: increased abiotic stress tolerance, increased yield, increased biomass, and altered root architecture, wherein the method comprises growing a plant from a seed comprising a recombinant DNA construct, wherein the recombinant DNA construct comprises a polynucleotide operably linked to at least one heterologous regulatory element, wherein the polynucleotide encodes a polypeptide having an amino acid sequence of at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, based on the Clustal V or the Clustal W method of alignment, using the respective default parameters, when compared to SEQ ID NOS: 1, 4-27, wherein the plant exhibits at least one trait selected from the group consisting of: increased nitrogen stress tolerance, increased yield, increased biomass, and altered root architecture, when compared to a control plant not comprising the recombinant DNA construct. The polypeptide may be over-expressed in at least one tissue of the plant, or during at least one condition of abiotic stress, or both. The plant may be selected from the group consisting of: maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, millet, sugar cane and switchgrass.

The practice of the present disclosure 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.

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.

It is understood, as those skilled in the art will appreciate, that the disclosure encompasses more than the specific exemplary sequences. Alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not affect the functional properties of the encoded polypeptide, are well known in the art. For example, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N terminal and C terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

The protein disclosed herein may also be a protein which comprises an amino acid sequence comprising deletion, substitution, insertion and/or addition of one or more amino acids in an amino acid sequence selected from the group consisting of SEQ ID NO: 1 or variants thereof. The substitution may be conservative, which means the replacement of a certain amino acid residue by another residue having similar physical and chemical characteristics. Non-limiting examples of conservative substitution include replacement between aliphatic group-containing amino acid residues such as Ile, Val, Leu or Ala, and replacement between polar residues such as Lys-Arg, Glu-Asp or Gln-Asn replacement.

Proteins derived by amino acid deletion, substitution, insertion and/or addition can be prepared when DNAs encoding their wild-type proteins are subjected to, for example, well-known site-directed mutagenesis (see, e.g., *Nucleic Acid Research*, Vol.

10, No. 20, p.6487-6500, 1982, which is hereby incorporated by reference in its entirety). As used herein, the term "one or more amino acids" is intended to mean a possible number of amino acids which may be deleted, substituted, inserted and/or added by site-directed mutagenesis.

5 Site-directed mutagenesis may be accomplished, for example, as follows using a synthetic oligonucleotide primer that is complementary to single-stranded phage DNA to be mutated, except for having a specific mismatch (i.e., a desired mutation). Namely, the above synthetic oligonucleotide is used as a primer to cause synthesis of a complementary strand by phages, and the resulting duplex DNA is then used to
10 transform host cells. The transformed bacterial culture is plated on agar, whereby plaques are allowed to form from phage-containing single cells. As a result, in theory, 50% of new colonies contain phages with the mutation as a single strand, while the remaining 50% have the original sequence. At a temperature which allows hybridization with DNA completely identical to one having the above desired mutation,
15 but not with DNA having the original strand, the resulting plaques are allowed to hybridize with a synthetic probe labeled by kinase treatment. Subsequently, plaques hybridized with the probe are picked up and cultured for collection of their DNA.

Techniques for allowing deletion, substitution, insertion and/or addition of one or more amino acids in the amino acid sequences of biologically active peptides such as
20 enzymes while retaining their activity include site-directed mutagenesis mentioned above, as well as other techniques such as those for treating a gene with a mutagen, and those in which a gene is selectively cleaved to remove, substitute, insert or add a selected nucleotide or nucleotides, and then ligated.

The protein disclosed herein may also be a protein which is encoded by a
25 nucleic acid comprising a nucleotide sequence comprising deletion, substitution, insertion and/or addition of one or more nucleotides in a nucleotide sequence selected from the group consisting of sequences encoding SEQ ID NOS: 1, 4-27. Nucleotide deletion, substitution, insertion and/or addition may be accomplished by site-directed mutagenesis or other techniques as mentioned above.

30 The protein disclosed herein may also be a protein which is encoded by a nucleic acid comprising a nucleotide sequence hybridizable under stringent conditions with the complementary strand of a nucleotide sequence selected from the group consisting of sequences encoding SEQ ID NOS: 1, 4-27.

The term "under stringent conditions" means that two sequences hybridize under
35 moderately or highly stringent conditions. More specifically, moderately stringent conditions can be readily determined by those having ordinary skill in the art, e.g.,

depending on the length of DNA. The basic conditions are set forth by Sambrook et al., Molecular Cloning: A Laboratory Manual, third edition, chapters 6 and 7, Cold Spring Harbor Laboratory Press, 2001 and include the use of a prewashing solution for nitrocellulose filters 5xSSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0), hybridization conditions of about 50% formamide, 2xSSC to 6xSSC at about 40-50 °C (or other similar hybridization solutions, such as Stark's solution, in about 50% formamide at about 42 °C) and washing conditions of, for example, about 40-60 °C, 0.5-6xSSC, 0.1% SDS. Preferably, moderately stringent conditions include hybridization (and washing) at about 50 °C and 6xSSC. Highly stringent conditions can also be readily determined by those skilled in the art, e.g., depending on the length of DNA.

Generally, such conditions include hybridization and/or washing at higher temperature and/or lower salt concentration (such as hybridization at about 65 °C, 6xSSC to 0.2xSSC, preferably 6xSSC, more preferably 2xSSC, most preferably 0.2xSSC), compared to the moderately stringent conditions. For example, highly stringent conditions may include hybridization as defined above, and washing at approximately 65-68 °C, 0.2xSSC, 0.1% SDS. SSPE (1xSSPE is 0.15 M NaCl, 10 mM NaH₂PO₄, and 1.25 mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15 M NaCl and 15 mM sodium citrate) in the hybridization and washing buffers; washing is performed for 15 minutes after hybridization is completed.

It is also possible to use a commercially available hybridization kit which uses no radioactive substance as a probe. Specific examples include hybridization with an ECL direct labeling & detection system. Stringent conditions include, for example, hybridization at 42 °C for 4 hours using the hybridization buffer included in the kit, which is supplemented with 5% (w/v) Blocking reagent and 0.5 M NaCl, and washing twice in 0.4% SDS, 0.5xSSC at 55 °C for 20 minutes and once in 2xSSC at room temperature for 5 minutes.

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 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* (Yamamoto, 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 codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present disclosure 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 and herein incorporated by reference). Thus, the maize preferred codon for a particular amino acid might be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants is listed in Table 4 of Murray, *et al.*, *supra*.

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. Heterologous may also indicate that a particular nucleic acid is foreign to its location in the genome as compared to its native location in the genome. 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 disclosure, 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. Preferably, host cells are monocotyledonous or dicotyledonous plant cells, including but not limited to maize, sorghum, sunflower, soybean, wheat, alfalfa, rice, cotton, canola, barley, millet and tomato. A particularly preferred monocotyledonous host cell is a maize 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).

5 The terms “isolated” refers 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
10 “heterologous” nucleic acids. Unless otherwise stated, the term “nitrate uptake-associated nucleic acid” means a nucleic acid comprising a polynucleotide (“nitrate uptake-associated polynucleotide”) encoding a full length or partial length nitrate uptake-associated polypeptide.

As used herein, “nucleic acid” includes reference to a deoxyribonucleotide or
15 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,
20 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*
25 *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
30 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
35 (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny of same. Plant cell,

as used herein includes, without limitation, 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 disclosure, 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*. A particularly preferred plant is *Zea mays*.

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 effect 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. Suitable constitutive promoters include for example, Ubiquitin promoters, actin promoters, and GOS2 promoter (de Pater et al (1992), The Plant Journal, 2: 837–844).

As used herein “recombinant” includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention or may have reduced or eliminated expression of a native gene. 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.

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. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition or spontaneous mutation.

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, 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. The preferred program to use 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 and hereby incorporated by reference). 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).

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%.

10 Orthologs and Paralogs

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.

Variant Nucleotide Sequences in the non-coding regions

The nitrate uptake-associated nucleotide sequences are used to generate variant nucleotide sequences having the nucleotide sequence of the 5'-untranslated region, 3'-untranslated region or promoter region that is approximately 70%, 75%, 80%, 85%, 90% and 95% identical to the original nucleotide sequence of the corresponding SEQ ID NO: 1. These variants are then associated with natural variation in the germplasm for component traits related to grain quality and/or grain yield. The associated variants are used as marker haplotypes to select for the desirable traits.

Variant Amino Acid Sequences of OsBG1-associated Polypeptides

Variant amino acid sequences of OsBG1-associated 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 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. These variants are then associated with natural variation in the germplasm for component traits related to grain quality and/or grain yield. The associated variants are used as marker haplotypes to select for the desirable traits.

5

Synthetic Methods for Constructing Nucleic Acids

The isolated nucleic acids of the present disclosure 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 US Patent Number 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.

20

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' 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 disclosure provides 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences.

30

Plant Transformation Methods

Numerous methods for introducing foreign genes into plants are known and can be used to insert a nitrate uptake-associated polynucleotide into a plant host, including

35

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, pp. 67-88 (1993). The methods chosen vary with the host plant, and include chemical
 5 transfection methods such as calcium phosphate, microorganism-mediated gene transfer such as *Agrobacterium* (Horsch *et al.*, (1985) *Science* 227:1229-31), electroporation, micro-injection and biolistic bombardment.

Expression cassettes and vectors and *in vitro* culture methods for plant cell or tissue transformation and regeneration of plants are known and available. See, e.g.,
 10 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,
 15 targeted for gene modification. Suitable methods of transforming plant cells include microinjection (Crossway, *et al.*, (1986) *Biotechniques* 4:320-334 and US Patent Number 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.*, US Patent Number
 20 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 Microprojectile Bombardment". pp. 197-213 in *Plant Cell, Tissue and Organ Culture, Fundamental Methods*, eds. Gamborg and Phillips. Springer-Verlag Berlin Heidelberg New York, 1995; US Patent Number 5,736,369 (meristem); Weissinger, *et al.*, (1988) *Ann. Rev.*
 25 *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); WO 91/10725 (maize); Klein, *et al.*, (1988) *Plant Physiol.* 91:440-444 (maize); Fromm, *et al.*, (1990)
 30 *Biotechnology* 8:833-839 and Gordon-Kamm, *et al.*, (1990) *Plant Cell* 2:603-618 (maize); Hooydaas-Van Slogteren and 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. Chapman, *et al.*, pp. 197-209. Longman, NY (pollen); Kaeppler, *et al.*, (1990) *Plant Cell Reports* 9:415-
 35 418; and Kaeppler, *et al.*, (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated

transformation); US Patent Number 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 (US Patent Number 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), all of which are herein incorporated by reference.

15 *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 US Patent Number 4,658,082; US Patent Application Serial Number 913,914, filed October 1, 1986, as referenced in US Patent Number 5,262,306, issued November 16, 1993 and Simpson, *et al.*, (1986) *Plant Mol. Biol.* 6:403-15 (also referenced in the '306 patent), all incorporated by reference in their entirety.

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; US Patent Number 4,658,082; Simpson, *et al.*, *supra*; and US Patent Application Serial Numbers 913,913 and 913,914, both filed October 1, 1986, as referenced in US Patent Number 5,262,306, issued November 16, 1993, the entire disclosures therein incorporated herein by reference.

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.

1. *Polynucleotide-Based Methods:*

In some embodiments of the present disclosure, a plant is transformed with an expression cassette that is capable of expressing a polynucleotide that inhibits the expression of OsBG1 of the disclosure. 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 disclosure, an expression cassette capable of expressing a polynucleotide that inhibits the expression of at least one nitrate uptake-associated polypeptide is an expression cassette capable of producing an RNA molecule that inhibits the transcription and/or translation of at least one nitrate uptake-associated polypeptide of the disclosure. 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 OsBG1 are given below.

i. Sense Suppression/Cosuppression

In some embodiments of the disclosure, inhibition of the expression of OsBG1 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 a messenger RNA encoding OsBG1 in the "sense" orientation. 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 nitrate uptake-associated polypeptide expression.

The polynucleotide used for cosuppression may correspond to all or part of the sequence encoding the nitrate uptake-associated polypeptide, all or part of the 5' and/or 3' untranslated region of *OsBG1* transcript or all or part of both the coding sequence and the untranslated regions of a transcript encoding *OsBG1*. In some embodiments where the polynucleotide comprises all or part of the coding region for the nitrate uptake-associated 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, US Patent Number 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 US Patent Numbers 5,034,323, 5,283,184 and 5,942,657, each of which is herein incorporated by reference. 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, US Patent Publication Number 2002/0048814, herein incorporated by reference. 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, US Patent Numbers 5,283,184 and 5,034,323, herein incorporated by reference.

ii. Antisense Suppression

In some embodiments of the disclosure, inhibition of the expression of the nitrate uptake-associated 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 a messenger RNA encoding the nitrate uptake-associated 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 of nitrate uptake-associated polypeptide expression.

iii. Double-Stranded RNA Interference

In some embodiments of the disclosure, inhibition of the expression of *OsBG1* 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.

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 expression cassettes may be used for the sense and antisense sequences. 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 and WO 99/49029, WO 99/53050, WO 99/61631 and WO 00/49035, each of which is herein incorporated by reference.

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

In some embodiments of the disclosure, inhibition of the expression of *OsBG1* 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 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 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. 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 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 US Patent Application Publication Number 2003/0175965, each of which is herein incorporated by reference. A transient assay for the efficiency of hpRNA constructs to silence gene expression *in vivo* has been described by Panstruga, *et al.*, (2003) *Mol. Biol. Rep.* 30:135-140, herein incorporated by reference.

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) *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* 30:289-295 and US Patent Application Publication Number 2003/0180945, each of which is herein incorporated by reference.

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 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; Aufsaetz, *et al.*, (2002) *Proc. Nat'l. Acad. Sci.* 99(4):16499-16506; Sijen, *et al.*, *Curr. Biol.* (2001) 11:436-440), herein incorporated by reference.

v. *Amplicon-Mediated Interference*

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 nitrate uptake-associated polypeptide). 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 US Patent Number 6,646,805, each of which is herein incorporated by reference.

vi. *Ribozymes*

In some embodiments, the polynucleotide expressed by the expression cassette of the disclosure is catalytic RNA or has ribozyme activity specific for the messenger RNA. This method is described, for example, in US Patent Number 4,987,071, herein incorporated by reference.

vii. *Small Interfering RNA or Micro RNA*

In some embodiments of the disclosure, inhibition of the expression of OsBG1 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 efficient at inhibiting the expression of endogenous genes. See, for example Javier, *et al.*, (2003) *Nature* 425:257-263, herein incorporated by reference.

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 structure containing a 22-nucleotide sequence that is complementary to another endogenous gene (target sequence). For suppression of nitrate uptake-associated expression, the 22-nucleotide sequence is selected from a nitrate uptake-associated transcript sequence and contains 22 nucleotides of said
5 nitrate uptake-associated 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 the RNA interference they induce is inherited by subsequent generations of plants.

10 2. *Polypeptide-Based Inhibition of Gene Expression*

In one embodiment, the polynucleotide encodes a zinc finger protein that binds to a gene encoding OsBG1, resulting in reduced expression of the gene. In particular embodiments, the zinc finger protein binds to a regulatory region of a nitrate uptake-associated gene. In other embodiments, the zinc finger protein binds to a messenger
15 RNA encoding OsBG1 and prevents its translation. Methods of selecting sites for targeting by zinc finger proteins have been described, for example, in US Patent Number 6,453,242 and methods for using zinc finger proteins to inhibit the expression of genes in plants are described, for example, in US. Patent Application Publication Number 2003/0037355, each of which is herein incorporated by reference.

20 3. *Polypeptide-Based Inhibition of Protein Activity*

In some embodiments of the disclosure, the polynucleotide encodes an antibody that binds to polypeptide of the disclosure. The expression of antibodies in plant cells and the inhibition of molecular pathways by expression and binding of antibodies to
25 proteins in plant cells are well known in the art. See, for example, Conrad and Sonnewald, (2003) *Nature Biotech.* 21:35-36, incorporated herein by reference.

4. *Gene Disruption*

In some embodiments of the present disclosure, the activity of OsBG1 is
30 reduced or eliminated by disrupting the gene encoding the nitrate uptake-associated polypeptide. The gene encoding the nitrate uptake-associated 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
35 have reduced nitrogen utilization activity.

i. *Transposon Tagging*

In one embodiment of the disclosure, transposon tagging is used to reduce or eliminate the nitrate uptake-associated activity of one or more nitrate uptake-associated polypeptide. Transposon tagging comprises inserting a transposon within an endogenous nitrate uptake-associated gene to reduce or eliminate expression of the nitrate uptake-associated polypeptide. "nitrate uptake-associated gene" is intended to mean the gene that encodes OsBG1 according to the disclosure.

In this embodiment, the expression of one or more nitrate uptake-associated polypeptide is reduced or eliminated by inserting a transposon within a regulatory region or coding region of the gene encoding the nitrate uptake-associated polypeptide. A transposon that is within an exon, intron, 5' or 3' untranslated sequence, a promoter or any other regulatory sequence of a nitrate uptake-associated gene may be used to reduce or eliminate the expression and/or activity of the encoded nitrate uptake-associated polypeptide.

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; 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 US Patent Number 5,962,764, each of which is herein incorporated by reference.

ii. *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 disclosure. 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, each of which is herein incorporated by reference. In addition, a fast and automatable method for screening for chemically induced mutations, TILLING (Targeting Induced Local Lesions

In Genomes), using denaturing HPLC or selective endonuclease digestion of selected PCR products is also applicable to the instant disclosure. See, McCallum, *et al.*, (2000) *Nat. Biotechnol.* 18:455-457, herein incorporated by reference.

5 Mutations that impact gene expression or that interfere with the function (enhanced nitrogen utilization activity) 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 nitrate uptake-associated polypeptides suitable for mutagenesis with the goal to eliminate nitrate uptake-associated activity have been
10 described. Such mutants can be isolated according to well-known procedures, and mutations in different nitrate uptake-associated loci can be stacked by genetic crossing. See, for example, Gruis, *et al.*, (2002) *Plant Cell* 14:2863-2882.

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

The disclosure encompasses additional methods for reducing or eliminating the activity of one or more nitrate uptake-associated polypeptide. 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
20 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, US Patent Numbers 5,565,350; 5,731,181; 5,756,325; 5,760,012; 5,795,972 and 5,871,984, each of which are herein incorporated by reference. See also, WO 98/49350, WO 99/07865,
25 WO 99/25821 and Beetham, *et al.*, (1999) *Proc. Natl. Acad. Sci. USA* 96:8774-8778, each of which is herein incorporated by reference.

vi. Modulating Reproductive Tissue Development

Methods for modulating reproductive tissue development are provided. In one
30 embodiment, methods are provided to modulate floral development in a plant. By "modulating floral development" is intended any alteration in a structure of a plant's reproductive tissue as compared to a control plant in which the activity or level of the nitrate uptake-associated polypeptide has not been modulated. "Modulating floral development" further includes any alteration in the timing of the development of a
35 plant's reproductive tissue (i.e., a delayed or an accelerated timing of floral development) when compared to a control plant in which the activity or level of the

nitrate uptake-associated polypeptide has not been modulated. Macroscopic alterations may include changes in size, shape, number, or location of reproductive organs, the developmental time period that these structures form or the ability to maintain or proceed through the flowering process in times of environmental stress.

5 Microscopic alterations may include changes to the types or shapes of cells that make up the reproductive organs.

In general, methods to modify or alter the host endogenous genomic DNA are available. This includes altering the host native DNA sequence or a pre-existing transgenic sequence including regulatory elements, coding and non-coding sequences.

10 These methods are also useful in targeting nucleic acids to pre-engineered target recognition sequences in the genome. As an example, the genetically modified cell or plant described herein, is generated using "custom" or engineered endonucleases such as meganucleases produced to modify plant genomes (see e.g., WO 2009/114321; Gao et al. (2010) *Plant Journal* 1:176-187). Another site-directed engineering is
15 through the use of zinc finger domain recognition coupled with the restriction properties of restriction enzyme. See e.g., Urnov, et al., (2010) *Nat Rev Genet.* 11(9):636-46; Shukla, et al., (2009) *Nature* 459 (7245):437-41. A transcription activator-like (TAL) effector-DNA modifying enzyme (TALE or TALEN) is also used to engineer changes in plant genome. See e.g., US20110145940, Cermak et al., (2011) *Nucleic Acids Res.*
20 39(12) and Boch et al., (2009), *Science* 326(5959): 1509-12. Site-specific modification of plant genomes can also be performed using the bacterial type II CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated) system. See e.g., Belhaj et al., (2013), *Plant Methods* 9: 39; The CRISPR/Cas system allows targeted cleavage of genomic DNA guided by a customizable small noncoding RNA.
25 Based on the disclosure of the BG1 coding sequences, polypeptide sequences of the orthologs/homologs and the genomic DNA sequences, site-directed mutagenesis can be readily performed to generate plants expressing a higher level of the endogenous BG1 polypeptide or an ortholog thereof.

Antibodies to a BG1 polypeptide disclosed herein or the embodiments or to
30 variants or fragments thereof are also encompassed. The antibodies of the disclosure include polyclonal and monoclonal antibodies as well as fragments thereof which retain their ability to bind to BG1 polypeptide disclosed herein. An antibody, monoclonal antibody or fragment thereof is said to be capable of binding a molecule if it is capable of specifically reacting with the molecule to thereby bind the molecule to the antibody,
35 monoclonal antibody or fragment thereof. The term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as fragments or binding

regions or domains thereof (such as, for example, Fab and F(ab).sub.2 fragments) which are capable of binding hapten. Such fragments are typically produced by proteolytic cleavage, such as papain or pepsin. Alternatively, hapten-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry. Methods for the preparation of the antibodies of the present disclosure are generally known in the art. For example, see, *Antibodies, A Laboratory Manual*, Ed Harlow and David Lane (eds.) Cold Spring Harbor Laboratory, N.Y. (1988), as well as the references cited therein. Standard reference works setting forth the general principles of immunology include: Klein, J. *Immunology: The Science of Cell-Noncell Discrimination*, John Wiley & Sons, N.Y. (1982); Dennett, *et al.*, *Monoclonal Antibodies, Hybridoma: A New Dimension in Biological Analyses*, Plenum Press, N.Y. (1980) and Campbell, "Monoclonal Antibody Technology," In *Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 13, Burdon, *et al.*, (eds.), Elsevier, Amsterdam (1984). See also, US Patent Numbers 4,196,265; 4,609,893; 4,713,325; 4,714,681; 4,716,111; 4,716,117 and 4,720,459. PtIP-50 polypeptide or PtIP-65 polypeptide antibodies or antigen-binding portions thereof can be produced by a variety of techniques, including conventional monoclonal antibody methodology, for example the standard somatic cell hybridization technique of Kohler and Milstein, (1975) *Nature* 256:495. Other techniques for producing monoclonal antibody can also be employed such as viral or oncogenic transformation of B lymphocytes. An animal system for preparing hybridomas is a murine system. Immunization protocols and techniques for isolation of immunized splenocytes for fusion are known in the art. Fusion partners (e.g., murine myeloma cells) and fusion procedures are also known. The antibody and monoclonal antibodies of the disclosure can be prepared by utilizing a BG1 polypeptide disclosed herein as antigens.

A kit for detecting the presence of a BG1 polypeptide disclosed herein or detecting the presence of a nucleotide sequence encoding a BG1 polypeptide disclosed herein, in a sample is provided. In one embodiment, the kit provides antibody-based reagents for detecting the presence of a BG1 polypeptide disclosed herein in a tissue sample. In another embodiment, the kit provides labeled nucleic acid probes useful for detecting the presence of one or more polynucleotides encoding BG1 polypeptide disclosed herein. The kit is provided along with appropriate reagents and controls for carrying out a detection method, as well as instructions for use of the kit.

As discussed above, one of skill will recognize the appropriate promoter to use to modulate floral development of the plant. Exemplary promoters for this embodiment

include constitutive promoters, inducible promoters, shoot-preferred promoters and inflorescence-preferred promoters.

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 disclosure can be used in combination ("stacked") with other polynucleotide sequences of interest in order to create plants with a desired phenotype.

This disclosure 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 disclosure may be practiced without departing from the spirit and the scope of the disclosure as herein disclosed and claimed.

EXAMPLES

EXAMPLE 1 - *Bg1-D* mutant exhibits increased grain size phenotype

To gain more insight into the genetic networks controlling grain size in rice, a number of mutants with various seed size were found in a large population of T-DNA rice mutants which were obtained by transforming rice (*Oryza sativa* L. cv. Nipponbare) by *Agrobacterium*-mediated transformation (Ma et al., 2009, J. Genet. Genomics, 36, 267-276). One of them, big grain1 (*bg1-D*), a dominant rice mutant with significantly increased grain size and grain weight was identified and characterized. Compared with wild-type Nipponbare (*Oryza Sativa* L. spp. Japonica), *bg1-D* showed increased plant stature and enlarged organ size during both the vegetative and reproductive stages (Fig. 1). In the seventh day after germination, an accelerated growth was observed in *bg1-D* mutant which has significantly elongated leaf sheath compared with the wild-type plants

(Fig. 1A and Fig. 13). The morphological differences between wild-type and *bg1-D* become more apparent after they enter the state of reproductive and vegetative phase. The leaves of *bg1-D* were larger than those of wild-type plants and *bg1-D* mutant showed an altered internode elongation which resulted in increased plant height (Fig. 1. B and C). Interestingly, *bg1-D* displayed longer panicle and larger grain phenotype (Fig. 1. D, E and F), which traits are closely linked with rice yield. Examination of the cross-section of rice hull before flowering revealed that an increasing number and volume of parenchyma cells appeared in the *bg1-D* (Fig. 2), giving rise to an increased grain width. Longitudinal sections of spikelet hull observed by scanning electron microscopy (SEM) showed the epidermis cells in *bg1-D* was longer than that of wild-type (Fig. 14). Further, grain filling rate by measuring fresh and dry weights was also compared. Both fresh and dry weight of *bg1-D* was significantly higher than those of wild-type at 15 d after fertilization, and the differences reached a maximum at about 25 d after fertilization (Fig. 15). Based on these results, it is likely that the increase in grain size and grain weight in *bg1-D* mutant mainly resulted from increases in grain width, length and grain filling rate. Agronomic traits were compared between mature plants between *bg1-D* and wild-type plants are presented Table A1.

Plant materials and growth conditions: The *Bg1-D* mutant was initially identified from our T-DNA insertion population in Nipponbare (ssp japonica) background. All mature plants were grown in a paddy field in Beijing or Hainan under natural conditions. For plant hormone treatment, germinated seeds were placed on bottomless 96-well plates floating on 1/2 MS and moved to growth chambers (28 °C) with 12 h light/12 h dark cycle. Seedlings were sampled and treated after 7 d.

Table A1. The agronomic traits of wild-type and Bg1-D plants.

Agronomic traits	Wild-Type	<i>bg1-D</i>	30
Plant height (cm)	94.3±3.8	110.5 ± 5.3***	
Tiller number	12.0±2.0	11.0 ± 3.0	
Panicle length (cm)	19.6±0.9	23.2 ± 1.2***	
Grain No. per panicle	105.9±11.7	109.3 ± 10.8	
Grain length (mm)	7.4±0.2	8.1±0.2**	
Grain width (mm)	3.2±0.1	3.5±0.1**	
1000-grain weight (g)	21.9±0.3	29.3 ± 0.4***	

Agronomic traits analyses were done on mature plants. Results represent means ± SD of populations of the size indicated in parentheses, (n=30). Asterisks indicate the

significance of differences between wild-type and Bg1-D plants determined by Student's t test. ** $0.001 \leq P < 0.01$, *** $P < 0.001$.

Table A2. Yield test in a paddy is shown below.

Traits	NP	OE-5 (NP)	OE-7 (NP)
Pannicles per plot	3450.3 ± 59.4	3223.6 ± 41.5	3159.5 ± 34.1
Grains per panicle	87.4 ± 4.6	83.7 ± 3.5	82.2 ± 4.3
1000-grain weight (g)	22.3 ± 0.4	27.3 ± 0.6	29.2 ± 0.5
Theoretical yield per plot (kg)	6.7 ± 0.7	7.4 ± 0.5	7.6 ± 0.6
Actual yield per plot (kg)	5.3 ± 0.5	6.1 ± 0.4	6.4 ± 0.5

The planting density was 20 cm × 20 cm, the area per plot was 15 m². Data were calculated from plants in block region with three replications under natural condition in Beijing, China.

EXAMPLE 2 – Molecular cloning and characterization of BG1

By genetics analysis, the phenotype of *bg1-D* co-segregated with the T-DNA insertion, and the segregation ratio of T1 generation is approximately 3:1, reflecting T-DNA inserted at a single locus in *bg1-D* mutant. In addition, the progeny of *bg1-D* showed big grain phenotype whether the *bg1-D* acted as male parent or female parent, suggesting *Bg1* is a dominant mutant through activation of BG1 (Table S1).

The T-DNA 3'-flanking sequence was cloned by Site-Finding PCR method based on thermal asymmetric interlaced PCR (TAIL-PCR) using known methods. Sequence analysis revealed that a gene, Os03g0175800, is located 0.8 kb downstream of the T-DNA insertion. The expression level of Os03g0175800 increased at about 10 times more than the level in the wild-type plants (Fig. 3). The relative expression level of adjacent genes of T-DNA insertion site and no significant difference existed between wild-type and *bg1-D* mutant. Based on these results, it was hypothesized that Os03g0175800 is the gene responsible for the *bg1-D* mutant and designated this gene as *BG1*. The full-length cDNA for *BG1* was cloned by 5' and 3' RACE kit. It is showed that the BG1 full-length cDNA is 1162 bp long and contains a 933 bp open reading frame (ORF) which encodes 330 amino acid. BLAST search of the protein databases reveals that BG1 has similarity to proteins from other higher plants. Phylogenetic analysis showed that BG1 and its 22 structural homologs can be classified into two subgroups: BG1 dicot-specific groups and BG1 monocot-specific groups.

In Arabidopsis, there are four BG1-like proteins displaying high sequence similarity with BG1, some residues are highly conserved in the N-terminus and C-terminus region of BG1 between rice and Arabidopsis (Fig. 16). In monocot, BG1 is closely related to Bradi1g72920 from Brachypodium, GRMZM2G007134 and GRMZM2G438606 from maize, suggesting BG1 is evolutionally conserved across plant species. BG1 may function as a novel plant specific regulator for grain size.

The expression pattern of BG1 was examined by real-time PCR and promoter GUS assay, the result showed that BG1 was ubiquitously expressed in all tissues, including root, culm, leaf, sheath and panicles. But it was strongly expressed in the young panicles (1-5 cm length) and culm, and the expression level gradually decreased as the panicles matured. Histochemical analysis of BG1PRO:GUS transgenic plants revealed that GUS activity was preferentially detected in the vascular tissues of the tissues examined (Fig. 5).

Cloning of the BG1 gene: Rice genomic DNA was isolated using known techniques. The 3'-flanking region of the T-DNA insertion in *Bg1-D* was identified by means of SiteFinding PCR method based on thermal asymmetric interlaced PCR (TAIL-PCR). The primers used for SiteFinding PCR were listed in Table S2. Full-length cDNA of BG1 was amplified by Invitrogen 5' and 3' RACE system.

Table S1. Phenotypic measurements of transgenic positive and negative plants in the T₁ generation of *Bg1-D*

Bgl (+) and *Bgl* (-) represent transgenic positive and negative plants in T₁ generation, respectively. Values are means $\pm$ SE.

Genotype	Number of plants	Plant height (cm)	Grain length (mm)	Grain width (mm)	1000-grain weight (g)
Wild-type	20	94.3 $\pm$ 3.8	7.35 $\pm$ 0.18	3.23 $\pm$ 0.08	21.90 $\pm$ 0.3
<i>Bgl-D</i> (T ₂)	20	110.5 $\pm$ 5.3	8.47 $\pm$ 0.25	3.78 $\pm$ 0.17	29.3 $\pm$ 0.40
<i>Bgl-D</i> (+)	53	105.1 $\pm$ 8.5	8.16 $\pm$ 0.51	3.64 $\pm$ 0.35	26.89 $\pm$ 0.55
<i>Bgl-D</i> (-)	16	95.2 $\pm$ 5.4	7.43 $\pm$ 0.15	3.34 $\pm$ 0.13	22.34 $\pm$ 0.5

Table S2. Primer used for Site-Finding PCR and RACE (SEQ ID NOS: 28-38, in consecutive order below).

Primer name	Sequence (5'-3')
SF1	CACGACACGCTACTCAACACACCACCTCGCACAGCGTCCTANNNNNNGCATG
SFP1	CACGACACGCTACTCAACAC
SFP2	ACTCAACACACCACCTCGCACAGC
GSP1	ACAATGCTGAGGGATTCAAATTCTACCCA
GSP2	TACTGAATTAACGCCGAATTGAATTC
GeneRacer 5' primer	CGACTGGAGCACGAGGACACTGA
GeneRacer RNA Oligo	CGACUGGAGCACGAGGACACUGAC
Sequence	AUGGACUGAAGGAGUAGAAA
BG1-1R	CTTGGCCGCCGCAGCCACCGCCTCCA
BG1-2R	AGAACCCGCCGTAGCTAGAACACT
BG1-3R	GCGAAGATGGAGTTGAGGAA

RNA extraction and quantitative real-time RT-PCR: Total RNAs were isolated from various rice tissues using the TRIzol reagent (Invitrogen). After RNase free DNase (Promega) treatment, 2 µg RNA was reverse transcribed using oligo (dT) primer and AMV reverse transcriptase (Promega). Quantitative real-time PCR (qRT-PCR) experiments were performed by using SsoFast EvaGreen Supermix (Biorad) with Bio Rad CFX96 real-time PCR machine. The rice ACTIN1 gene (OsACT1) was used as the internal control. Gene-specific primers used in qRT-PCR are listed in Table S3.

Vector construction and plant transformation: BG1 full-length coding sequences were amplified from cDNA of WT plants and cloned into pCAMBIA 2300-Actin to generate pCAMBIA 2300-Actin:BG1 overexpression constructs. A 3.6 kb genomic fragment containing a 2 kb native promoter and the entire BG1 coding region with 3' UTR were cloned into pCAMBIA 2300 vector to create pCAMBIA 2300-ProBG1:BG1 overexpression constructs. To generate BG1-RNAi construct, a 440 bp gene-specific fragment (SEQ ID NO: 75) of BG1 coding sequence was amplified and cloned into pCAMBIA 2300-Actin vector by the method described previously (Lin et al., 2012, Plant Physiol. 158, 451-464). For promoter-GUS assay, a 2-kb genomic fragment upstream of the BG1 gene translation start codon was amplified and cloned into the pCAMBIA2391Z. For subcellular localization, BG1 was in-frame fused with C-terminal GFP tag driven by 35S promoter. Upon sequence confirmation, all of the resulting constructs were transformed into the rice genome by known Agrobacterium tumefaciens-mediated transformation methods. The transgenic plants were then transferred to the field for normal growth and seed harvesting. The primer sequences for vector construction are listed in Table S4.

Table S3. Primers used for real-time PCR (SEQ ID NOS: 39-64, in consecutive order, with the forward and reverse primers for each of the gene indicated below).

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
<i>BG1</i>	GATGGAGAGCGACGAGGAC	GCAATGGCGGCGAAGTTC
<i>Ubiquitin</i>	ACCACTTCGACCGCCACTACT	ACGCCTAAGCCTGCTGGTT
<i>Actin</i>	TGGCATCTCTCAGCACATTCC	TGCACAATGGATGGGTCAGA
<i>CDKA1</i>	GGTTTGGACCTTCTCTCTAAAATGC	AGAGCCTGTCTAGCTGTGATCCTT
<i>E2F2</i>	TGTTGGTGGCTGCCGATAT	CGCCAGGTGCACCCTTT
<i>H1</i>	GCAAGGCACCTGCAGCTT	AGGCAGCCTTTGTACAGATCCT
<i>CYCD1;1</i>	CCGTCGCTCGCATTTCGTA	TCATGTGCCTGGTGGTGA
<i>CYCT1</i>	GCTTCTTTCAGAGGGTCAT	GCATTTCAAGTACACGAGGT
<i>OsEXPA5</i>	AAGGCTGTGGCTTGATTGACA	TTAGGCCCAATTTTGCTATTTTG
<i>OsEXPA10</i>	TGACCAACTACAACGTGGTCCC	GCCAGTGTATGTTTTGCCGAAG
<i>OsEXPB3</i>	CTTTGAGTGGTTGGAGTGGTGG	GCAGCCTTCTTGGAGATGGAA
<i>OsEXPB4</i>	GTCGGTCTGTGTTGCGATTG	CCTCCATTTCACACAGCTT
<i>OsEXPB7</i>	ACGGTGATCATCACGGACAT	TCGAAGTGGTACAGCGACACT

5

Table S4. Primers used for vector construction. (SEQ ID NOS: 65-74, in consecutive order, with the forward and reverse primers as indicated below)
(Labeled sequence indicates enzyme cleavage site).

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
BG1OX	<u>TCTAGA</u> ATGGAGAGGTGGGCGGCCCAAG	<u>CTGCAGT</u> CAGACAACCTCTGGCGCTCCGT
BG1RNAi	<u>CTCGAG</u> CAGCCGTCTTCTCATCCAC	<u>GGATCC</u> GCGAAGATGGAGTTGAGGAA
BG1GUS	<u>GGATCC</u> GCTTGTCATTGCTGTACATG	<u>GAATTC</u> AGGAAACAAACGCCAAGAAG
BG1EGFP	<u>TCTAGA</u> ATGGAGAGGTGGGCGGCCCAAG	<u>ACTAGT</u> GACAACCTCTGGCGCTCCGTCCAT
BG1DNAOX	<u>GAATTC</u> GCTTGTCATTGCTGTACATGCTGT	<u>GGATCC</u> CACACACATGTTCCCATTAGGTTT

10

Histological analysis and microscopy observation: Cross-sections of spikelet hulls were analyzed by light microscopy (BX51; Olympus). Spikelet hulls were fixed with formalin: glacial acetic acid: 70% ethanol (1:1:18) and then dehydrated in a graded ethanol series. Fixed tissues were embedded in Paraplast Plus chips (Sigma), and cut using a microtome into 10 µm thick sections and then affixed to poly-Lys-coated slides (Sigma), followed by dewaxed in xylene, gradely rehydrated and dried before toluidine blue staining for light microscopy. Cells number and cell area in the outer parenchyma layer of the spikelet hulls were measured by Olympus stream software. Longitudinal sections of spikelet hull were analyzed by scanning electron microscopy (S-3000N; Hitachi), The sample pretreatment was done as described previously (Li et al., 2010, Cell Res. 20, 838-849). Cell lengths were measured both in WT and in *Bg1-D*.

EXAMPLE 3- BG1 is a novel positive regulator controlling grain size

To further validate BG1's function as a regulator of grain size and whether overexpression of BG1 is necessary for the increased grain size phenotype, we transformed wild-type rice with an expression vector that drives BG1 cDNA under the control of the rice ACTIN1 promoter. As show in Fig. 6, the BG1 overexpressed plants showed various degrees of grain size increasing phenotypes and of panicle elongation (Fig. 6A, B, C). The magnitude of the change in the phenotypes was apparently correlated with the level of BG1 expression (Fig. 6D). The 1000-grain weight of BG1-overexpression lines increased dramatically, resulting from the increase of grain length and grain width in transgenic plants compared with the wild-type (Fig. 6E, F, G). On the contrary, suppression of BG1 by RNAi led to obvious phenotypes in rice. The plant height is shorter and the length of panicle and grain size were greatly decreased (Fig. 7A, B, C, D), which lead to 1000-grain weight reduction compared with the wild-type plants. (Fig. 7F). All these phenotype, to some extent, was consistent with the result that BG1 is strongly expressed in culm and panicles. It was concluded that BG1 is the gene responsible for the *bg1-D* dominant mutant. In addition, transformed wild-type Arabidopsis (Col-0) expressing OsBG1 driven by cauliflower mosaic virus (CaMV) 35S promoter showed that overexpression of OsBG1 in Arabidopsis also result in increased seed size, indicating OsBG1 may share some common regulatory mechanism between monocots and dicots (Fig. 17).

EXAMPLE 4 - BG1 localized to the plasma membrane

To study the function of BG1 on grain size regulation, the subcellular localization of BG1 was investigated. Prediction by web-based tools described previously showed that BG1 protein contains no putative sequence motifs that define a specific subcellular localization or extracellular secretion. Then, the ORF of BG1 was fused with enhanced green fluorescent protein (BG1-eGFP) to generate pC2300-35S:BG1-eGFP vector. The transgenic plants showed the same phenotype as the *bg1-D* mutant, such as increased grain size and plant height, indicating that the fused BG1-eGFP did have the same function as the BG1 alone. Transgenic young seedling's root was observed and the fluorescent of BG1-eGFP is predominantly localized to the plasma membrane (Fig. 8A), consisting with our transient expression assay of BG1-eGFP in rice protoplast. To further confirm this result, the microsome of transgenic plants with BG1-eGFP was separated, it was observed that the BG-eGFP can be detected by GFP antibody in the microsomal pellet, but only very faint signal was detected in the soluble fraction. Two-phase partitioning assay further confirm BG1-GFP protein was highly enriched in the

plasma membrane-enriched fractions, but not in the other membrane's fractions (Fig. 8B). All these results indicated the BG1 is a novel protein which predominantly localized to the plasma membrane.

Subcellular localization of BG1: The BG1 cDNA sequence was cloned in frame with the eGFP protein under the control the 35S promoter. For confocal microscopy, the localization patterns of BG1:eGFP was examined by imaging root tips of 7-day-old seedlings of BG1:eGFP transgenic lines and rice protoplast using a confocal laser scanning microscope (Leica TCS SP5). eGFP were excited at 488 nm wavelengths, and the emission filters were 500-530 nm. Preparation and efficiency transfection of rice protoplasts were performed according to the method described previously with minor modifications (Zhang et al., 2011, Plant Methods 7.21). Microsomal fractionations and two-phase membrane partitioning experiments were conducted as previously described protocol. Equal amount of proteins of the total microsomal and the PM-enriched upper phase were analyzed by SDS-PAGE and immunoblotting with anti-GFP (Roche), anti-PIP2 (Agrisera), anti-PEPC (Agrisera) and anti-LHCII (Agrisera) antibody

EXAMPLE 5 - BG1 is a primary auxin response gene

Phytohormones play a very vital role in the growth and development of plants. To determine whether BG1 can response to exogenous plant hormone, the relative mRNA level of BG1 was investigated by real-time-PCR under different treatment. The result showed that BG1 can be specifically induced by IAA after 2-hour application, but not by other phytohormones (Fig. 9A). Time-course IAA induction in wild-type plants revealed BG1 mRNA level rapidly increased within 30 mins of IAA application and the maximal level is approximately 8-fold greater than the level seen in uninduced plants at 4-hour later, then the expression declined and gradually to the primitive level (Fig. 9B). The induction of BG1 mRNA by IAA occurs in the presence of the protein synthesis inhibitor cycloheximide (CHX), suggesting the accumulation of BG1 transcripts was independent of protein synthesis (Fig. 9C). Moreover, another two auxin compounds, naphthylacetic acid (NAA) and 2,4-dichlorophenoxyacetic acid (2-4-D) can also induce the expression of BG1. However, tryptophan, a molecule that is structurally similar to IAA but without auxin activity, fails to induce BG1 mRNA accumulation (Fig. 9D). These results are consistent with that two auxin responsive element (TGTCTC) were found within 1500 bp upstream of the BG1 translational start site. Based on these results, it was predicted that BG1 is novel a primary auxin-response gene. To test the sensitivity of exogenous IAA, leaf sheath of 7-day-old seedlings of WT and *Bg1-D* were treated with 10-3 μ M IAA, *Bg1-D* mutant showed elongated leaf sheath compared with that of

wild-type plants, suggesting the activation of BG1 led to increased IAA sensitivity (Fig. 9E).

EXAMPLE 6 - *Bg1-D* mutant exhibit altered IAA distribution and transport

To further elucidate the mechanisms responsible for the phenotypes conferred by the BG1 overexpression, the effects of the increased BG1 activity on auxin transport was observed. When DR5-GUS plants were crossed to *Bg1-D*, the expression pattern of GUS was altered in the double mutant. GUS expression was mainly accumulated at the root apex area in wild-type plants. However, in DR5-GUS/*Bg1-D* double mutant, the GUS expression was more intense and covered a broader area in the root elongation zone (Fig. 10B). Furthermore, when the basipetal IAA was measured directly, it was found that that approximately a 50% increase in basipetal IAA transport was observed in the coleoptile of *Bg1-D* dark-grown seedlings, indicating that the enhanced basipetal PAT resulted in altered endogenous IAA distribution in the *Bg1-D* mutant plant (Fig. 10A). Consistently, *Bg1-D* exhibited enhanced resistance to NPA, an exogenous IAA transport inhibitor, revealing that BG1 can specifically response to IAA coupling with enhanced polar auxin transport (Fig. 18).

Auxin transport assay: The polar auxin transport assays were performed using 5-day-old coleoptiles of dark-grown rice seedlings as described previously with minor modifications (Li et al., 2007, Cell Res. 17, 402-410; Qi et al., 2008, Plant Physiol. 147, 1947-1959), 2.0 cm length apical coleoptile segments (0.2 cm away from the tip was cutted off) were preincubated in 1/2 MS medium with shaking at 50 rpm for 2 h. The apical ends of coleoptile segment were then submerged in 15 μ L of 1/2 MS liquid medium containing 0.2% phytoigel, 500 nM 3H-IAA and 500 nM free IAA in a 1.5 mL microcentrifuge tube in the dark at room temperature for 3 h. NPA and 3H-BA were added as the IAA transport control. After 3 h transport assay, 0.5 cm from the unsubmerged ends of the segments were excised and washed three times in 1/2 MS liquid medium. The segments were then grouped in sets of five and incubated in 2.0 mL of universal scintillation fluid for 18 h. The radioactivity was counted with a liquid scintillation counter (1450 MicroBeta Trix; Perkin-Elmer). Based on the orientation of the coleoptile segments submerged in the buffer, acropetal auxin transport was also measured as a negative control.

Example 7: Organ size-associated genes was upregulated in *Bg1-D* mutant

Several genes have been characterized in controlling organ size in Arabidopsis, especially the genes responsible for cell proliferation and cell expansion, such as AINEGUMENTA (ANT), AtEXP10, ARGOS, CycD3;1, AtGIF1 and AtGRF1. To investigate the relationship of BG1 with the genes regulating the plant cell proliferation and cell expansion in rice, the transcript levels of core cell cycle and cell expansion regulators were analyzed by using quantitative RT-PCR in both mutant and wild-type plants. Genes regulating G1 to S1 phase of cell cycle and cell expansion genes were upregulated significantly in the young panicles of *Bg1-D* compared to the wild-type plants (Fig. 11). These results indicated that BG1 may play a new positive regulatory role in the control of cell proliferation and cell expansion by regulating auxin signal transduction or transport.

EXAMPLE 8 – Increased expression of BG1 increases biomass and grain yield

Since activation of BG1 in *bg1-D* leads to increased grain size and grain weight, it was tested whether biomass and grain yield could be improved by BG1 overexpression. When grown in paddy fields, transgenic rice plants overexpressing BG1 showed a significant increase in dry weight of mature organs. In contrast, two independent BG1-RNAi transgenic plant (Ri1 and Ri 2) showed decreased biomass (Fig. 19). To avoid the ectopic over-expression of BG1, we overexpressed BG1 in wild-type plants using BG1 genomic DNA driven by its native promoter. Compared to the non-transgenic control plants, the transgenic rice overexpressing BG1 showed increased grain size and larger panicle without other significant penalty trait (Fig. 12 A, B and C). The 1000-grain weight of transgenic lines OE-5, OE-7 increased 15.48% and 25.38%, and the grain yield per plant increased 6.69%, and 16.65% respectively. (Fig. 12 D and E). Yield tests revealed plants with increased BG1 expression have about 15% to 20% increase in grain yield compared with that of Nipponbare (Table A2). Taken together, these results demonstrated that BG1 overexpression in rice is positively correlated with an increase in biomass and grain yield.

Protein sequence and phylogenetic analysis: BG1 sequence of rice and Arabidopsis were obtained from the Rice Annotation Project Database (<http://rapdb.dna.affrc.go.jp/>) and the Arabidopsis Information Resource (<http://www.arabidopsis.org/index.jsp>), respectively. Protein sequences were aligned using ClustalX 1.83, and the conserved residues were displayed by GeneDoc (<http://www.nrbsc.org/gfx/genedoc/>). For phylogenetic analysis of BG1 homologues in plants, a total of 23 unique protein sequences were obtained through National Center for Biotechnology Information (NCBI) BLAST search and aligned using a ClustalW

program. Phylogenetic tree was constructed using MEGA 4.0 based on the neighbor-joining method. Topological robustness was assessed by bootstrap analysis with 500 replicates.

5 Global Identity (% identity) – **Table 2**

Gap creation penalty: 8 and Gap extension penalty: 2

SEQ ID NO	17	12	14	13	27	22	23	24	19	18	11	21	20	15	16	25	26
1	70	62	72	62	29	32	31	32	29	22	51	35	32	47	46	72	50
17	-	55	65	55	28	32	31	33	28	18	44	33	32	44	42	68	46
12	-	-	70	99	26	29	28	28	28	21	43	31	29	39	55	66	43
14	-	-	-	70	29	30	32	31	28	18	46	32	31	44	53	77	48
13	-	-	-	-	27	30	28	29	28	19	43	30	29	38	54	66	43
27	-	-	-	-	-	34	38	39	36	19	23	41	43	24	19	27	23
22	-	-	-	-	-	-	38	38	50	26	25	40	36	25	18	31	24
23	-	-	-	-	-	-	-	94	42	21	21	54	53	25	18	33	27
24	-	-	-	-	-	-	-	-	42	20	21	55	55	25	17	32	26
19	-	-	-	-	-	-	-	-	-	23	23	43	42	23	14	29	24
18	-	-	-	-	-	-	-	-	-	-	21	23	21	19	10	18	21
11	-	-	-	-	-	-	-	-	-	-	-	27	21	60	32	48	64
21	-	-	-	-	-	-	-	-	-	-	-	-	73	26	21	34	29
20	-	-	-	-	-	-	-	-	-	-	-	-	-	21	19	31	22
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	30	48	79
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	50	32
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	49
26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Global Similarity (% similarity) - **Table 3**

SEQ ID NO	17	12	14	13	27	22	23	24	19	18	11	21	20	15	16	25	26
1	76	69	77	69	37	48	43	45	43	34	58	47	43	56	50	77	58
17	-	63	71	63	37	46	43	43	39	31	54	46	43	50	47	73	52
12	-	-	76	99	35	45	43	42	39	37	53	43	39	47	59	73	52
14	-	-	-	76	38	45	45	46	42	31	53	47	42	51	56	80	54
13	-	-	-	-	35	45	42	42	39	34	53	42	39	46	59	73	52
27	-	-	-	-	-	44	45	46	43	26	34	48	47	33	27	36	31
22	-	-	-	-	-	-	48	49	62	39	42	55	50	40	27	47	40
23	-	-	-	-	-	-	-	96	55	30	34	64	65	35	27	44	36
24	-	-	-	-	-	-	-	-	55	29	34	65	67	36	26	45	37
19	-	-	-	-	-	-	-	-	-	39	34	59	56	34	23	43	35
18	-	-	-	-	-	-	-	-	-	-	31	31	32	32	17	31	30

SEQ ID NO	17	12	14	13	27	22	23	24	19	18	11	21	20	15	16	25	26
11	-	-	-	-	-	-	-	-	-	-	-	39	31	67	39	57	71
21	-	-	-	-	-	-	-	-	-	-	-	-	78	39	30	48	41
20	-	-	-	-	-	-	-	-	-	-	-	-	-	32	26	41	31
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	37	54	83
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	53	39
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	55
26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

The disclosure has been described with reference to various specific and preferred 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 disclosure.

5

WHAT IS CLAIMED IS:

1. A method of improving an agronomic characteristic of a plant, the method comprising modulating the expression of (i) a polynucleotide encoding an amino acid sequence comprising SEQ ID NO: 1 or an amino acid sequence that is at least 95% identical to one of SEQ ID NO: 1 (ii) a polynucleotide that hybridizes under stringent hybridization conditions to a fragment of polynucleotide comprising SEQ ID NO: 2, wherein the fragment comprises at least 100 contiguous nucleotides of SEQ ID NO: 2 (iii) a polynucleotide that encodes an amino acid sequence that is at least 90% identical to SEQ ID NO: 1, (iv) a polynucleotide encoding a polypeptide comprising one or more deletions or insertions or substitutions of amino acids compared to SEQ ID NO: 1.
2. The method of claim 1, wherein the expression of the polynucleotide encoding a polypeptide having at least 95% identity to SEQ ID NO: 1 is increased by transforming the plant with a recombinant polynucleotide operably linked to a heterologous promoter.
3. The method of claim 1, wherein the expression of an endogenous polynucleotide encoding a polypeptide having at least 95% identity to SEQ ID NO: 1 is increased by upregulating a regulatory element operably associated with the endogenous polynucleotide.
4. The method of claim 1, wherein the expression of the polynucleotide is increased by expressing the polynucleotide under a regulatory element of SEQ ID NO: 3.
5. The method of claim 1, wherein the agronomic characteristic is selected from the group consisting of (i) an increase in grain size, (ii) an increase in grain weight, (iii) an increase in panicle length, (iv) an increase in grain yield (v) an increase in grain filling rate, (vi) an increase in biomass.
6. The method of claim 1, wherein the agronomic performance is an increase in plant biomass during vegetative and/or reproductive stages.
7. The method of claim 1, wherein the grain weight is increased in relation to a control plant not having an increased expression of the polynucleotide.
8. The method of claim 1, wherein the plant is a monocot.
9. The method of claim 1, wherein the plant is rice or maize.
10. The method of claim 1, wherein the plant is a dicot.
11. The method of claim 1, wherein the plant is soybean.
12. A method of improving yield of a plant, the method comprising increasing the expression of a polynucleotide that encodes a polypeptide comprising the amino

acid sequence selected from the group consisting of SEQ ID NOS: 1, 4-27 or an allelic variant thereof.

13. A method of improving rice grain yield, the method comprising the expression of a polynucleotide that encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a variant thereof.
14. A method of marker assisted selection of a plant for improved yield, the method comprising:
 - a. performing marker-assisted selection of plants that have one or more variations in genomic regions encoding a protein comprising SEQ ID NO: 1 or a variant thereof or a regulatory sequence thereof; and
 - b. identifying the plant that exhibits higher yield.
15. A method of identifying one or more alleles in a population of rice plants that are associated with increased grain yield, the method comprising:
 - a. evaluating in a population of rice plants for one or more allelic variations in (i) a genomic region, the genomic region encoding a polypeptide or (ii) the regulatory region controlling the expression of the polypeptide, wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 1 or a sequence that is 95% identical to SEQ ID NO: 1;
 - b. obtaining phenotypic values of increased yield for the one or more rice plants in the population;
 - c. associating the allelic variations in the genomic region with the phenotype; and
 - d. identifying the one or more alleles that are associated with increased yield.
16. An isolated polynucleotide (i) encoding an amino acid sequence comprising one of SEQ ID NOS: 1, 4-27 or an amino acid sequence that is at least 95% identical to one of SEQ ID NOS: 1, 4-27 (ii) hybridizing under stringent hybridization conditions to a fragment of polynucleotide selected from the group consisting of SEQ ID NO: 2, wherein the fragment comprises at least 100 contiguous nucleotides of SEQ ID NO: 2 (iii) that encodes an amino acid sequence that is at least 90% identical to SEQ ID NO: 1, (iv) a polynucleotide encoding a polypeptide comprising one or more deletions or insertions or substitution of amino acids compared to SEQ ID NO: 1, wherein the polynucleotide encodes a polypeptide involved in the regulation of grain width, weight or yield.

17. A recombinant expression cassette, comprising the polynucleotide of claim 16, wherein the polynucleotide is operably linked to a regulatory element, wherein the expression cassette is functional in a plant cell.
18. A host cell comprising the expression cassette of claim 17.
- 5 19. A transgenic plant comprising the recombinant expression cassette of claim 18.
20. A transgenic plant part comprising a plant regulatory element that operably regulates the expression of a polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO: 1 or a variant or an ortholog thereof, wherein the regulatory element is heterologous to the polynucleotide.
- 10 21. The polynucleotide of claim 20, wherein the polynucleotide that regulates the expression comprises a sequence selected from the group consisting of SEQ ID NOS: 7-8 or a functional promoter fragment thereof.
22. A method of marker assisted selection of a maize plant for improved yield, the method comprising:
 - 15 a. detecting one or more genetic variations in a genomic region comprising a polynucleotide encoding a protein selected from the group consisting of SEQ ID NOS: 12-14 or a variant thereof or in a regulatory sequence controlling the expression of the polynucleotide thereof, in a plurality of maize plants; and
 - 20 b. associating the one or more variations with increased yield and thereby selecting the maize plant that has the one or more variations associated with higher yield.
23. A method of identifying one or more alleles associated with increased yield in a population of maize plants, the method comprising:
 - 25 a. evaluating in a population of maize plants one or more genetic variations in (i) a genomic region encoding a polypeptide or (ii) a regulatory region controlling the expression of the polypeptide, wherein the polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NOS: 12-14 or a sequence that is 95% identical to SEQ ID NOS: 12-
30 14;
 - b. obtaining yield data for one or more maize plants in the population;
 - c. associating the one or more genetic variations in the genomic region encoding the polypeptide or in the regulatory region controlling the
35 expression of the polypeptide with yield, thereby identifying one or more alleles associated with increased yield.

24. The method of claim 23, wherein the one or more genetic variations is in the coding region of the polynucleotide.
25. The method of claim 23, wherein the regulatory region is a promoter element.
26. The method of claim 23, wherein the yield is grain yield or seed yield.

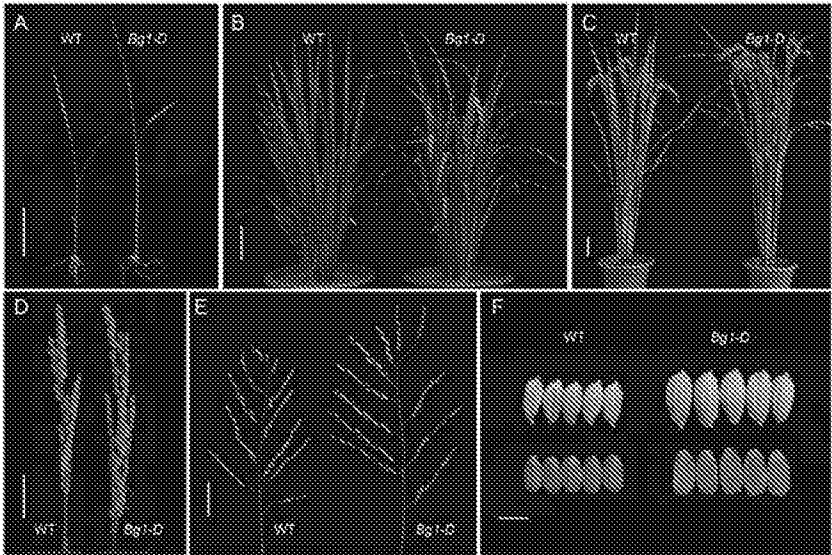


FIG. 1

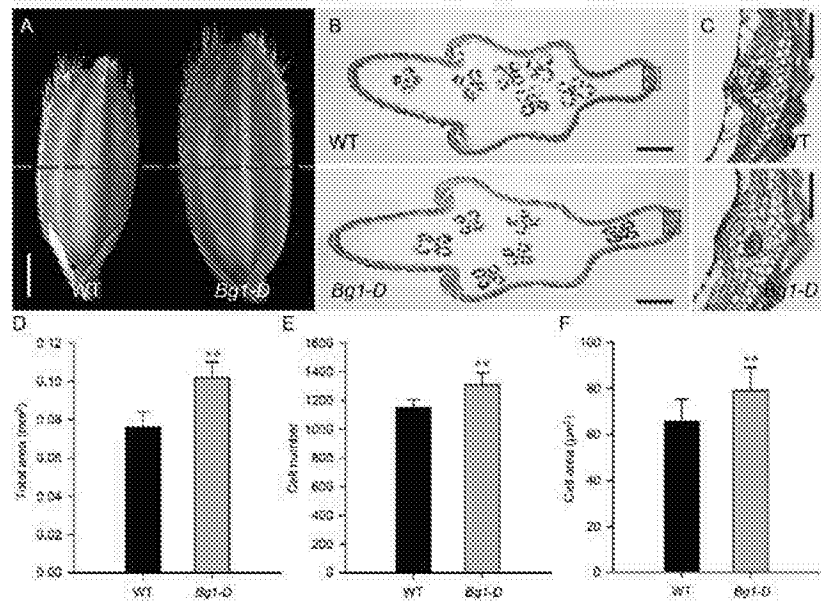
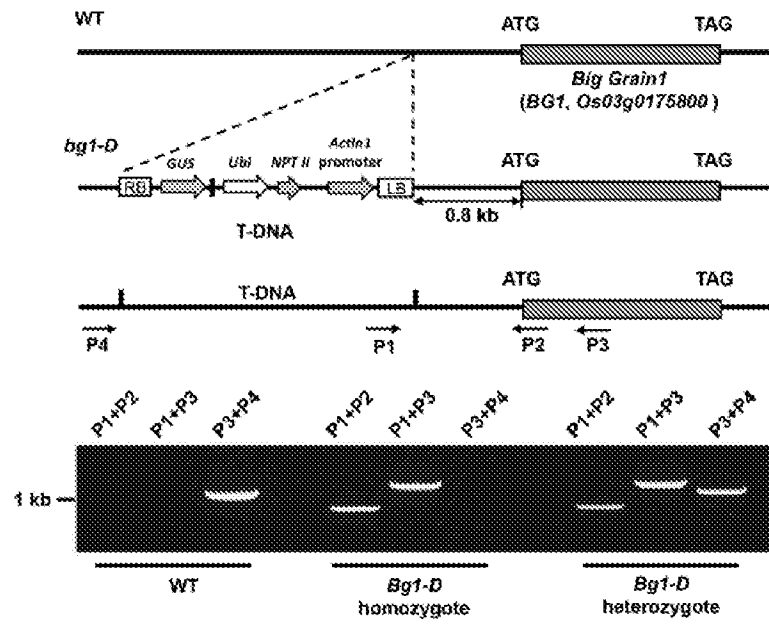


FIG. 2

A.



B.

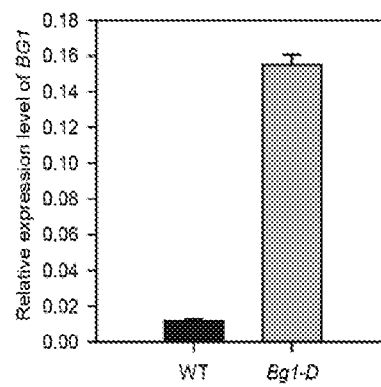
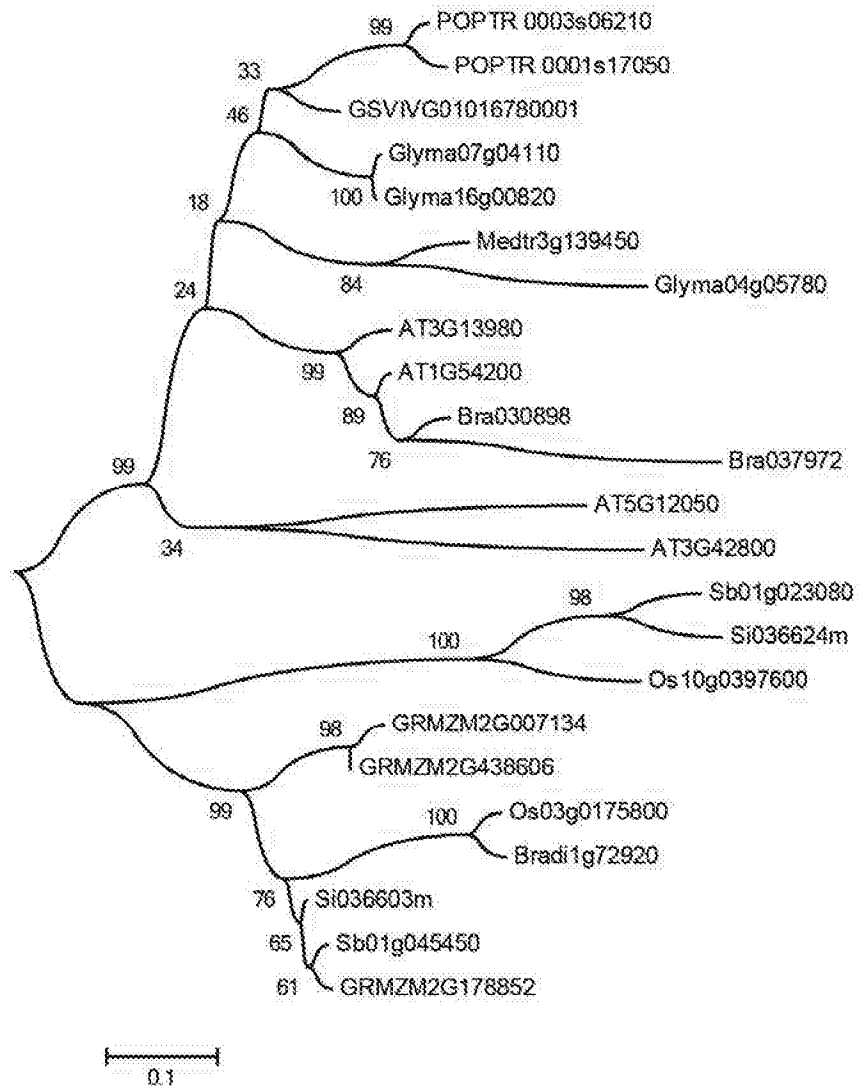


FIG. 3

**FIG. 4**

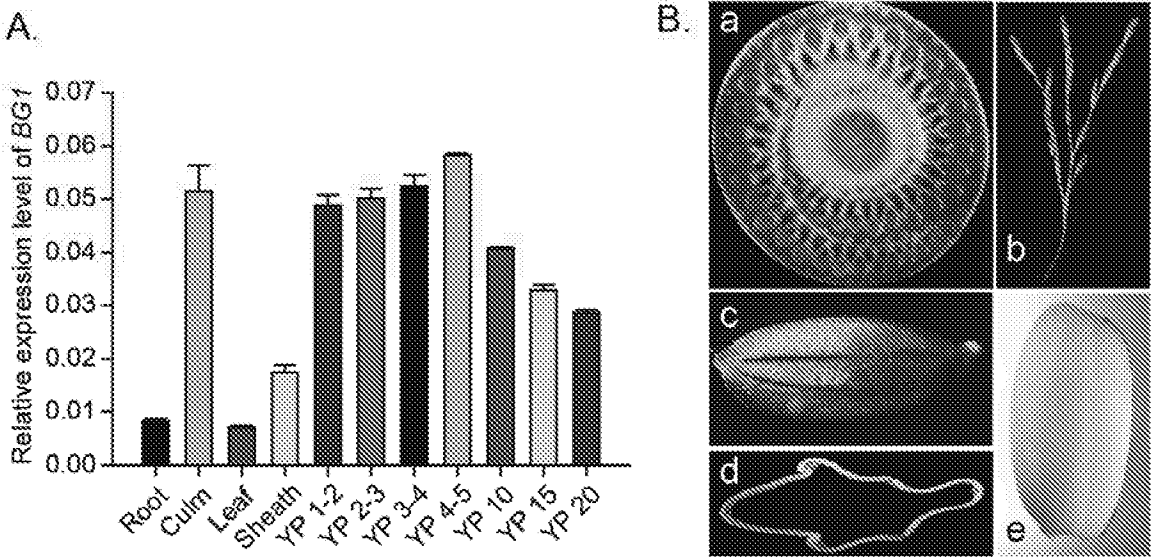


FIG. 5

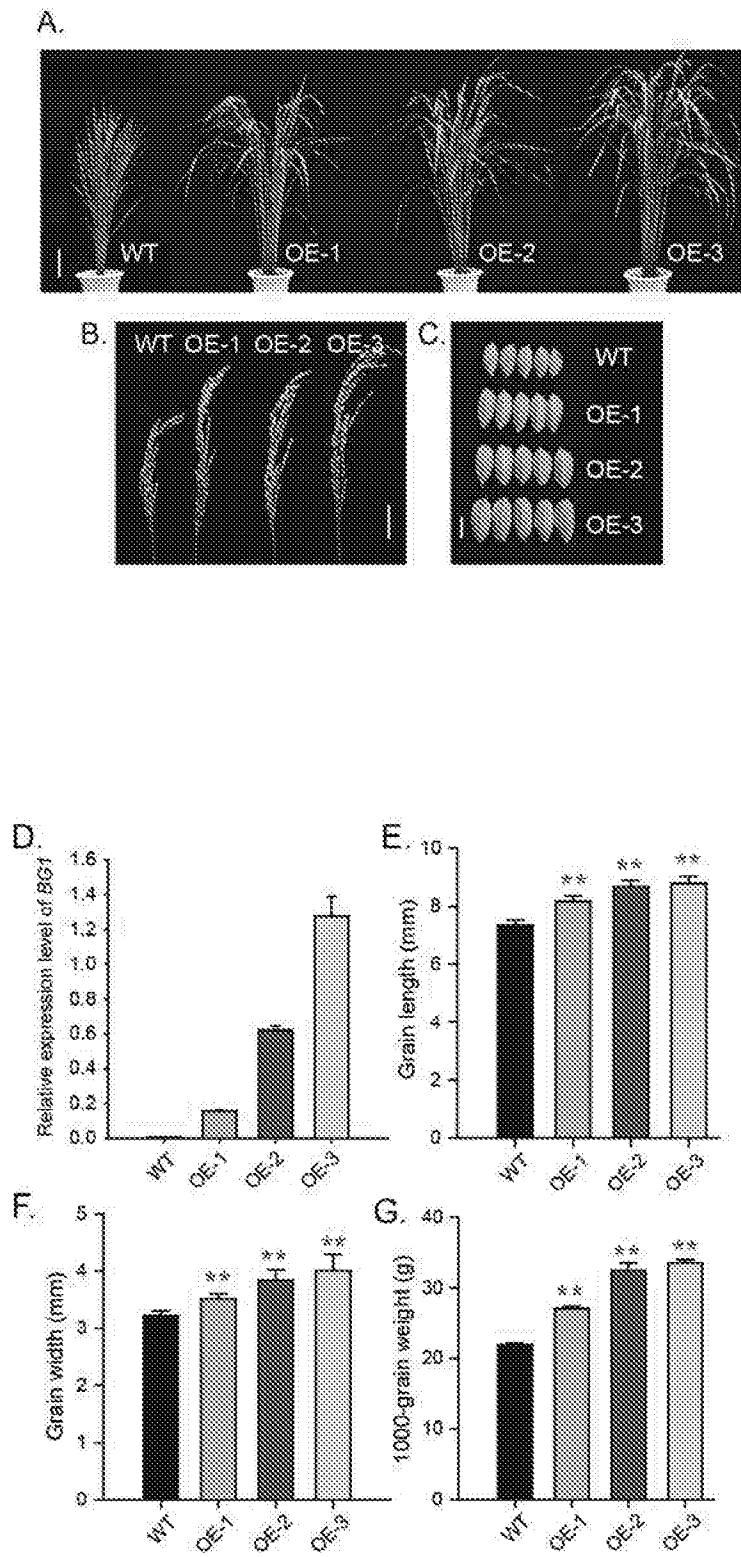
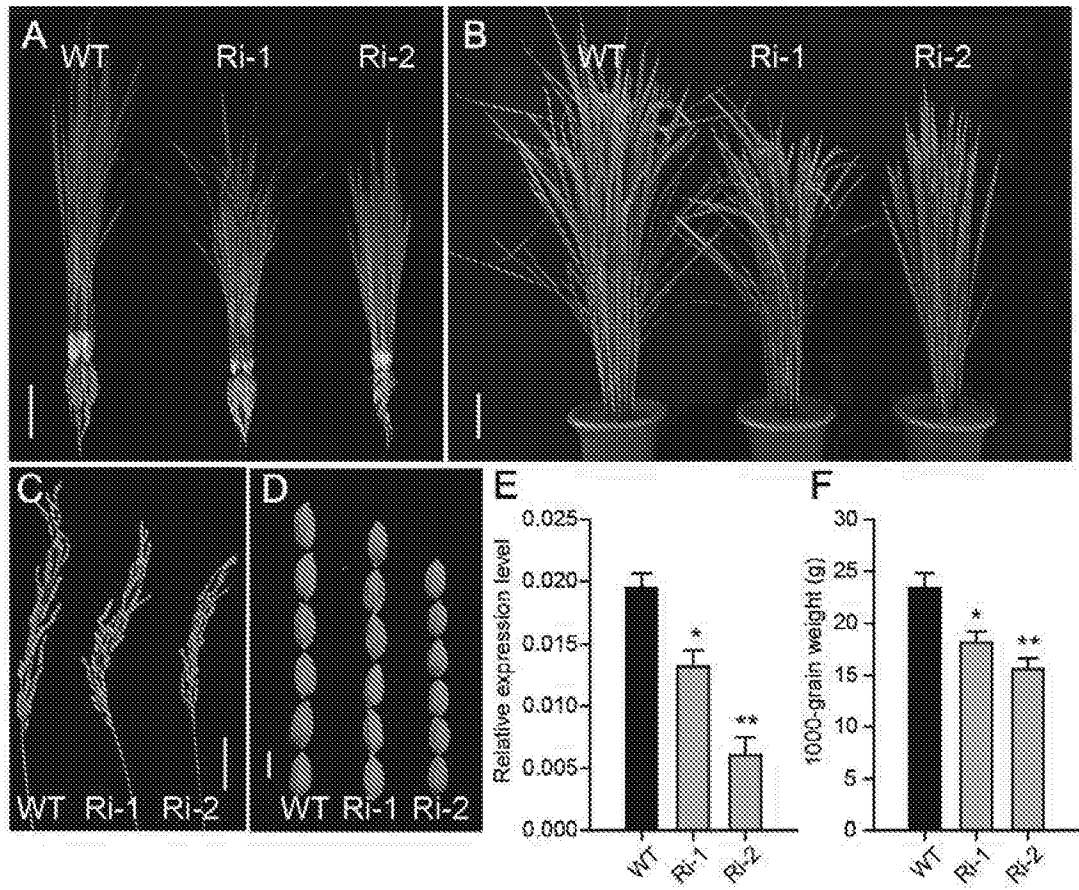


FIG. 6

**FIG. 7**

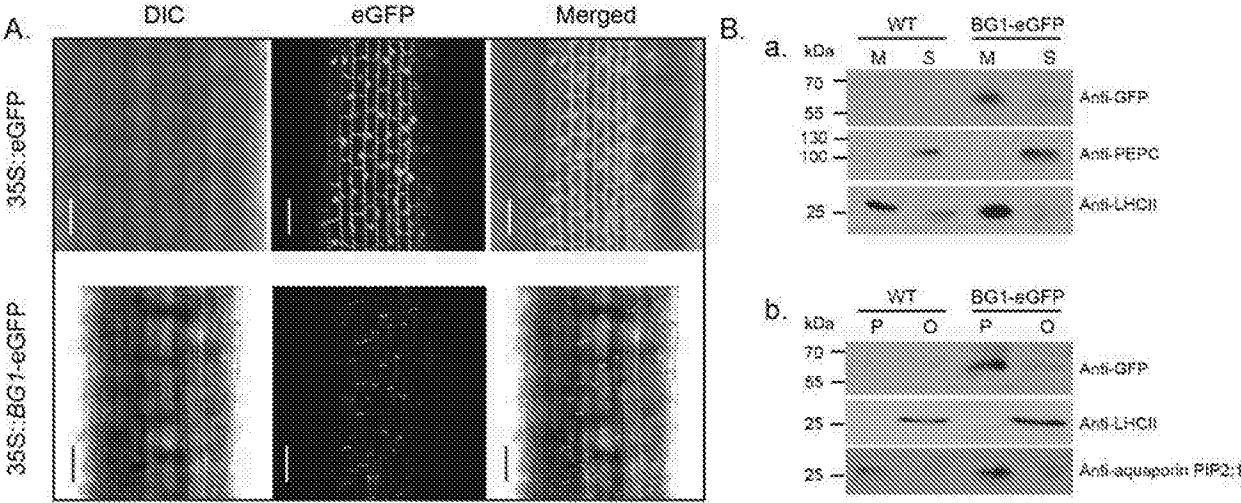


FIG. 8

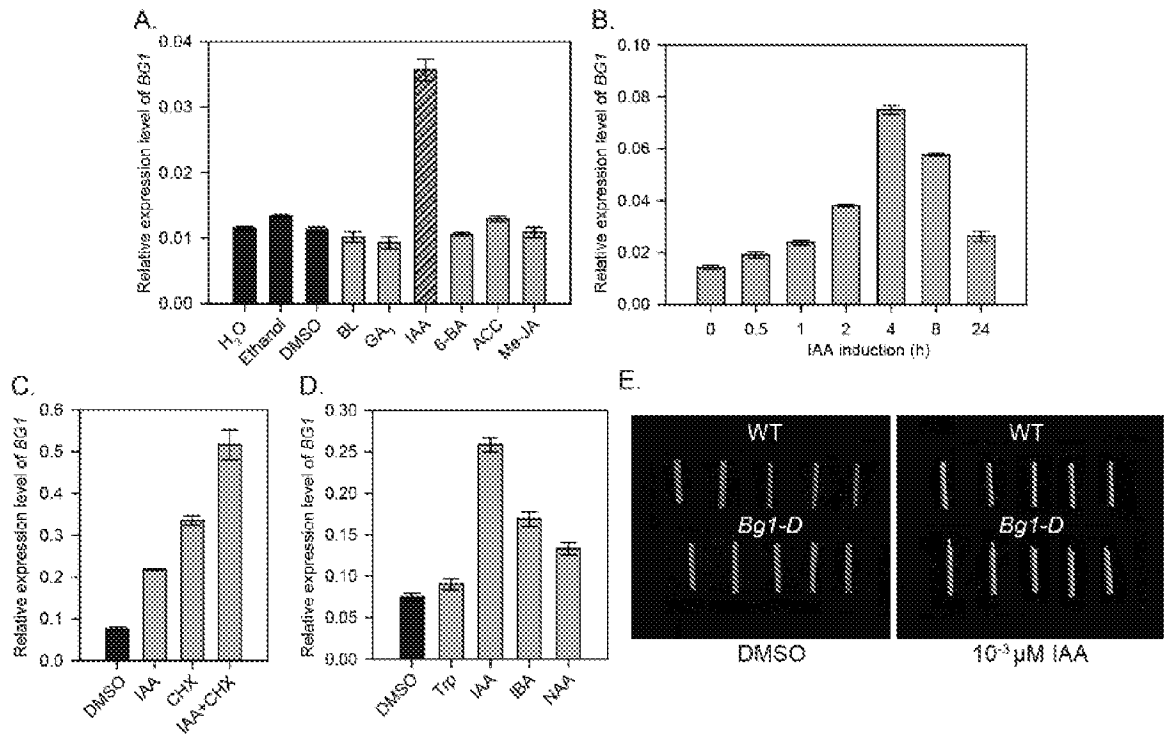


FIG. 9

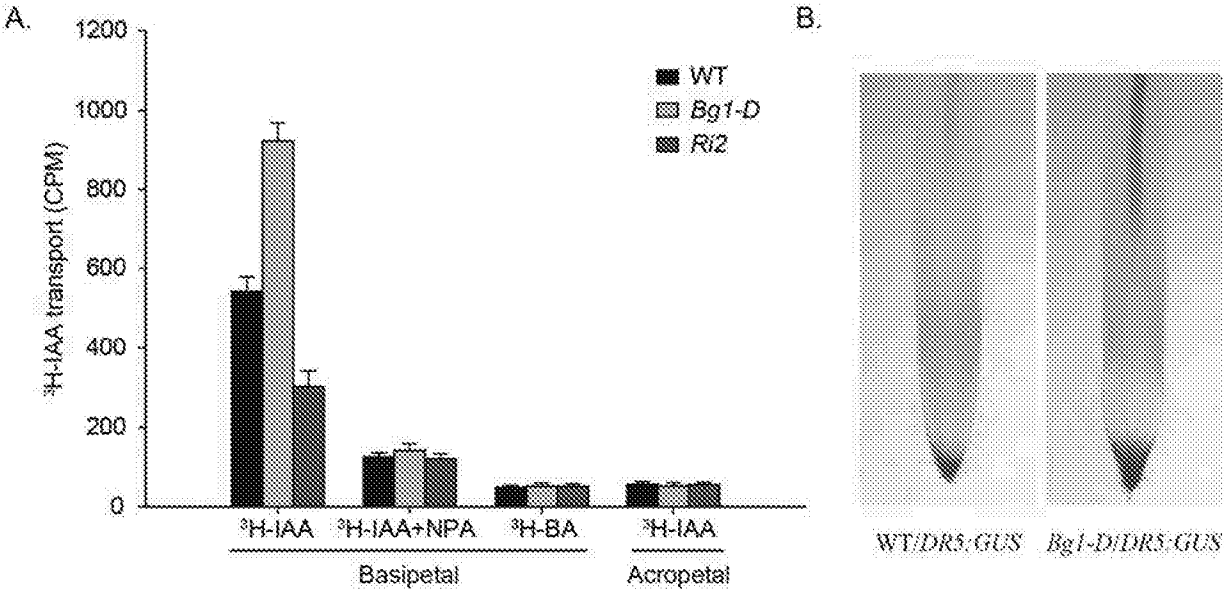


FIG. 10

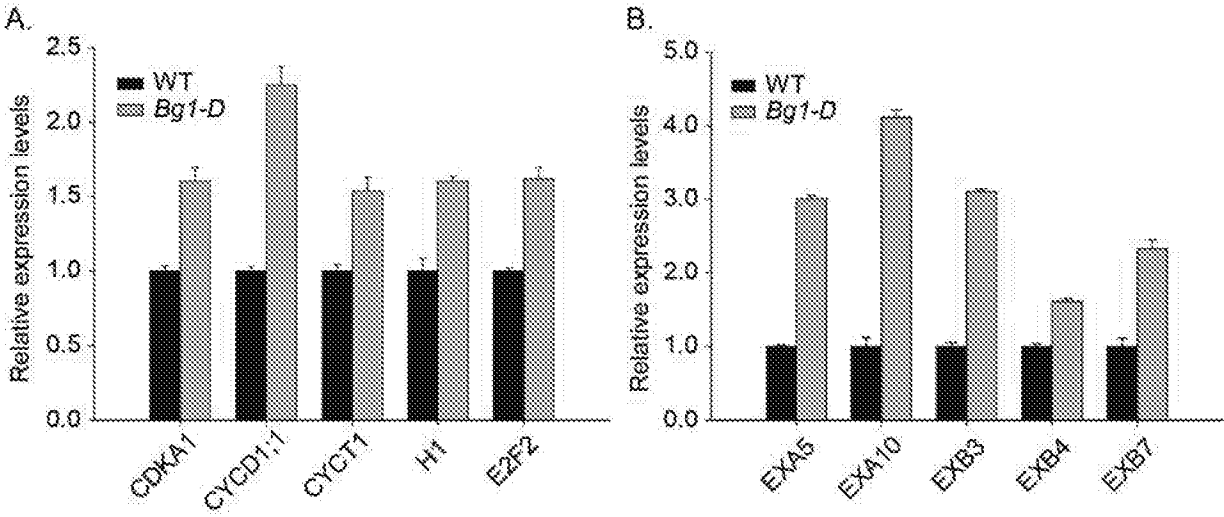


FIG. 11

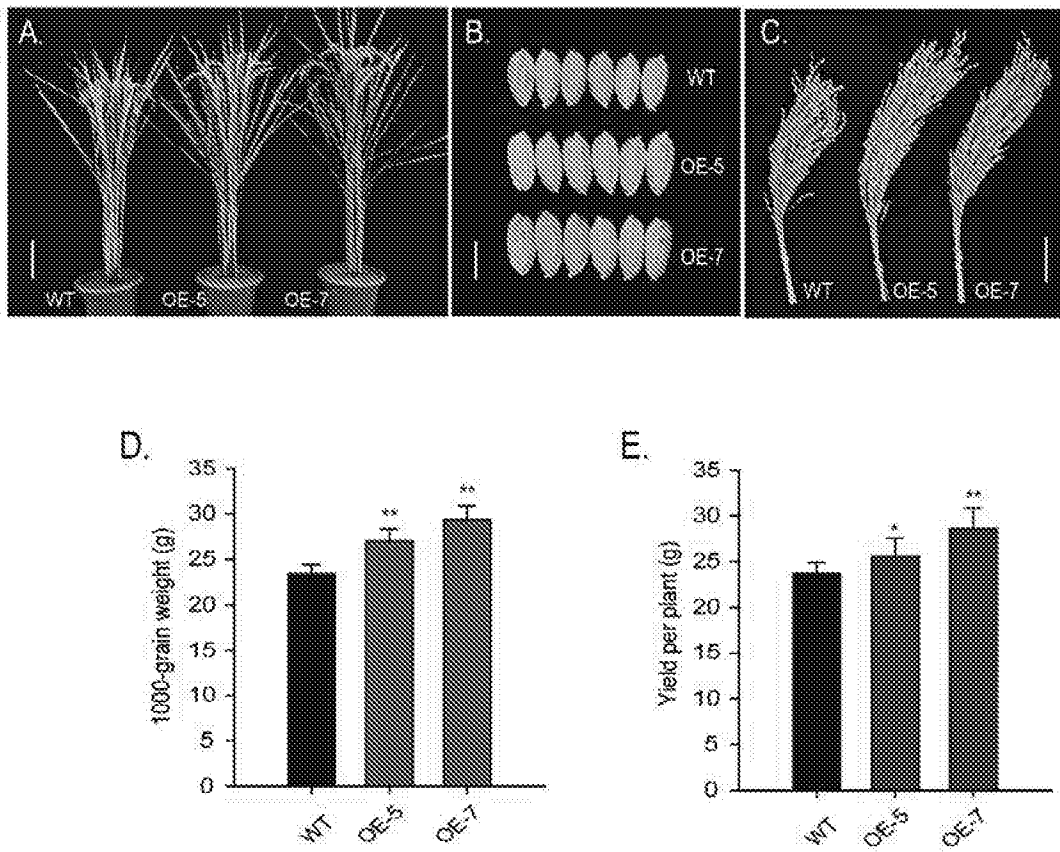


FIG. 12

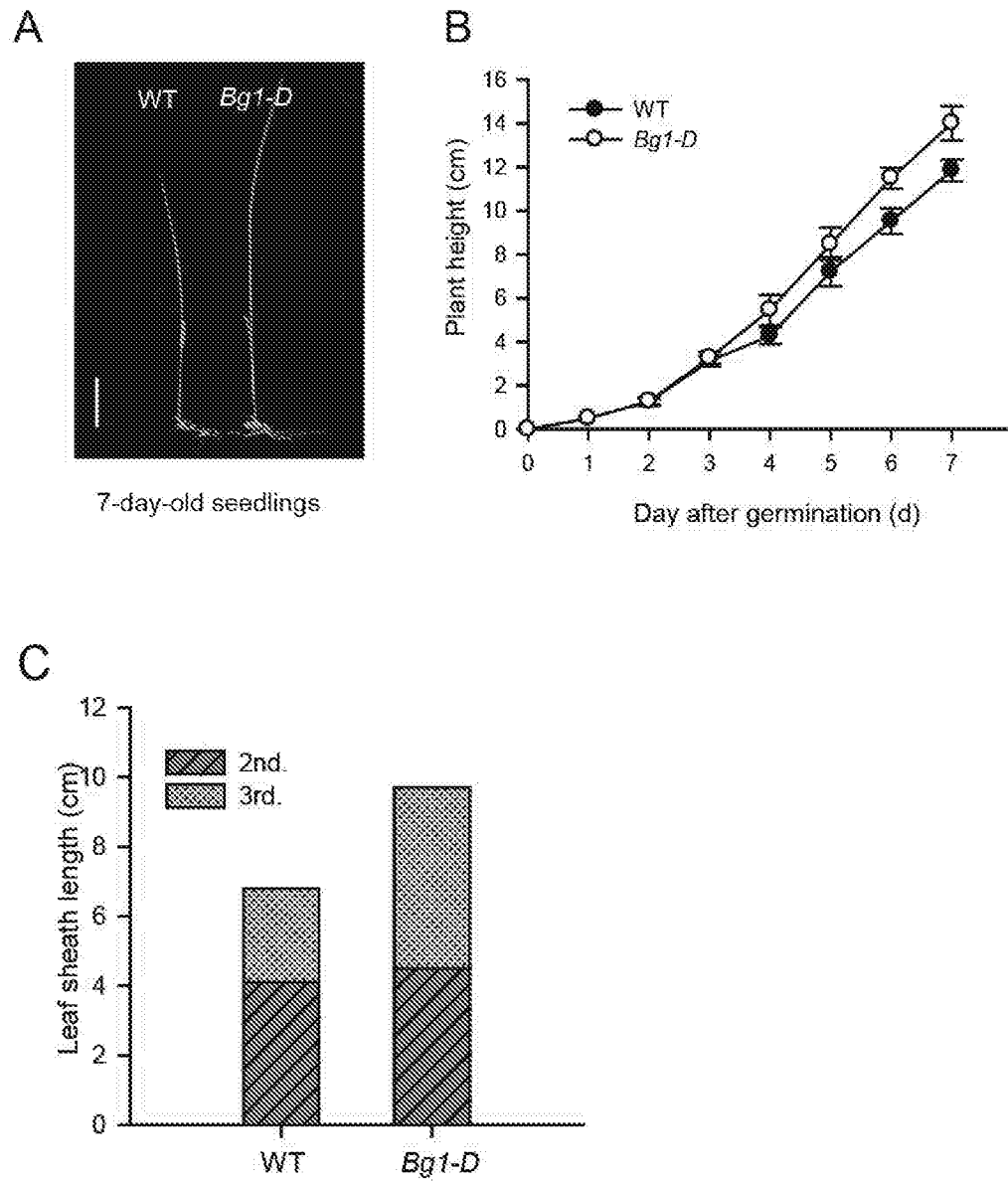
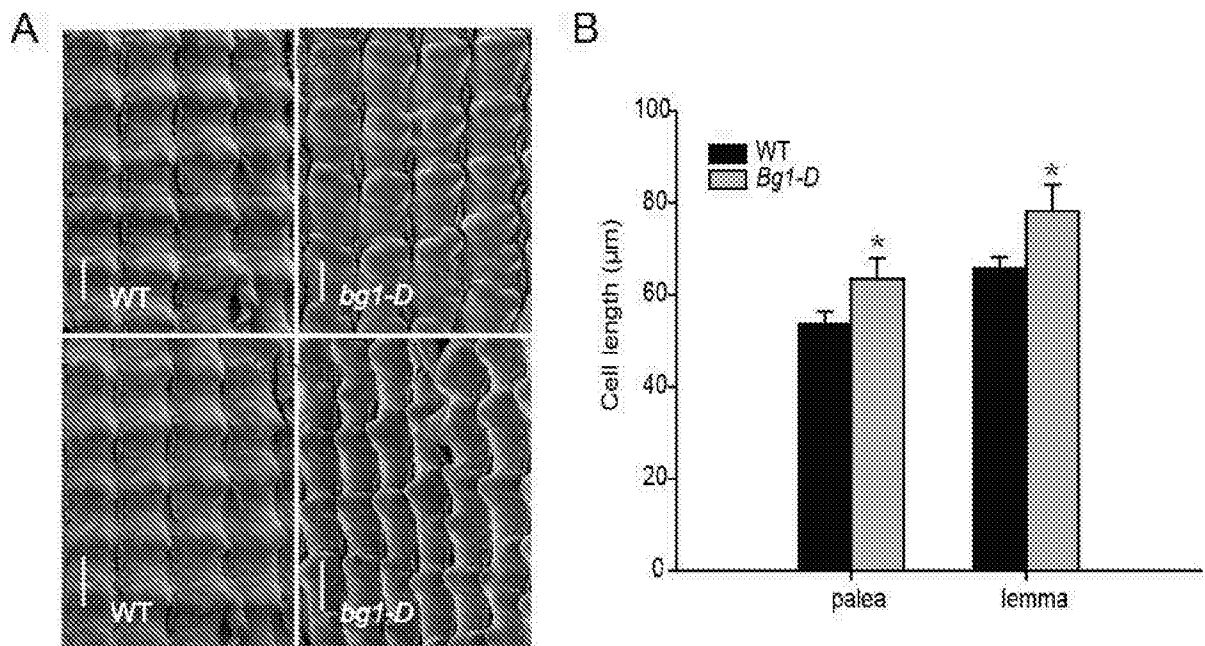
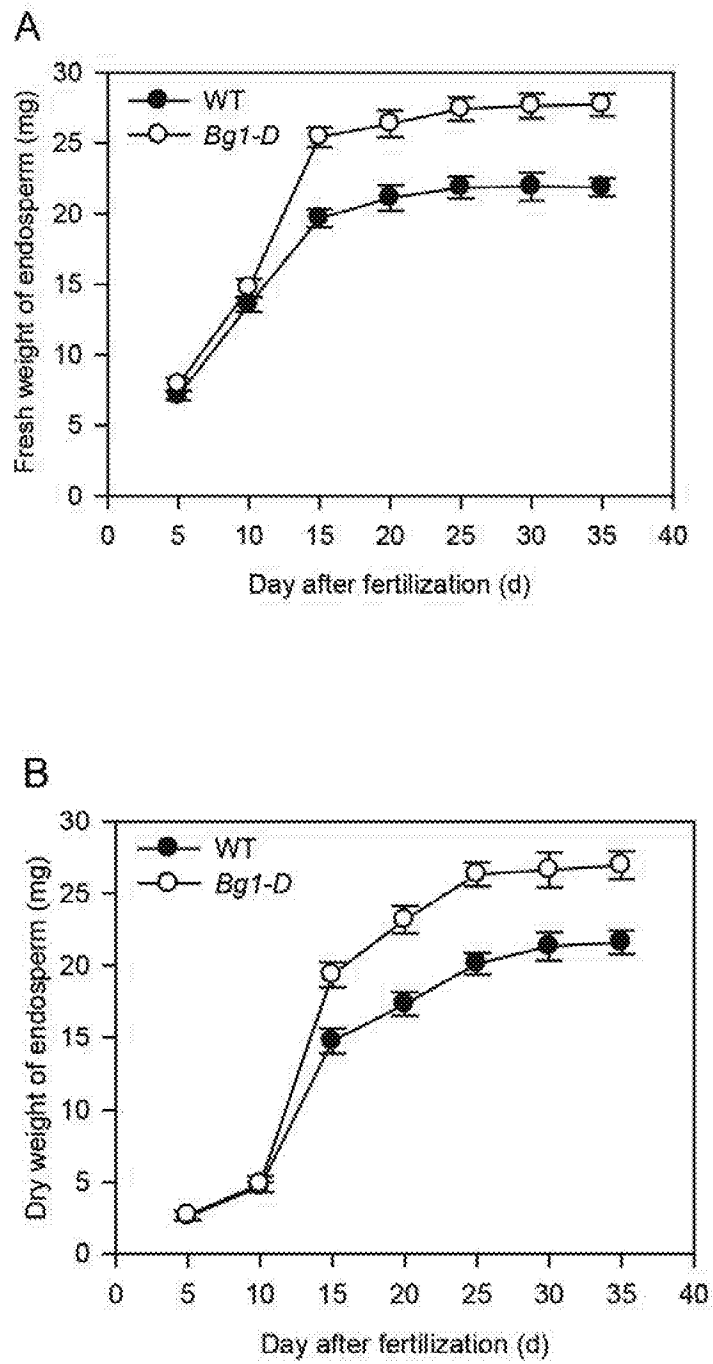


FIG. 13

**FIG. 14**

**FIG. 15**

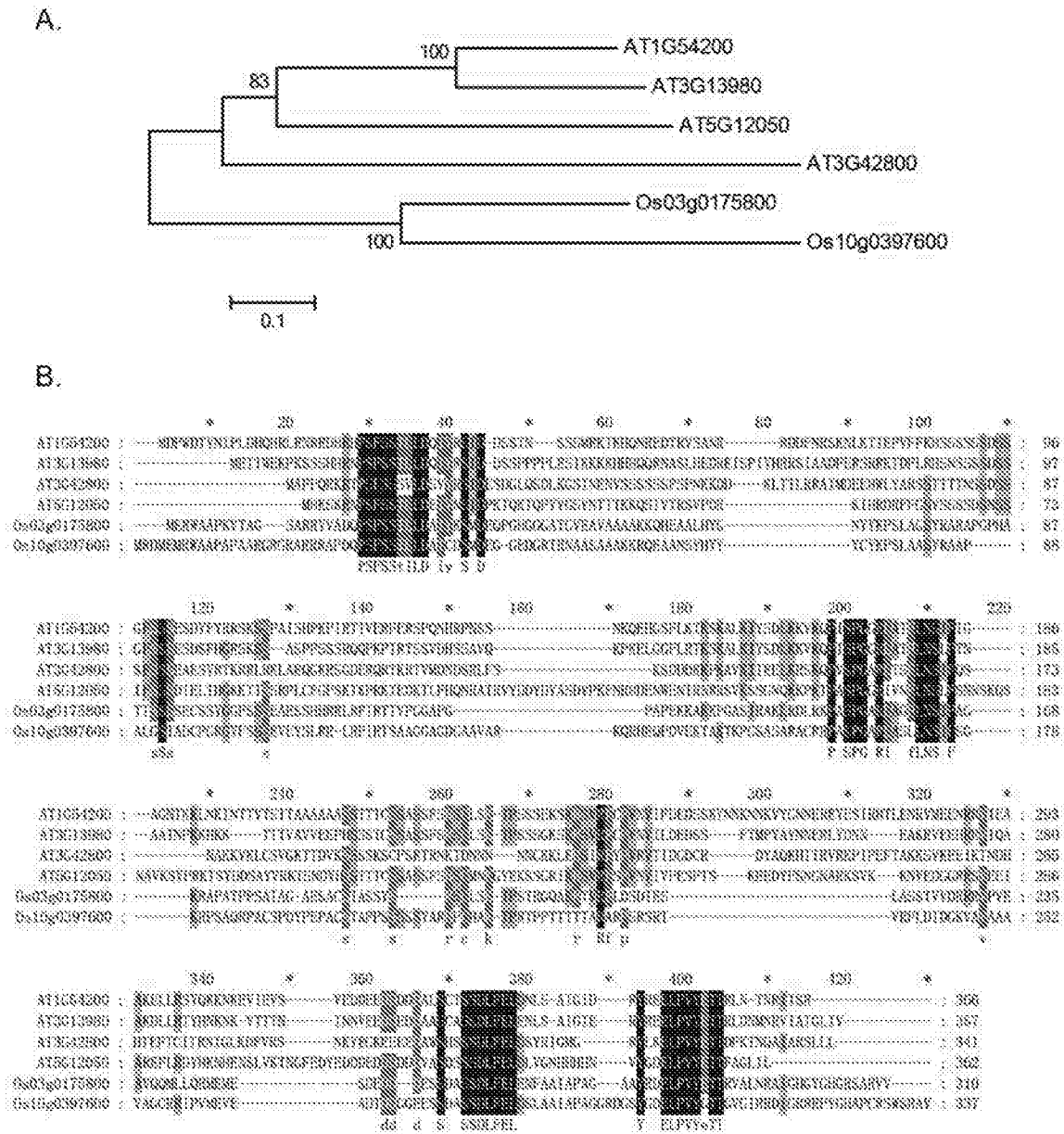


FIG. 16

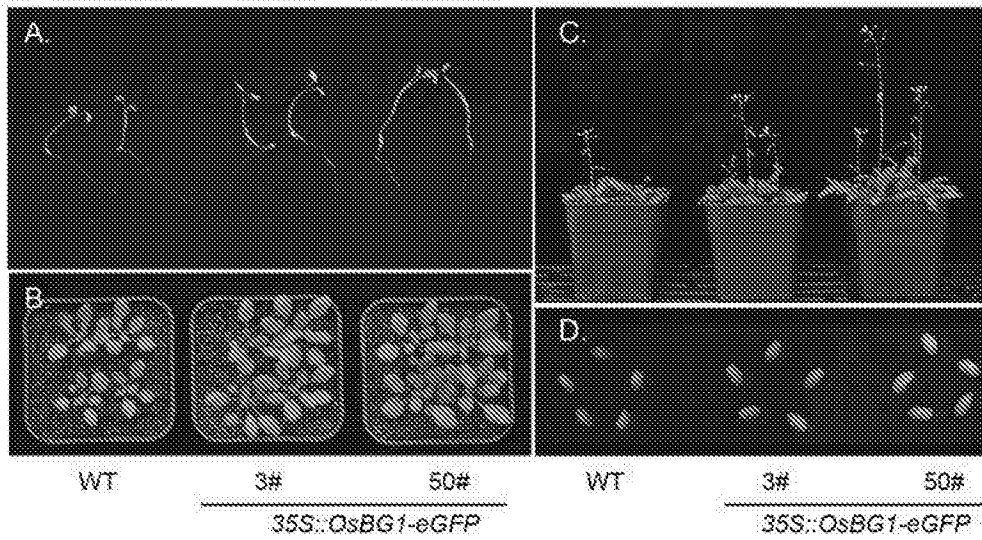


FIG. 17

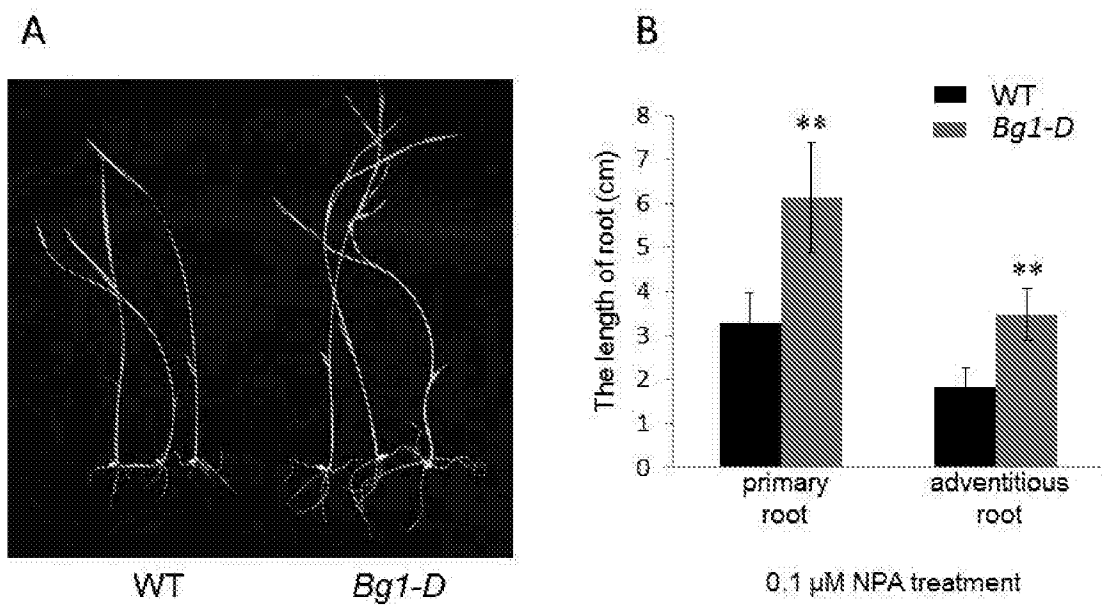
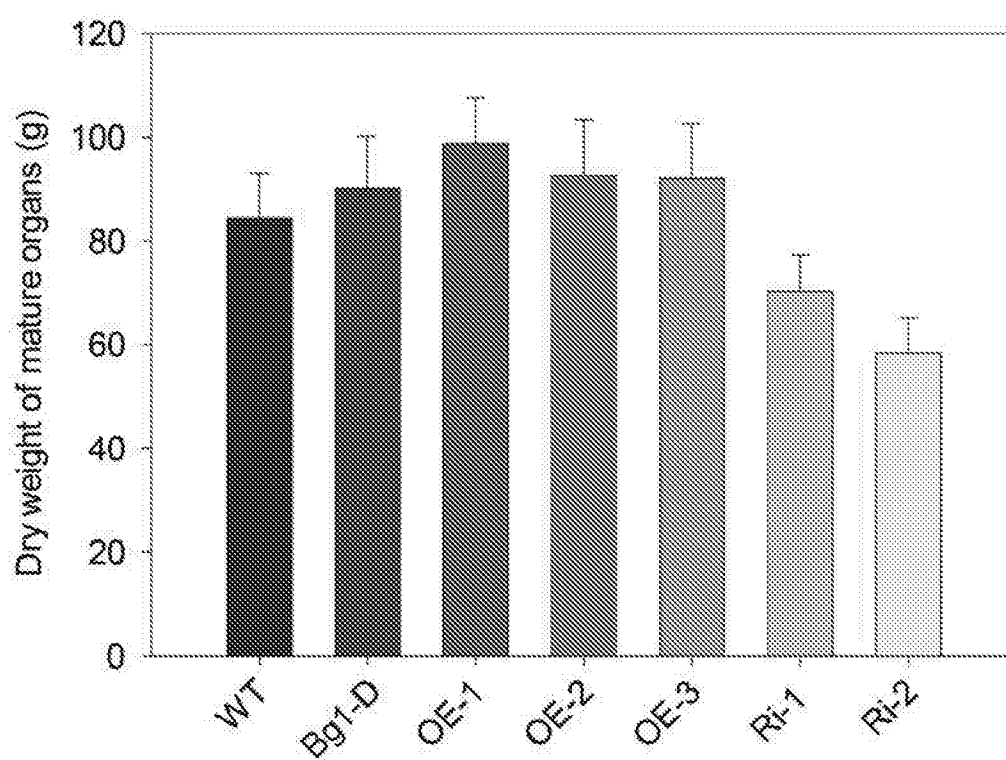


FIG. 18

**FIG. 19**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/083982

A. CLASSIFICATION OF SUBJECT MATTER

C07K 14/415(2006.01)i; C12N 15/29(2006.01)i; C12N 15/113(2010.01)i; C12N 15/82(2006.01)i; A01H 5/00(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)

C07K, C12N, A01H

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

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

CNABS;CPRSABS;CNTXT;DWPI;SIPOABS;EPTXT;WOTXT;USTXT;JPTXT;CNKI;PUBMED and keywords:rice, Oryza sativa L., maize, Zea mays L., BG1, big grain, yield, grain, etc. GENBANK;EMBL:SEQ ID Nos:1-75.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 103408648 A (INST GENETICS & DEV BIOLOG CAS) 27 November 2013 (2013-11-27) description, paragraph 0051-0053 and 0153, claims 1-10 and SEQ ID Nos: 1-6	1-26
X	"Genbank Accession No: NP_001049135.1" , 08 June 2010 (2010-06-08), see sequence	1-3, 8-11, 16-20
A	CN 101161675 A (SHANGHAI INST BIOL SCIENCES) 16 April 2008 (2008-04-16) the whole document	1-26
A	"Genbank Accession No: NP_001146458.1" , 16 June 2012 (2012-06-16), see sequence	1-26
A	"Genbank Accession No: DAA43664.1" , 14 November 2012 (2012-11-14), see sequence	1-26



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	"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
"E"	earlier application or patent but published on or after the international filing date	"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
"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)	"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
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

04 November 2014

Date of mailing of the international search report

19 November 2014

Name and mailing address of the ISA/CN

STATE INTELLECTUAL PROPERTY OFFICE OF THE
P.R.CHINA(ISA/CN)
6,Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088 China

Authorized officer

WANG,Ying

Facsimile No. (86-10)62019451

Telephone No. (86-10)62089434

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2014/083982

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	103408648	A	27 November 2013	None		
CN	101161675	A	16 April 2008	CN	101161675	B 04 July 2012