



(43) International Publication Date
17 February 2005 (17.02.2005)

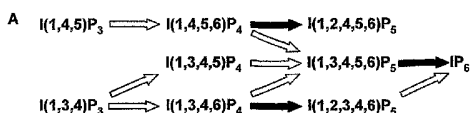
PCT

(10) International Publication Number
WO 2005/014794 A2

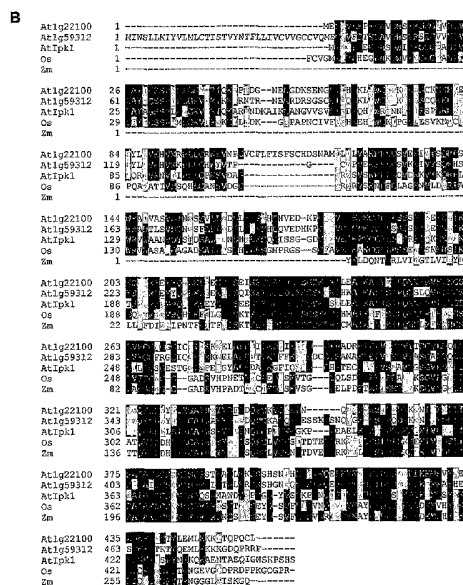
- | | |
|---|--|
| <p>(51) International Patent Classification⁷: C12N</p> <p>(21) International Application Number:
PCT/US2004/025756</p> <p>(22) International Filing Date: 9 August 2004 (09.08.2004)</p> <p>(25) Filing Language: English</p> <p>(26) Publication Language: English</p> <p>(30) Priority Data:
60/493,267 7 August 2003 (07.08.2003) US</p> <p>(71) Applicant (for all designated States except US): DUKE UNIVERSITY [US/US]; Office of Science & Technology, P.O. Box 90083, Durham, NC 27708 (US).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): YORK, John, D.</p> | <p>[US/US]; 3709 Sweeten Creek Road, Chapel Hill, NC 27514 (US). STEVENSON-PAULIK, Jill [US/US]; 215 King George Loop, Cary, NC 27511-6338 (US).</p> <p>(74) Agent: TAYLOR, Arles, A., Jr.; Jenkins, Wilson & Taylor, P.A., Suite 1400, University Tower, 3100 Tower Boulevard, Durham, NC 27707 (US).</p> <p>(81) Designated States (<i>unless otherwise indicated, for every kind of national protection available</i>): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.</p> <p>(84) Designated States (<i>unless otherwise indicated, for every kind of regional protection available</i>): ARIPO (BW, GH,</p> |
|---|--|

[Continued on next page]

(54) Title: GENERATION OF LOW PHYTATE PLANTS BY MOLECULAR DISRUPTION OF INOSITOL POLYPHOSPHATE KINASES



(57) Abstract: The present subject matter provides novel inositol phosphate kinase polypeptides (Ipk1 and Ipk2) and nucleic acid molecules encoding the same. The present subject matter also provides methods of using the novel polypeptides and 5 nucleic acid molecules. In some embodiments, the methods provide steps for producing a low phytate plant.



C

	M	leaf	stem	root	silique	flower
<i>AtIPK1</i>	+	+	+	+	+	+
<i>ACT2</i>	+	+	+	+	+	+



GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DESCRIPTION

GENERATION OF LOW PHYTATE PLANTS BY MOLECULAR DISRUPTION OF
INOSITOL POLYPHOSPHATE KINASES

5

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/493,267, filed August 7, 2003; the disclosure of which is incorporated herein by reference in its entirety.

10

GOVERNMENT INTEREST

This invention was made with Government support under Grant No.RO1 #HL-55672-07 and RO1 #HL-55672-05 awarded by National Heart, Lung, and Blood Institute, National Institutes of Health. The Government has certain rights in the invention.

15

TECHNICAL FIELD

The presently disclosed subject matter pertains to nucleic acid molecules encoding inositol polyphosphate kinase polypeptides. The presently disclosed subject matter also relates to methods of using the nucleic acid molecules and/or polypeptides and mutants thereof in transgenic plants to confer desirable agronomic traits, and more specifically to produce transgenic plants having reduced phytate levels, and/or increase the levels of non-phytate phosphorous in the transgenic plants. Low phytate content is a preferred trait in plants used for food or animal feed.

20

25

TABLE OF ABBREVIATIONS

(1,3,4,6)P ₄	-	inositol 1,3,4,6 tetrakisphosphate
5-ptase	-	inositol polyphosphate 5-phosphatase
AMV	-	Alfalfa Mosaic Virus
ATCC	-	America Type Culture Collection
BiP	-	binding polypeptide
bp	-	basepair(s)

30

	CAB	-	chlorophyll a/b binding
	CaMV	-	cauliflower mosaic virus
			carboxylase/oxygenase
	cDNA	-	complementary DNA
5	CDPK	-	calcium-dependent protein kinase
	cM	-	centimorgan(s)
	CMV	-	cytomegalovirus
	ColE1	-	a synthetic <i>E. coli</i> origin of replication
	DEAE	-	di-ethyl-amino-ethyl
10	DHFR	-	dihydrofolate reductase
	dsRNA	-	double-stranded RNA
	EDTA	-	ethylene diamine tetraacetic acid
	ELISA	-	enzyme-linked immunosorbent assay
	EMCV	-	encephalomyocarditis virus
15	EPSP	-	5-enol-pyruvyl shikimate-3-phosphate
	ER	-	endoplasmic reticulum
	EST	-	expressed sequence tag
	GS2	-	glutamine synthase 2
	GST	-	glutathione S-transferase
20	GUS	-	β -glucuronidase
	HPLC	-	high performance liquid chromatography
	HSV	-	herpes simplex virus
	I(1,2,3,4,6)P ₅	-	inositol 1,2,3,4,6 pentakisphosphate
	I(1,2,4,5,6)P ₅	-	inositol 1,2,4,5,6 pentakisphosphate
25	I(1,3,4)P ₃	-	inositol 1,3,4 trisphosphate
	I(1,3,4,5)P ₄	-	inositol 1,3,4,5 tetrakisphosphate
	I(1,3,4,5,6)P ₅	-	inositol 1,3,4,5,6 pentakisphosphate
	I(1,4,5)P ₃	-	inositol 1,4,5 trisphosphate
	I(1,4,5,6)P ₄	-	inositol 1,4,5,6 tetrakisphosphate
30	IP	-	inositol polyphosphate
	IP ₃	-	inositol trisphosphate
	IP ₄	-	inositol tetrakisphosphate

	IP ₅	-	inositol pentakisphosphate
	IP ₆	-	inositol hexakisphosphate
	kb	-	kilobase(s)
	k _{cat}	-	catalytic constant
5	K _m	-	Michaelis Constant
	MCMV	-	Maize Chlorotic Mottle Virus
	MDMV	-	Maize Dwarf Mosaic Virus
	MTL	-	metallothionein-like
	ORF	-	open reading frame
10	PCR	-	polymerase chain reaction
	PEG	-	polyethylene glycol
	PEPC	-	phosphoenol carboxylase
	pgk	-	phosphoglycerate kinase
	Pi	-	inorganic phosphate
15	PMI	-	phosphomannose isomerase
	Protox	-	protoporphyrinogen oxidase
	QTL	-	quantitative trait locus/loci
	RFLP	-	restriction fragment length polymorphism
	RNAi	-	RNA interference
20	RUBISCO	-	ribulose-1,5-bisphosphate
	SELEX	-	Systematic Evolution of Ligands by Exponential Enrichment
	Sf9	-	<i>Spodoptera frugiperda</i> cell line
	TBE	-	tris-borate-EDTA
25	TEV	-	Tobacco Etch Virus
	tk	-	thymidine kinase
	TMTD	-	tetramethylthiuram disulfide
	TMV	-	Tobacco Mosaic Virus
30	<u>Amino Acid Abbreviations, Codes, and Functionally Equivalent Codons</u>		

<u>Amino Acid</u>	<u>3-Letter</u>	<u>1-Letter</u>	<u>Codons</u>
Alanine	Ala	A	GCA GCC GCG GCU

	Arginine	Arg	R	AGA AGG CGA CGC CGG CGU
	Asparagine	Asn	N	AAC AAU
	Aspartic Acid	Asp	D	GAC GAU
	Cysteine	Cys	C	UGC UGU
5	Glutamic acid	Glu	E	GAA GAG
	Glutamine	Gln	Q	CAA CAG
	Glycine	Gly	G	GGA GGC GGG GGU
	Histidine	His	H	CAC CAU
	Isoleucine	Ile	I	AUA AUC AUU
10	Leucine	Leu	L	UUA UUG CUA CUC CUG CUU
	Lysine	Lys	K	AAA AAG
	Methionine	Met	M	AUG
	Phenylalanine	Phe	F	UUC UUU
	Proline	Pro	P	CCA CCC CCG CCU
15	Serine	Ser	S	ACG AGU UCA UCC UCG UCU
	Threonine	Thr	T	ACA ACC ACG ACU
	Tryptophan	Trp	W	UGG
	Tyrosine	Tyr	Y	UAC UAU
20	Valine	Val	V	GUA GUC GUG GUU

BACKGROUND

As a relatively cheap source of energy-dense and nutrient-rich foodstuff, cereal grains contribute more than 50% of worldwide energy intake and thus offer the most important food supply to the world population. The high nutritional value of seeds comes from the deposition of starch, lipids, proteins, and essential minerals during seed development. In this process, there is a corresponding accumulation of phytate, or *myo*-inositol hexakisphosphate (IP₆), which typically comprises >1% of the dry weight and is responsible for at least two-thirds of total seed phosphate.

Although the role of phytate reserves in the plant seed is not understood, its abundance in grain feed is known to cause nutritional and environmental problems. Phytate is largely undigestible by monogastric animals and it chelates essential minerals such as iron, zinc, and calcium, rendering them unavailable for absorption. In

developing countries, high phytate, grain-based diets are feared to exacerbate iron and zinc malnutrition and substitution of low-phytate maize for normal maize in the diet shows promise in improving zinc and iron absorption in human subjects. In developed countries where livestock is fed primarily grain-based feed, the excreted phytate
5 contributes to environmental phosphorus pollution by washing into surface waters where it accelerates eutrophication. The problem of unavailable seed phosphate in animal feed is usually overcome by supplementation with bioavailable forms of phosphate or with fungal phytase, an enzyme that degrades phytate.

Various other strategies have been used to deal with the high phytate content of
10 feed, including the production of transgenic animals (Golovan et al., 2001) and plants (Holm et al., 2002) that express increased levels of phytase. In another approach, to avoid phytate synthesis in the plant seed altogether, there have been low-phytate mutants generated by random mutagenesis in maize (Raboy et al. 2000), barley (Larson et al., 1998; and Hatzack et al., 2000), rice (Larson et al., 2000), and soybean
15 (Wilcox et al., 2000; and Hitz et al., 2002). In most cases, the site of the mutation is not known, although based on inositol phosphate analyses, many of the mutations are predicted to occur in early steps of phytate synthesis. Since precursors of phytate are known to participate in important events such as cytosolic and nuclear signaling, it is predicted that plants with mutations early in the phytate pathway could have pleiotropic
20 defects that would lead to decreased productivity.

Therefore, there is a need for the identification of genes encoding peptides involved late in the phytate production pathway that can be manipulated to decrease phytate production in food and animal feed plants. This and other needs in the art are addressed by the present disclosure.

SUMMARY

This Summary lists several embodiments of the presently disclosed subject matter, and in many cases lists variations and permutations of these embodiments. This Summary is merely exemplary of the numerous and varied embodiments. Mention
30 of one or more representative features of a given embodiment is likewise exemplary. Such an embodiment can typically exist with or without the feature(s) mentioned; likewise, those features can be applied to other embodiments of the presently disclosed

subject matter, whether listed in this Summary or not. To avoid excessive repetition, this Summary does not list or suggest all possible combinations of such features.

The presently disclosed subject matter provides nucleic acids and polypeptides involved with inositol phosphorylation and phytate production. In one embodiment, an isolated nucleic acid molecule encoding an inositol phosphate kinase polypeptide is provided. The nucleic acid molecule can be selected from the group consisting of:

- (a) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of one of even numbered SEQ ID NOs: 2-10 or a polypeptide at least 40% identical to even numbered SEQ ID NOs: 2-10 and having inositol phosphate kinase activity;
- (b) a nucleic acid molecule comprising a nucleic acid sequence of one of odd numbered SEQ ID NOs: 1-9;
- (c) a nucleic acid molecule that has a nucleic acid sequence at least 90% identical to the nucleic acid sequence of the nucleic acid molecule of (a) or (b);
- (d) a nucleic acid molecule that hybridizes to (a) or (b) under stringent hybridization conditions;
- (e) a nucleic acid molecule comprising a nucleic acid sequence complementary to (a); and
- (f) a nucleic acid molecule comprising a nucleic acid sequence that is the full reverse complement of (a).

In one embodiment, a vector is provided comprising the nucleic acids disclosed herein. In another embodiment, an expression cassette comprising at least one of the nucleic acids disclosed herein is provided. In some embodiments, the expression cassette is operably linked to a promoter. In some embodiments the promoter is a plant promoter. In some embodiments the promoter is a constitutive promoter. In some embodiments the promoter is a tissue-specific or a cell type-specific promoter. In some embodiments, the tissue-specific or cell type-specific promoter directs expression of the expression cassette in a location selected from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, seed, and combinations thereof.

In another embodiment, an isolated inositol phosphate kinase polypeptide having

inositol phosphate kinase activity is provided. The polypeptide is selected from the group consisting of:

- (a) an amino acid sequence of one of even numbered SEQ ID NOs: 2-10;
- (b) an amino acid sequence that is at least 40% identical to (a);
- 5 (c) an amino acid sequence encoded by a nucleotide sequence substantially identical to a nucleotide sequence of one of odd numbered SEQ ID NOs: 1-9; and
- (d) an amino acid sequence encoded by a nucleic acid molecule capable of hybridizing under stringent conditions to a nucleic acid molecule of one of
- 10 odd numbered SEQ ID NOs: 1-9 or to a sequence fully complementary thereto.

A method for producing the polypeptides disclosed herein is also provided. The method comprises growing cells comprising an expression cassette under suitable growth conditions, the expression cassette comprising a nucleic acid molecule

15 disclosed herein; and isolating the polypeptide from the cells.

In yet another embodiment of the presently disclosed subject matter, an isolated polypeptide having or comprising at least about 90% amino acid identity with a polypeptide as disclosed in Examples 6-11 of the subject application, over an entire sequence or functional fragment thereof, wherein the polypeptide functions as an

20 inositol phosphate kinase is provided.

In yet another embodiment of the presently disclosed subject matter, a transgenic plant cell comprising a homozygous disruption in at least one endogenous inositol phosphate kinase gene homologous to the nucleic acid molecules disclosed herein, wherein the disruption substantially inhibits the expression of a functional

25 inositol phosphate polypeptide is provided. A transgenic plant comprising at least one of the transgenic plant cells is further provided herein. In some embodiments, the transgenic plant comprises a decreased level of phytic acid and/or an increased level of non-phytic acid when compared to a non-transformed parental plant. In some embodiments, the homozygous disruption is tissue-specific. In some embodiments, the

30 tissue-specific disruption is in a tissue selected from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, seed, and combinations thereof. In some embodiments, the tissue is seed and the seed

comprises a decreased level of phytic acid and/or an increased level of non-phytic acid when compared to a non-transformed parental plant.

In some embodiments, the transgenic plant is a plant selected from the group consisting of *Arabidopsis thaliana*, corn (*Zea mays*), *Brassica* sp., alfalfa (*Medicago sativa*), rice (*Oryza sativa* ssp.), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*),
 5 peanut (*Arachis hypogaea*), cotton, sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilis*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*),
 10 cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed (*Lemna*), and barley. In some embodiments, the plant is selected from the group consisting of a vegetable, an ornamental, and a conifer. In some embodiments, the vegetable is selected from the group consisting of tomatoes, lettuce, guar, locust
 20 bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, green bean, lima bean, pea, and members of the genus *Cucumis*. In some embodiments, the ornamental is selected from the group consisting of impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo,
 25 Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, azalea, hydrangea, hibiscus, rose, tulip, daffodil, petunia, carnation, poinsettia, and chrysanthemum. In some embodiments, the conifer is selected from the group consisting of loblolly pine, slash pine, ponderosa pine, lodgepole pine, Monterey pine, Douglas-fir, Western hemlock, Sitka spruce, redwood,
 30 silver fir, balsam fir, Western red cedar, and Alaska yellow-cedar. In some embodiments, the homozygous disruption to the at least one endogenous inositol phosphate kinase gene is to at least a gene encoding an Ipk1 polypeptide and a gene

encoding an lpk2 polypeptide

In still another embodiment of the presently disclosed subject matter, a method of modulating production of phytate in a plant is provided. The method comprises, modulating the enzymatic activity of at least one inositol phosphate kinase polypeptide
5 selected from the group consisting of lpk1 and lpk2.

In another embodiment of the presently disclosed subject matter, a method of producing a plant with low levels of phytate is provided. The method comprises modulating in the plant the enzymatic activity of at least one inositol phosphate kinase polypeptide selected from the group consisting of lpk1 and lpk2.

10 In another embodiment of the presently disclosed subject matter, a method of screening a plurality of compounds for a modulator of the enzymatic activity of an inositol phosphate kinase polypeptide is provided. The method comprises in some embodiments:

- (a) providing a library of test compounds;
- 15 (b) contacting an inositol phosphate kinase polypeptide selected from the group consisting of lpk1 and lpk2 with each test compound;
- (c) detecting an interaction between a test compound and the inositol phosphate kinase polypeptide;
- (d) identifying a test compound that interacts with the inositol
20 phosphate kinase polypeptide; and
- (e) isolating the test compound that interacts with the inositol phosphate kinase polypeptide, whereby a plurality of compounds is screened for a modulator of inositol phosphate kinase polypeptide enzymatic activity.

25 Some of the objects of the presently disclosed subject matter having been stated hereinabove, and which are addressed in whole or in part by the present subject matter, other objects will become evident as the description proceeds when taken in connection with the accompanying drawings as best described hereinbelow.

30 BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A through 1C disclose the identification and sequence alignment of plant IP₅ 2-kinases.

Figure 1A is a schematic depicting the combined IP kinase activities of AtIpK1 (black arrows), AtIpK2 (grey arrows), and I(1,3,4)P₃ 5/6 kinase (white arrows) described herein can synthesize IP₆ *in vitro* through the indicated routes from I(1,4,5)P₃ and I(1,3,4)P₃.

Figure 1B is a multi-sequence alignment of the family of IpK1 proteins. GenBank accession numbers for the plant sequences are: OsIpK1 – AK102842; ZmIpK1 – AY104429; AtIpK1 – AY093362; At1g22100 – NM_102060; At1g59312 – NM_104639; HsIpK1 – AF520811.

Figure 1C is digital image of gel electrophoresis showing that *AtIPK1* is expressed throughout Arabidopsis tissues. Semi-quantitative, non-competitive RT-PCR was performed from 250 ng total RNA from the indicated tissues. As a control for RNA quantification and comparison, actin (*ACT2*) was amplified.

Figures 2A-2C depict that AtIpK1 is an IP₄ and IP₅ 2-kinase.

Figure 2A is a data readout of an I(1,3,4,5,6)P₅ substrate assay. 10 μM of [³H]-I(1,3,4,5,6)P₅ (7000 CPM) was incubated with 0.7 pmol purified recombinant GST-AtIpK1 for 30 min at 37°C. The reaction products were resolved by HPLC on a Partisphere strong anion exchange (SAX) HPLC column. IPs were identified based on comparison to known standards.

Figure 2B is a data readout of an I(1,3,4,6)P₄ substrate assay. 10 μM was incubated with 0.7 pmol GST-AtIpK1 and then analyzed by HPLC.

Figure 2C is a data readout of an I(1,4,5,6)P₄ substrate assay. [³H]-I(1,4,5,6)P₄ and ³H-I(1,3,4,5,6)P₅ was incubated with 0.7 pmol GST-AtIpK1β for 30 min at 37 °C and then separated by HPLC. Both substrates were phosphorylated to produce I(1,2,4,5,6)P₅ and IP₆, respectively.

Figures 3A-3D depict dose dependence of specific IP species formation. GST-AtIpK1 and GST-AtIpK2β were incubated with 10 μM [³H]-I(1,4,5)P₃.

Figure 3A is a data readout of an experiment wherein 1.6 pmol AtIpK2β and 1.3 pmol AtIpK1 were incubated with I(1,4,5)P₃ (top panel) for 2 hr at 37°C. The reaction (lower panel) was stopped by boiling for 1 min and then was separated by Partisphere SAX HPLC.

Figure 3B is a data readout of an experiment wherein 3.2 pmol AtIpk2 β and 1.3 pmol AtIpk1 was incubated with I(1,4,5)P₃ as above for 1 hr at 37°C, then stopped and analyzed by HPLC.

Figure 3C is a data readout of an experiment wherein 8 pmol AtIpk2 β and 1.3 pmol AtIpk1 were incubated with I(1,4,5)P₃ as above for 1 hr at 37°C, and then analyzed by HPLC.

Figure 3D is a data readout showing *in vitro* synthesis of IP₆ from I(1,3,4)P₃ by I(1,3,4)P₃ 5/6-kinase, AtIpk2, and AtIpk1. 10 μ M [³H]-I(1,3,4)P₃ (3,500 CPM) (top panel) was incubated with 2-4 pmol each of purified recombinant mammalian I(1,3,4)P₃ 5/6-kinase, AtIpk2 β , and AtIpk1 for 1 hr at 37° C (lower panel).

Figures 4A-4C show the molecular characterization of an *AtIPK1* T-DNA insertional mutant.

Figure 4A is a schematic of the T-DNA insertional element of SALK_065337 mapped 77 nt upstream of the stop codon of *AtIPK1*.

Figure 4B is a digital image of a Northern blot analysis of seedling RNA from wild type and *atipk1-1* Arabidopsis (ecotype Columbia-0) mutants probed with ³²P-labeled *AtIPK1* cDNA or actin (*ACT2*) cDNA.

Figure 4C is a digital image of a Southern blot of T-DNA insertion. Ten μ g of genomic DNA was digested with SpeI. Left panel is the Southern blot probed with the 5' UTR of *AtIPK1*, which yields a 6.4 kb band and indicates complete digestion of genomic DNA. The right panel is the same Southern probed with the sequence for the T-DNA. A single T-DNA element would yield a 6.8 kb band. The band identified here is greater than 15 kb, indicating at least three T-DNA's are inserted into the same site of the *AtIPK1* gene. Lane 1 is wild type genomic DNA. Lanes 2-5 are the genomic DNA from five independent T5 *atipk1-1* plants.

Figures 5A-5C show the effect of T-DNA insertion on IP synthesis.

Figure 5A is a data readout of an experiment wherein wild type and *atipk1-1* T₄ seeds were germinated in liquid MS salts with 0.4 mCi/ml [³H]-myo-inositol for 6 days and inositol phosphates were harvested and analyzed by Partisphere SAX HPLC. IP_{3x} is an unknown species.

Figure 5B is a data readout of an experiment wherein individual mature siliques (containing embryos in the late bent-cotyledon stage from wild type (top panel) and

atipk1-1 plants (lower panel) were labeled with [^3H]-myo-inositol as described in the methods and the IPs were harvested and analyzed by Partisphere SAX HPLC.

Figure 5C is a data readout of an experiment wherein individual immature seeds containing embryos in the late bent cotyledon stage of development were dissected from siliques that had been labeled for 2 days with 2 mCi/ml [^3H]-myo-inositol. The IPs were harvested from the seeds and analyzed by HPLC as described.

Figures 6A and 6B are data readouts of a comparison of IP₆ synthesis in developing Arabidopsis seeds. Individual seeds were dissected from wild type (Figure 6A) and *atipk1-1* (Figure 6B) siliques that had been labeled with [^3H]-myo-inositol for two days. IPs were extracted and analyzed by Partisphere SAX HPLC and identified based on known standards. Top panels are the IP profiles from seeds with embryos in the early-bent cotyledon stage and the bottom panels are from seeds with embryos in the late-bent cotyledon stage.

Figures 7A and 7B are data readouts of an identification of accumulated IP species in the *atipk1-1* mutant. [^3H]-IP extracts from *atipk1-1* siliques were incubated with 1.3 pmol GST-AtIpK1 (Figure 7A), or the I(3,4,5,6)P₄ 1-kinase (Figure 7B) for 60 min at 37 °C. The reaction products were separated by Partisphere SAX HPLC. Since IP₅ was phosphorylated by AtIpK1, its identity is I(1,3,4,5,6)P₅. Since IP₄ was phosphorylated by AtIpK1 to make IP₅x and the I(3,4,5,6)P₄ 1-kinase to make I(1,3,4,5,6)P₅, it is mainly comprised of I(3,4,5,6)P₄. Neither IP₄ nor IP₅ were phosphorylated by active AtIpK2 β , indicating that IP₄x is not I(1,4,5,6)P₄, I(1,3,4,5)P₄, or I(1,3,4,6)P₄, which are all substrates of AtIpK2 β (Stevenson-Paulik et al., 2002).

Figure 8 is a data readout from the non-radioactive mass IP analysis of mature desiccated seed extracts. Wild type (top panel) or T₄ *atipk1-1* (bottom panel) seed extracts were separated by an IonPac AS7 anion-exchange column and detected by metal-dye chelation and measured at 550 nm. Data points are plotted as the negative absorbance values.

Figures 9A and 9B show the affect of *AtIPK1* T-DNA insertion on vegetative growth. T₄ seeds were germinated on a mixture of sand and vermiculite through sub irrigation until the cotyledons had emerged. Plants were then top-fed half-strength Hoagland's solution daily and were grown at 20°C under a 14-hr light (240 $\mu\text{mol}/\text{m}^2 \text{ s}^{-1}$) and 10-hr dark cycle in the Duke University National Phytotron.

Figure 9A is a series of photographs of a top view of 4 week-old plant that is heterozygous for the *AtIPK1* T-DNA insertion (left panel, left plant), wild type (left panel, right plant), *atipk1-1* (middle panel), *atipk1-1* plant constitutively expressing *AtIPK1* (right panel).

5 Figure 9B is a data readout and a series of photographs of a complementation of IP synthesis and growth phenotype in *atipk1-1* by constitutive expression of *AtIPK1*. *AtIPK1* expression is driven by the constitutive 35S promoter. Left panels are HPLC chromatograms of IP profiles from [³H]-inositol-labeled 5-day-old seedlings. Right panel are leaves from 5-week-old plants grown in half-strength Hoagland's solution as described. Top panel and leaf row are wild type plants transformed with the empty pBART expression vector. Middle panel and leaf row are *atipk1-1* plants transformed with the pBART expression vector harboring the *AtIPK1* gene. Bottom panel and leaf row are *atipk1-1* plants transformed with the empty pBART expression vector.

15 Figures 10A and 10B show a conditional growth defect and altered phosphate levels of *atipk1-1*.

Figure 10A is a photograph of leaves from 5-week-old wild type and *atipk1-1* plants grown as described above in half-strength Hoagland's solution with variable phosphate concentrations as indicated.

20 Figure 10B is a bar graph of inorganic phosphate concentrations of the leaves from plants grown in conditions described in Figure 10A.

Figures 11A and 11B show root hair growth in response to variable phosphate levels.

25 Figure 11A is a set of representative images of main root and root hairs from wild type (left panel) and *atipk1-1* plants grown vertically on MS-agar plates containing 10 μ M phosphate. Bar indicates 0.2 mm.

Figure 11B is a bar graph showing root hair lengths of wild type and *atipk1-1* plants grown vertically on MS-agar plates containing various phosphate concentrations as indicated. n=12 for each genotype. * indicates p < 0.001.

Figure 12A is a schematic of T-DNA insertion into *atipk2 β* locus.

30 Figure 12B is a Northern Blot analysis of mRNA isolated from wild-type, *atipk2 β -1*, or *atipk1 β -1* plants.

Figure 13 is a schematic of IP determination methods used in Example 5.

Figures 14A and 14B are spectral data and a schematic, respectively, of a non-radioactive method of measuring Phytate (IP_6) levels in mature seeds derived from wild-type or kinase mutant plants.

Figure 15 is a series of photographs of *ipk1/ipk2* double mutant plants.

5 Figure 16 is a bar graph of total seed phytate (IP_6) levels calculated from *ipk1*, *ipk2* and *ipk1/ipk2* double mutant seeds.

Figures 17A and 17B are HPLC data and a bar graph, respectively, of total seed extracts as described in Example 5.

10 Figures 18-23 depict various alignments between translated reading frames from various plants and the genes disclosed herein, as outlined in Examples 6-11.

DETAILED DESCRIPTION

The presently disclosed subject matter will be now be described more fully hereinafter with reference to the accompanying Examples, in which representative
15 embodiments of the presently disclosed subject matter are shown. The presently disclosed subject matter can, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the presently disclosed subject matter to those skilled in
20 the art.

All publications, patent applications, patents, and other references cited herein are incorporated by reference in their entireties.

The presently disclosed subject matter relates, in part, to newly identified polynucleotides and polypeptides; variants and derivatives of these polynucleotides and
25 polypeptides; processes for making these polynucleotides and these polypeptides, and their variants and derivatives; molecules that bind these polynucleotides and polypeptides including antagonists; methods of screening for these molecules; transgenic plants incorporating these polynucleotides and polypeptides as well as transgenic plants having endogenous homologues of these polynucleotides disrupted;
30 and uses of these polynucleotides, polypeptides, variants, derivatives and antagonists. In particular, in these and in other regards, the presently disclosed subject matter relates to polynucleotides and polypeptides of the phytate metabolic pathway, most

particularly with the enzymes inositol phosphate kinase 1 (Ipk 1) and inositol phosphate kinase 2 (Ipk 2) and genes encoding same.

Described herein is the molecular and biochemical characterization of the machinery plants need to produce phytate, inositol hexakisphosphate (IP₆). As phytate is a major antinutrient in grains used for feeding livestock, thus it has been a long-standing goal to generate a plant seed that lacks phytate. Disclosed herein is disruption of a single gene, IPK1, which is required for the last step in the biosynthetic pathway, resulting in seeds with greater than a 90% reduction in phytate levels. Disruption of Ipk2 β , an inositol polyphosphate kinase that functions as the penultimate step in the synthesis of phytate, also results in low phytate seeds. Furthermore, simultaneous disruption of both AtIpk1 and AtIpk2b, as described herein, results in the 95% ablation of phytate and other inositol polyphosphates in seeds with minimal yield effects.

In one example of the presently disclosed subject matter, the gene and gene product of inositol polyphosphate 2-kinase, designated AtIPK1, from *Arabidopsis thaliana* has been identified, isolated, and characterized. As the enzyme that catalyzes the last step in inositol hexakisphosphate (IP₆ or phytate) biosynthesis, Ipk1, is an attractive target for the genetic manipulation of phytate synthesis in plant seeds. To generate a low-phytate plant, a mutant *Arabidopsis thaliana* line with a foreign DNA element (T-DNA) stably inserted in the AtIPK1b, gene, has been identified and obtained, and shown herein to interrupt transcription of the gene by at least 90%. The mutation also causes a greater than 90% loss of phytate in seeds (developing and mature), and whole seedlings. This approach to generating a low-phytate plant by disrupting the last step in phytate synthesis is novel and can be applied to agriculturally relevant plants to alleviate the negative impact of high phytate grain diets on animal nutrition and environmental phosphorus pollution.

Additionally, gene and gene products of two related inositol polyphosphate 6-/3-/5-kinases, designated AtIPK2 α and AtIPK2 β , from *Arabidopsis thaliana* have been identified, isolated, and characterized and are disclosed herein. As partially redundant enzymes that catalyze the penultimate step in phytate biosynthesis both AtIPK2 α and AtIpk2 β are attractive targets for the genetic manipulation of phytate synthesis in plant seeds. To generate a low-phytate plant, a mutant *Arabidopsis thaliana* line with a

foreign DNA element (T-DNA) stably inserted in the AtIPK1b, gene, has been identified and obtained, which interrupts transcription of the gene by at least 90%. The mutation causes a significant loss of phytate in seeds (developing and mature), but not seedlings, thereby appearing seed specific. The mutant atipk2 β plant has no detectable adverse growth phenotype and seed yields are normal. This approach to generating a low-phytate plant by disrupting the penultimate step in phytate synthesis is novel and may be applied to agriculturally relevant plants to alleviate the negative impact of high phytate grain diets on animal nutrition and environmental phosphorus pollution.

Plants with T-DNA disruptions in both AtIPK1, and AtIPK2 β are also disclosed herein and produced through breeding of single mutant lines described above. The double mutant plants are fertile, germinate normally and have seeds of nearly normal size. Examination of seed inositol phosphates demonstrates a 95% reduction of phytate with a minimal accumulation of IP₅, IP₄, IP₃ and IP₂. Collectively, these phenotypes represent an improvement over other described low phytate plants, especially in the area of IP₆ reduction and seed yield.

Identification of the gene sequence and corresponding amino acid sequence of the enzyme involved in the last step of phytate synthesis in *Arabidopsis* is useful to identify orthologs in other plant species. Furthermore, now identification of the biosynthetic pathway of IP₆ production, provides an advantage over random mutagenesis strategies in that a targeted approach in engineering altered gene expression (through antisense or post-transcriptional gene-silencing technologies) at multiple steps of the pathway and at specific developmental stages and tissues where phytate synthesis is greatest is possible. By spatiotemporally controlling the gene inactivation, a plant can be engineered to produce seed with reduced phytate, but avoid the loss of inositol phosphate production in other tissues where it may be vital for normal plant growth.

I. General Considerations

A goal of functional genomics is to identify genes controlling expression of organismal phenotypes, and functional genomics employs a variety of methodologies including, but not limited to, bioinformatics, gene expression studies, gene and gene

product interactions, genetics, biochemistry, and molecular genetics. For example, bioinformatics can assign function to a given gene by identifying genes in heterologous organisms with a high degree of similarity (homology) at the amino acid or nucleotide level. Studies of the expression of a gene at the mRNA or polypeptide levels can
5 assign function by linking expression of the gene to an environmental response, a developmental process, or a genetic (mutational) or molecular genetic (gene overexpression or underexpression) perturbation. Expression of a gene at the mRNA level can be ascertained either alone (for example, by Northern analysis) or in concert with other genes (for example, by microarray analysis), whereas expression of a gene
10 at the polypeptide level can be ascertained either alone (for example, by native or denatured polypeptide gel or immunoblot analysis) or in concert with other genes (for example, by proteomic analysis). Knowledge of polypeptide/polypeptide and polypeptide/DNA interactions can assign function by identifying polypeptides and nucleic acid sequences acting together in the same biological process. Genetics can
15 assign function to a gene by demonstrating that DNA lesions (mutations) in the gene have a quantifiable effect on the organism, including, but not limited to, its development; hormone biosynthesis and response; growth and growth habit (plant architecture); mRNA expression profiles; polypeptide expression profiles; ability to resist diseases; tolerance of abiotic stresses (for example, drought conditions); ability to acquire
20 nutrients; photosynthetic efficiency; altered primary and secondary metabolism; and the composition of various plant organs. Biochemistry can assign function by demonstrating that the polypeptide(s) encoded by the gene, typically when expressed in a heterologous organism, possesses a certain enzymatic activity, either alone or in combination with other polypeptides. Molecular genetics can assign function by
25 overexpressing or underexpressing the gene in the native plant or in heterologous organisms, and observing quantifiable effects as disclosed in functional assignment by genetics above.

Sequence homology (sequence identity) of unknown sequences to known sequences provides suggestion of a possible function of the polypeptide encoded by
30 the gene. In functional genomics, any or all of the above approaches are utilized, often in concert, to confirm suggested function provided by sequence homology to known genes, allowing functions to be assigned to genes across any of a number of

organismal phenotypes.

It is recognized by those skilled in the art that these different methodologies can each provide data as evidence for the function of a particular gene, and that such evidence is stronger with increasing amounts of data used for functional assignment: in one embodiment from a single methodology, in another embodiment from two methodologies, and in still another embodiment from more than two methodologies. In addition, those skilled in the art are aware that different methodologies can differ in the strength of the evidence provided for the assignment of gene function. Typically, but not always, a datum of biochemical, genetic, or molecular genetic evidence is considered stronger than a datum of bioinformatic or gene expression evidence. Finally, those skilled in the art recognize that, for different genes, a single datum from a single methodology can differ in terms of the strength of the evidence provided by each distinct datum for the assignment of the function of these different genes.

The objective of crop trait functional genomics is to identify crop trait genes of interest, for example, genes capable of conferring useful agronomic traits in crop plants either through the upregulation or downregulation of the genes. Such agronomic traits include, but are not limited to, enhanced yield, whether in quantity or quality; enhanced nutrient acquisition and metabolic efficiency; enhanced or altered nutrient composition of plant tissues used for food, feed, fiber, or processing; enhanced utility for agricultural or industrial processing; enhanced resistance to plant diseases; enhanced tolerance of adverse environmental conditions (abiotic stresses) including, but not limited to, drought, excessive cold, excessive heat, or excessive soil salinity or extreme acidity or alkalinity; and alterations in plant architecture or development, including changes in developmental timing. The deployment of such identified trait genes by either transgenic or non-transgenic approaches can materially improve crop plants for the benefit of agriculture.

Cereals are the most important crop plants on the planet in terms of both human and animal consumption. Genomic synteny (conservation of gene order within large chromosomal segments) is observed in rice, maize, wheat, barley, rye, oats, and other agriculturally important monocots including sorghum (see *e.g.*, Kellogg, 1998; Song *et al.*, 2001, and references therein), which facilitates the mapping and isolation of orthologous genes from diverse cereal species based on the sequence of a single

cereal gene. Rice has the smallest (about 420 Mb) genome among the cereal grains, and has recently been a major focus of public and private genomic and EST sequencing efforts. See Goff *et al.*, 2002.

Phytate (inositol hexakisphosphate; IP₆), is present in all eukaryotic cells and is highly abundant in plant seeds where it compromises animal nutrition and the environment. As a molecular approach for reducing seed phytate and studying IP₆ signaling function in plants, disclosed herein is the gene encoding the terminal kinase in Arabidopsis phytate synthesis, inositol polyphosphate 2-kinase (*AtIPK1*). *AtIPK1* specifically phosphorylates the D-2 position of the inositol ring and uses I(1,4,5,6)P₄, I(1,3,4,6)P₄, I(3,4,5,6)P₄, and I(1,3,4,5,6)P₅ as substrates. Together with *AtIPK2*, an IP₃/IP₄ 6-/3-/5-kinase, *AtIPK1* can generate multiple inositol polyphosphate products including IP₆ from an I(1,4,5)P₃ precursor. A partial disruption of *AtIPK1* causes an 80-90% loss of IP₆ in seeds and all tissues tested. The low-phytate *atipk1* mutant plants are conditionally growth compromised and have defective root phosphate sensation; however no defects in germination are observed.

Described herein are the identification, isolation, and characterization of the gene and gene products of two inositol polyphosphate 6-/3-/5-kinases, designated *AtIPK2α* and *AtIPK2β*, from *Arabidopsis thaliana*.

As partially redundant enzymes that catalyze the penultimate step in inositol hexakisphosphate (IP₆ or phytate) biosynthesis both *AtIPK2α* and *AtIPK2β* are attractive targets for the genetic manipulation of phytate synthesis in plant seeds. To generate a low-phytate plant, a mutant *Arabidopsis thaliana* line with a foreign DNA element (T-DNA) stably inserted in the *AtIPK1β* gene has been identified and obtained, which interrupts transcription of the gene by at least 90%. The mutation causes a significant loss of phytate in seeds (developing and mature), but not seedlings, thereby appearing seed specific. The mutant *atipk2β* plant has no detectable adverse growth phenotype and seed yields are normal.

Also, plants have been generated with T-DNA disruptions in both *AtIPK1*, and *AtIPK2β* through breeding of single mutant lines described above. The double mutant plants are fertile, germinate normally and have seeds of nearly normal size. Of significant interest, examination of seed inositol phosphates demonstrates a 95%

reduction of phytate with a minimal accumulation of IP₅, IP₄, IP₃ and IP₂.

Collectively, these phenotypes represent a significant improvement over other described low phytate plants, especially in the area of IP₆ reduction and seed yield. Targeting the last two enzymatic steps of phytate synthesis has advantages over
5 targeting earlier steps, such as D-myo-inositol 3-phosphate synthase (sometimes referred to as L-myo-inositol 1-phosphate synthase) or inositol 1,3,4-trisphosphate 5/6-kinase, both in reported efficacy of phytate reduction and markedly reduced adverse growth effects on plant growth and seed yield.

The presently disclosed subject matter indicates a signaling role for phytate in
10 the regulation of phosphate biology and demonstrate that inositol phosphate kinases, *AtIPK1* and *Atlpk2*, are commercially relevant targets for generating low-phytate cereals.

II. Definitions

15 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the presently disclosed subject matter pertains. For clarity of the present specification, certain definitions are presented hereinbelow.

Following long-standing patent law convention, the terms "a" and "an" mean "one
20 or more" when used in this application, including in the claims.

As used herein, the terms "associated with" and "operatively linked" refer to two nucleic acid sequences that are related physically or functionally. For example, a promoter or regulatory DNA sequence is said to be "associated with" a DNA sequence that encodes an RNA or a polypeptide if the two sequences are operatively linked, or
25 situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence.

As used herein, the term "chimera" refers to a polypeptide that comprises domains or other features that are derived from different polypeptides or are in a position relative to each other that is not naturally occurring.

30 As used herein, the term "chimeric construct" refers to a recombinant nucleic acid molecule in which a promoter or regulatory nucleic acid sequence is operatively linked to, or associated with, a nucleic acid sequence that codes for an mRNA or which

is expressed as a polypeptide, such that the regulatory nucleic acid sequence is able to regulate transcription or expression of the associated nucleic acid sequence. The regulatory nucleic acid sequence of the chimeric construct is not normally operatively linked to the associated nucleic acid sequence as found in nature.

5 As used herein, the terms "coding sequence" and "open reading frame" (ORF) are used interchangeably and refer to a nucleic acid sequence that is transcribed into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA, or antisense RNA. In one embodiment, the RNA is then translated *in vivo* or *in vitro* to produce a polypeptide.

10 As used herein, the term "complementary" refers to two nucleotide sequences that comprise antiparallel nucleotide sequences capable of pairing with one another upon formation of hydrogen bonds between the complementary base residues in the antiparallel nucleotide sequences. As is known in the art, the nucleic acid sequences of two complementary strands are the reverse complement of each other when each is viewed in the 5' to 3' direction.

15 As is also known in the art, two sequences that hybridize to each other under a given set of conditions do not necessarily have to be 100% fully complementary. As used herein, the terms "fully complementary" and "100% complementary" refer to sequences for which the complementary regions are 100% in Watson-Crick base-pairing, *i.e.*, that no mismatches occur within the complementary regions. However, as
20 is often the case with recombinant molecules (for example, cDNAs) that are cloned into cloning vectors, certain of these molecules can have non-complementary overhangs on either the 5' or 3' ends that result from the cloning event. In such a situation, it is understood that the region of 100% or full complementarity excludes any sequences that are added to the recombinant molecule (typically at the ends) solely as a result of,
25 or to facilitate, the cloning event. Such sequences are, for example, polylinker sequences, linkers with restriction enzyme recognition sites, etc.

30 As used herein, the terms "domain" and "feature", when used in reference to a polypeptide or amino acid sequence, refers to a subsequence of an amino acid sequence that has a particular biological function. Domains and features that have a particular biological function include, but are not limited to, ligand binding, nucleic acid binding, catalytic activity, substrate binding, and polypeptide-polypeptide interacting domains. Similarly, when used herein in reference to a nucleic acid sequence, a

"domain", or "feature" is that subsequence of the nucleic acid sequence that encodes a domain or feature of a polypeptide.

As used herein, the term "enzymatic activity" or "enzyme activity", for example the activity of the novel functional inositol phosphate kinases disclosed herein, refers to the ability of an enzyme to catalyze the conversion of a substrate into a product. In the case of the inositol phosphate kinases disclosed herein, the enzyme activity is the addition of phosphate groups at specific locations on the backbone of a specific inositol substrate species. A substrate for the enzyme can comprise the natural substrate of the enzyme but also can comprise analogues of the natural substrate, which can also be converted by the enzyme into a product or into an analogue of a product. The activity of the enzyme is measured for example by determining the amount of product in the reaction after a certain period of time, or by determining the amount of substrate remaining in the reaction mixture after a certain period of time. The activity of the enzyme can also be measured by determining the amount of an unused co-factor of the reaction remaining in the reaction mixture after a certain period of time or by determining the amount of used co-factor in the reaction mixture after a certain period of time. The activity of the enzyme can also be measured by determining the amount of a donor of free energy or energy-rich molecule (e.g. ATP, phosphoenolpyruvate, acetyl phosphate, or phosphocreatine) remaining in the reaction mixture after a certain period of time or by determining the amount of a used donor of free energy or energy-rich molecule (e.g. ADP, pyruvate, acetate, or creatine) in the reaction mixture after a certain period of time.

As used herein, the term "expression cassette" refers to a nucleic acid molecule capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operatively linked to the nucleotide sequence of interest which is operatively linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually encodes a polypeptide of interest but can also encode a functional RNA of interest, for example antisense RNA or a non-translated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest can be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette can also

be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. Typically, however, the expression cassette is heterologous with respect to the host; *i.e.*, the particular DNA sequence of the expression cassette does not occur naturally in the host cell and was introduced into the host cell or an ancestor of the host cell by a transformation event. The expression of the nucleotide sequence in the expression cassette can be under the control of a constitutive promoter or of an inducible promoter that initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism such as a plant, the promoter can also be specific to a particular tissue, organ, or stage of development, such as for example, specific for seed tissue.

As used herein, the term "fragment" refers to a sequence that comprises a subset of another sequence. When used in the context of a nucleic acid or amino acid sequence, the terms "fragment" and "subsequence" are used interchangeably. A fragment of a nucleic acid sequence can be any number of nucleotides that is less than that found in another nucleic acid sequence, and thus includes, but is not limited to, the sequences of an exon or intron, a promoter, an enhancer, an origin of replication, a 5' or 3' untranslated region, a coding region, and a polypeptide binding domain. It is understood that a fragment or subsequence can also comprise less than the entirety of a nucleic acid sequence, for example, a portion of an exon or intron, promoter, enhancer, etc. Similarly, a fragment or subsequence of an amino acid sequence can be any number of residues that is less than that found in a naturally occurring polypeptide, and thus includes, but is not limited to, domains, features, repeats, etc. Also similarly, it is understood that a fragment or subsequence of an amino acid sequence need not comprise the entirety of the amino acid sequence of the domain, feature, repeat, etc. A fragment can also be a "functional fragment", in which the fragment retains a specific biological function of the nucleic acid sequence or amino acid sequence of interest. For example, a functional fragment of a transcription factor can include, but is not limited to, a DNA binding domain, a transactivating domain, or both. Similarly, a functional fragment of a receptor tyrosine kinase includes, but is not limited to a ligand binding domain, a kinase domain, an ATP binding domain, and combinations thereof.

As used herein, the term "gene" is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include, but are not limited to, nucleic acid molecules that are coding sequences and/or the regulatory sequences required for their expression. Genes can also include non-expressed DNA segments that, for example, form recognition sequences for a polypeptide. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and can include sequences designed to have desired parameters.

The terms "heterologous", "recombinant", and "exogenous", when used herein to refer to a nucleic acid sequence (e.g. a DNA sequence) or a gene, refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling or other recombinant techniques. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position or form within the host cell in which the element is not ordinarily found. Similarly, when used in the context of a polypeptide or amino acid sequence, an exogenous polypeptide or amino acid sequence is a polypeptide or amino acid sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, exogenous DNA segments can be expressed to yield exogenous polypeptides.

A "homologous" nucleic acid (or amino acid) sequence is a nucleic acid (or amino acid) sequence naturally associated with a host cell into which it is introduced.

As used herein, the term "disruption" refers to partial or complete reduction of the expression of at least a portion of a nucleic acid or a polypeptide encoded by one or more endogenous genes of a single cell, selected cells, or all of the cells of a plant. The plant can have a "heterozygous disruption," wherein one allele of one or more endogenous genes have been disrupted. Alternatively, the plant can have a "homozygous disruption," wherein both alleles of one or more endogenous genes have been disrupted.

The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. The phrase "bind(s) substantially" refers to complementary hybridization
5 between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

As used herein, the term "inhibitor" or "antagonist" refers to a chemical substance that inactivates or decreases the biological activity of a polypeptide such as a
10 biosynthetic and catalytic activity, receptor, signal transduction polypeptide, structural gene product, or transport polypeptide. The term "herbicide" (or "herbicidal compound") is used herein to define an inhibitor applied to a plant at any stage of development, whereby the herbicide inhibits the growth of the plant or kills the plant.

As used herein, the term "isolated", when used in the context of an isolated DNA
15 molecule or an isolated polypeptide, is a DNA molecule or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or polypeptide can exist in a purified form or can exist in a non-native environment such as, for example, in a transgenic host cell.

As used herein, the term "mature polypeptide" refers to a polypeptide from which
20 the transit peptide, signal peptide, and/or propeptide portions have been removed.

As used herein, the term "minimal promoter" refers to the smallest piece of a promoter, such as a TATA element, that can support any transcription. A minimal promoter typically has greatly reduced promoter activity in the absence of upstream or downstream activation. In the presence of a suitable transcription factor, a minimal
25 promoter can function to permit transcription.

As used herein, the term "modified enzyme activity" refers to enzyme activity that is different from that which naturally occurs in a plant (*i.e.* enzyme activity that occurs naturally in the absence of direct or indirect manipulation of such activity by man). A
"modulator of enzymatic activity" is a compound that modifies the activity of an enzyme.

As used herein, the term "native" refers to a gene that is endogenous, *i.e.* naturally present, in the genome of an untransformed plant cell. Similarly, when used in

the context of a polypeptide, a "native polypeptide" is a polypeptide that is encoded by a native gene of an untransformed plant cell's genome.

As used herein, the term "naturally occurring" refers to an object that is found in nature as distinct from being artificially produced by man. For example, a polypeptide or nucleotide sequence that is present in an organism (including a virus) in its natural state, which has not been intentionally modified or isolated by man in the laboratory, is naturally occurring. As such, a polypeptide or nucleotide sequence is considered "non-naturally occurring" if it is encoded by or present within a recombinant molecule, even if the amino acid or nucleic acid sequence is identical to an amino acid or nucleic acid sequence found in nature.

As used herein, the term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof ("nucleic acid molecules") in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g. degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions can be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, 1991; Ohtsuka *et al.*, 1985; Rossolini *et al.*, 1994). The terms "nucleic acid", "nucleic acid molecule", "polynucleotide" or "nucleic acid sequence" can also be used interchangeably with gene, cDNA, and mRNA encoded by a gene.

As used herein, the term "orthologs" refers to genes in different species that encode protein that perform the same biological function. For example, the inositol phosphate kinase 1 and 2 genes from, for example, Arabidopsis, maize, wheat, sorghum and rice, are orthologs. Typically, orthologous nucleic acid sequences are characterized by a degree of sequence similarity (in some examples, at least about 40% identity over a complete polypeptide sequence, in some examples, at least about 80% sequence identity over at least shared functional domains). A nucleic acid sequence of an ortholog in one species (for example, Arabidopsis) can be used to

isolate the nucleic acid sequence of the ortholog in another species (for example, maize) using standard molecular biology techniques. This can be accomplished, for example, using techniques described in more detail below (see *also* Sambrook & Russell, 2001 for a discussion of hybridization conditions that can be used to isolate
5 closely related sequences).

As used herein, the phrase "percent identical" or "percent identity" in the context of two nucleic acid or polypeptide sequences, refers to two or more sequences or subsequences that have in one embodiment 40%, in one embodiment 45%, in one embodiment 50%, in one embodiment 55%, in one embodiment 60%, in another
10 embodiment 75%, in another embodiment 70%, in another embodiment 80%, in another embodiment 90%, in another embodiment 95%, and in still another embodiment at least 99% nucleotide or amino acid residue identity, respectively, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. The percent identity
15 exists in one embodiment over a region of the sequences, for example a functional domain, that is at least about 50 residues in length, in another embodiment over a region of at least about 100 residues, and in another embodiment, the percent identity exists over at least about 150 residues. In still another embodiment, the percent identity exists over the entire length of the sequences.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for
20 the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, for example, by the local homology algorithm disclosed in Smith & Waterman, 1981, by the homology alignment algorithm disclosed in Needleman & Wunsch, 1970, by the search for
30 similarity method disclosed in Pearson & Lipman, 1988, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Package, available from Accelrys, Inc., San Diego, California, United States

of America), or by visual inspection. See generally, Ausubel *et al.*, 1994.

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, 1990. Software for performing BLAST analysis is publicly available over the internet through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence.

T is referred to as the neighborhood word score threshold. See generally, Altschul *et al.*, 1990. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M = 5, N = -4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. See Henikoff & Henikoff, 1992.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see *e.g.*, Karlin & Altschul, 1993). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the

reference nucleic acid sequence is in one embodiment less than about 0.1, in another embodiment less than about 0.01, and in still another embodiment less than about 0.001.

As used herein, the term "phytic acid" refers to *myo*-inositol hexaphosphoric acid, also referred to herein as "IP₆". As a salt with cations, phytic acid is usually referred to as "phytate", however the two terms are used interchangeably herein.

As used herein, the term "shuffled nucleic acid" refers to a recombinant nucleic acid molecule in which the nucleotide sequence comprises a plurality of nucleotide sequence fragments, wherein at least one of the fragments corresponds to a region of a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, and wherein at least two of the plurality of sequence fragments are in an order, from 5' to 3', which is not an order in which the plurality of fragments naturally occur in a nucleic acid.

The term "substantially identical", in the context of two nucleotide or amino acid sequences, can refer to two or more sequences or subsequences that have in one embodiment at least about 40% nucleotide or amino acid identity, at least about 45% nucleotide or amino acid identity, at least about 50% nucleotide or amino acid identity, at least about 55% nucleotide or amino acid identity, at least about 60% nucleotide or amino acid identity, in another embodiment at least about 65% nucleotide or amino acid identity, in another embodiment at least about 70% nucleotide or amino acid identity, in another embodiment at least about 75% nucleotide or amino acid identity, in another embodiment at least about 80% nucleotide or amino acid identity, in another embodiment at least about 85% nucleotide or amino acid identity, in another embodiment at least about 90% nucleotide or amino acid identity, and in yet another embodiment at least about 95% nucleotide or amino acid identity, when compared and aligned for maximum correspondence, as measured using one of the above-referenced sequence comparison algorithms or by visual inspection. In one example, the substantial identity exists in nucleotide or amino acid sequences of at least 50 residues, in another example in nucleotide or amino acid sequence of at least about 100 residues, in another example in nucleotide or amino acid sequences of at least about 150 residues, and in yet another example in nucleotide or amino acid sequences comprising complete coding sequences or complete amino acid sequences. In one example, the substantial identity exists in nucleotide or amino acid sequences of a

conserved motif of the polypeptides compared. In one embodiment, for example, the conserved motif of Ipk2 is PxxxDxKxG, wherein 'x' is any amino acid, and polypeptides compared and each having this conserved motif would be considered substantially identical. Likewise, in another embodiment, the art recognized conserved motif of Ipk1
5 (as discussed in York et al. Science 1999; Odom et al. 2000 J. Biol. Chem.; and Stevenson-Paulik et al. J. Biol. Chem. 2002), if shared by compared polypeptides, then the polypeptides would be considered substantially identical.

In one aspect, polymorphic sequences can be substantially identical sequences. The term "polymorphic" refers to the two or more genetically determined alternative
10 sequences or alleles in a population. An allelic difference can be as small as one base pair. Nonetheless, one of ordinary skill in the art would recognize that the polymorphic sequences correspond to the same gene.

Another indication that two nucleotide sequences are substantially identical is that the two molecules specifically or substantially hybridize to each other under
15 conditions of medium or high stringency. In the context of nucleic acid hybridization, two nucleic acid sequences being compared can be designated a "probe sequence" and a "target sequence". A "probe sequence" is a reference nucleic acid molecule, and a "target sequence" is a test nucleic acid molecule, often found within a heterogeneous population of nucleic acid molecules. A "target sequence" is synonymous with a "test
20 sequence".

An exemplary nucleotide sequence employed for hybridization studies or assays includes probe sequences that are complementary to or mimic in one embodiment at least an about 14 to 40 nucleotide sequence of a nucleic acid molecule of the presently disclosed subject matter. In one example, probes comprise 14 to 20 nucleotides, or
25 even longer where desired, such as 30, 40, 50, 60, 100, 200, 300, or 500 nucleotides or up to the full length (for example, the full complement) of any of the nucleic acid sequence set forth in the odd numbered SEQ ID NOs: 1-9. Such fragments can be readily prepared by, for example, directly synthesizing the fragment by chemical synthesis, by application of nucleic acid amplification technology, or by introducing
30 selected sequences into recombinant vectors for recombinant production.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern

blot analysis are both sequence- and environment-dependent. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, 1993. Generally, high stringency hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Typically, under "highly stringent conditions" a probe will hybridize specifically to its target subsequence, but to no other sequences. Similarly, medium stringency hybridization and wash conditions are selected to be more than about 5°C lower than the T_m for the specific sequence at a defined ionic strength and pH. Exemplary medium stringency conditions include hybridizations and washes as for high stringency conditions, except that the temperatures for the hybridization and washes are in one embodiment 8°C, in another embodiment 10°C, in another embodiment 12°C, and in still another embodiment 15°C lower than the T_m for the specific sequence at a defined ionic strength and pH.

The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of highly stringent hybridization conditions for Southern or Northern Blot analysis of complementary nucleic acids having more than about 100 complementary residues is overnight hybridization in 50% formamide with 1 mg of heparin at 42°C. An example of highly stringent wash conditions is 15 minutes in 0.1x standard saline citrate (SSC), 0.1% (w/v) SDS at 65°C. Another example of highly stringent wash conditions is 15 minutes in 0.2x SSC buffer at 65°C (see Sambrook and Russell, 2001 for a description of SSC buffer and other stringency conditions). Often, a high stringency wash is preceded by a lower stringency wash to remove background probe signal. An example of medium stringency wash conditions for a duplex of more than about 100 nucleotides is 15 minutes in 1X SSC at 45°C. Another example of medium stringency wash for a duplex of more than about 100 nucleotides is 15 minutes in 4-6X SSC at 40°C. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1M Na^+ ion, typically about 0.01 to 1M Na^+ ion concentration (or other salts) at pH 7.0-8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2-fold (or higher) than

that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

The following are examples of hybridization and wash conditions that can be used to clone homologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the presently disclosed subject matter: a probe nucleotide sequence hybridizes in one example to a target nucleotide sequence in 7% sodium dodecyl sulfate (NaDS), 0.5M NaPO₄, 1 mm ethylene diamine tetraacetic acid (EDTA) at 50°C followed by washing in 2X SSC, 0.1% NaDS at 50°C; in another example, a probe and target sequence hybridize in 7% NaDS, 0.5 M NaPO₄, 1 mm EDTA at 50°C followed by washing in 1X SSC, 0.1% NaDS at 50°C; in another example, a probe and target sequence hybridize in 7% NaDS, 0.5 M NaPO₄, 1 mm EDTA at 50°C followed by washing in 0.5X SSC, 0.1% NaDS at 50°C; in another example, a probe and target sequence hybridize in 7% NaDS, 0.5 M NaPO₄, 1 mm EDTA at 50°C followed by washing in 0.1X SSC, 0.1% NaDS at 50°C; in yet another example, a probe and target sequence hybridize in 7% NaDS, 0.5 M NaPO₄, 1 mm EDTA at 50°C followed by washing in 0.1X SSC, 0.1% NaDS at 65°C. In one embodiment, hybridization conditions comprise hybridization in a roller tube for at least 12 hours at 42°C.

As used herein, the term "purified", when applied to a nucleic acid or polypeptide, denotes that the nucleic acid or polypeptide is essentially free of other cellular components with which it is associated in the natural state. It can be in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A polypeptide that is the predominant species present in a preparation is substantially purified. The term "purified" denotes that a nucleic acid or polypeptide gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or polypeptide is in one embodiment at least about 50% pure, in another embodiment at least about 85% pure, and in still another embodiment at least about 99% pure.

Two nucleic acids are "recombined" when sequences from each of the two nucleic acids are combined in a progeny nucleic acid. Two sequences are "directly" recombined when both of the nucleic acids are substrates for recombination. Two

sequences are "indirectly recombined" when the sequences are recombined using an intermediate such as a cross-over oligonucleotide. For indirect recombination, no more than one of the sequences is an actual substrate for recombination, and in some cases, neither sequence is a substrate for recombination.

5 As used herein, the term "regulatory elements" refers to nucleotide sequences involved in controlling the expression of a nucleotide sequence. Regulatory elements can comprise a promoter operatively linked to the nucleotide sequence of interest and termination signals. Regulatory sequences also include enhancers and silencers. They also typically encompass sequences required for proper translation of the nucleotide
10 sequence.

As used herein, the term "significant increase" refers to an increase in activity (for example, enzymatic activity) that is larger than the margin of error inherent in the measurement technique, in one embodiment an increase by about 2-fold or greater over a baseline activity (for example, the activity of the wild-type enzyme in the
15 presence of the inhibitor), in another embodiment an increase by about 5-fold or greater, and in still another embodiment an increase by about 10-fold or greater.

As used herein, the terms "significantly less", "significantly inhibited" and "significantly reduced" refer to a result (for example, an amount of a product of an enzymatic reaction) that is reduced by more than the margin of error inherent in the
20 measurement technique, in one embodiment a decrease by about 2-fold or greater with respect to a baseline activity (for example, the activity of the wild-type enzyme in the absence of the inhibitor), in another embodiment, a decrease by about 5-fold or greater, and in still another embodiment a decrease by about 10-fold or greater.

As used herein, the terms "specific binding" and "immunological cross-reactivity"
25 refer to an indicator that two molecules are substantially identical. An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with, or specifically binds to, the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example,
30 where the two polypeptides differ only by conservative substitutions.

The phrase "specifically (or selectively) binds to an antibody," or "specifically (or selectively) immunoreactive with," when referring to a polypeptide or peptide, refers to a

binding reaction which is determinative of the presence of the polypeptide in the presence of a heterogeneous population of polypeptides and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular polypeptide and do not bind in a significant amount to other polypeptides present in the sample. Specific binding to an antibody under such conditions can require an antibody that is selected for its specificity for a particular polypeptide. For example, antibodies raised to the polypeptide with the amino acid sequence encoded by any of the nucleic acid sequences of the presently disclosed subject matter can be selected to obtain antibodies specifically immunoreactive with that polypeptide and not with other polypeptides except for polymorphic variants. A variety of immunoassay formats can be used to select antibodies specifically immunoreactive with a particular polypeptide. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a polypeptide. See Harlow & Lane, 1988, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

As used herein, the term "subsequence" refers to a sequence of nucleic acids or amino acids that comprises a part of a longer sequence of nucleic acids or amino acids (e.g., polypeptide), respectively.

As used herein, the term "substrate" refers to a molecule that an enzyme naturally recognizes and converts to a product in the biochemical pathway in which the enzyme naturally carries out its function; or is a modified version of the molecule, which is also recognized by the enzyme and is converted by the enzyme to a product in an enzymatic reaction similar to the naturally-occurring reaction.

As used herein, the term "suitable growth conditions" refers to growth conditions that are suitable for a certain desired outcome, for example, the production of a recombinant polypeptide or the expression of a nucleic acid molecule.

As used herein, the term "transformation" refers to a process for introducing heterologous DNA into a plant cell, plant tissue, or plant. Transformed plant cells, plant tissue, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof.

As used herein, the terms "transformed", "transgenic", and "recombinant" refer to a host organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A "non-transformed," "non-transgenic", or "non-recombinant" host refers to a wild-type organism, *e.g.*, a bacterium or plant, which does not contain the heterologous nucleic acid molecule.

As used herein, the term "viability" refers to a fitness parameter of a plant. Plants are assayed for their homozygous performance of plant development, indicating which polypeptides are essential for plant growth.

III. Nucleic Acid Molecules and Polypeptides

The presently disclosed subject matter encompasses the identification and isolation of cDNAs encoding genes of interest in the expression of inositol kinase phosphates. Inositol kinase phosphates are a class of enzymes that facilitate production of inositol hexakisphosphate (IP₆), also known as phytate, along the phytate biosynthetic pathway through the serial addition of phosphates at specific carbons to *myo*-inositol. Altering the expression of genes related to these traits can be used to improve or modify plants, and in particular improve or modify plants producing seed utilized as food or animal feed. Examples describe the isolated genes of interest and methods of analyzing the alteration of expression and their effects on the plant characteristics.

III.A. Nucleic Acid Molecules

Embodiments of the presently disclosed subject matter encompass isolated nucleic acid molecules corresponding to genes that encode inositol phosphate kinases. In particular, the presently disclosed subject matter encompasses isolated nucleic acids encoding inositol phosphate kinase 2 (Ipk2) and inositol phosphate kinase 1 (Ipk1).

As discussed herein, odd numbered SEQ ID NOs: 1-9 are nucleotide sequences

encoding inositol phosphate kinases from *Arabidopsis thaliana* that have been isolated and identified using the methods and compositions disclosed herein.

As discussed herein, even numbered SEQ ID NOs: 2-10 are inositol phosphate kinase polypeptide sequences encoded by the immediately preceding nucleotide sequence. For example, SEQ ID NO: 2 is the polypeptide encoded by the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO: 4 is the polypeptide encoded by the nucleotide sequence of SEQ ID NO: 3, etc.

In one embodiment, an isolated nucleic acid molecule of the presently disclosed subject matter comprises a nucleic acid encoding a polypeptide comprising an amino acid sequence of one of even numbered SEQ ID NOs: 2-10. In another embodiment, an isolated nucleic acid molecule of the presently disclosed subject matter comprises a nucleic acid encoding a polypeptide comprising an amino acid sequence at least 40% identical, in some embodiments at least about 45% amino acid identity, in some embodiments at least about 50% amino acid identity, in some embodiments at least about 55% amino acid identity, in some embodiments at least about 60% amino acid identity, in some embodiments at least about 65% amino acid identity, in some embodiments at least about 70% amino acid identity, in some embodiments at least about 75% amino acid identity, in some embodiments at least about 80% amino acid identity, in some embodiments at least about 85% amino acid identity, in some embodiments at least about 90% amino acid identity, in some embodiments at least about 95% amino acid identity, and in some embodiments at least about 95% amino acid identity, when compared and aligned for maximum correspondence, as measured using one of the above-referenced sequence comparison algorithms or by visual inspection, to a complete sequence, functional domain, or other subsequence of one of even numbered SEQ ID NOs: 2-10 and having inositol phosphate kinase activity.

In one embodiment, an isolated nucleic acid molecule of the presently disclosed subject matter comprises a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence as set forth in odd numbered sequences SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof. In another embodiment, an isolated nucleic acid molecule of the presently disclosed subject matter comprises a nucleotide sequence having substantial identity to a nucleotide sequence that

hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence as set forth in odd numbered sequences SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof. Another embodiment of the presently disclosed subject matter encompasses an isolated nucleic acid molecule comprising a nucleotide sequence that is complementary to, or the reverse complement of, a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof.

Still another embodiment of the presently disclosed subject matter encompasses an isolated nucleic acid molecule comprising a nucleotide sequence that is complementary to, or the reverse complement of, a nucleotide sequence that has substantial identity to, or is capable of hybridizing specifically to, a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof.

In a representative embodiment, the substantial identity is at least about 60% identity, in another embodiment at least about 65% identity, in another embodiment at least about 70% identity, in another embodiment at least about 75% identity, in another embodiment about 80% identity, in another embodiment at least about 85% identity, in another embodiment about 90% identity, and in still another embodiment at least about 95% identity to the nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof.

In another embodiment, the nucleotide sequence having substantial identity comprises an allelic variant of the nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof. In yet another embodiment, the nucleotide sequence having substantial identity comprises a naturally occurring variant.

In another embodiment, the nucleotide sequence having substantial identity comprises a polymorphic variant of the nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C

in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof.

In another embodiment, the nucleic acid having substantial identity comprises a deletion or insertion of at least one nucleotide. In one embodiment, the deletion or insertion comprises more than half the nucleotides. In one embodiment, the deletion or insertion comprises less than about thirty nucleotides. In another embodiment, the deletion or insertion comprises less than about five nucleotides. In another embodiment, the sequence of the isolated nucleic acid having substantial identity comprises a substitution in at least one codon. In one embodiment, the substitution is conservative.

In another embodiment, the isolated nucleic acid comprises a plurality of regions having a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or an exon, domain, or feature thereof.

In one embodiment, the sequence having substantial identity to the nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof, is from a plant. In one embodiment, the plant is a dicot. In another embodiment, the plant is a gymnosperm. In another embodiment, the plant is a monocot. In one embodiment, the monocot is a cereal. In one embodiment, the cereal can be, for example, maize, wheat, barley, oats, rye, millet, sorghum, triticale, secale, einkorn, spelt, emmer, teff, milo, flax, gamma grass, *Tripsacum* sp., or *teosinte*. In another embodiment, the cereal is rice.

In one embodiment, the nucleic acid is expressed in a specific location or tissue of a plant. In one embodiment, the location or tissue includes, but is not limited to, epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, and combinations thereof. In another embodiment, the location or tissue is a seed.

In one embodiment, the nucleic acid encodes a polypeptide involved in a function including, but not limited to, inositol phosphate kinase activity.

Embodiments of the presently disclosed subject matter further relate to an

isolated polynucleotide comprising a nucleotide sequence of at least 10 bases, which sequence is identical, complementary (for example, fully complementary), or substantially identical to a region of any sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence of odd numbered sequences of SEQ ID NOs: 1-9, and wherein the polynucleotide is adapted for any of numerous uses.

In one embodiment, the polynucleotide is used as a chromosomal marker. In another embodiment, the polynucleotide is used as a marker for restriction fragment length polymorphism (RFLP) analysis. In another embodiment, the polynucleotide is used as a marker for quantitative trait-linked breeding. In another embodiment, the polynucleotide is used as a marker for marker-assisted breeding. In another embodiment, the polynucleotide is used as a bait sequence in a two-hybrid system to identify sequence-encoding polypeptides interacting with the polypeptide encoded by the bait sequence. In another embodiment, the polynucleotide is used as a diagnostic indicator for genotyping or identifying an individual or population of individuals. In yet another embodiment, the polynucleotide is used for genetic analysis to identify boundaries of genes or exons.

Embodiments of the presently disclosed subject matter also relate to a shuffled nucleic acid molecule comprising a plurality of nucleotide sequence fragments, wherein at least one of the fragments corresponds to a region of a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, and wherein at least two of the plurality of sequence fragments are in an order, from 5' to 3', which is not an order in which the plurality of fragments naturally occur. In one embodiment, all of the fragments in a shuffled nucleic acid comprising a plurality of nucleotide sequence fragments are from a single gene. In another embodiment, the plurality of fragments is derived from at least two different genes. In one embodiment, the shuffled nucleic acid is operatively linked to a promoter sequence. In another embodiment, the shuffled nucleic acid comprises a chimeric polynucleotide comprising a promoter sequence operatively linked to the shuffled nucleic acid. In still another embodiment, the shuffled nucleic acid is contained within a host cell.

III.B. Identifying, Cloning, and Sequencing cDNAs

The cloning and sequencing of the cDNAs of the presently disclosed subject matter is accomplished using techniques known in the art. See *generally*, Sambrook & Russell, 2001; Silhavy *et al.*, 1984; Ausubel *et al.*, 1994; Reiter *et al.*, 1992; Schultz *et al.*, 1998.

The isolated nucleic acids and polypeptides of the presently disclosed subject matter are usable over a range of plants – monocots and dicots – in particular monocots such as sorghum, rice, wheat, barley, and maize. In one embodiment, the monocot is a cereal. A more inclusive, but not intended to be limiting, list of plant genera relevant to the presently disclosed subject matter include *Arabidopsis thaliana*, corn (*Zea mays*), *Brassica* sp., alfalfa (*Medicago sativa*), rice (*Oryza sativa* ssp.), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanut (*Arachis hypogaea*), cotton, sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilis*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed (*Lemna*), and barley. Still other plants relevant to the presently disclosed subject matter include vegetables, ornamentals, and conifers. Relevant vegetables include, but are not limited to, tomatoes, lettuce, guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, green bean, lima bean, pea, and members of the genus *Cucumis*. Relevant ornamentals include, but are not limited to, impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, azalea, hydrangea, hibiscus, rose, tulip,

daffodil, petunia, carnation, poinsettia, and chrysanthemum. Relevant conifers include, but are not limited to, loblolly pine, slash pine, ponderosa pine, lodgepole pine, Monterey pine, Douglas-fir, Western hemlock, Sitka spruce, redwood, silver fir, balsam fir, Western red cedar, and Alaska yellow-cedar.

5 The presently disclosed subject matter also provides a method for genotyping a plant or plant part comprising a nucleic acid molecule of the presently disclosed subject matter. Optionally, the plant is a monocot such as, but not limited to, grains such as corn, sorghum, rice or wheat. Genotyping provides a methodology for distinguishing homologs of a chromosome pair and can be used to differentiate segregants in a plant
10 population. Molecular marker methods can be used in phylogenetic studies, characterizing genetic relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, mapping based cloning, and the study of quantitative inheritance (see Clark, 1997; Paterson, 1996).

 The method for genotyping can employ any number of molecular marker
15 analytical techniques including, but not limited to, restriction length polymorphisms (RFLPs). As is well known in the art, RFLPs are produced by differences in the DNA restriction fragment lengths resulting from nucleotide differences between alleles of the same gene. Thus, the presently disclosed subject matter provides a method for following segregation of a gene or nucleic acid of the presently disclosed subject matter
20 or chromosomal sequences genetically linked by using RFLP analysis. Linked chromosomal sequences are in one embodiment within 50 centimorgans (cM), in another embodiment within 40 cM, in another embodiment within 30 cM, in another embodiment within 20 cM, in another embodiment within 10 cM, and in still other embodiments within 5, 3, 2, or 1 cM of the nucleic acid of the presently disclosed
25 subject matter.

 Embodiments of the presently disclosed subject matter also relate to an isolated nucleic acid molecule comprising a nucleotide sequence, its complement (for example, its full complement), or its reverse complement (for example, its full reverse complement), the nucleotide sequence encoding a polypeptide (for example, a
30 biologically active polypeptide or biologically active fragment). In one embodiment, the nucleotide sequence encodes a polypeptide that is an ortholog of a polypeptide comprising a polypeptide sequence listed in even numbered sequences of SEQ ID

NOs: 2-10, or a fragment, domain, repeat, feature, or chimera thereof. In another embodiment, the nucleotide sequence encodes a polypeptide that is an ortholog of a polypeptide comprising a polypeptide sequence having at least about 40% identity to a polypeptide sequence listed in even numbered sequences of SEQ ID NOs: 2-10, or a
5 fragment, domain, repeat, feature, or chimera thereof. In another embodiment, the nucleotide sequence encodes a polypeptide that is an ortholog of a polypeptide comprising a polypeptide sequence encoded by a nucleotide sequence identical to or having substantial identity to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof, or a sequence
10 complementary thereto. In another embodiment, the nucleotide sequence encodes a polypeptide comprising a polypeptide sequence encoded by a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or to a sequence complementary thereto. In
15 still another embodiment, the nucleotide sequence encodes a functional fragment of a polypeptide of the presently disclosed subject matter.

In one embodiment, the isolated nucleic acid comprises a polypeptide-encoding sequence. In one embodiment, the polypeptide-encoding sequence encodes a polypeptide that is an ortholog of a polypeptide comprising a polypeptide sequence
20 listed in even numbered sequences of SEQ ID NOs: 2-10, or a fragment thereof. Representative orthologs are disclosed in Examples 6-11 herein below. In some embodiments, an isolated nucleic acid of the presently disclosed subject encodes a polypeptide has or comprises at least about having substantial identity (in some embodiments 90% amino acid identity) with a polypeptide as disclosed in Examples 6-
25 11, over an entire sequence or functional fragment thereof. In one embodiment, the polypeptide functions as an inositol phosphate kinase. More specifically, in some embodiments, the polypeptide is an lpk2 polypeptide and functions as a IP₃/IP₄/IP₅ 6-/3-/5-kinase generating multiple IP₄, IP₅ and IP₆ products. More specifically, in some other embodiments, the polypeptide is an lpk1 polypeptide and functions as an
30 I(1,4,5,6)P₄, I(1,3,4,6)P₄ and I(1,3,4,5,6)P₅ 2-kinase generating I(1,2,4,5,6)P₅, I(1,2,3,4,6)P₅ and phytate (IP₆).

In another embodiment, the polypeptide is a plant polypeptide. In one

embodiment, the plant is a dicot. In another embodiment, the plant is a gymnosperm. In another embodiment, the plant is a monocot. In one embodiment, the monocot is a cereal. In one embodiment, the cereal includes, but is not limited to, maize, wheat, barley, oats, rye, millet, sorghum, triticale, secale, einkorn, spelt, emmer, teff, miloflax, 5 gamma grass, *Tripsacum*, and *teosinte*. In another embodiment, the cereal is sorghum.

In one embodiment, the polypeptide functions as an inositol phosphate kinase.

Embodiments of the presently disclosed subject matter also relate to an isolated nucleic acid molecule comprising a nucleotide sequence, its complement (for example, 10 its full complement), or its reverse complement (for example, its full reverse complement), encoding a polypeptide selected from a group comprising one or more of:

- (a) a polypeptide sequence encoded by a nucleotide sequence that hybridizes under conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in 15 odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof, or a sequence complementary thereto; and
- (b) a functional fragment of (a).

In one embodiment, the polypeptide having substantial identity comprises an allelic variant of a polypeptide that is an ortholog of a polypeptide having an amino acid 20 sequence listed in even numbered sequences of SEQ ID NOs: 2-10, or a fragment, domain, repeat, feature, or chimera thereof. In one embodiment, the isolated nucleic acid comprises a plurality of regions from the polypeptide sequence encoded by a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide 25 sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof, or a sequence complementary thereto.

In another embodiment, the sequence of the isolated nucleic acid encodes a polypeptide useful for generating an antibody having immunoreactivity against a polypeptide encoded by a nucleotide sequence that hybridizes under stringent 30 conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a fragment, domain, or feature thereof.

III.C. Polypeptides

The presently disclosed subject matter further relates to isolated polypeptides that are orthologs of the polypeptides comprising the amino acid sequences set forth in even numbered SEQ ID NOs: 2-10, including biologically active polypeptides. In one
5 embodiment, the polypeptide comprises a polypeptide sequence of an ortholog of a polypeptide listed in even numbered sequences of SEQ ID NOs: 2-10. In another embodiment, the polypeptide comprises a functional fragment or domain of an ortholog of a polypeptide comprising a polypeptide sequence listed in even numbered sequences of SEQ ID NOs: 2-10. In yet another embodiment, the polypeptide
10 comprises a chimera of an ortholog of the polypeptide sequence listed in even numbered sequences of SEQ ID NOs: 2-10, where the chimera can comprise functional polypeptide motifs, including domains, repeats, post-translational modification sites, or other features.

Representative orthologs are disclosed in Examples 6-11 herein below. In some
15 embodiments, a polypeptide of the presently disclosed subject matter has or comprises substantial identity (in some embodiments at least about 90% amino acid identity) with a polypeptide as disclosed in Examples 6-11, over an entire sequence or functional fragment thereof. In one embodiment, the polypeptide functions as an inositol phosphate kinase. More specifically, in some embodiments, the polypeptide is an Ipk2
20 polypeptide and functions as a $IP_3/IP_4/IP_5$ 6-/3-/5-kinase generating multiple IP_4 , IP_5 and IP_6 products. More specifically, in some other embodiments, the polypeptide is an Ipk1 polypeptide and functions as an $I(1,4,5,6)P_4$, $I(1,3,4,6)P_4$ and $I(1,3,4,5,6)P_5$ 2-kinase generating $I(1,2,4,5,6)P_5$, $I(1,2,3,4,6)P_5$ and phytate (IP_6).

In another embodiment, the polypeptide is a plant polypeptide. In one
25 embodiment, the plant is a dicot. In another embodiment, the plant is a gymnosperm. In another embodiment, the plant is a monocot. In one embodiment, the monocot is a cereal. A more inclusive, but not limited to, list of plant genera relevant to the presently disclosed subject matter include *Arabidopsis thaliana*, corn (*Zea mays*), *Brassica* sp., alfalfa (*Medicago sativa*), rice (*Oryza sativa* ssp.), rye (*Secale cereale*), sorghum
30 (*Sorghum bicolor*, *Sorghum vulgare*), pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum*

aestivum), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanut (*Arachis hypogaea*), cotton, sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilis*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed (*Lemna*), and barley. Still other plants relevant to the presently disclosed subject matter include vegetables, ornamentals, and conifers. Relevant vegetables include, but are not limited to, tomatoes, lettuce, guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, green bean, lima bean, pea, and members of the genus *Cucumis*. Relevant ornamentals include, but are not limited to, impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, azalea, hydrangea, hibiscus, rose, tulip, daffodil, petunia, carnation, poinsettia, and chrysanthemum. Relevant conifers include, but are not limited to, loblolly pine, slash pine, ponderosa pine, lodgepole pine, Monterey pine, Douglas-fir, Western hemlock, Sitka spruce, redwood, silver fir, balsam fir, Western red cedar, and Alaska yellow-cedar.

In one embodiment, the polypeptide is expressed in a specific location or tissue of a plant. In one embodiment, the location or tissue includes, but is not limited to, epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, and combinations thereof. In another embodiment, the location or tissue is a seed.

In one embodiment, the polypeptide functions as an inositol phosphate kinase. More specifically, in some embodiments, the polypeptide is an Ipk2 polypeptide and functions as a IP₃/IP₄/IP₅ 6-/3-/5-kinase generating multiple IP₄, IP₅ and IP₆ products. More specifically, in some other embodiments, the polypeptide is an Ipk1 polypeptide and functions as an I(1,4,5,6)P₄, I(1,3,4,6)P₄ and I(1,3,4,5,6)P₅ 2-kinase generating I(1,2,4,5,6)P₅, I(1,2,3,4,6)P₅ and phytate (IP₆).

In one embodiment, isolated polypeptides comprise the amino acid sequences of orthologs of the polypeptides comprising the amino acid sequences set forth in even numbered SEQ ID NOs: 2-10, and variants having conservative amino acid modifications. The term "conservative modified variants" refers to polypeptides that can
5 be encoded by nucleic acid sequences having degenerate codon substitutions wherein at least one position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, 1991; Ohtsuka *et al.*, 1985; Rossolini *et al.*, 1994). Additionally, one skilled in the art will recognize that individual substitutions, deletions, or additions to a nucleic acid, peptide, polypeptide, or
10 polypeptide sequence that alters, adds, or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservative modification" where the modification results in the substitution of an amino acid with a chemically similar amino acid. Conservative modified variants provide similar biological activity as the unmodified polypeptide. Conservative substitution tables listing functionally similar
15 amino acids are known in the art. See Creighton, 1984.

The term "conservatively modified variant" also refers to a peptide having an amino acid residue sequence substantially identical to a sequence of a polypeptide of the presently disclosed subject matter in which one or more residues have been conservatively substituted with a functionally similar residue. Examples of conservative
20 substitutions include the substitution of one non-polar (hydrophobic) residue such as isoleucine, valine, leucine or methionine for another; the substitution of one polar (hydrophilic) residue for another such as between arginine and lysine, between glutamine and asparagine, between glycine and serine; the substitution of one basic residue such as lysine, arginine or histidine for another; or the substitution of one acidic
25 residue, such as aspartic acid or glutamic acid for another.

Amino acid substitutions, such as those which might be employed in modifying the polypeptides described herein, are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. An analysis of the size, shape and type of the amino acid
30 side-chain substituents reveals that arginine, lysine and histidine are all positively charged residues; that alanine, glycine and serine are all of similar size; and that phenylalanine, tryptophan and tyrosine all have a generally similar shape. Therefore,

based upon these considerations, arginine, lysine and histidine; alanine, glycine and serine; and phenylalanine, tryptophan and tyrosine; are defined herein as biologically functional equivalents. Other biologically functionally equivalent changes will be appreciated by those of skill in the art.

5 In making biologically functional equivalent amino acid substitutions, the hydrophobic index of amino acids can be considered. Each amino acid has been assigned a hydrophobic index on the basis of their hydrophobicity and charge characteristics, these are: isoleucine (+ 4.5); valine (+ 4.2); leucine (+ 3.8); phenylalanine (+ 2.8); cysteine (+ 2.5); methionine (+ 1.9); alanine (+ 1.8); glycine (-
10 0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydrophobic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte & Doolittle, 1982,
15 incorporated herein by reference). It is known that certain amino acids can be substituted for other amino acids having a similar hydrophobic index or score and still retain a similar biological activity. Substitutions of amino acids involve amino acids for which the hydrophobic indices are in one embodiment within ± 2 of the original value, in another embodiment within ± 1 of the original value, and in still another embodiment
20 within ± 0.5 of the original value in making changes based upon the hydrophobic index.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Pat. No. 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its
25 immunogenicity and antigenicity, *i.e.* with a biological property of the protein. It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent protein.

As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 $\pm$
30 1); glutamate (+3.0 $\pm$ 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 $\pm$ 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3);

phenylalanine (-2.5); tryptophan (-3.4).

Substitutions of amino acids involve amino acids for which the hydrophilicity values are in one embodiment within ± 2 of the original value, in another embodiment within ± 1 of the original value, and in still another embodiment within ± 0.5 of the original value in making changes based upon similar hydrophilicity values.

While discussion has focused on functionally equivalent polypeptides arising from amino acid changes, it will be appreciated that these changes can be effected by alteration of the encoding DNA, taking into consideration also that the genetic code is degenerate and that two or more codons can code for the same amino acid.

In an alternative embodiment, the sequence having substantial identity contains a deletion or insertion of at least one amino acid. In another embodiment, the deletion or insertion is of less than about ten amino acids. In still another embodiment, the deletion or insertion is of less than about three amino acids.

In one embodiment, the sequence having substantial identity encodes a substitution in at least one amino acid.

Embodiments of the presently disclosed subject matter also provide an isolated polypeptide comprising a polypeptide sequence selected from the group consisting of:

- (e) an amino acid sequence of one of even numbered SEQ ID NOs: 2-10;
- (f) an amino acid sequence that is at least 40% identical to (a);
- (g) an amino acid sequence encoded by a nucleotide sequence substantially identical to a nucleotide sequence of one of odd numbered SEQ ID NOs: 1-9;
- (h) an amino acid sequence encoded by a nucleic acid molecule capable of hybridizing under stringent conditions to a nucleic acid molecule of one of odd numbered SEQ ID NOs: 1-9 or to a sequence fully complementary thereto; and
- (i) a functional fragment of (a), (b), (c) or (d).

In one embodiment, a polypeptide having substantial identity to a polypeptide sequence listed in even numbered SEQ ID NO: 2-10, or a domain or feature thereof, is an allelic variant of the polypeptide sequence listed in even numbered SEQ ID NO: 2-10. In another embodiment, a polypeptide having substantial identity to a polypeptide sequence listed in even numbered SEQ ID NO: 2-10, or a domain or feature thereof, is

a naturally occurring variant of the polypeptide sequence listed in even numbered SEQ ID NO: 2-10. In another embodiment, a polypeptide having substantial identity to a polypeptide sequence listed in even numbered SEQ ID NO: 2-10, or a domain or feature thereof, is a polymorphic variant of the polypeptide sequence listed in even
5 numbered SEQ ID NO: 2-10.

In one embodiment, the polypeptide is an ortholog of a polypeptide comprising one of the amino acid sequences listed in even numbered SEQ ID NO: 2-10. In another embodiment, the polypeptide is a functional fragment or domain of an ortholog of a polypeptide comprising one of the amino acid sequences listed in even numbered
10 SEQ ID NOs: 2-10. In yet another embodiment, the polypeptide is a chimera, where the chimera comprises a functional polypeptide domain, including, but not limited to, a domain, a repeat, a post-translational modification site, and combinations thereof.

Representative orthologs are disclosed in Examples 6-11 herein below. In some embodiments, a polypeptide of the presently disclosed subject matter has or comprises
15 substantial identity (in some embodiments at least about 90% amino acid identity) with a polypeptide as disclosed in Examples 6-11, over an entire sequence or functional fragment thereof. In one embodiment, the polypeptide functions as an inositol phosphate kinase. More specifically, in some embodiments, the polypeptide is an Ipk2 polypeptide and functions as a dual-specificity IP_3/IP_4 6-/3-kinase generating IP_5 . More
20 specifically, in some other embodiments, the polypeptide is an Ipk1 polypeptide and functions as an $I(1,3,4,5,6)P_5$ 2-kinase generating phytate (IP_6).

In one embodiment, the polypeptide is a plant polypeptide. In one embodiment, the plant is a dicot. In another embodiment, the plant is a gymnosperm. In another embodiment, the plant is a monocot. In one embodiment, the monocot is a cereal. A
25 more inclusive, but not intended to be limiting, list of plant genera relevant to the presently disclosed subject matter include *Arabidopsis thaliana*, corn (*Zea mays*), *Brassica* sp., alfalfa (*Medicago sativa*), rice (*Oryza sativa* ssp.), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat
30 (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanut (*Arachis hypogaea*), cotton, sweet potato (*Ipomoea*

batatus), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilane*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya
5 (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed (*Lemna*), and barley. Still other plants relevant to the presently disclosed subject matter include vegetables, ornamentals, and conifers. Relevant vegetables include, but are not limited to, tomatoes, lettuce, guar, locust bean,
10 fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, green bean, lima bean, pea, and members of the genus *Cucumis*. Relevant ornamentals include, but are not limited to, impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura,
15 Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, azalea, hydrangea, hibiscus, rose, tulip, daffodil, petunia, carnation, poinsettia, and chrysanthemum. Relevant conifers include, but are not limited to, loblolly pine, slash pine, ponderosa pine, lodgepole pine, Monterey pine, Douglas-fir, Western hemlock, Sitka spruce, redwood, silver fir, balsam fir, Western red
20 cedar, and Alaska yellow-cedar.

In one embodiment, the polypeptide is expressed in a specific location or tissue of a plant. In one embodiment, the location or tissue includes, but is not limited to, epidermis, vascular tissue, meristem, cambium, cortex, or pith. In another embodiment, the location or tissue is leaf or sheath, root, flower, and developing ovule or seed. In
25 another embodiment, the location or tissue can be, for example, epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, or flower. In yet another embodiment, the location or tissue is a seed.

In one embodiment, the polypeptide sequence is encoded by a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M
30 NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered SEQ ID NO: 1-9 or a fragment, domain, or feature thereof or a sequence complementary thereto, wherein the nucleotide sequence includes a deletion

or insertion of at least one nucleotide. In one embodiment, the deletion or insertion is of less than about thirty nucleotides. In another embodiment, the deletion or insertion is of less than about five nucleotides. In another embodiment, the polypeptide sequence encoded by a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered SEQ ID NO: 1-9, or a fragment, domain, or feature thereof or a sequence complementary thereto, includes a substitution of at least one codon. In one embodiment, the substitution is conservative.

In another embodiment, the polypeptide sequences having substantial identity to the polypeptide sequence listed in even numbered SEQ ID NO: 2-10, or a fragment, domain, repeat, feature, or chimera thereof, includes a deletion or insertion of at least one amino acid.

The polypeptides of the presently disclosed subject matter, fragments thereof, or variants thereof, can comprise any number of contiguous amino acid residues from a polypeptide of the presently disclosed subject matter, wherein the number of residues is selected from the group of integers consisting of from 10 to the number of residues in a full-length polypeptide of the presently disclosed subject matter. In one embodiment, the portion or fragment of the polypeptide is a functional polypeptide. The presently disclosed subject matter includes active polypeptides having specific activity of at least in one embodiment 20%, in another embodiment 30%, in another embodiment 40%, in another embodiment 50%, in another embodiment 60%, in another embodiment 70%, in another embodiment 80%, in another embodiment 90%, and in still another embodiment 95% that of the native (non-synthetic) endogenous polypeptide. Further, the substrate specificity (k_{cat}/K_m) can be substantially identical to the native (non-synthetic), endogenous polypeptide. Typically the K_m will be at least in one embodiment 30%, in another embodiment 40%, in another embodiment 50% of the native, endogenous polypeptide; and in another embodiment at least 60%, in another embodiment 70%, in another embodiment 80%, and in yet another embodiment 90% of the native, endogenous polypeptide. Methods of assaying and quantifying measures of activity and substrate specificity are well known to those of skill in the art.

The isolated polypeptides of the presently disclosed subject matter can elicit production of an antibody specifically reactive to a polypeptide of the presently

disclosed subject matter when presented as an immunogen. Therefore, the polypeptides of the presently disclosed subject matter can be employed as immunogens for constructing antibodies immunoreactive to a polypeptide of the presently disclosed subject matter for such purposes including, but not limited to, immunoassays or polypeptide purification techniques. Immunoassays for determining binding are well known to those of skill in the art and include, but are not limited to, enzyme-linked immunosorbent assays (ELISAs) and competitive immunoassays.

Embodiments of the presently disclosed subject matter also relate to chimeric polypeptides encoded by the isolated nucleic acid molecules of the present disclosure including a chimeric polypeptide containing a polypeptide sequence encoded by an isolated nucleic acid containing a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered SEQ ID NO: 1-9, or an exon, domain, or feature thereof;
- (b) a nucleotide sequence complementary (for example, fully complementary) to (a); and
- (c) a nucleotide sequence which is the reverse complement (for example, full reverse complement) of (a);
- (d) or a functional fragment thereof.

IV. Controlling and Altering the Expression of Nucleic Acid Molecules

IV.A. General Considerations

One aspect of the presently disclosed subject matter provides compositions and methods for altering or modulating (*i.e.* increasing or decreasing) the level of nucleic acid molecules and/or polypeptides of the presently disclosed subject matter in plants. In particular, the nucleic acid molecules and polypeptides of the presently disclosed subject matter are expressed constitutively, temporally, or spatially (*e.g.* at developmental stages), in certain tissues, and/or quantities, which are uncharacteristic of non-recombinantly engineered plants. Therefore, the presently disclosed subject matter provides utility in such exemplary applications as altering the specified

characteristics identified above.

The isolated nucleic acid molecules of the presently disclosed subject matter are useful for expressing a polypeptide of the presently disclosed subject matter in a recombinantly engineered cell such as a bacterial, yeast, insect, mammalian, or plant cell. Expressing cells can produce the polypeptide in a non-natural condition (e.g. in quantity, composition, location and/or time) because they have been genetically altered to do so. Those skilled in the art are knowledgeable in the numerous expression systems available for expression of nucleic acids encoding a polypeptide of the presently disclosed subject matter.

Embodiments of the presently disclosed subject matter provide an expression cassette comprising a promoter sequence operatively linked to an isolated nucleic acid as disclosed herein above, the isolated nucleic acid encoding a polypeptide as disclosed herein above. In some embodiments the nucleic acid comprise one of:

- (a) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of one of even numbered SEQ ID NOs: 2-10 or a polypeptide at least 40% identical to even numbered SEQ ID NOs: 2-10 and having inositol phosphate kinase activity;
- (b) a nucleic acid molecule comprising a nucleic acid sequence of one of odd numbered SEQ ID NOs: 1-9;
- (c) a nucleic acid molecule that has a nucleic acid sequence that is substantially identical to the nucleic acid sequence of the nucleic acid molecule of (a) or (b);
- (d) a nucleic acid molecule that hybridizes to (a) or (b) under stringent hybridization conditions;
- (e) a nucleic acid molecule comprising a nucleic acid sequence complementary to (a); and
- (f) a nucleic acid molecule comprising a nucleic acid sequence that is the full reverse complement of (a).

Further encompassed within the presently disclosed subject matter is a recombinant vector comprising an expression cassette according to the embodiments of the presently disclosed subject matter. Also encompassed are transgenic plant cells comprising expression cassettes according to the present disclosure, and transgenic

plants comprising these plant cells. In one embodiment, the plant is a dicot. In another embodiment, the plant is a gymnosperm. In another embodiment, the plant is a monocot. In one embodiment, the monocot is a cereal. A more inclusive, but not intended to be limiting, list of plant genera relevant to the presently disclosed subject matter include *Arabidopsis thaliana*, corn (*Zea mays*), *Brassica* sp., alfalfa (*Medicago sativa*), rice (*Oryza sativa* ssp.), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanut (*Arachis hypogaea*), cotton, sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilane*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed (*Lemna*), and barley. Still other plants relevant to the presently disclosed subject matter include vegetables, ornamentals, and conifers. Relevant vegetables include, but are not limited to, tomatoes, lettuce, guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, green bean, lima bean, pea, and members of the genus *Cucumis*. Relevant ornamentals include, but are not limited to, impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, azalea, hydrangea, hibiscus, rose, tulip, daffodil, petunia, carnation, poinsettia, and chrysanthemum. Relevant conifers include, but are not limited to, loblolly pine, slash pine, ponderosa pine, lodgepole pine, Monterey pine, Douglas-fir, Western hemlock, Sitka spruce, redwood, silver fir, balsam fir, Western red cedar, and Alaska yellow-cedar.

In one embodiment, the expression cassette is expressed throughout the plant. In another embodiment, the expression cassette is expressed in a specific location or tissue of a plant. In one embodiment, the location or tissue includes, but is not limited to, epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, and combinations thereof. In another embodiment, the location or tissue is a seed.

In one embodiment, the expression cassette is involved in a function including, but not limited to, inositol phosphate kinase activity resulting in phytate production. In one embodiment, the chimeric polypeptide functions as an inositol phosphate kinase, or replaces the endogenous inositol phosphate kinase polypeptide as a non-functional or reduced function homolog.

Embodiments of the presently disclosed subject matter also relate to an expression vector comprising an isolated nucleic acid as disclosed herein above, the isolated nucleic acid encoding a polypeptide as disclosed herein above. In some embodiments the nucleic acid comprises one of: (a) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of one of even numbered SEQ ID NOs: 2-10 or a polypeptide at least 40% identical to even numbered SEQ ID NOs: 2-10 and having inositol phosphate kinase activity; (b) a nucleic acid molecule comprising a nucleic acid sequence of one of odd numbered SEQ ID NOs: 1-9; (c) a nucleic acid molecule that has a nucleic acid sequence that is substantially identical to the nucleic acid sequence of the nucleic acid molecule of (a) or (b); (d) a nucleic acid molecule that hybridizes to (a) or (b) under stringent hybridization conditions; (e) a nucleic acid molecule comprising a nucleic acid sequence complementary to (a); and (f) a nucleic acid molecule comprising a nucleic acid sequence that is the full reverse complement of (a).

In one embodiment, the expression vector comprises one or more elements including, but not limited to, a promoter-enhancer sequence, a selection marker sequence, an origin of replication, an epitope tag-encoding sequence, and an affinity purification tag-encoding sequence. In one embodiment, the promoter-enhancer sequence comprises, for example, the cauliflower mosaic virus (CaMV) 35S promoter, the CaMV 19S promoter, the tobacco PR-1a promoter, the ubiquitin promoter, or the phaseolin promoter. In one embodiment, the promoter is operable in plants, and in another embodiment, the promoter is a constitutive or inducible promoter. In one

embodiment, the selection marker sequence encodes an antibiotic resistance gene. In one embodiment the affinity purification tag sequence encodes a polyamino acid sequence or a polypeptide. In one embodiment, the polyamino acid sequence comprises polyhistidine. In one embodiment, the polypeptide is chitin-binding domain or glutathione-S-transferase. In another embodiment, the affinity purification tag sequence comprises an intein encoding sequence.

In one embodiment, the expression vector comprises a eukaryotic expression vector, and in another embodiment, the expression vector comprises a prokaryotic expression vector. In one embodiment, the eukaryotic expression vector comprises a tissue-specific promoter. In another embodiment, the expression vector is operable in plants.

Embodiments of the presently disclosed subject matter also relate to a cell comprising a nucleic acid construct comprising an expression vector and a nucleic acid comprising a nucleic acid encoding a polypeptide that is an ortholog of a polypeptide as listed in even numbered sequences of SEQ ID NOs: 2-10, or a nucleic acid sequence that hybridizes under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a subsequence thereof, in combination with a heterologous sequence. Representative orthologs are disclosed in Examples 6-11 herein below. In some embodiments, a polypeptide of the presently disclosed subject matter has or comprises substantial identity (in some embodiments at least about 90% amino acid identity) with a polypeptide as disclosed in Examples 6-11, over an entire sequence or functional fragment thereof. In one embodiment, the polypeptide functions as an inositol phosphate kinase. More specifically, in some embodiments, the polypeptide is an Ipk2 polypeptide and functions as a dual-specificity IP₃/IP₄ 6-/3-kinase generating IP₅. More specifically, in some other embodiments, the polypeptide is an Ipk1 polypeptide and functions as an I(1,3,4,5,6)P₅ 2-kinase generating phytate (IP₆).

In one embodiment, the cell is a bacterial cell, a fungal cell, a plant cell, or an animal cell. In one embodiment, the polypeptide is expressed in a specific location or tissue of a plant. In one embodiment, the location or tissue includes, but is not limited to, epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, and

combinations thereof. In an alternative embodiment, the location or tissue is a seed. In one embodiment, the polypeptide functions as an inositol phosphate kinase.

Prokaryotic cells including, but not limited to, *Escherichia coli* and other microbial strains known to those in the art, can be used as host cells. Methods for expressing polypeptides in prokaryotic cells are well known to those in the art and can be found in many laboratory manuals such as Sambrook & Russell, 2001. A variety of promoters, ribosome binding sites, and operators to control expression are available to those skilled in the art, as are selectable markers such as antibiotic resistance genes. The type of vector is chosen to allow for optimal growth and expression in the selected cell type.

A variety of eukaryotic expression systems are available such as, for example, yeast, insect cell lines, plant cells, and mammalian cells. Expression and synthesis of heterologous polypeptides in yeast is well known (see Sherman *et al.*, 1982). Yeast strains widely used for production of eukaryotic polypeptides are *Saccharomyces cerevisiae* and *Pichia pastoris*, and vectors, strains, and protocols for expression are available from commercial suppliers (e.g., Invitrogen Corp., Carlsbad, California, United States of America).

Mammalian cell systems can be transformed with expression vectors for production of polypeptides. Suitable host cell lines available to those in the art include, but are not limited to, the HEK293, BHK21 and CHO cell lines. Expression vectors for these cells can include expression control sequences such as an origin of replication, a promoter, (e.g., the CMV promoter, a Herpes Simplex Virus thymidine kinase (HSV-tk) promoter or phosphoglycerate kinase (pgk) promoter), an enhancer, and polypeptide processing sites such as ribosome binding sites, RNA splice sites, polyadenylation sites, and transcription terminator sequences. Other animal cell lines useful for the production of polypeptides are available commercially or from depositories such as the American Type Culture Collection (Manassas, Virginia, United States of America).

Expression vectors for expressing polypeptides in insect cells are usually derived from baculovirus or other viruses known in the art. A number of suitable insect cell lines are available including, but not limited to, mosquito larvae, silkworm, armyworm (for example, *Spodoptera frugiperda*), moth, and *Drosophila* cell lines.

Methods of transforming animal and lower eukaryotic cells are known.

Numerous methods can be used to introduce exogenous DNA into eukaryotic cells including, but not limited to, calcium phosphate precipitation, fusion of the recipient cell with bacterial protoplasts containing the DNA, treatment of the recipient cells with liposomes containing the DNA, DEAE dextran, electroporation, biolistics, and
5 microinjection of the DNA directly into the cells. Transformed cells are cultured using means well known in the art (see Kuchler, 1997).

Once a polypeptide of the presently disclosed subject matter is expressed it can be isolated and purified from the expressing cells using methods known to those skilled in the art. The purification process can be monitored using Western blot techniques,
10 radioimmunoassay, or other standard immunoassay techniques. Polypeptide purification techniques are commonly known and used by those skilled in the art (see Scopes, 1982; Deutscher *et al.*, 1990).

Embodiments of the presently disclosed subject matter provide a method for producing a recombinant polypeptide in which the expression vector comprises one or
15 more elements including, but not limited to, a promoter-enhancer sequence, a selection marker sequence, an origin of replication, an epitope tag-encoding sequence, and an affinity purification tag-encoding sequence. In one embodiment, the nucleic acid construct comprises an epitope tag-encoding sequence and the isolating step employs an antibody specific for the epitope tag. In another embodiment, the nucleic acid
20 construct comprises a polyamino acid-encoding sequence and the isolating step employs a resin comprising a polyamino acid binding substance, in one embodiment where the polyamino acid is polyhistidine and the polyamino acid binding resin is nickel-charged agarose resin. In yet another embodiment, the nucleic acid construct comprises a polypeptide-encoding sequence and the isolating step employs a resin
25 comprising a polypeptide binding substance. In one embodiment, the polypeptide is a chitin-binding domain and the resin contains chitin-sepharose.

The polypeptides of the presently disclosed subject matter can be synthesized using non-cellular synthetic methods known to those in the art. Techniques for solid phase synthesis are disclosed in Barany & Merrifield, 1980; Merrifield *et al.*, 1963;
30 Stewart & Young, 1984.

The presently disclosed subject matter further provides a method for modifying (*i.e.* increasing or decreasing) the concentration or composition or activity of a

polypeptide of the presently disclosed subject matter in a plant or part thereof. Modification can be effected by increasing or decreasing the concentration, the composition, and/or the activity (*i.e.* the ratio of the polypeptides of the presently disclosed subject matter) in a plant. The method comprises introducing into a plant cell
5 an expression cassette comprising a nucleic acid molecule of the presently disclosed subject matter as disclosed above to obtain a transformed plant cell or tissue, and culturing the transformed plant cell or tissue. The nucleic acid molecule can be under the regulation of a constitutive or inducible promoter. The method can further comprise inducing or repressing expression of a nucleic acid molecule of a sequence in the plant
10 for a time sufficient to modify the concentration and/or composition in the plant or plant part. Inducing or repressing the nucleic acids disclosed herein encoding lpk1 or 2 correlates with an observable modification in the amount of phytate produced in the plant or plant part relative to a non-transformed parental plant.

A plant or plant part having modified expression of a nucleic acid molecule of the
15 presently disclosed subject matter can be analyzed and selected using methods known to those skilled in the art including, but not limited to, Southern blotting, DNA sequencing, or PCR analysis using primers specific to the nucleic acid molecule and detecting amplicons produced therefrom.

In general, a concentration or composition is increased or decreased by at least
20 in one embodiment 5%, in another embodiment 10%, in another embodiment 20%, in another embodiment 30%, in another embodiment 40%, in another embodiment 50%, in another embodiment 60%, in another embodiment 70%, in another embodiment 80%, and in still another embodiment greater than 90% relative to a non-transformed native control plant, plant part, or cell lacking the expression cassette.

25 IV.B. Alteration of Expression of Nucleic Acid Molecules

The alteration in expression of the nucleic acid molecules of the presently disclosed subject matter can be achieved, for example, in one of the following ways:

IV.B.1. "Sense" Suppression

Alteration of the expression of a nucleotide sequence of the presently disclosed
30 subject matter, in one embodiment reduction of its expression, is obtained by "sense" suppression (referenced in *e.g.* Jorgensen *et al.*, 1996). In this case, the entirety or a portion of a nucleotide sequence of the presently disclosed subject matter is comprised

in a DNA molecule. The DNA molecule can be operatively linked to a promoter functional in a cell comprising the target gene, in one embodiment a plant cell, and introduced into the cell, in which the nucleotide sequence is expressible. The nucleotide sequence is inserted in the DNA molecule in the "sense orientation", meaning that the coding strand of the nucleotide sequence can be transcribed. In one embodiment, the nucleotide sequence is fully translatable and all the genetic information comprised in the nucleotide sequence, or portion thereof, is translated into a polypeptide. In another embodiment, the nucleotide sequence is partially translatable and a short peptide is translated. In one embodiment, this is achieved by inserting at least one premature stop codon in the nucleotide sequence, which brings translation to a halt. In another embodiment, the nucleotide sequence is transcribed but no translation product is made. This is usually achieved by removing the start codon, *i.e.* the "ATG", of the polypeptide encoded by the nucleotide sequence. In a further embodiment, the DNA molecule comprising the nucleotide sequence, or a portion thereof, is stably integrated in the genome of the plant cell. In another embodiment, the DNA molecule comprising the nucleotide sequence, or a portion thereof, is comprised in an extrachromosomally replicating molecule.

In transgenic plants containing one of the DNA molecules disclosed immediately above, the expression of the nucleotide sequence corresponding to the nucleotide sequence comprised in the DNA molecule can be reduced. The nucleotide sequence in the DNA molecule in one embodiment is at least 70% identical to the nucleotide sequence the expression of which is reduced, in another embodiment is at least 80% identical, in another embodiment is at least 90% identical, in another embodiment is at least 95% identical, and in still another embodiment is at least 99% identical.

25 IV.B.2. "Antisense" Suppression

In another embodiment, the alteration of the expression of a nucleotide sequence of the presently disclosed subject matter, for example the reduction of its expression, is obtained by "antisense" suppression. The entirety or a portion of a nucleotide sequence of the presently disclosed subject matter is comprised in a DNA molecule. The DNA molecule can be operatively linked to a promoter functional in a plant cell, and introduced in a plant cell, in which the nucleotide sequence is expressible. The nucleotide sequence is inserted in the DNA molecule in the "antisense

orientation", meaning that the reverse complement (also called sometimes non-coding strand) of the nucleotide sequence can be transcribed. In one embodiment, the DNA molecule comprising the nucleotide sequence, or a portion thereof, is stably integrated in the genome of the plant cell. In another embodiment the DNA molecule comprising the nucleotide sequence, or a portion thereof, is comprised in an extrachromosomally replicating molecule. Several publications describing this approach are cited for further illustration (Green *et al.*, 1986; van der Krol *et al.*, 1991; Powell *et al.*, 1989; Ecker & Davis, 1986).

In transgenic plants containing one of the DNA molecules disclosed immediately above, the expression of the nucleotide sequence corresponding to the nucleotide sequence comprised in the DNA molecule can be reduced. The nucleotide sequence in the DNA molecule is in one embodiment at least 70% identical to the nucleotide sequence the expression of which is reduced, in another embodiment at least 80% identical, in another embodiment at least 90% identical, in another embodiment at least 95% identical, and in still another embodiment at least 99% identical.

IV.B.3. Homologous Recombination

In another embodiment, at least one genomic copy corresponding to a nucleotide sequence of the presently disclosed subject matter is modified in the genome of the plant by homologous recombination as further illustrated in Paszkowski *et al.*, 1988.

This technique uses the ability of homologous sequences to recognize each other and to exchange nucleotide sequences between respective nucleic acid molecules by a process known in the art as homologous recombination. Homologous recombination can occur between the chromosomal copy of a nucleotide sequence in a cell and an incoming copy of the nucleotide sequence introduced in the cell by transformation.

Specific modifications are thus accurately introduced in the chromosomal copy of the nucleotide sequence. In one embodiment, the regulatory elements of the nucleotide sequence of the presently disclosed subject matter are modified. Such regulatory elements are easily obtainable by screening a genomic library using the nucleotide sequence of the presently disclosed subject matter, or a portion thereof, as a probe.

The existing regulatory elements are replaced by different regulatory elements, thus altering expression of the nucleotide sequence, or they are mutated or deleted, thus abolishing the expression of the nucleotide sequence. In another embodiment, the

nucleotide sequence is modified by deletion of a part of the nucleotide sequence or the entire nucleotide sequence, or by mutation. Expression of a mutated polypeptide in a plant cell is also provided in the presently disclosed subject matter. Recent refinements of this technique to disrupt endogenous plant genes have been disclosed (Kempin *et al.*, 1997 and Miao & Lam, 1995) and are generally known in the art.

In one embodiment, a mutation in the chromosomal copy of a nucleotide sequence is introduced by transforming a cell with a chimeric oligonucleotide composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends. An additional feature of the oligonucleotide is for example the presence of 2'-O-methylation at the RNA residues. The RNA/DNA sequence is designed to align with the sequence of a chromosomal copy of a nucleotide sequence of the presently disclosed subject matter and to contain the desired nucleotide change. For example, this technique is further illustrated in U.S. Patent No. 5,501,967 and Zhu *et al.*, 1999.

IV.B.4. Ribozymes

In a further embodiment, an RNA coding for a polypeptide of the presently disclosed subject matter is cleaved by a catalytic RNA, or ribozyme, specific for such RNA. The ribozyme is expressed in transgenic plants and results in reduced amounts of RNA coding for the polypeptide of the presently disclosed subject matter in plant cells, thus leading to reduced amounts of polypeptide accumulated in the cells. This method is further illustrated in U.S. Patent No. 4,987,071.

IV.B.5. Dominant-Negative Mutants

In another embodiment, the activity of a polypeptide encoded by the nucleotide sequences of the presently disclosed subject matter is changed. This is achieved by expression of dominant negative mutants of the polypeptides in transgenic plants, leading to the loss of activity of the endogenous polypeptide.

IV.B.6. Aptamers

In a further embodiment, the activity of polypeptide of the presently disclosed subject matter is inhibited by expressing in transgenic plants nucleic acid ligands, so-called aptamers, which specifically bind to the polypeptide. Aptamers can be obtained by the SELEX (Systematic Evolution of Ligands by Exponential Enrichment) method. In the SELEX method, a candidate mixture of single stranded nucleic acids having regions

of randomized sequence is contacted with the polypeptide and those nucleic acids having an increased affinity to the target are partitioned from the remainder of the candidate mixture. The partitioned nucleic acids are amplified to yield a ligand-enriched mixture. After several iterations a nucleic acid with optimal affinity to the polypeptide is
5 obtained and is used for expression in transgenic plants. This method is further illustrated in U.S. Patent No. 5,270,163.

IV.B.7. Zinc Finger Polypeptides

A zinc finger polypeptide that binds a nucleotide sequence of the presently disclosed subject matter or to its regulatory region can also be used to alter expression
10 of the nucleotide sequence. In alternative embodiments, transcription of the nucleotide sequence is reduced or increased. Zinc finger polypeptides are disclosed in, for example, Beerli *et al.*, 1998, or in WO 95/19431, WO 98/54311, or WO 96/06166, all incorporated herein by reference in their entirety.

IV.B.8. dsRNA

Alteration of the expression of a nucleotide sequence of the presently disclosed subject matter can also be obtained by double stranded RNA (dsRNA) interference (RNAi) as disclosed, for example, in WO 99/32619, WO 99/53050, or WO 99/61631, all
15 incorporated herein by reference in their entirety. RNAi as used herein encompasses both plasmid-based and genomic-based strategies.

In one embodiment, the alteration of the expression of a nucleotide sequence of the presently disclosed subject matter, in one embodiment the reduction of its expression, is obtained by dsRNA interference. The entirety, or in one embodiment a portion, of a nucleotide sequence of the presently disclosed subject matter, can be comprised in a DNA molecule. The size of the DNA molecule is in one embodiment
20 from 100 to 1000 nucleotides or more; the optimal size to be determined empirically. Two copies of the identical DNA molecule are linked, separated by a spacer DNA molecule, such that the first and second copies are in opposite orientations. In one embodiment, the first copy of the DNA molecule is the reverse complement (also known as the non-coding strand) and the second copy is the coding strand; in another
25 embodiment, the first copy is the coding strand, and the second copy is the reverse complement. The size of the spacer DNA molecule is in one embodiment 200 to 10,000 nucleotides, in another embodiment 400 to 5000 nucleotides, and in yet another
30

embodiment 600 to 1500 nucleotides in length. The spacer is in one embodiment a random piece of DNA, in another embodiment a random piece of DNA without homology to the target organism for dsRNA interference, and in still another embodiment a functional intron that is effectively spliced by the target organism. The two copies of the DNA molecule separated by the spacer are operatively linked to a promoter functional in a plant cell, and introduced in a plant cell in which the nucleotide sequence is expressible. In one embodiment, the DNA molecule comprising the nucleotide sequence, or a portion thereof, is stably integrated in the genome of the plant cell. In another embodiment, the DNA molecule comprising the nucleotide sequence, or a portion thereof, is comprised in an extrachromosomally replicating molecule. Several publications describing this approach are cited for further illustration (Waterhouse *et al.*, 1998; Chuang & Meyerowitz, 2000; Smith *et al.*, 2000).

In transgenic plants containing one of the DNA molecules disclosed immediately above, the expression of the nucleotide sequence corresponding to the nucleotide sequence comprised in the DNA molecule is in one embodiment reduced. In one embodiment, the nucleotide sequence in the DNA molecule is at least 70% identical to the nucleotide sequence the expression of which is reduced, in another embodiment it is at least 80% identical, in another embodiment it is at least 90% identical, in another embodiment it is at least 95% identical, and in still another embodiment it is at least 99% identical.

IV.B.9. Insertion of a DNA Molecule (Insertional Mutagenesis)

In one embodiment, a DNA molecule is inserted into a chromosomal copy of a nucleotide sequence of the presently disclosed subject matter, or into a regulatory region thereof. In one embodiment, such DNA molecule comprises a transposable element capable of transposition in a plant cell, such as, for example, Ac/Ds, Em/Spm, mutator. Alternatively, in another embodiment, the DNA molecule comprises a T-DNA border of an *Agrobacterium* T-DNA. The DNA molecule can also comprise a recombinase or integrase recognition site that can be used to remove part of the DNA molecule from the chromosome of the plant cell. Methods of insertional mutagenesis using T-DNA, transposons, oligonucleotides, or other methods known to those skilled in the art are also encompassed. Methods of using T-DNA and transposon for insertional mutagenesis are disclosed in Winkler & Feldmann, 1989, and Martienssen, 1998,

incorporated herein by reference in their entireties.

IV.B.10. Deletion Mutagenesis

In yet another embodiment, a mutation of a nucleic acid molecule of the presently disclosed subject matter is created in the genomic copy of the sequence in the cell or plant by deletion of a portion of the nucleotide sequence or regulator sequence. Methods of deletion mutagenesis are known to those skilled in the art. See e.g., Miao & Lam, 1995.

In yet another embodiment, a deletion is created at random in a large population of plants by chemical mutagenesis or irradiation and a plant with a deletion in a gene of the presently disclosed subject matter is isolated by forward or reverse genetics. Irradiation with fast neutrons or gamma rays is known to cause deletion mutations in plants (Silverstone *et al.*, 1998; Bruggemann *et al.*, 1996; Redei & Koncz, 1992). Deletion mutations in a gene of the presently disclosed subject matter can be recovered in a reverse genetics strategy using PCR with pooled sets of genomic DNAs as has been shown in *C. elegans* (Liu *et al.*, 1999). A forward genetics strategy involves mutagenesis of a line bearing a trait of interest followed by screening the M2 progeny for the absence of the trait. Among these mutants would be expected to be some that disrupt a gene of the presently disclosed subject matter. This could be assessed by Southern blotting or PCR using primers designed for a gene of the presently disclosed subject matter with genomic DNA from these mutants.

IV.B.11. Overexpression in a Plant Cell

In yet another embodiment, a nucleotide sequence of the presently disclosed subject matter encoding a polypeptide is over-expressed. Examples of nucleic acid molecules and expression cassettes for over-expression of a nucleic acid molecule of the presently disclosed subject matter are disclosed above. Methods known to those skilled in the art of over-expression of nucleic acid molecules are also encompassed by the presently disclosed subject matter.

In one embodiment, the expression of the nucleotide sequence of the presently disclosed subject matter is altered in every cell of a plant. This can be obtained, for example, through homologous recombination or by insertion into a chromosome. This can also be obtained, for example, by expressing a sense or antisense RNA, zinc finger polypeptide or ribozyme under the control of a promoter capable of expressing the

sense or antisense RNA, zinc finger polypeptide, or ribozyme in every cell of a plant. Constitutive, inducible, tissue-specific, or developmentally-regulated expression are also within the scope of the presently disclosed subject matter and result in a constitutive, inducible, tissue-specific, or developmentally-regulated alteration of the expression of a nucleotide sequence of the presently disclosed subject matter in the plant cell. Constructs for expression of the sense or antisense RNA, zinc finger polypeptide, or ribozyme, or for over-expression of a nucleotide sequence of the presently disclosed subject matter, can be prepared and transformed into a plant cell according to the teachings of the presently disclosed subject matter, for example, as disclosed herein.

IV.C. Construction of Plant Expression Vectors

Coding sequences intended for expression in transgenic plants can be first assembled in expression cassettes operatively linked to a suitable promoter expressible in plants. The expression cassettes can also comprise any further sequences required or selected for the expression of the transgene. Such sequences include, but are not limited to, transcription terminators, extraneous sequences to enhance expression such as introns, vital sequences, and sequences intended for the targeting of the gene product to specific organelles and cell compartments. These expression cassettes can then be easily transferred to the plant transformation vectors disclosed below. The following is a description of various components of typical expression cassettes.

IV.C.1. Promoters

The selection of the promoter used in expression cassettes can determine the spatial and temporal expression pattern of the transgene in the transgenic plant. Selected promoters can express transgenes in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific tissues or organs (seeds, roots, leaves, or flowers, for example) and the selection can reflect the desired location for accumulation of the gene product. Alternatively, the selected promoter can drive expression of the gene under various inducing conditions. Promoters vary in their strength; *i.e.*, their abilities to promote transcription. Depending upon the host cell system utilized, any one of a number of suitable promoters can be used, including the gene's native promoter. The following are non-limiting examples of promoters that can be used in expression cassettes.

IV.C.1.a. Constitutive Expression: the Ubiquitin Promoter

Ubiquitin is a gene product known to accumulate in many cell types and its promoter has been cloned from several species for use in transgenic plants (e.g. sunflower - Binet *et al.*, 1991; maize - Christensen *et al.*, 1989; and *Arabidopsis* - Callis *et al.*, 1990; Norris *et al.*, 1993). The maize ubiquitin promoter has been developed in transgenic monocot systems and its sequence and vectors constructed for monocot transformation are disclosed in the patent publication EP 0 342 926 (to Lubrizol) which is herein incorporated by reference. Taylor *et al.*, 1993, describes a vector (pAHC25) that comprises the maize ubiquitin promoter and first intron and its high activity in cell suspensions of numerous monocotyledons when introduced via microprojectile bombardment. The *Arabidopsis* ubiquitin promoter is suitable for use with the nucleotide sequences of the presently disclosed subject matter. The ubiquitin promoter is suitable for gene expression in transgenic plants, both monocotyledons and dicotyledons. Suitable vectors are derivatives of pAHC25 or any of the transformation vectors disclosed herein, modified by the introduction of the appropriate ubiquitin promoter and/or intron sequences.

IV.C.1.b. Constitutive Expression: the CaMV 35S Promoter

Construction of the plasmid pCGN1761 is disclosed in the published patent application EP 0 392 225, which is hereby incorporated by reference. pCGN1761 contains the "double" CaMV 35S promoter and the *tml* transcriptional terminator with a unique *EcoRI* site between the promoter and the terminator and has a pUC-type backbone. A derivative of pCGN1761 is constructed which has a modified polylinker that includes *NotI* and *XhoI* sites in addition to the existing *EcoRI* site. This derivative is designated pCGN1761ENX. pCGN1761ENX is useful for the cloning of cDNA sequences or coding sequences (including microbial ORF sequences) within its polylinker for the purpose of their expression under the control of the 35S promoter in transgenic plants. The entire 35S promoter-coding sequence-*tml* terminator cassette of such a construction can be excised by *HindIII*, *SphI*, *Sall*, and *XbaI* sites 5' to the promoter and *XbaI*, *BamHI* and *BglI* sites 3' to the terminator for transfer to transformation vectors such as those disclosed below. Furthermore, the double 35S promoter fragment can be removed by 5' excision with *HindIII*, *SphI*, *Sall*, *XbaI*, or *PstI*, and 3' excision with any of the polylinker restriction sites (*EcoRI*, *NotI* or *XhoI*) for replacement with another promoter. If desired, modifications around the cloning sites

can be made by the introduction of sequences that can enhance translation. This is particularly useful when overexpression is desired. For example, pCGN1761ENX can be modified by optimization of the translational initiation site as disclosed in Example 37 of U.S. Patent No. 5,639,949, incorporated herein by reference.

5 IV.C.1.c. Constitutive Expression: the Actin Promoter

Several isoforms of actin are known to be expressed in most cell types and consequently the actin promoter can be used as a constitutive promoter. In particular, the promoter from the rice *Act1* gene has been cloned and characterized (McElroy *et al.*, 1990). A 1.3 kilobase (kb) fragment of the promoter was found to contain all the
10 regulatory elements required for expression in rice protoplasts. Furthermore, numerous expression vectors based on the *Act1* promoter have been constructed specifically for use in monocotyledons (McElroy *et al.*, 1991). These incorporate the *Act1*-intron 1, *Adhl* 5' flanking sequence (from the maize alcohol dehydrogenase gene) and *Adhl*-intron 1 and sequence from the CaMV 35S promoter. Vectors showing highest expression were
15 fusions of 35S and *Act1* intron or the *Act1* 5' flanking sequence and the *Act1* intron. Optimization of sequences around the initiating ATG (of the β -glucuronidase (GUS) reporter gene) also enhanced expression. The promoter expression cassettes disclosed in McElroy *et al.*, 1991, can be easily modified for gene expression and are particularly suitable for use in monocotyledonous hosts. For example, promoter-
20 containing fragments are removed from the McElroy constructions and used to replace the double 35S promoter in pCGN1761ENX, which is then available for the insertion of specific gene sequences. The fusion genes thus constructed can then be transferred to appropriate transformation vectors. In a separate report, the rice *Act1* promoter with its first intron has also been found to direct high expression in cultured barley cells
25 (Chibbar *et al.*, 1993).

IV.C.1.d. Inducible Expression: PR-1 Promoters

The double 35S promoter in pCGN1761ENX can be replaced with any other promoter of choice that will result in suitably high expression levels. By way of example, one of the chemically regulatable promoters disclosed in U.S. Patent No. 5,614,395, such as the tobacco PR-1a promoter, can replace the double 35S promoter.
30 Alternately, the *Arabidopsis* PR-1 promoter disclosed in Lebel *et al.*, 1998, can be used. The promoter of choice can be excised from its source by restriction enzymes, but can

alternatively be PCR-amplified using primers that carry appropriate terminal restriction sites. Should PCR-amplification be undertaken, the promoter can be re-sequenced to check for amplification errors after the cloning of the amplified promoter in the target vector. The chemically/pathogen regulatable tobacco PR-1a promoter is cleaved from
5 plasmid pCIB1004 (for construction, see example 21 of EP 0 332 104, which is hereby incorporated by reference) and transferred to plasmid pCGN1761ENX (Uknes *et al.*, 1992). pCIB1004 is cleaved with *Nco*I and the resulting 3' overhang of the linearized fragment is rendered blunt by treatment with T4 DNA polymerase. The fragment is then
10 cleaved with *Hind*III and the resultant PR-1a promoter-containing fragment is gel purified and cloned into pCGN1761ENX from which the double 35S promoter has been removed. This is accomplished by cleavage with *Xho*I and blunting with T4 polymerase, followed by cleavage with *Hind*III, and isolation of the larger vector-terminator containing
15 fragment into which the pCIB1004 promoter fragment is cloned. This generates a pCGN1761ENX derivative with the PR-1a promoter and the *tml* terminator and an intervening polylinker with unique *Eco*RI and *Not*I sites. The selected coding sequence can be inserted into this vector, and the fusion products (*i.e.* promoter-gene-terminator) can subsequently be transferred to any selected transformation vector, including those
20 disclosed herein. Various chemical regulators can be employed to induce expression of the selected coding sequence in the plants transformed according to the presently disclosed subject matter, including the benzothiadiazole, isonicotinic acid, and salicylic acid compounds disclosed in U.S. Patent Nos. 5,523,311 and 5,614,395.

IV.C.1.e. Inducible Expression: an Ethanol-Inducible Promoter

A promoter inducible by certain alcohols or ketones, such as ethanol, can also be used to confer inducible expression of a coding sequence of the presently disclosed
25 subject matter. Such a promoter is for example the *alcA* gene promoter from *Aspergillus nidulans* (Caddick *et al.*, 1998). In *A. nidulans*, the *alcA* gene encodes alcohol dehydrogenase I, the expression of which is regulated by the AlcR transcription factors in presence of the chemical inducer. For the purposes of the presently disclosed subject matter, the CAT coding sequences in plasmid palcA:CAT comprising a *alcA*
30 gene promoter sequence fused to a minimal 35S promoter (Caddick *et al.*, 1998) are replaced by a coding sequence of the presently disclosed subject matter to form an

expression cassette having the coding sequence under the control of the *alcA* gene promoter. This is carried out using methods known in the art.

IV.C.1.f. Inducible Expression: a Glucocorticoid-Inducible Promoter

Induction of expression of a nucleic acid sequence of the presently disclosed
5 subject matter using systems based on steroid hormones is also provided. For
example, a glucocorticoid-mediated induction system is used (Aoyama & Chua, 1997)
and gene expression is induced by application of a glucocorticoid, for example a
synthetic glucocorticoid, for example dexamethasone, at a concentration ranging in one
embodiment from 0.1 mM to 1 mM, and in another embodiment from 10 mM to 100
10 mM. For the purposes of the presently disclosed subject matter, the luciferase gene
sequences Aoyama & Chua are replaced by a nucleic acid sequence of the presently
disclosed subject matter to form an expression cassette having a nucleic acid sequence
of the presently disclosed subject matter under the control of six copies of the GAL4
upstream activating sequences fused to the 35S minimal promoter. This is carried out
15 using methods known in the art. The trans-acting factor comprises the GAL4 DNA-
binding domain (Keegan *et al.*, 1986) fused to the transactivating domain of the herpes
viral polypeptide VP16 (Triezenberg *et al.*, 1988) fused to the hormone-binding domain
of the rat glucocorticoid receptor (Picard *et al.*, 1988). The expression of the fusion
polypeptide is controlled either by a promoter known in the art or disclosed herein. A
20 plant comprising an expression cassette comprising a nucleic acid sequence of the
presently disclosed subject matter fused to the 6x GAL4/minimal promoter is also
provided. Thus, tissue-specificity of the fusion polypeptide is achieved leading to
inducible tissue-specificity of the nucleic acid sequence to be expressed.

IV.C.1.g. Root Specific Expression

25 Another pattern of gene expression is root expression. A suitable root promoter
is the promoter of the maize metallothionein-like (MTL) gene disclosed in de Framond,
1991, and also in U.S. Patent No. 5,466,785, each of which is incorporated herein by
reference. This "MTL" promoter is transferred to a suitable vector such as
pCGN1761ENX for the insertion of a selected gene and subsequent transfer of the
30 entire promoter-gene-terminator cassette to a transformation vector of interest.

IV.C.1.h. Wound-Inducible Promoters

Wound-inducible promoters can also be suitable for gene expression. Numerous

such promoters have been disclosed (e.g. Xu *et al.*, 1993; Logemann *et al.*, 1989; Rohrmeier & Lehle, 1993; Firek *et al.*, 1993; Warner *et al.*, 1993) and all are suitable for use with the presently disclosed subject matter. Logemann *et al.* describe the 5' upstream sequences of the dicotyledonous potato *wun1* gene. Xu *et al.* show that a wound-inducible promoter from the dicotyledon potato (*pin2*) is active in the monocotyledon rice. Further, Rohrmeier & Lehle describe the cloning of the maize *Wip1* cDNA that is wound induced and which can be used to isolate the cognate promoter using standard techniques. Similarly, Firek *et al.* and Warner *et al.* have disclosed a wound-induced gene from the monocotyledon *Asparagus officinalis*, which is expressed at local wound and pathogen invasion sites. Using cloning techniques well known in the art, these promoters can be transferred to suitable vectors, fused to the genes pertaining to the presently disclosed subject matter, and used to express these genes at the sites of plant wounding.

IV.C.1.i. Pith-Preferred Expression

PCT International Publication WO 93/07278, which is herein incorporated by reference, describes the isolation of the maize *trpA* gene, which is preferentially expressed in pith cells. The gene sequence and promoter extending up to -1726 basepairs (bp) from the start of transcription are presented. Using standard molecular biological techniques, this promoter, or parts thereof, can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a foreign gene in a pith-preferred manner. In fact, fragments containing the pith-preferred promoter or parts thereof can be transferred to any vector and modified for utility in transgenic plants.

IV.C.1.j. Leaf-Specific Expression

A maize gene encoding phosphoenol carboxylase (PEPC) has been disclosed by Hudspeth & Grula, 1989. Using standard molecular biological techniques, the promoter for this gene can be used to drive the expression of any gene in a leaf-specific manner in transgenic plants.

IV.C.1.k. Pollen-Specific Expression

WO 93/07278 describes the isolation of the maize calcium-dependent protein kinase (CDPK) gene that is expressed in pollen cells. The gene sequence and promoter extend up to 1400 bp from the start of transcription. Using standard molecular

biological techniques, this promoter or parts thereof can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a nucleic acid sequence of the presently disclosed subject matter in a pollen-specific manner.

5 IV.C.1.I. Seed-Specific Expression

In some embodiments of the presently disclosed subject matter, a lack of expression or reduced expression in seeds can be desired. This can optionally be accomplished by knocking expression in the plant and then re-establishing expression in one or more other tissues (including all tissues) except seeds.

10 Example 5 below describes an ipk1/ipk2 beta double knockout plant. The double mutant grows about the same as the single mutant disclosed in the Examples. Thus, the seed specific knockdown of all IPs is not detrimental to seed yield and plant growth. The term "seed specific" is used because analysis of the plant tissue from the atlpk2 beta mutant shows no defect in IP metabolism whereas the seeds do. Further the ipk1
15 ipk2 beta double mutant appears to synthesize IPs except for IP5 in the tissue.

The growth issue and yield problems of ipk1 mutant and ipk1/ipk2 beta double mutant plants are expected to be corrected with the tissue (but not seed) expression of ipk1. Complementation of the ipk1 mutant plant with pBART lpk1 complements the growth. The ipk1/ipk2 beta double mutant can be complemented with lpk1 under one
20 or more tissue specific promoters as disclosed herein above, e.g. one or more promoters that can drive expression in one or more tissues except seeds.

IV.C.2. Transcriptional Terminators

A variety of transcriptional terminators are available for use in expression cassettes. These are responsible for termination of transcription and correct mRNA
25 polyadenylation. Appropriate transcriptional terminators are those that are known to function in plants and include the CaMV 35S terminator, the *tm1* terminator, the nopaline synthase terminator, and the pea *rbcS* E9 terminator. These can be used in both monocotyledons and dicotyledons. In addition, a gene's native transcription terminator can be used.

30 IV.C.3. Sequences for the Enhancement or Regulation of Expression

Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes

of the presently disclosed subject matter to increase their expression in transgenic plants.

Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize *Adhl* gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, 1987). In the same experimental system, the intron from the maize *bronze1* gene had a similar effect in enhancing expression. Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV; the "W-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (see *e.g.* Gallie *et al.*, 1987; Skuzeski *et al.*, 1990). Other leader sequences known in the art include, but are not limited to, picornavirus leaders, for example, EMCV (encephalomyocarditis virus) leader (5' noncoding region; see Elroy-Stein *et al.*, 1989); potyvirus leaders, for example, from Tobacco Etch Virus (TEV; see Allison *et al.*, 1986); Maize Dwarf Mosaic Virus (MDMV; see Kong & Steinbiss 1998); human immunoglobulin heavy-chain binding polypeptide (BiP) leader (Macejak & Sarnow, 1991); untranslated leader from the coat polypeptide mRNA of alfalfa mosaic virus (AMV; RNA 4; see Jobling & Gehrke, 1987); tobacco mosaic virus (TMV) leader (Gallie *et al.*, 1989); and Maize Chlorotic Mottle Virus (MCMV) leader (Lommel *et al.*, 1991). See also, Della-Cioppa *et al.*, 1987.

In addition to incorporating one or more of the aforementioned elements into the 5' regulatory region of a target expression cassette of the presently disclosed subject matter, other elements can also be incorporated. Such elements include, but are not limited to, a minimal promoter. By minimal promoter it is intended that the basal promoter elements are inactive or nearly so in the absence of upstream or downstream activation. Such a promoter has low background activity in plants when there is no transactivator present or when enhancer or response element binding sites are absent.

One minimal promoter that is particularly useful for target genes in plants is the Bz1 minimal promoter, which is obtained from the *bronze1* gene of maize. The Bz1 core promoter is obtained from the "myc" mutant Bz1-luciferase construct pBz1LucR98 via cleavage at the *NheI* site located at positions -53 to -58 (Roth *et al.*, 1991). The derived Bz1 core promoter fragment thus extends from positions -53 to +227 and includes the Bz1 intron-1 in the 5' untranslated region. Also useful for the presently disclosed subject matter is a minimal promoter created by use of a synthetic TATA element. The TATA element allows recognition of the promoter by RNA polymerase factors and confers a basal level of gene expression in the absence of activation (*see generally*, Mukumoto *et al.*, 1993; Green, 2000).

IV.C.4. Targeting of the Gene Product Within the Cell

Various mechanisms for targeting gene products are known to exist in plants and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various polypeptides that is cleaved during chloroplast import to yield the mature polypeptides (*see e.g.*, Comai *et al.*, 1988). These signal sequences can be fused to heterologous gene products to affect the import of heterologous products into the chloroplast (Van den Broeck *et al.*, 1985). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO) polypeptide, the chlorophyll a/b binding (CAB) polypeptide, the 5-enol-pyruvyl shikimate-3-phosphate (EPSP) synthase enzyme, the GS2 polypeptide and many other polypeptides which are known to be chloroplast localized. *See also*, the section entitled "Expression With Chloroplast Targeting" in Example 37 of U.S. Patent No. 5,639,949, herein incorporated by reference.

Other gene products can be localized to other organelles such as the mitochondrion and the peroxisome (*e.g.* Unger *et al.*, 1989). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous gene products to these organelles. Examples of such sequences are the nuclear-encoded ATPases and specific aspartate amino transferase isoforms for mitochondria. Targeting cellular polypeptide bodies has been disclosed by Rogers *et al.*, 1985.

In addition, sequences have been characterized that control the targeting of gene products to other cell compartments. Amino terminal sequences are responsible for targeting to the endoplasmic reticulum (ER), the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, 1990). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al.*, 1990).

By the fusion of the appropriate targeting sequences disclosed above to transgene sequences of interest it is possible to direct the transgene product to any organelle or cell compartment. For chloroplast targeting, for example, the chloroplast signal sequence from the RUBISCO gene, the CAB gene, the EPSP synthase gene, or the GS2 gene is fused in frame to the amino terminal ATG of the transgene. The signal sequence selected can include the known cleavage site, and the fusion constructed can take into account any amino acids after the cleavage site that are required for cleavage.

In some cases this requirement can be fulfilled by the addition of a small number of amino acids between the cleavage site and the transgene ATG or, alternatively, replacement of some amino acids within the transgene sequence. Fusions constructed for chloroplast import can be tested for efficacy of chloroplast uptake by *in vitro* translation of *in vitro* transcribed constructions followed by *in vitro* chloroplast uptake using techniques disclosed by Bartlett *et al.*, 1982 and Wasmann *et al.*, 1986. These construction techniques are well known in the art and are equally applicable to mitochondria and peroxisomes.

The above-disclosed mechanisms for cellular targeting can be utilized not only in conjunction with their cognate promoters, but also in conjunction with heterologous promoters so as to effect a specific cell-targeting goal under the transcriptional regulation of a promoter that has an expression pattern different from that of the promoter from which the targeting signal derives.

IV.D. Construction of Plant Transformation Vectors

Numerous transformation vectors available for plant transformation are known to those of ordinary skill in the plant transformation art, and the genes pertinent to the presently disclosed subject matter can be used in conjunction with any such vectors. The selection of vector will depend upon the selected transformation technique and the target species for transformation. For certain target species, different antibiotic or

herbicide selection markers might be employed. Selection markers used routinely in transformation include the *nptII* gene, which confers resistance to kanamycin and related antibiotics (Messing & Vieira, 1982; Bevan *et al.*, 1983); the *bar* gene, which confers resistance to the herbicide phosphinothricin (White *et al.*, 1990; Spencer *et al.*, 1990); the *hph* gene, which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, 1984); the *dhfr* gene, which confers resistance to methotrexate (Bourouis & Jarry, 1983); the EPSP synthase gene, which confers resistance to glyphosate (U.S. Patent Nos. 4,940,935 and 5,188,642); and the mannose-6-phosphate isomerase gene, which provides the ability to metabolize mannose (U.S. Patent Nos. 5,767,378 and 5,994,629).

IV.D.1. Vectors Suitable for *Agrobacterium* Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, 1984). Below, the construction of two typical vectors suitable for *Agrobacterium* transformation is disclosed.

IV.D.1.a. pCIB200 and pCIB2001

The binary vectors pCIB200 and pCIB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and are constructed in the following manner. pTJS75kan is created by *NarI* digestion of pTJS75 (Schmidhauser & Helinski, 1985) allowing excision of the tetracycline-resistance gene, followed by insertion of an *AccI* fragment from pUC4K carrying an NPTII sequence (Messing & Vieira, 1982; Bevan *et al.*, 1983; McBride & Summerfelt, 1990). *XhoI* linkers are ligated to the *EcoRV* fragment of PCIB7 which contains the left and right T-DNA borders, a plant selectable *nos/nptII* chimeric gene and the pUC polylinker (Rothstein *et al.*, 1987), and the *XhoI*-digested fragment are cloned into *Sall*-digested pTJS75kan to create pCIB200 (see also EP 0 332 104, example 19). pCIB200 contains the following unique polylinker restriction sites: *EcoRI*, *SstI*, *KpnI*, *BglII*, *XbaI*, and *Sall*. pCIB2001 is a derivative of pCIB200 created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001 are *EcoRI*, *SstI*, *KpnI*, *BglII*, *XbaI*, *Sall*, *MluI*, *BclI*, *AvrII*, *Apal*, *HpaI*, and *StuI*. pCIB2001, in addition to containing these unique restriction sites, also has plant and bacterial kanamycin selection, left and right

T-DNA borders for *Agrobacterium*-mediated transformation, the RK2-derived *trfA* function for mobilization between *E. coli* and other hosts, and the *OriT* and *OriV* functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

5 IV.D.1.b. pCIB10 and Hygromycin Selection Derivatives Thereof

The binary vector pCIB10 contains a gene encoding kanamycin resistance for selection in plants, T-DNA right and left border sequences, and incorporates sequences from the wide host-range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is disclosed by Rothstein *et al.*, 1987. Various
10 derivatives of pCIB10 can be constructed which incorporate the gene for hygromycin B phosphotransferase disclosed by Gritz & Davies, 1983. These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

IV.D.2. Vectors Suitable for non-*Agrobacterium* Transformation

15 Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector, and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones disclosed above that contain T-DNA sequences. Transformation techniques that do not rely on *Agrobacterium* include transformation via particle
20 bombardment, protoplast uptake (e.g. polyethylene glycol (PEG) and electroporation), and microinjection. The choice of vector depends largely on the species being transformed. Below, the construction of typical vectors suitable for non-*Agrobacterium* transformation is disclosed.

IV.D.2.a. pCIB3064

25 pCIB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide BASTA[®] (glufosinate ammonium or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter in operational fusion to the *E. coli* β -glucuronidase (GUS) gene and the CaMV 35S transcriptional terminator and is disclosed in the PCT International Publication
30 WO 93/07278. The 35S promoter of this vector contains two ATG sequences 5' of the start site. These sites are mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites *SspI* and *PvuII*. The new restriction

sites are 96 and 37 bp away from the unique *Sall* site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 is designated pCIB3025. The GUS gene is then excised from pCIB3025 by digestion with *Sall* and *SacI*, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 is
5 obtained from the John Innes Centre, Norwich, England, and the 400 bp *SmaI* fragment containing the *bar* gene from *Streptomyces viridochromogenes* is excised and inserted into the *HpaI* site of pCIB3060 (Thompson *et al.*, 1987). This generated pCIB3064, which comprises the *bar* gene under the control of the CaMV 35S promoter and terminator for herbicide selection, a gene for ampicillin resistance (for selection in *E.*
10 *coli*) and a polylinker with the unique sites *SphI*, *PstI*, *HindIII*, and *BamHI*. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

IV.D.2.b. pSOG19 and pSOG35

pSOG35 is a transformation vector that utilizes the *E. coli* dihydrofolate
15 reductase (DHFR) gene as a selectable marker conferring resistance to methotrexate. PCR is used to amplify the 35S promoter (-800 bp), intron 6 from the maize Adh1 gene (-550 bp), and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250-bp fragment encoding the *E. coli* dihydrofolate reductase type II gene is also amplified by PCR and these two PCR fragments are assembled with a *SacI*-*PstI* fragment from
20 pB1221 (BD Biosciences Clontech, Palo Alto, California, United States of America) that comprises the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generates pSOG19 that contains the 35S promoter in fusion with the intron 6 sequence, the GUS leader, the DHFR gene, and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader
25 sequence from Maize Chlorotic Mottle Virus (MCMV) generates the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *HindIII*, *SphI*, *PstI*, and *EcoRI* sites available for the cloning of foreign substances.

IV.D.3. Vector Suitable for Chloroplast Transformation

For expression of a nucleotide sequence of the presently disclosed subject
30 matter in plant plastids, plastid transformation vector pPH143 (PCT International Publication WO 97/32011, example 36) can be used. The nucleotide sequence is inserted into pPH143 thereby replacing the protoporphyrinogen oxidase (Protox) coding

sequence. This vector can then be used for plastid transformation and selection of transformants for spectinomycin resistance. Alternatively, the nucleotide sequence can be inserted in pPH143 so that it replaces the *aadH* gene. In this case, transformants are selected for resistance to PROTOX inhibitors.

5 IV.E. Transformation

Once a nucleic acid sequence of the presently disclosed subject matter has been cloned into an expression system, it is transformed into a plant cell. The receptor and target expression cassettes of the presently disclosed subject matter can be introduced into the plant cell in a number of art-recognized ways. Methods for regeneration of plants are also well known in the art. For example, Ti plasmid vectors have been utilized for the delivery of foreign DNA, as well as direct DNA uptake, liposomes, electroporation, microinjection, and microprojectiles. In addition, bacteria from the genus *Agrobacterium* can be utilized to transform plant cells. Below are descriptions of representative techniques for transforming both dicotyledonous and monocotyledonous plants, as well as a representative plastid transformation technique.

15 IV.E.1. Transformation of Dicotyledons

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques that do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation-mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are disclosed in Paszkowski *et al.*, 1984; Potrykus *et al.*, 1985; Reich *et al.*, 1986; and Klein *et al.*, 1987. In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

25 *Agrobacterium*-mediated transformation is a useful technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. *Agrobacterium* transformation typically involves the transfer of the binary vector carrying the foreign DNA of interest (e.g. pCIB200 or pCIB2001) to an appropriate *Agrobacterium* strain which can depend on the complement of *vir* genes carried by the host *Agrobacterium* strain either on a co-resident Ti plasmid or chromosomally (e.g. strain CIB542 for pCIB200 and pCIB2001 (Uknes *et al.*, 1993). The transfer of the recombinant binary vector to *Agrobacterium* is accomplished by a

triparental mating procedure using *E. coli* carrying the recombinant binary vector, a helper *E. coli* strain that carries a plasmid such as pRK2013 and which is able to mobilize the recombinant binary vector to the target *Agrobacterium* strain. Alternatively, the recombinant binary vector can be transferred to *Agrobacterium* by DNA transformation (Höfgen & Willmitzer, 1988).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

Another approach to transforming plant cells with a gene involves propelling inert or biologically active particles at plant tissues and cells. This technique is disclosed in U.S. Patent Nos. 4,945,050; 5,036,006; and 5,100,792; all to Sanford et al. Generally, this procedure involves propelling inert or biologically active particles at the cells under conditions effective to penetrate the outer surface of the cell and afford incorporation within the interior thereof. When inert particles are utilized, the vector can be introduced into the cell by coating the particles with the vector containing the desired gene. Alternatively, the target cell can be surrounded by the vector so that the vector is carried into the cell by the wake of the particle. Biologically active particles (e.g., dried yeast cells, dried bacterium, or a bacteriophage, each containing DNA sought to be introduced) can also be propelled into plant cell tissue.

IV.E.2. Transformation of Monocotyledons

Transformation of most monocotyledon species has now also become routine. Exemplary techniques include direct gene transfer into protoplasts using PEG or electroporation, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (*i.e.* co-transformation), and both these techniques are suitable for use with the presently disclosed subject matter. Co-transformation can have the advantage of avoiding complete vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the selectable marker in subsequent generations, should this be regarded as desirable. However, a disadvantage of the use of co-transformation is the less than 100% frequency with

which separate DNA species are integrated into the genome (Schocher *et al.*, 1986).

Patent Applications EP 0 292 435, EP 0 392 225, and WO 93/07278 describe techniques for the preparation of callus and protoplasts from an elite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration
5 of maize plants from transformed protoplasts. Gordon-Kamm *et al.*, 1990 and Fromm *et al.*, 1990 have published techniques for transformation of A188-derived maize line using particle bombardment. Furthermore, WO 93/07278 and Koziel *et al.*, 1993 describe techniques for the transformation of elite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length
10 excised from a maize ear 14-15 days after pollination and a PDS-1000He Biolistic particle delivery device (DuPont Biotechnology, Wilmington, Delaware, United States of America) for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated transformation has
15 been disclosed for *Japonica*-types and *Indica*-types (Zhang *et al.*, 1988; Shimamoto *et al.*, 1989; Datta *et al.*, 1990) of rice. Both types are also routinely transformable using particle bombardment (Christou *et al.*, 1991). Furthermore, WO 93/21335 describes techniques for the transformation of rice via electroporation. Casas *et al.*, 1993 discloses the production of transgenic sorghum plants by microprojectile bombardment.

Patent Application EP 0 332 581 describes techniques for the generation,
20 transformation, and regeneration of *Pooideae* protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat transformation has been disclosed in Vasil *et al.*, 1992 using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil *et al.*, 1993 and Weeks *et al.*, 1993 using particle
25 bombardment of immature embryos and immature embryo-derived callus.

A representative technique for wheat transformation, however, involves the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, embryos (0.75-1 mm in length) are plated onto MS medium with 3%
30 sucrose (Murashige & Skoog, 1962) and 3 mg/l 2,4-dichlorophenoxyacetic acid (2,4-D) for induction of somatic embryos, which is allowed to proceed in the dark. On the chosen day of bombardment, embryos are removed from the induction medium and

placed onto the osmoticum (*i.e.* induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 hours and are then bombarded. Twenty embryos per target plate are typical, although not critical. An appropriate gene-carrying plasmid (such as pCIB3064 or pSG35) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with the DuPont BIOLISTICS® helium device using a burst pressure of about 1000 pounds per square inch (psi) using a standard 80 mesh screen. After bombardment, the embryos are placed back into the dark to recover for about 24 hours (still on osmoticum). After 24 hours, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l BASTA® in the case of pCIB3064 and 2 mg/l methotrexate in the case of pSOG35). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contain half-strength MS, 2% sucrose, and the same concentration of selection agent.

Transformation of monocotyledons using *Agrobacterium* has also been disclosed. See WO 94/00977 and U.S. Patent No. 5,591,616, both of which are incorporated herein by reference. See also Negrotto *et al.*, 2000, incorporated herein by reference. Zhao *et al.*, 2000 specifically discloses transformation of sorghum with *Agrobacterium*. See also U.S. Patent No. 6,369,298.

Rice (*Oryza sativa*) can be used for generating transgenic plants. Various rice cultivars can be used (Hiei *et al.*, 1994; Dong *et al.*, 1996; Hiei *et al.*, 1997). Also, the various media constituents disclosed below can be either varied in quantity or substituted. Embryogenic responses are initiated and/or cultures are established from mature embryos by culturing on MS-CIM medium (MS basal salts, 4.3 g/liter; B5 vitamins (200 x), 5 ml/liter; Sucrose, 30 g/liter; proline, 500 mg/liter; glutamine, 500 mg/liter; casein hydrolysate, 300 mg/liter; 2,4-D (1 mg/ml), 2 ml/liter; pH adjusted to 5.8 with 1 N KOH; Phytigel, 3 g/liter). Either mature embryos at the initial stages of culture response or established culture lines are inoculated and co-cultivated with the *Agrobacterium tumefaciens* strain LBA4404 (*Agrobacterium*) containing the desired

vector construction. *Agrobacterium* is cultured from glycerol stocks on solid YPC medium (plus 100 mg/L spectinomycin and any other appropriate antibiotic) for about 2 days at 28°C. *Agrobacterium* is re-suspended in liquid MS-CIM medium. The *Agrobacterium* culture is diluted to an OD₆₀₀ of 0.2-0.3 and acetosyringone is added to a
5 final concentration of 200 µM. Acetosyringone is added before mixing the solution with the rice cultures to induce *Agrobacterium* for DNA transfer to the plant cells. For inoculation, the plant cultures are immersed in the bacterial suspension. The liquid bacterial suspension is removed and the inoculated cultures are placed on co-cultivation medium and incubated at 22°C for two days. The cultures are then
10 transferred to MS-CIM medium with ticarcillin (400 mg/liter) to inhibit the growth of *Agrobacterium*. For constructs utilizing the PMI selectable marker gene (Reed *et al.*, 2001), cultures are transferred to selection medium containing mannose as a carbohydrate source (MS with 2% mannose, 300 mg/liter ticarcillin) after 7 days, and cultured for 3-4 weeks in the dark. Resistant colonies are then transferred to
15 regeneration induction medium (MS with no 2,4-D, 0.5 mg/liter IAA, 1 mg/liter zeatin, 200 mg/liter TIMENTIN®, 2% mannose, and 3% sorbitol) and grown in the dark for 14 days. Proliferating colonies are then transferred to another round of regeneration induction media and moved to the light growth room. Regenerated shoots are transferred to GA7 containers with GA7-1 medium (MS with no hormones and 2%
20 sorbitol) for 2 weeks and then moved to the greenhouse when they are large enough and have adequate roots. Plants are transplanted to soil in the greenhouse (T₀ generation) grown to maturity and the T₁ seed is harvested.

IV.E.3. Transformation of Plastids

25 As one example of plastid transformation, seeds of *Nicotiana tabacum* c.v. 'Xanthi nc' are germinated seven per plate in a 1" circular array on T agar medium and bombarded 12-14 days after sowing with 1 µm tungsten particles (M10, Biorad, Hercules, California, United States of America) coated with DNA from plasmids pPH143 and pPH145 essentially as disclosed (Svab & Maliga, 1993). Bombarded seedlings are
30 incubated on T medium for two days after which leaves are excised and placed abaxial side up in bright light (350-500 µmol photons/m²/s) on plates of RMOP medium (Svab *et al.*, 1990) containing 500 µg/ml spectinomycin dihydrochloride (Sigma, St. Louis,

Missouri, United States of America). Resistant shoots appearing underneath the bleached leaves three to eight weeks after bombardment are subcloned onto the same selective medium, allowed to form callus, and secondary shoots isolated and subcloned. Complete segregation of transformed plastid genome copies
5 (homoplasmy) in independent subclones is assessed by standard techniques of Southern blotting (Sambrook & Russell, 2001). *Bam*HI/*Eco*RI-digested total cellular DNA (Mettler, 1987) is separated on 1% Tris-borate-EDTA (TBE) agarose gels, transferred to nylon membranes (Amersham Biosciences, Piscataway, New Jersey, United States of America) and probed with ³²P-labeled random primed DNA sequences
10 corresponding to a 0.7 kb *Bam*HI/*Hind*III DNA fragment from pC8 containing a portion of the *rps7/12* plastid targeting sequence. Homoplasmic shoots are rooted aseptically on spectinomycin-containing MS/IBA medium (McBride *et al.*, 1994) and transferred to the greenhouse.

15 V. Plants, Breeding, and Seed Production

V.A. Plants

The presently disclosed subject matter also provides plant cells and plants comprising the disclosed compositions. The presently disclosed subject matter further provides plant cells and plants comprising homozygous disruptions in endogenous
20 nucleic acids homologous to the novel nucleic acid molecules disclosed herein.

In one embodiment, the plant is characterized by a modification of a phenotype or measurable characteristic of the plant, the modification being attributable to the expression cassette comprising a novel nucleic acid disclosed herein.

In one embodiment, the modification involves modulation of the expression of a
25 functional inositol phosphate kinase polypeptide. In some embodiments, the functional inositol phosphate kinase polypeptide is *lpk1* or *2*. In some embodiments, the modification results in a change in the level of phytate and/or non-phytate phosphorous when compared to a non-transformed parental plant. In some embodiments, the modulation of functional inositol phosphate kinase polypeptide is an inhibition of the
30 expression of the polypeptide resulting in a subsequent decreased level of phytate and/or a subsequent increase in non-phytate phosphorous. In some embodiments, the

phytate levels are decreased by at least 90% when compared to a non-transformed parental plant.

In another embodiment, the modification includes overexpression, underexpression, antisense modulation, sense suppression, inducible expression, inducible repression, or inducible modulation of a gene.

V.B. Breeding

The plants obtained via transformation with a nucleic acid sequence of the presently disclosed subject matter can be any of a wide variety of plant species, including monocots and dicots; however, the plants used in the method for the presently disclosed subject matter are selected in one embodiment from the list of agronomically important target crops set forth hereinabove. The expression of a gene of the presently disclosed subject matter in combination with other characteristics important for production and quality can be incorporated into plant lines through breeding. Breeding approaches and techniques are known in the art. See e.g., Welsh, 1981; Wood, 1983; Mayo, 1987; Singh, 1986; Wricke & Weber, 1986.

The genetic properties engineered into the transgenic seeds and plants disclosed above are passed on by sexual reproduction or vegetative growth and can thus be maintained and propagated in progeny plants. Generally, the maintenance and propagation make use of known agricultural methods developed to fit specific purposes such as tilling, sowing, or harvesting. Specialized processes such as hydroponics or greenhouse technologies can also be applied. As the growing crop is vulnerable to attack and damage caused by insects or infections as well as to competition by weed plants, measures are undertaken to control weeds, plant diseases, insects, nematodes, and other adverse conditions to improve yield. These include mechanical measures such as tillage of the soil or removal of weeds and infected plants, as well as the application of agrochemicals such as herbicides, fungicides, gametocides, nematocides, growth regulants, ripening agents, and insecticides.

Use of the advantageous genetic properties of the transgenic plants and seeds according to the presently disclosed subject matter can further be made in plant breeding, which aims at the development of plants with improved properties such as improved nutritional value, tolerance of pests, herbicides, or abiotic stress, increased yield, or improved structure causing less loss from lodging or shattering. The various

breeding steps are characterized by well-defined human intervention such as selecting the lines to be crossed, directing pollination of the parental lines, or selecting appropriate progeny plants.

Depending on the desired properties, different breeding measures are taken.

5 The relevant techniques are well known in the art and include, but are not limited to, hybridization, inbreeding, backcross breeding, multiline breeding, variety blend, interspecific hybridization, aneuploid techniques, etc. Hybridization techniques can also include the sterilization of plants to yield male or female sterile plants by mechanical, chemical, or biochemical means. Cross-pollination of a male sterile plant with pollen of
10 a different line assures that the genome of the male sterile but female fertile plant will uniformly obtain properties of both parental lines. Thus, the transgenic seeds and plants according to the presently disclosed subject matter can be used for the breeding of improved plant lines that, for example, increase the effectiveness of conventional methods such as herbicide or pesticide treatment or allow one to dispense with said
15 methods due to their modified genetic properties. Alternatively new crops with improved stress tolerance can be obtained, which, due to their optimized genetic "equipment", yield harvested product of better quality than products that were not able to tolerate comparable adverse developmental conditions (for example, drought).

V.C. Seed Production

20 Some embodiments of the presently disclosed subject matter also provide seed and isolated product from plants that comprise an expression cassette comprising a promoter sequence operatively linked to an isolated nucleic acid as disclosed herein above, the isolated nucleic acid encoding a polypeptide as disclosed herein above. In some embodiments the nucleic acid comprises one of: (a) a nucleic acid molecule
25 encoding a polypeptide comprising an amino acid sequence of one of even numbered SEQ ID NOs: 2-10 or a polypeptide at least 40% identical to even numbered SEQ ID NOs: 2-10 and having inositol phosphate kinase activity; (b) a nucleic acid molecule comprising a nucleic acid sequence of one of odd numbered SEQ ID NOs: 1-9; (c) a nucleic acid molecule that has a nucleic acid sequence that is substantially identical to
30 the nucleic acid sequence of the nucleic acid molecule of (a) or (b); (d) a nucleic acid molecule that hybridizes to (a) or (b) under stringent hybridization conditions; (e) a nucleic acid molecule comprising a nucleic acid sequence complementary to (a); and (f)

a nucleic acid molecule comprising a nucleic acid sequence that is the full reverse complement of (a).

Embodiments of the presently disclosed subject matter also relate to seed and isolated products produced by expression of an isolated nucleic acid comprising one of:

- 5 (a) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of one of even numbered SEQ ID NOs: 2-10 or a polypeptide at least 40% identical to even numbered SEQ ID NOs: 2-10 and having inositol phosphate kinase activity; (b) a nucleic acid molecule comprising a nucleic acid sequence of one of odd numbered SEQ ID NOs: 1-9; (c) a nucleic acid molecule that has a nucleic acid sequence that is
10 substantially identical to the nucleic acid sequence of the nucleic acid molecule of (a) or (b); (d) a nucleic acid molecule that hybridizes to (a) or (b) under stringent hybridization conditions; (e) a nucleic acid molecule comprising a nucleic acid sequence complementary to (a); and a nucleic acid molecule comprising a nucleic acid sequence that is the full reverse complement of (a).

- 15 Embodiments further include seed and isolated products resulting from a homozygous disruption of an endogenous gene homologous to a nucleic acid molecule described herein.

- In one embodiment, the product is produced in a plant. In another embodiment, the product is produced in cell culture. In another embodiment, the product is produced
20 in a cell-free system. In one embodiment, the product comprises an inositol phosphate kinase polypeptide. In another embodiment, the polypeptide is Ipk1 or 2 and homologs and orthologs thereof. In another embodiment, the product is a polypeptide comprising an amino acid sequence listed in even numbered sequences of SEQ ID NOs: 2-10, or orthologs thereof, or a polypeptide encoded by a nucleic acid sequence that hybridizes
25 under stringent conditions of hybridization of 45°C in 1 M NaCl, followed by a final washing step at 50°C in 0.1 M NaCl to a nucleotide sequence listed in odd numbered sequences of SEQ ID NOs: 1-9, or a subsequence thereof. Representative orthologs are disclosed in Examples 6-11 herein below. In some embodiments, a polypeptide of the presently disclosed subject matter has or comprises substantial identity (in some
30 embodiments at least about 90% amino acid identity) with a polypeptide as disclosed in Examples 6-11, over an entire sequence or functional fragment thereof. In one embodiment, the polypeptide functions as an inositol phosphate kinase. More

specifically, in some embodiments, the polypeptide is an lpk2 polypeptide and functions as a dual-specificity IP_3/IP_4 6-/3-kinase generating IP_5 . More specifically, in some other embodiments, the polypeptide is an lpk1 polypeptide and functions as an $I(1,3,4,5,6)P_5$ 2-kinase generating phytate (IP_6).

5 In seed production, germination quality and uniformity of seeds are essential product characteristics. As it is difficult to keep a crop free from other crop and weed seeds, to control seedborne diseases, and to produce seed with good germination, fairly extensive and well-defined seed production practices have been developed by seed producers who are experienced in the art of growing, conditioning, and marketing
10 of pure seed. Thus, it is common practice for the farmer to buy certified seed meeting specific quality standards instead of using seed harvested from his own crop. Propagation material to be used as seeds is customarily treated with a protectant coating comprising herbicides, insecticides, fungicides, bactericides, nematocides, molluscicides, or mixtures thereof. Customarily used protectant coatings comprise
15 compounds such as captan, carboxin, thiram (tetramethylthiuram disulfide; TMTD®; available from R. T. Vanderbilt Company, Inc., Norwalk, Connecticut, United States of America), methalaxyl (APRON XL®; available from Syngenta Corp., Wilmington, Delaware, United States of America), and pirimiphos-methyl (ACTELLIC®; available from Agrilience, LLC, St. Paul, Minnesota, United States of America). If desired, these
20 compounds are formulated together with further carriers, surfactants, and/or application-promoting adjuvants customarily employed in the art of formulation to provide protection against damage caused by bacterial, fungal, or animal pests. The protectant coatings can be applied by impregnating propagation material with a liquid formulation or by coating with a combined wet or dry formulation. Other methods of application are also
25 possible such as treatment directed at the buds or the fruit.

VI. Other Applications

The presently disclosed subject matter also provides a method for producing a polypeptide as disclosed herein, the method comprising (a) growing cells comprising an
30 expression cassette under suitable growth conditions, the expression cassette comprising a nucleic acid molecule encoding the polypeptide; and (b) isolating the polypeptide from the cells.

The presently disclosed subject matter also provides a method for modulating production of phytate in a plant, the method comprising modulating the enzymatic activity of at least one inositol phosphate kinase polypeptide, wherein the polypeptide is lpk1 and/or lpk2. In one embodiment, the method reduces the enzymatic activity. In another embodiment, the enzymatic activity is reduced by inhibiting expression of a functional form of at least one inositol phosphate kinase polypeptide, using techniques as disclosed herein. In some embodiments, inhibiting expression of the functional polypeptide results in a decrease in the phytate content of the plant and/or an increase in the non-phytate phosphorous content, in some embodiments by at least 90%. In one embodiment, the polypeptide expression is modulated in a predetermined location or tissue of a plant. In one embodiment, the location or tissue is selected from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, seed, and combinations thereof. In one embodiment, the tissue is seed.

The presently disclosed subject matter also provides a method of producing a plant with low phytate levels. The method comprises modulating in the plant the enzymatic activity of at least one inositol phosphate kinase polypeptide, wherein the polypeptide is lpk1 and/or 2. In one embodiment, the method reduces the enzymatic activity. In another embodiment, the enzymatic activity is reduced by inhibiting expression of a functional form of at least one inositol phosphate kinase polypeptide, using techniques as disclosed herein. In some embodiments, inhibiting expression of the functional polypeptide results in a significant decrease in the phytate content of the plant and/or a significant increase in the non-phytate phosphorous content, in some embodiments by at least 90%. In one embodiment, the polypeptide expression is modulated in a predetermined location or tissue of a plant. In one embodiment, the location or tissue is selected from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, seed, and combinations thereof. In one embodiment, the tissue is seed.

The presently disclosed subject matter also provides a method for decreasing the expression of an isolated nucleic acid molecule as disclosed herein in a plant, the method selected from the group consisting of (a) expressing in said plant a molecule of the presently disclosed subject matter or a portion thereof in "sense" orientation; (b) expressing in said plant a molecule of the presently disclosed subject matter or a

portion thereof in "antisense" orientation; (c) expressing in said plant a ribozyme capable of specifically cleaving a messenger RNA transcript encoded by an endogenous gene corresponding to an isolated nucleic acid molecule of the presently disclosed subject matter; (d) expressing in a plant an aptamer specifically directed to a polypeptide encoded by an isolated nucleic acid molecule of the presently disclosed subject matter; (e) expressing in a plant a mutated or a truncated form of an isolated nucleic acid molecule of the presently disclosed subject matter; (f) modifying by homologous recombination in a plant at least one chromosomal copy of the gene corresponding to an isolated nucleic acid molecule of the presently disclosed subject matter; (g) modifying by homologous recombination in a plant at least one chromosomal copy of the regulatory elements of a gene corresponding to an isolated nucleic acid molecule of the presently disclosed subject matter; and (h) expressing in said plant an isolated nucleic acid molecule of the presently disclosed subject matter or a portion thereof in the "sense" and "antisense" orientation.

The presently disclosed subject matter also provides a method for screening a plurality of compounds for a modulator of the enzymatic activity of a polypeptide disclosed herein. The method comprises:

- (f) providing a library of test compounds;
- (g) contacting an inositol phosphate kinase polypeptide selected from the group consisting of Ipk1 and Ipk2 with each test compound;
- (h) detecting an interaction between a test compound and the inositol phosphate kinase polypeptide;
- (i) identifying a test compound that interacts with the inositol phosphate kinase polypeptide; and
- (j) isolating the test compound that interacts with the inositol phosphate kinase polypeptide, whereby a plurality of compounds is screened for a modulator of inositol phosphate kinase polypeptide enzymatic activity.

In some embodiments, the test compounds are bound to a substrate. In some embodiments, the test compounds are also synthesized directly on the substrate.

In some embodiments, yeast knockouts for Ipk 1 and/or 2 are complemented with plant orthologs of interest. Yeast strains used rely on the plant orthologs to survive.

As such, test compounds can be screened as lpk inhibitors by contacting the compound with the yeast knockouts complemented with the lpk of interest. Compounds that inhibit the lpk of interest will adversely affect the yeast.

In some embodiments, the modulator of the inositol phosphate kinase polypeptide enzymatic activity is an antagonist of the inositol phosphate kinase polypeptide enzymatic activity. An antagonist is a compound that decreases the natural biological functions of the enzymes.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing phytate biosynthetic enzyme-induced activities, thereby preventing the action of the enzyme by excluding the enzyme from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such as binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules.

EXAMPLES

The following Examples have been included to illustrate representative modes of the presently disclosed subject matter. In light of the present disclosure and the general level of skill in the art, those of skill will appreciate that the following Examples are intended to be exemplary only in that numerous changes, modification, and alterations can be employed without departing from the spirit and scope of the invention.

Example 1

AtIPK1 encodes an IP₄ and IP₅ 2-kinase

In order to identify the Arabidopsis ortholog of yeast lpk1 the Arabidopsis genome database was searched for candidate genes with sequence similarity to the known IP₅ 2-kinases from *S. cerevisiae*, *S. pombe*, and *C. albicans*. E. B. Ives, J.

Nichols, S. R. Went, J. D. York, *J. Biol. Chem.* 275, 36575 (2000). Three putative IP₅ 2-kinase genes at loci At1g22100, At5g42810, and At1g59312, were identified and had regions of similarity to the known human and fungal IP₅-2 kinases and to a rice and maize sequence (Figure 1B). Of the three Arabidopsis genes, only At5g42810, designated *AtIPK1*, appeared to be expressed in major tissues throughout the plant (Figure 1C). Expression of the other two putative Arabidopsis *IPK1* genes could not be detected and there are currently no ESTs available for these in the database, indicating that they are pseudogenes, or are selectively expressed in a spatio-temporal manner thereby limiting their detection by Northern and reverse-transcriptase PCR.

To verify that the putative *AtIPK1* gene encoded a functional IP₅ 2-kinase, purified recombinant AtIpK1 was tested *in vitro* for the ability to phosphorylate [³H]-I(1,3,4,5,6)P₅. As shown in Figure 2A, in the presence of ATP, recombinant AtIpK1 completely phosphorylated IP₅ to IP₆. It was also found that AtIpK1 could phosphorylate I(1,3,4,6)P₄ (Figure 2B) and I(1,4,5,6)P₄ (Figure 2C) on the D-2 position to produce I(1,2,3,4,6)P₅ and I(1,2,4,5,6)P₅, respectively. In an effort to recapitulate IP₆ synthesis *in vitro* from an I(1,4,5)P₃ precursor, nearly equal amounts of AtIpK2β (1.6 pmol) and AtIpK1 (1.3 pmol) were initially used, and surprisingly it was found that I(1,2,4,5,6)P₅ was the only major product (Figure 3A). Interestingly, the end product of the reaction could be changed to IP₆ with different relative doses of AtIpK2β. For instance, when 2.5 times as much AtIpK2β as AtIpK1 (3.2 pmol and 1.3 pmol, respectively) was incubated with I(1,4,5)P₃, I(1,2,4,5,6)P₅ formation was diminished and IP₆ was produced (Figure 3B). When the ratio of AtIpK2β (8 pmol) to AtIpK1 (1.3 pmol) was increased further to approximately 6:1, the final product was only IP₆ (Figure 3C). The differential product formation based on the relative dose of each enzyme may be due to different substrate affinities and rates of catalysis. The K_m of I(1,4,5,6)P₄ for AtIpK1 (7.6 μM) is approximately equal to its K_m of I(1,3,4,5,6)P₅ (7.1 μM) and one-half of the K_m value of I(1,4,5,6)P₄ for AtIpK2β (15 μM) (J. Stevenson-Paulik, A. R. Odom, J. D. York, *J. Biol. Chem.* 277, 42711 (2002)).

Since AtIpK1 has a higher apparent affinity for I(1,4,5,6)P₄ than does AtIpK2β, this could explain why, when the enzymes are present at nearly equal concentrations, the majority of the I(1,4,5,6)P₄ synthesized by AtIpK2β is converted to I(1,2,4,5,6)P₅ by AtIpK1. This result also indicates that although AtIpK2β performs two phosphorylation

events to convert $I(1,4,5)P_3$ to $I(1,3,4,5,6)P_5$, it does so in a non-processive manner in which the $I(1,4,5,6)P_4$ intermediate is released in between phosphorylation steps. When the ratio of $AtIPK2\beta$ to $AtIPK1$ is increased from 1:1 to 6:1, $AtIPK2\beta$ appears to out-compete $AtIPK1$ for $I(1,4,5,6)P_4$ phosphorylation and produces $I(1,3,4,5,6)P_5$ which is rapidly converted to IP_6 by $AtIPK1$. Another explanation for the dose-dependent changes in end-product formation is that $I(1,2,4,5,6)P_5$ inhibits $AtIPK2\beta$ and this can be overcome with an increased dose of $AtIPK2\beta$ to the reaction. Together these data indicate that IP_6 can be synthesized from an $I(1,4,5)P_3$ precursor by the activity of just two Arabidopsis gene-products and that the relative dose of each enzyme could influence the major IP formed. Additionally, it has been found that, along with the activity of a previously characterized $I(1,3,4)P_3$ 5/6-kinase (M.P. Wilson, P.W. Majerus, *Biochem. Biophys. Res. Commun.*, 232, 678-681 (1997); M. P. Wilson, P. W. Majerus. *J. Biol. Chem.* 271, 11904 (1996)), these enzymes can convert $I(1,3,4)P_3$ to IP_6 *in vitro* (Figure 3D) and that these three gene-products can synthesize up to 7 different IP species as outlined in Figure 1A.

Example 2

AtIPK1 is necessary for IP_6 synthesis in vivo

To study the effect of a loss of *AtIPK1* function on IP_6 production in the plant, publicly available populations of Arabidopsis T-DNA insertional mutants were searched for a disruption of *AtIPK1*, and identified a mutant (SALK_065337; designated here as *atipk1-1*) from the Salk Institute Genome Analysis Laboratory (SIGnAL) population of mapped insertions, with an insertion 77 nucleotides upstream of the stop codon in the last exon of the open reading frame (Figure 4A). The disruption causes a greater than 70% reduction in transcription of the *AtIPK1* gene as indicated by Northern blot analysis (Figure 4B). Southern blot analysis revealed that there were at least three T-DNA insertions in tandem in the same site of the *AtIPK1* gene (Figure 4C).

To assess the effect of the disruption of *AtIPK1* expression on IP_6 synthesis, IP_6 synthesis in fourth generation (T_4) seed of the T-DNA transformed line was monitored. Seeds were germinated in liquid Murashige and Skoog (MS) media containing 400 $\mu\text{Ci/ml}$ [^3H]-myo-inositol. The soluble inositol phosphates were harvested from 6 day-old seedlings and analyzed by Partisphere SAX HPLC. Under these conditions, the IP_6 content of wild type seedlings was 9.6% of the total labeled IP species. Conversely, the

IP₆ content of *atipk1-1* seedlings was 0.71% of the total labeled IP species, representing a 93% reduction from wild type (Figure 5A). The decrease in IP₆ was accompanied by a corresponding accumulation of IP₅ (3.4% of total), IP₄ (1.3% of total) and IP₃ (1.3% of total) species, which are undetectable in wild type seedlings. Of interest, the I(1,2,4,5,6)P₅ peak was no longer present in the *atipk1-1* seedlings, which is consistent with the *in vitro* data that this IP is *AtIPK1*-dependent.

To determine if there is an effect of the loss-of-function of the *AtIPK1* gene on inositol phosphate production in seed development, the siliques of T₄ *atipk1-1* plants were labeled and compared to wild type. Similar to seedlings, there was an 86% reduction of IP₆, a loss of I(1,2,4,5,6)P₅, and a correlative increase in IP₅, IP₄ and IP₃ in *atipk1-1* as compared to wild type siliques (Figure 5B). To take a closer look at individually developing seeds, the immature seeds from the labeled siliques were dissected and the IP profile was analyzed (Figure 5C). The reduction of IP₆ was 90% in *atipk1-1* as compared to wild type. In performing these labeling studies, it was found that although IP₆ synthesis occurred in the silique tissue, the incorporation of [³H]-inositol into IP₆ was barely detectable in the developing seeds of wild type and *atipk1-1* until the final stages of embryogenesis, in the late bent-cotyledon stage (Figure 6), just prior to desiccation where the majority of the endosperm has been consumed and the bulk of the developing seed is comprised of the embryo itself. This could indicate that the majority of IP₆ synthesis occurs in the last stages of seed development, or it could indicate that inositol uptake into the seed is not efficient until the siliques and seeds are developmentally mature.

To identify the species of IP₄ and IP₅ that accumulates in the *ipk1-1* mutant, *in vitro* kinase experiments were performed with recombinant *Atlpk1*, *Atlpk2*, and I(3,4,5,6)P₄ 1-kinase to determine if the peaks were substrates for these kinases. As indicated in Figure 7A, the IP₅ peak is completely phosphorylated by *Atlpk1*, indicating that it is I(1,3,4,5,6)P₅. In this reaction the IP₄ peak was also completely phosphorylated by *Atlpk1*, indicating that it is an IP₄ with a 2-position hydroxyl. Active *Atlpk2β* was unable to phosphorylate the IP₃, IP₄, or IP₅ peaks in the extract indicating that the IP₃ is not I(1,4,5)P₃, and that the IP₄ is not I(1,4,5,6)P₄, I(1,3,4,6)P₄, or I(1,3,4,5)P₅. However, the I(3,4,5,6)P₄ 1-kinase was able to phosphorylate the majority of the IP₄ peak, indicating that the peak is mostly I(3,4,5,6)P₄ (Figure 7B). The nature of the IP₃x

species was not identified, other than to determine that it is not I(1,4,5)P₃. The accumulated I(3,4,5,6)P₄ could result from the build-up of IP₆ precursors due to a change in equilibrium, or the dephosphorylation of the accumulated I(1,3,4,5,6)P₅ by an IP₅ 1-ptase which has been previously described to exist in other systems (M. W. Ho, X. Yang, M. A. Carew, T. Zhang, L. Hua, Y. U. Kwon, S. K. Chung, S. Adelt, G. Vogel, A. M. Rileey, B. V. Potter, S. B. Shears. *Curr Biol.* 12, 477 (2002)) and whose cognate gene has been identified in Arabidopsis (M.P. Wilson, P.W. Majerus, *Biochem. Biophys. Res. Commun.*, 232, 678-681 (1997)).

To measure mass levels of IPs in mature, desiccated seeds, a non-radioactive method to analyze bulk seed extracts was used. As indicated in Figure 8, IP₆ comprises 100% of the detected IPs in wild type mature seeds and equals approximately 22.3 +/- 0.4 nmol IP₆/mg seed. By comparison, the *atipk1-1* seeds contain 3.9 +/- 0.9 nmol IP₆/mg, 29.7 +/- 0.6 nmol IP₅/mg and 3.3 +/- 0.3 nmol IP₄/mg seed. Additionally, there is a 2-fold increase in the concentration of Pi in the *atipk1-1* (15.4 +/- 1.2 nmol/mg) seeds over wild type (7.5 +/- 1.7 nmol/mg). These data reveal that not only are the final seed concentrations of IP₆ reduced by over 80% in mature *atipk1-1* seeds, but also that there is a net increase of total inositol phosphates in the mutant. These data were further corroborated by results indicating that recombinant AtIpK1 can complement phytate production in IpK1 null yeast.

Example 3

***atipk1-1* has conditional growth effects and intracellular Pi levels**

To assess the effect of a disruption of IP₆ synthesis on plant growth, seeds were germinated in a mixture of sand and vermiculite, watered daily with half-strength Hoagland's solution, and monitored the growth over the life cycle of the plant. No discernible differences could be seen in the timing of germination or in the appearance of the cotyledons of wild type, heterozygotes, or homozygotes for the T-DNA insertion. After three weeks under these conditions, the rosette leaves of the *atipk1-1* plants became abaxially curled and this epinastic phenotype was more pronounced at four weeks where the leaves appeared to have necrotic margins and were approximately a third of the size of wild type and heterozygotes. This growth phenotype as well as the IP production was rescued by constitutive expression of *AtIPK1* (Figures 9A and 9B).

Because the phenotype appeared to be the result of some toxicity from the Hoagland's solution, each nutrient was systematically removed from the solution and it was found that only the reduction of phosphate could alleviate the leaf epinasty and necrosis. At Pi concentrations below 1 mM, the leaves of the *atipk1-1* mutants were not epinastic and the mass differential between the wild type and *atipk1-1* leaves was reduced. At Pi concentrations of 1 mM or higher, the leaves of the *atipk1-1* plants were epinastic and were one-third to one-quarter the size of wild type (Figure 10A). The leaf epinasty also corresponded with an increased intracellular concentration of Pi in the *atipk1-1* mutants (Figure 10B).

Thus, the *atipk1-1* plants are unable to maintain normal intracellular phosphate concentrations and appear to suffer from phosphate toxicity when administered phosphate levels that are typically optimal for plant growth. This increase in leaf intracellular phosphate cannot be explained by a redistribution of the phosphate from the IP₆ pool to the inorganic phosphate pool. In wild type leaves, one finds that Pi concentrations are typically 5-10 nmol/mg fresh weight while the IP₆ concentration is less than 0.250 nmol/mg. This difference represents at least a 20 to 40-fold higher level of Pi than IP₆ in leaves, which would indicate that the contribution of the redistributed phosphate that resulted from the change in IP₆, IP₅ and IP₄ concentrations in *atipk1-1* would be insignificant and would not account for a 2 to 3-fold increase in intracellular Pi levels. This argument suggests that the altered concentrations of IP₆, IP₅, and/or IP₄ in the *atipk1-1* mutant could cause an aberration in signaling that disrupts normal phosphate metabolism, such as uptake, distribution, or retention.

Example 4

***atipk1-1* roots have a decreased sensitivity to environmental phosphate**

Because the *atipk1-1* mutant appears to have altered phosphate homeostasis, it was tested to see if it also had an altered sensitivity to extracellular Pi levels. Plants respond to a sensed change in phosphate by increasing or decreasing the surface area of the roots to alter uptake. One of the ways that a change in surface area is achieved is by altering the length of the root hairs. Thus, for instance, when phosphate is limiting, Arabidopsis seedlings produce longer root hairs to scavenge more Pi (Lopez-Bucio et

al., 2003). Seeds from wild type, *atipk1-1*, and *atipk1-1* plants transformed with *pAtIPK1* were plated onto MS agar containing variable concentrations of Pi and grown in a vertical position under constant light. After 12 days, the roots were imaged and the root hair length was measured. It was found that the *atipk1-1* roots had longer root hairs than wild type at lower concentrations of Pi and a decreased ability to sense the increase in Pi as indicated by the inability to form shorter root hairs at higher Pi concentrations (Figure 11). The Pi insensitive root hair growth could be rescued in the *atipk1-1* mutant by complementation of *AtIPK1*, indicating that the effect is due to the loss of *AtIPK1*. This result suggests an inability of the mutants to sense the extracellular phosphate levels as readily as wild type.

Discussion of Examples 1-4

Given the negative effect of phytate accumulation in seeds on animal nutrition and the environment and the general lack of understanding of its role in plant biology, it is of great interest to delineate phytate biosynthesis and function. The results of Examples 1-4 support the notion that there is a connection between phosphate homeostasis and inositol phosphate production. Multiple low-phytate mutants have been generated by random mutagenesis in various crop plants and these mutants have a reported molar equivalent increase of Pi that compensates for the loss of phosphate from the IP pool. S. Larson, K. Young, A. Cook, T. Blake, V. Raboy, *Theor. Appl. Genet.*, 97, 141-146 (1998); F. Hatzack, K.S. Johansen, S.K. Rasmussen, *J. Agric. Food Chem.* 48, 6074-6080, (2000); S. Larson, J. Rutger, K. Young, V. Raboy, *Crop Sci.* 40, 1397-1405, (2000); J. R. Wilcox, G. S. Premachandra, K. S. Young, V. Raboy, *Crop Sci.* 40, 1601 (2000); W. D. Hitz, T. J. Carlson, P.S. Kerr, S. A. Sebastian, *Plant Physiol.* 128, 650 (2002); M. S. Otegui, R. Capp, L. A. Staehelin, *The Plant Cell* 14, 1311 (2002). The present data differ from these studies in that even though there is a net increase of IPs in the *atipk1-1* mutant, there is still an increase in Pi, which may indicate that perturbation of the IP₄, IP₅, and/or IP₆ levels causes a defect in phosphate metabolism and signaling resulting in a detrimental effect on growth when the plants are provided with adequate levels of environmental phosphate. The *lpk1* and 2 double knockout mutants described herein blow further reduces phytate to 7% that of the wild-type organisms, and further, other inositol phosphates do not accumulate either.

IP₆ has been implicated in the ABA-induced regulation of stomatal pore closure, thus controlling water conservation in leaves (J. D. York, A. R. Odom, R. Murphy, E. B. Ives, S. R. Went, *Science* 285, 96 (1999)). It was tested if the *atipk1-1* leaves had altered guard cell responses to ABA or increased water loss due to improper regulation of stomatal closure, but no differences between wild type and the mutant were found. This either indicates that IP₆ is not a mediator of this process or that the residual 10% of IP₆ in the mutant cells is adequate to maintain this signaling pathway. IP₆ was shown to be necessary for efficient mRNA export in *S. cerevisiae* (Y. Feng, S. R. Went, P. W. Majerus, *Proc. Natl. Acad. Sci. USA* 98, 875 (2001)) and human embryonic cells (X. Shen, H., Xiao, R. Ranallo, W-H., Wu, C., *Science* 299, 112 (2003)) and to be associated with the mammalian DNA-PK complex as a necessity for the non-homologous end-joining activity of the complex (L. A. Hanakahi, S. C. West, *Embo J.* 21, 2038 (2002); Y. Ma, M. R. Lieber. *J. Biol. Chem.* 277, 10756 (2002); R. F. Irvine, M. J. Schell. *Nature Reviews* 2, 327 (2001)). Since some of these functions appear to be conserved across species, the *atipk1-1* plants might suffer from a defect in mRNA translocation and/or an inability to repair DNA damage, although altered mRNA export or any significant differences in the sensitivity of *atipk1-1* seed germination or seedling growth to methyl methane sulfonate, a strong mutagen that induces double strand breaks, were not detected.

The effect of the *AtIPK1* loss-of-function on vegetative growth could be the result of a specific defect such as an increased cellular Pi concentration, or a pleiotropic one such as altered concentrations of multiple IP species, each of which might have specific regulatory functions. For instance, the phenotype might not be due to the loss of IP₆, but actually a loss of I(1,2,4,5,6)P₅ or another IP either produced by *AtIPK1* or an IP for which IP₆ is a precursor, such as a pyrophosphorylated inositol species. Additionally, since IP₃x, I(3,4,5,6)P₄, and I(1,3,4,5,6)P₅, accumulate in these tissues the aberrant growth of the mutant plants could be due to the gain-of-function of these IP species. Hence, one hypothesis is that the perturbation of synthesis of higher IPs in the *atipk1-1* mutant causes an uncontrolled response to environmental phosphate. Also, proper control of transcription and chromatin remodeling of the *PHO5* locus depends on *IPK2*, the kinase that generates IP₄ and IP₅ in yeast (L. A. Hanakahi, M. Barlet-Jones, C. Chappell, D. Pappin, S. C. West, *Cell* 102, 721 (2000)). *PHO5* is induced during

phosphate starvation and encodes a secreted acid phosphatase that hydrolyzes phosphate from extracellular organic sources. Thus, a link is suggested between the maintenance of intracellular phosphate levels and IP production. Indeed, the present findings support the notion that there is communication between inositol phosphate production and phosphate homeostasis within the cell.

The present findings that the IP₆ content in seeds can be reduced by more than 80% without compromising seed development, germination, and seedling development challenges the hypothesis that IP₆ itself is an essential nutrient store of phosphate in the seed. Thus, specific inactivation of the *IPK1* gene in developing seeds appears to spare the defects in vegetative growth and signaling and thus, this gene is a viable target for engineering low-phytate crops to enhance the nutritional value of feed grains and reduce phytate-induced phosphorus pollution.

Materials and Methods for Examples 1-4

Inositol phosphate kinase assays - Unlabeled IPs were purchased from Cell Signals, Inc. Tritiated [³H]-I(1,4,5)P₃ was purchased from Perkin Elmer Life Sciences. [³H]-I(1,3,4,6)P₄ was synthesized as previously described (B. Q. Phillippy, A. H. Ullah, K. C. Ehrlich. *J. Biol. Chem.* 269, 28393 (1994)). [³H]-I(1,4,5,6)P₄ was synthesized by incubating 10 μM [³H]-I(1,4,5)P₃ with 1.6 pmol GST-AtIpK2β for 10 min at 37°C. [³H]-I(1,3,4,5,6)P₅ was synthesized by incubating 10 μM [³H]-I(1,4,5)P₃ with 1.6 pmol GST-AtIpK2β for 30 min. The reaction was stopped by heat-inactivation at 100°C for 1 min. 1.3 pmol AtIpK1 was then incubated with 10 μM [³H]-I(1,3,4,6)P₄, [³H]-I(1,4,5,6)P₄, and [³H]-I(1,3,4,5,6)P₅ for 30 min at 37°C in a buffer containing 50 mM Hepes, pH 7.5, 50 mM KCl, 2 mM ATP, 10 mM MgCl₂. The reaction products were separated and analyzed by HPLC over a Partisphere strong anion exchange (SAX) column and a linear gradient from 10 mM to 1.7 M NH₄H₂PO₄ (pH 3.5) over 12 min, followed by elution for 25 min with 1.7 M NH₄H₂PO₄. IPs were identified based on comparison to the elution of known IP species.

Semi-quantitative RT-PCR. This was performed exactly as previously described (Stevenson-Paulik et al., 2002). The primers used were designed to amplify the 5' half of the open reading frame (kb) of *AtIPK1* and the entire open reading frame of *ACT2* (kb) (GenBank accession U41998). *AtIPK1* primers used were constructed as follows.

AtIPK1 sense primer: 5' GGA GAT GAT TTT GGA GGA GAA AGA TGC AT 3'

AtIPK1 antisense primer: 5' ATG CAT TTT GAA ACG GCT TAC GCT TGT TTT G 3'

ACT2 sense primer: 5'ATG GCT GAG GCT GAT GAT ATT CAA C 3'

ACT2 antisense primer: 5' TGT GAA CGA TTC CTG GAC CTG CCT C 3'

- 5 *[³H] myo-inositol labeling of Arabidopsis seedlings and siliques.* Individual wild type and *atipk1-1* seeds were stratified at 4°C for 2 days and then germinated in 25-50 µl liquid Murashige and Skoog (MS) salts media with 0.4 mCi/ml [³H]myo-inositol for 6 days at 20°C and constant light. The seedlings were washed twice with cold dH₂O and then disrupted by agitation twice for 2 min with acid-washed glass beads in the presence of 100 µl 0.5 N HCl, 125 µl 2M KCl, 125 µl CHCl₃, 372 µl CHCl₃:MeOH (1:2). Disrupted tissue was centrifuged for 5 min and the soluble layer was removed and analyzed by Partisphere SAX HPLC as described above. Siliques were labeled by standing vertically in tubes containing 50 µl MS salts and 100 µCi [³H] myo-inositol for 2 days at 20°C, constant light. IPs were harvested either from whole siliques or from individually dissected developing seeds exactly as described for seedlings.

- 15 *Non-radioactive, mass seed IP analysis.* For each sample, approximately 50 µl seeds were weighed and 10 volumes of 0.4 N HCl and 50 µl acid-washed glass beads (425-600 microns) were combined for pulverization by a "bead beater" for 4 minutes. The seeds were then boiled for 5 minutes and beaten again for 4 minutes. The seed extract was then centrifuged for 10 minutes at maximum speed and the supernatant was frozen on dry ice, thawed to room temperature and centrifuged again. The supernatant was loaded onto an IonPac AS7 anion exchange column (Dionex, Sunnyvale, California, United States of America) equilibrated with 10 mM methyl piperazine pH 4.0 (Buffer A). The IPs were eluted with a linear gradient from 0 to 100% buffer B (1 M NaNO₃ pH 4.0) with a flow rate of 0.3 ml/min using an HP190 pump. Eluate was mixed with color reagent (0.015% [w/v] FeCl₃; 0.15% [w/v] sulfosalicylic acid) and detected with a photodiode array detector at an absorbance of 550 nm.

- 20 *IPlant growth conditions.* Plants were grown in a mixture of vermiculite and sand in a controlled growth chamber with constant temperature of 21°C. The plants were exposed to 14 hr light (135-150 µmol/m² s⁻¹) and 10 hr dark cycles and were watered with half-strength Hoagland's solution every morning and with dH₂O every evening.

30 *Pi assay.* Leaf samples were homogenized in 0.4 N HCl, centrifuged, and the

supernatant was incubated 1:1 with color reagent (3 M H₂SO₄:2.5%NH₄Mo: 10% ascorbate: dH₂O; 1:1:1:1). A standard curve was generated with K₂HPO₄ and found to be linear between 0-100 nmol phosphate.

5

Example 5

Double knockout eliminates most IP production including IP₄, IP₅ and IP₆

This Example describes an ipk1/ipk2 beta double knockout plant. The double mutant grows about the same as the single mutant disclosed herein above. Thus, the seed specific knockdown of all IPs is not detrimental to seed yield and plant growth. The term "seed specific" is used because analysis of the plant tissue from the atlpk2 beta mutant shows no defect in IP metabolism whereas the seeds do. Further the ipk1 ipk2 beta double mutant appears to synthesize IPs except for IP₅ in the tissue.

The growth issue and yield problems of ipk1 mutant and ipk1/ipk2 beta double mutant plants are expected to be corrected with the tissue (but not seed) expression of ipk1. Complementation of the ipk1 mutant plant with pBART Ipk1 complements the growth. The ipk1/ipk2 beta double mutant can be complemented with Ipk1 under one or more tissue specific promoters as disclosed herein above, e.g. one or more promoters that can drive expression in one or more tissues except seeds.

Figure 12A is a schematic of T-DNA insertion into *atlpk2β* locus. *AtIPK2* is a single exon and the T-DNA inserts at nucleotide 400 of the 901 nucleotide open reading frame (ORF). Figure 12B depicts a Northern blot analysis of mRNA isolated from wild-type, *atipk2β* -1, or *atipk1β* -1 plants confirms that the T-DNA insertion disrupts over 90% of the expression of *atlpk2β* levels. Probing for *atlpk2β* shows that its expression is unaffected by the loss of *atlpk1β* and *atlpk2β*. Figure 13 is a schematic of IP determination methods used.

Using non-radioactive method of measuring Phytate (IP₆) levels in mature seeds derived from wild-type or kinase mutant plants (single deletion) shows that both Ipk2 and Ipk1 result in lowered phytate in seeds. See Figures 14A and 14B. Given there are two genes encoding Ipk2 it is likely that the alpha gene partially compensates for the loss of beta. A double knockout of Ipk2 alpha and beta should ablate IP₃, IP₄, IP₅

and IP₆ synthesis. IP analysis from the *lpk1* and *lpk2* beta double mutant is shown in Figures 14A and 14B.

The *ipk1/ipk2* double mutant plants are viable, are partially growth compromised and develop normally as compared to wild type. See Figure 15, where this is shown as *ipk1-1* complemented with *plpk1*. Restoration of seed excluded expression of *plPK1* with a specific promoter will likely restore full growth of the double mutant.

Total seed phytate (IP₆) levels calculated from wild-type, *ipk1*, *ipk2*, and *ipk1 ipk2* double mutant seeds are 100%, 12%, 66% and 7%, respectively. See Figure 16.

HPLC analysis of total seed extracts is listed and quantified in Figures 17A and 17B. It is noted that deletion of both *ipk1* and *ipk2* results in lowering total IP production, specifically IP₅ and nearly ablating IP₆. Additionally, IP₂, IP₃ and IP₄ levels do not accumulate to high levels of phytate seen in wild-type seeds. The reduction of phytate and total IP molecules is greater than in the *ipk1-1* mutant alone. Wild-type HPLC analysis is provided in accordance with techniques disclosed herein above.

Example 6

***Atlpk2* beta TBLASTN of rice (*Oryza sativa*)**

The amino acid sequence of the *A. thaliana* *Atlpk2* beta gene product was compared to the *Oryza sativa* genomic sequence database available at the website of The Institute for Genomic Research using the program TBLASTN. The results are presented in Figure 18.

An unspliced genomic clone identified as "9630.t02959" in the *Oryza sativa* genomic database had high homology (Identities = 153/297 (51%), Positives = 195/297 (65%), Frame = +1).

Example 7

***Atlpk2* beta TBLASTN of Maize (*Zea mays*)**

The amino acid sequence of the *A. thaliana* *Atlpk2* beta gene product was compared to a *Zea mays* genomic sequence database available at the website of The Institute for Genomic Research using the program TBLASTN. The results are presented in Figure 19.

Several genomic clones were identified as showing high homology to the *A. thaliana* *Atlpk2* beta gene product. These included several translational frames of clones AZM4_23277 (Figure 19A), AZM4_23275 (Figures 19B-D), and AZM4_23276

(Figures 19E-G).

Example 8

Atlpk2 beta TBLASTN of All Plant Sequences at GENBANK®

The amino acid sequence of the *A. thaliana* Atlpk2 beta gene product was compared to the GENBANK® sequence database available at the website of The National Center for Biotechnology Information using the program TBLASTN. The results are presented in Figure 20.

Several homologous sequences were identified, including Accession No. AB010069.1 (Figure 20A), Accession No. AL163912.1 (Figure 20B), and Accession No. AP006476.1 (Figure 20C).

Example 9

Atlpk2 alpha TBLASTN of Maize (*Zea mays*) at TIGR

The amino acid sequence of the *A. thaliana* Atlpk2 alpha gene product was compared to a *Zea mays* genomic sequence database available at the website of The Institute for Genomic Research using the program TBLASTN. The results are presented in Figure 21.

This BLAST search identified three genes encoding lpk2 orthologs – note all three have PxxxDxKxG motif (shaded in Figure 21) found in all lpk2 family members. Actual DNA sequence and homology are shown separately for all three loci in Figure 21. These include subsequences of AZM4_23277 (Figures 21A and B), AZM4_23275 (Figures 21C and D), and AZM4_23276 (Figures 21E-G).

Example 10

Atlpk2 alpha TBLASTN of All Plant Sequences at GENBANK®

The amino acid sequence of the *A. thaliana* Atlpk2 alpha gene product was compared to the GENBANK® sequence database available at the website of The National Center for Biotechnology Information using the program TBLASTN. The results are presented in Figure 22.

Several homologous sequences were identified, including Accession No. AL163912 (Figure 22A), Accession No. AB010069.1 (Figure 22B), and Accession No. AP006476.1 (Figure 22C).

Example 11

TBLASTN of all Plant genomes at GENBANK®

This Example pertains to a TBLASTN search of all plant genomes (excluding *Arabidopsis*). The search obtained 1 hit from rice genome (*Oryza sativa*) – 1 gene product found with 42% overall identity to Atlpk1beta protein. The gene comprises at least 7 exons as outlined in Figure 23A. The sequences comprise signature motifs of
5 lpk1 protein sequences. Hits from wheat are also shown in Figure 23.

Figures 23B-F depict alignments of the Atlpk1beta protein sequence with various translational frames of GENBANK® Accession No. AL606608.3. Figures 23G and H depict alignments of the Atlpk1beta protein sequence with various translational frames of GENBANK® Accession No. BE498028.1 and BQ169513.1, clones from a wheat
10 cDNA library.

REFERENCES

The references listed below as well as all references cited in the specification are incorporated herein by reference to the extent that they supplement, explain, provide a background for or teach methodology, techniques, and/or compositions employed
15 herein.

Allison RF, Johnston RE & Dougherty WG (1986) The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology* 154:9-20.

20 Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic Local Alignment Search Tool. *J Mol Biol* 215:403-410.

Aoyama T & Chua N-H (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J* 11:605-612.

Barany G & Merrifield RB (1980) Solid-Phase Peptide Synthesis, in Peptides: Analysis, Synthesis, Biology, Vol. 2, Special Methods in Peptide Synthesis, Part A (Gross
25 E & Meienhofer J, eds) Academic Press, New York, New York, United States of America, pp. 3-284.

Bartlett SG, Grossman AR & Chua N-H (1982) in Methods in Chloroplast Molecular Biology, (Edelman M, Hallick RB & Chua N-H, eds.) Elsevier Biomedical Press,
30 New York, New York, United States of America, pp. 1081-1091.

Batzer MA, Carlton JE & Deininger PL (1991) Enhanced evolutionary PCR using oligonucleotides with inosine at the 3'-terminus. *Nucleic Acid Res.* 19:5081.

- Beerli RR, Segal DJ, Dreier B & Barbas CF 3rd. (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* 95:14628-14633.
- 5 Bevan M (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucl. Acids Res* 12:8711-21.
- Bevan M, Flavell R B & Chilton M D (1983) A chimeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature* 304:184-187.
- Binet MN Weil JH & Tessier LH (1991) Structure and expression of sunflower ubiquitin
10 genes. *Plant Mol Biol* 17:395-407.
- Bloching K & Diggelmann H (1984) Hygromycin B phosphotransferase as a selectable marker for DNA transfer experiments with higher eucaryotic cells. *Mol Cell Biol* 4:2929-2931.
- Bourouis M & Jarry B (1983) Vectors containing a prokaryotic dihydrofolate reductase
15 gene transform *Drosophila* cells to methotrexate-resistance. *EMBO J* 2:1099-1104.
- Bruggemann E, Handwerger K, Essex C & Storz G (1996) Analysis of fast neutron-generated mutants at the *Arabidopsis thaliana* HY4 locus. *Plant J* 10:755-760.
- Caddick MX, Greenland AJ, Jepson I, Krause KP, Qu N, Riddell KV, Salter MG, Schuch
20 W, Sonnewald U & Tomsett AB. (1998) An ethanol inducible gene switch for plants used to manipulate carbon metabolism. *Nat Biotechnol* 16:177-180.
- Callis J, Fromm M & Walbot V (1987) Introns increase gene expression in cultured maize cells. *Genes Dev.* 1:1183-1200.
- Callis J, Raasch JA & Vierstra RD (1990) Ubiquitin extension proteins of *Arabidopsis thaliana*. Structure, localization, and expression of their promoters in transgenic
25 tobacco. *J Biol Chem* 265:12486-12493.
- Casas AM, Kononowicz AK, Zehr UB, Tomes DT, Axtell JD, Butler LG, Bressan RA & Hasegawa PM (1993) Transgenic sorghum plants via microprojectile bombardment. *Proc Natl Acad Sci U S A* 90:11212-6.
- 30 Chibbar RN, Kartha KK, Datta RSS, Leung N, Caswell K, Mallard CS & Steinhauer L (1993) The effect of different promoter-sequences on transient expression of *gus* reporter gene in cultured barley (*Hordeum vulgare* L.) cells. *Plant Cell Rep*

12:506-509.

Christensen AH & Quail PH (1989) Sequence analysis and transcriptional regulation by heat shock of polyubiquitin transcripts from maize. *Plant Mol Biol* 12:619-632.

Christou P, Ford T & Kofron M (1991) Production of transgenic rice (*Oryza sativa* L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. *Bio/Technology* 9: 957-962.

Chuang CF & Meyerowitz EM (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 97:4985-90.

Clark MS (ed.) (1997) Plant Molecular Biology: A Laboratory Manual, Chapter 7, Springer-Verlag GmbH & Co. KG, Berlin, Germany.

Comai L, Larson-Kelly N, Kiser J, Mau CJ, Pokalsky AR, Shewmaker CK, McBride K, Jones A & Stalker DM (1988) Chloroplast transport of a ribulose biphosphate carboxylase small subunit-5-enolpyruvyl 3-phosphoshikimate synthase chimeric protein requires part of the mature small subunit in addition to the transit peptide. *J Biol Chem* 263:15104-15109.

Creighton TE (1984) Proteins, WH Freeman & Co., New York, New York, United States of America.

Datta SK, Peterhans A, Datta K & Potrykus I (1990) Genetically engineered fertile Indica-rice recovered from protoplasts. *Bio/Technology* 8:736-740.

de Framond AJ (1991) A metallothionein-like gene from maize (*Zea mays*). Cloning and characterization. *FEBS Lett* 290:103-6.

Della-Cioppa G, Kishore GM, Beachy RN & Fraley RT (1987) Protein trafficking in plant cells. *Plant Physiol* 84:965-968.

Deutscher MP, Simon MI & Abelson JN (eds) (1997) Guide to Protein Purification, *Meth Enzymol* Volume 182.

Dong, JJ, Teng, WM, Buchholz WG & Hall TC (1996). *Agrobacterium*-mediated transformation of javanica rice. *Mol Breeding* 2:267-276.

Ecker JR & Davis RW (1986) Inhibition of Gene Expression in Plant Cells by Expression of Antisense RNA. *Proc Natl Acad Sci USA* 83:5372-5376.

Elroy-Stein O, Fuerst TR & Moss B (1989) Cap-Independent Translation of mRNA

Conferred by Encephalomyocarditis Virus 5' Sequence Improves the Performance of the Vaccinia Virus/Bacteriophage T7 Hybrid Expression System. *Proc Natl Acad Sci USA* 86:6126-6130.

EP 0 292 435

5 EP 0 332 104

EP 0 332 581

EP 0 342 926

EP 0 392 225

10 Feng, Y, Wenthe, SR, Majerus, PW, Over expression of the inositol phosphatase SopB in human 293 cells stimulates cellular chloride influx and inhibits nuclear mRNA export. *Proc. Natl, Acad. Sci. USA* 98, 875 (2001).

Firek S, Draper J, Owen MR, Gandeche A, Cockburn B & Whitelam GC (1993) Secretion of a functional single-chain Fv protein in transgenic tobacco plants and cell suspension cultures. *Plant Mol Biol* 22:129-142.

15 Fromm ME, Morrish F, Armstrong C, Williams R, Thomas J & Klein TM. (1990) Inheritance and expression of chimeric genes in the progeny of transgenic maize plants. *Biotechnology (NY)* 8:833-839.

20 Gallie DR, Sleat DE, Watts JW, Turner PC & Wilson TM (1987) A comparison of eukaryotic viral 5'-leader sequences as enhancers of mRNA expression *in vivo*. *Nucl Acids Res* 15:8693-8711.

Gallie DR *et al.* (1989) in Molecular Biology of RNA (Cech TR, ed.), Alan R. Liss, Inc., New York, New York, United States of America, pp. 237-256.

GenBank accession U41998

25 Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl
30 T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A & Briggs, S (2002) A draft sequence of the rice genome (*Oryza sativa L ssp japonica*). *Science* 296:92-100 .

- Golovan, SP, Meidinger, RG, Ajakaiye, A., Cottrill, M., Wiederkehr, MZ, Barney, DJ, Plante, C., Pollard, JW, Fan, MZ, Hayes, MA, Laursen, J., Hjorth, JP, Hacker, RR, Phillips, JP, Forsberg, CW, Pigs expressing salivary phytase produce low-phosphorus manure, *Nature Biotech.*, 19, 741-745 (2001).
- 5 Gordon-Kamm WJ, Spencer TM, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams WR Jr, Willetts NG, Rice TB, Mackey CJ, Krueger RW, Kausch AP & Lemaux PG (1990) Transformation of Maize Cells and Regeneration of Fertile Transgenic Plants. *Plant Cell* 2:603-618.
- Green MR (2000) TBP-associated factors (TAFII)s: multiple, selective transcriptional
10 mediators in common complexes. *Trends Biochem Sci* 25:59-63.
- Green PJ, Pines O & Inouye M (1986) The role of antisense RNA in gene regulation. *Ann Rev Biochem* 55:569-597.
- Gritz L & Davies J (1983) Plasmid-encoded hygromycin B resistance: the sequence of
hygromycin B phosphotransferase gene and its expression in *Escherichia coli*
15 and *Saccharomyces cerevisiae*. *Gene* 25:179-188.
- Harlow E & Lane D (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, United States of America.
- Hanakahi, LA, Jones-Barlet, M, Chappel, C, Pappin, D, West, SC, Binding of inositol
phosphate to DNA-PK and stimulation of double-strand break repair. *Cell* 102,
20 721 (2000).
- Hanakahi, LA, West, SC, Specific interaction of IP6 with human Ku70/80, the DNA-binding subunit of DNA-PK *Embo J.* 21, 2038 (2002).
- Hatzack, F, Johansen, KS, Rasmussen, SK, Nutritionally relevant parameters in low-
phytate barley (*Hordeum vulgare* L) grain mutants *J. Agric. Food Chem.* 48,
25 6074-6080, (2000).
- Henikoff S & Henikoff JG (1992) Amino Acid Substitution Matrices from Protein Blocks. *Proc Natl Acad Sci U S A* 89:10915-10919.
- Hiei Y, Komari T & Kubo T (1997) Transformation of rice mediated by *Agrobacterium tumefaciens*. *Plant Mol Biol* 35:205-18.
- 30 Hiei Y, Ohta S, Komari T & Kumashiro Y (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J* 6(2):271-282.

- Hitz, WD, Carlson, TJ, Kerr, P.S., Sebastian, SA, Biochemical and Molecular Characterization of a Mutation That Confers a Decreased Raffinosaccharide and Phytic Acid Phenotype on Soybean Seeds. *Plant Physiology*. 128, 650 (2002).
- Ho, MW, Yang, X., Carew, MA, Zhang, T., Hua, L., Kwon, YU, Chung, SK, Adelt, S.,
5 Vogel, G., Riley, AM, Potter, BV, Shears, SB Regulation of Ins(3,4,5,6)P(4) signaling by a reversible kinase/phosphatase. *Curr Biol*. 12,477 (2002).
- Höfgen R & Willmitzer L (1988) Storage of competent cells for *Agrobacterium* transformation. *Nucl Acids Res* 16:9877.
- Holm, PB, Kristiansen, KN, Pedersen, HB, Transgenic approaches in consumed
10 cereals to improve iron and zinc content and bioavailability. Symposium: breeding: A new tool for fighting micronutrient malnutrition. Special issue. *Journal of Nutrition*, 132, 514S-516S (March 2002).
- Hudspeth RL & Grula JW (1989) Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis.
15 *Plant Molec Biol* 12:579-589.
- Irvine, RF, Schell, MJ Back in the water: the return of the inositol phosphates.. *Nature Reviews* 2, 327 (2001).
- Ives, B., Nichols, J., Went, SR, York, JD, Biochemical and Functional Characterization of Inositol 1,3,4,5,6-Pentakisphosphate 2-Kinases *J. Biol. Chem.* 275,36575
20 (2000).
- Jobling SA & Gehrke L (1987) Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. *Nature* 325:622-625.
- Jorgensen RA, Cluster PD, English J, Que Q & Napoli CA (1996) Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs.
25 antisense constructs and single-copy vs. complex T-DNA sequences. *Plant Mol Biol* 31:957-973.
- Karlin S & Altschul SF (1993) Applications and Statistics for Multiple High-Scoring Segments in Molecular Sequences. *Proc Natl Acad Sci U S A* 90:5873-5877.
- Keegan L, Gill G & Ptashne M (1986) Separation of DNA binding from the transcription-
30 activating function of a eukaryotic regulatory protein. *Science* 231:699-704.
- Kellogg EA (1998) Relationships of cereal crops and other grasses. *Proc Natl Acad Sci USA* 95:2005-2010

- Kempin SA, Liljegren SJ, Block LM, Rounsley SD, Yanofsky MF & Lam E (1997) Targeted disruption in *Arabidopsis*. *Nature* 389:802-803.
- Klein TM, Wolf ED, Wu R & Sanford JC (1987) High velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327:70-73.
- 5 Koehler SM & Ho TM (1990) Hormonal regulation, processing, and secretion of cysteine proteinases in barley aleurone layers. *Plant Cell* 2:769-783.
- Kong P & Steinbiss HH (1998) Complete nucleotide sequence and analysis of the putative polyprotein of maize dwarf mosaic virus genomic RNA (Bulgarian isolate). *Arch Virol* 143:1791-1799.
- 10 Koziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, Crossland L, Dawson J, Desai N, Hill M, Kadwell S, Launis L, Lewis K, Maddox D, McPherson K, Meghji MR, Merlin E, Rhodes R, Warren GW, Wright M & Evola SV (1993) Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. *Bio/Technology* 11:194-200.
- 15 Kuchler RJ (1997) Biochemical Methods in Cell Culture and Virology, Dowden, Hutchinson and Ross, Inc., Stroudsburg, Pennsylvania, United States of America
- Kyte J & Doolittle RF (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. *J Mol Biol* 157:105-132.
- Larson, S, Rutger, J, Young, K, Raboy, V, Isolation and Genetic Mapping of a Non-
- 20 Lethal Rice (*Oryza sativa* L.) *low phytic acid 1* Mutation *Crop Sci.* 40, 1397-1405, (2000)
- Larson, S, Young, K, Cook, A, Blake, T, Raboy, V, Linkage mapping of two mutations that reduce phytic acid content of barley grain. *Theoretical and Applied Genetics*, 97, 141-146 (1998).
- 25 Lebel E, Heifetz P, Thorne L, Uknes S, Ryals J & Ward E (1998) Functional analysis of regulatory sequences controlling PR-1 gene expression in *Arabidopsis*. *Plant J* 16:223-233.
- Liu LX, Spoerke JM, Mulligan EL, Chen J, Reardon B, Westlund B, Sun L, Abel K, Armstrong B, Hardiman G, King J, McCague L, Basson M, Clover R & Johnson
- 30 CD (1999) High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res* 9:859-867.
- Logemann J, Lipphardt S, Lorz H, Hauser I, Willmitzer L & Schell J (1989) 5' upstream

sequences from the *wun1* gene are responsible for gene activation by wounding in transgenic plants. *Plant Cell* 1:151-158.

Lommel SA, Kendall TL, Xiong Z & Nutter RC (1991) Identification of the maize chlorotic mottle virus capsid protein cistron and characterization of its subgenomic messenger RNA. *Virology* 81:382-385.

Ma, Y, Lieber, MR, Binding of Inositol Hexakisphosphate (IP₆) to Ku but Not to DNA-PK_{cs}, *J. Biol. Chem.* 277, 107256 (2002).

Macejak DG & Sarnow P (1991) Internal initiation of translation mediated by the 5' leader of a cellular mRNA. *Nature* 353:90-94.

Martienssen RA (1998) Functional genomics: probing plant gene function and expression with transposons. *Proc Natl Acad Sci USA* 95:2021-2026.

Mayo O (1987) The Theory of Plant Breeding, Second Edition, Clarendon Press, New York, New York, United States of America.

McBride KE & Summerfelt KR (1990) Improved binary vectors for *Agrobacterium*-mediated plant transformation. *Plant Mol Biol* 14: 269-276.

McBride KE, Schaaf DJ, Daley M & Stalker DM (1994) Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proc Natl Acad Sci USA* 91:7301-7305.

McElroy D, Blowers AD, Jenes B & Wu R (1991) Construction of expression vectors based on the rice actin 1 (Act1) 5' region for use in monocot transformation. *Mol. Gen. Genet.* 231:150-160.

McElroy D, Zhang W, Cao J & Wu R (1990) Isolation of an efficient actin promoter for use in rice transformation. *Plant Cell* 2:163-71.

Merrifield RB (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide *J. Am. Chem. Soc.* 85:2149-54.

Messing J & Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19:259-268.

Mettler IJ (1987) A simple and rapid method for minipreparation of DNA from tissue-cultured plant cells *Plant Mol Biol Reporter* 5:346-349.

Miao ZH & Lam E (1995) Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*. *Plant J* 7:359-365.

Mukumoto F, Hirose S, Imaseki H & Yamazaki K (1993) DNA sequence requirement of

a TATA element-binding protein from *Arabidopsis* for transcription *in vitro*. *Plant Mol Biol* 23: 995-1003.

Murashige T & Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiologia Plantarum* 15: 473-497 (1962)

5 Needleman SB & Wunsch CD (1970) A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two proteins. *J Mol Biol* 48:443-453.

Negrotto D, Jolley M, Beer S, Wenck AR & Hansen G (2000) The use of phosphomannose-isomerase as a selection marker to recover transgenic maize plant (*Zea mays* L.) via *Agrobacterium* transformation. *Plant Cell Reports* 19:798-803.

10 Norris SR, Meyer SE & Callis J (1993) The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. *Plant Mol Biol* 21:895-906.

15 Ohtsuka E, Matsuki S, Ikehara M, Takahashi Y & Matsubara K (1985) An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J. Biol. Chem.* 260:2605-2608.

Otegui, MS, Capp, R, Staehelin, LA, Developing seeds of *Arabidopsis* store different minerals in two types of vacuoles and in the ER *The Plant Cell* 14, 1311 (2002).

20 Ozturk ZN, Talame V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, Tuberosa R & Bohnert HJ (2002). Monitoring largescale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol. Biol* 48:551-573.

Parry MAJ, Androljic PJ, Khan S, Lea PJ & Keys AJ (2002) Rubisco activity: effects of drought stress. *Ann Bot* 89:833-839.

25 Paszkowski J, Bauer M, Bogucki A & Potrykus I (1988) Gene targeting in plants. *EMBO J* 7:4021-26.

Paszkowski J, Shillito RD, Saul M, Mandak V, Hohn T, Hohn B & Potrykus I (1984) Direct gene transfer to plants. *EMBO J* 3:2717-2722.

30 Paterson AH (1996) The DNA Revolution, chapter 2 in Genome Mapping in Plants, Paterson AH (ed), Academic Press/R.G. Lands Co., Austin, Texas, United States of America.

Paulik-Stevenson, J., Odom, AR, York, JD, Molecular and biochemical characterization of two plant inositol polyphosphate 6-/3-/5-kinases, *J. Biol. Chem.*, 277, 42711

(2002).

PCT International Publication WO 93/07278

PCT International Publication WO 93/21335

PCT International Publication WO 94/00977

5 PCT International Publication WO 95/19431

PCT International Publication WO 96/06166

PCT International Publication WO 97/32011

PCT International Publication WO 98/54311

PCT International Publication WO 99/32619

10 PCT International Publication WO 99/53050

PCT International Publication WO 99/61631

PCT International Publication WO 02/059324

Pearson WR & Lipman DJ (1988) Improved Tools for Biological Sequence Comparison.

Proc Natl Acad Sci U S A 85:2444-2448.

15 Phillippy, BQ, Ullah, AH, Ehrlich, KC, Purification and Some Properties of Inositol
1,3,4,5,6-Pentakisphosphate 2-Kinase from Immature Soybean Seeds *J. Biol.*
Chem. 269, 28393 (1994).

Picard D, Salser SJ & Yamamoto KR (1988) A movable and regulable inactivation
function within the steroid binding domain of the glucocorticoid receptor. *Cell* 54:

20 1073-1080.

Potrykus I, Paszkowski J, Saul MW, Petruska J & Shillito RD (1985) Molecular and
general genetics of a hybrid foreign gene introduced into tobacco by direct gene
transfer. *Mol Gen Genet* 199:169-177.

Powell PA, Stark DM, Sanders PR & Beachy RN (1989) Protection against Tobacco
25 Mosaic Virus in Transgenic Plants that Express Tobacco Mosaic Virus Antisense
RNA. *Proc. Natl. Acad. Sci. USA* 86:6949-6952.

Raboy, V., Gerbasi, PF, Young, KA, Stoneberg, SD, Pickett, SG, Bauman AT, Murthy,
PP, Sheridan, WF, Ertl, DS, *Plant Physiol.*, 124, 355-368 (2000).

Reed JN, Privalle LS, Powell MJ, Meghji M, Dawson J, Dunder EM, Suttie J, Wenck A,
30 Launis K, Kramer C, Chang Y, Hansen G & Wright M (2001) Phosphomannose
isomerase: An efficient selectable marker for plant transformation. *In Vitro Cell*
Dev Biol-Plant 37:127-132.

- Redei GP & Koncz C (1992) Classical mutagenesis, in Methods in Arabidopsis Research (Koncz C, Chua N-H & Schell J, eds.) World Scientific Press, River Edge, New Jersey, United States of America, pp. 16-82.
- Reich TJ, Iyer VN & Miki BL (1986) Efficient transformation of alfalfa protoplasts by the
5 intranuclear microinjection of Ti-plasmids. *Bio/Technology* 4:1001-1004.
- Reiter RS, Young RM & Scolnik PA (1992) Genetic Linkage of the *Arabidopsis* Genome: Methods for Mapping with Recombinant Inbreds and Random Amplified Polymorphic DNAs (RAPDs) in Methods in Arabidopsis Research, World Scientific Press, River Edge, New Jersey, United States of America.
- 10 Rogers JC, Dean D & Heck GR (1985) Aleurain: a barley thiol protease closely related to mammalian cathepsin H. *Proc. Natl. Acad. Sci. USA* 82:6512-6516.
- Rohrmeier T & Lehle L (1993) WIP1, a wound-inducible gene from maize with homology to Bowman-Birk proteinase inhibitors. *Plant Mol Biol* 22:783-792.
- Rossolini GM, Cresti S, Ingianni A, Cattani P, Riccio ML & Satta G (1994) Use of
15 deoxyinosine-containing primers vs degenerate primers for polymerase chain reaction based on ambiguous sequence information. *Mol Cell Probes* 8:91-98.
- Roth BA, Goff SA, Klein TM & Fromm ME (1991) C1- and R-dependent expression of the maize Bz1 gene requires sequences with homology to mammalian myb and myc binding sites. *Plant Cell* 3:317-325.
- 20 Rothstein SJ, Lahners KN, Lotstein RJ, Carozzi NB, Jayne SM & Rice DA (1987) Promoter cassettes, antibiotic-resistance genes, and vectors for plant transformation. *Gene* 53:153-161.
- Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B & Bennett J (2002). Proteomic analysis of rice leaves during drought stress and recovery. *Proteomics* 2:1131-
25 1145.
- Sambrook J & Russell DW (2001) Molecular Cloning : A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, United States of America.
- Schmidhauser TJ & Helinski DR (1985) Regions of broad-host-range plasmid RK2
30 involved in replication and stable maintenance in nine species of gram-negative bacteria. *J Bacteriol* 164:446-455.
- Schocher RJ, Shillito RD, Saul MW, Paszkowski J & Potrykus I (1986) Co-

transformation of foreign genes into plants. *Bio/Technology* 4:1093-1096.

Schultz B *et al.* (1998) T-DNA tagging in *Arabidopsis thaliana*: Cloning by gene disruption, in Plant Molecular Biology Manual (2nd edition, Gelvin SB, Schilperoort RA & Verma DPS, eds.) Kluwer Academic Publishers, New York, New York, United States of America.

Scopes R (1982) Protein Purification: Principles and Practice, Springer-Verlag, New York, New York, United States of America.

Shen, X, Xiao, H, Ranallo, R, W-H, Wu, C., Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates *Science* 299, 112 (2003)

Sherman F, Fink GR & Hicks J (1982) Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, United States of America

Shimamoto K, Terada R, Izawa T & Fujimoto H (1989) Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* 338:274-276.

Shinshi H, Neuhas JM, Ryals J & Meins F Jr (1990) Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. *Plant Mol Biol* 14:357-368.

Silhavy TJ, M.L. Berman, and L.W. Enquist (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, United States of America.

Silverstone AL, Ciampaglio CN & Sun T (1998) The *Arabidopsis* RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* 10:155-169.

Singh DP (1986) Breeding for Resistance to Diseases and Insect Pests, Springer-Verlag, New York, New York, United States of America.

Skuzeski JM, Nichols LM & Gesteland RF (1990) Analysis of leaky viral translation termination codons in vivo by transient expression of improved beta-glucuronidase vectors. *Plant Mol Biol* 15:65-79.

Smet FD, Mathys J, Marchal K, Thijs G, Moor BD & Moreau Y (2002) Adaptive Quality-based clustering of gene expression profiles. *Bioinformatics* 18:735-746.

Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG & Waterhouse PM (2000) Total silencing by intron-spliced hairpin RNAs. *Nature* 407:319-320.

- Smith TF & Waterman M (1981) Comparison of Biosequences. *Adv Appl Math* 2:482-489.
- Song R, Llaca V & Messing J (2001) Mosaic organization of orthologous sequences in grass genomes. *Genome Res* 12:1549-1555.
- 5 Spencer TM, Gordon-Kamm WJ, Daines RJ, Start W & Lemaux P (1990). *Theor Appl Genet* 79:625-631.
- Stewart JM & Young JD (1984) Solid Phase Peptide Synthesis, 2nd ed. Pierce Chemical Co., Rockford, Illinois, United States of America.
- Svab Z, Hajdukiewicz P & Maliga P (1990) Stable transformation of plastids in higher
10 plants. *Proc Natl Acad Sci USA* 87:8526-8530.
- Svab Z & Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc Natl Acad Sci USA* 90:913-917.
- Taylor M, Vasil V & Vasil IK (1993) Enhanced gus gene expression in cereal grass cell
15 suspensions and immature embryos using the maize ubiquitin-based plasmid pAHC25. *Plant Cell Rep* 12:491-495.
- Thompson CJ, Movva NR, Tizard R, Crameri R, Davies JE, Lauwereys M & Botterman J (1987) Characterization of the herbicide-resistance gene bar from *Streptomyces hygroscopicus*. *EMBO J* 6:2519-2523.
- Tijssen P (1993) Laboratory Techniques in Biochemistry and Molecular Biology-
20 Hybridization with Nucleic Acid Probes. Elsevier, New York, New York, United States of America.
- Triezenberg SJ, Kingsbury RC & McKnight SL (1988) Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev* 2:718-729.
- 25 Uknes S, Dincher S, Friedrich L, Negrotto D, Williams S, Thompson-Taylor H, Potter S, Ward E & Ryals J (1993) Regulation of pathogenesis-related protein-1a gene expression in tobacco. *Plant Cell* 5:159-169.
- Uknes S, Mauch-Mani B, Moyer M, Potter S, Williams S, Dincher S, Chandler D, Slusarenko A, Ward E & Ryals J (1992) Acquired resistance in *Arabidopsis*.
30 *Plant Cell* 4:645-656.
- Unger EA, Hand JM, Cashmore AR & Vasconcelos AC (1989) Isolation of a cDNA encoding mitochondrial citrate synthase from *Arabidopsis thaliana*. *Plant Mol Biol*

13:411-418.

U.S. Patent No. 4,554,101

U.S. Patent No. 4,940,935

U.S. Patent No. 4,945,050

5 U.S. Patent No. 4,987,071

U.S. Patent No. 5,036,006

U.S. Patent No. 5,100,792

U.S. Patent No. 5,188,642

U.S. Patent No. 5,270,163

10 U.S. Patent No. 5,466,785

U.S. Patent No. 5,501,967

U.S. Patent No. 5,523,311

U.S. Patent No. 5,591,616

U.S. Patent No. 5,614,395

15 U.S. Patent No. 5,639,949

U.S. Patent No. 5,767,378

U.S. Patent No. 5,994,629

U.S. Patent No. 6,369,298

U.S. Patent Publication No. 2003/30009011

20 Van den Broeck G, Timko MP, Kausch AP, Cashmore AR, Van Montagu M & Herrera-Estrella L (1985) Targeting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase. *Nature* 313:358-363.

25 van der Krol A *et al.*, In: Antisense nucleic acids and proteins (Joseph M & van der Krol A, eds.) Marcel Dekker Inc, New York, New York, United States of America, pp. 125-141.

Vasil V, Castillo AM, Fromm ME & Vasil IK (1992) Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. *Bio/Technology* 10:667-674.

30 Vasil V, Srivastava V, Castillo AM, Fromm MR & Vasil IK (1993) Rapid production of transgenic plants by direct bombardment of cultured immature embryos. *Bio/Technology* 11:1553-1558.

- Warner SA, Scott R, Draper J (1993) Isolation of an asparagus intracellular PR gene (AoPR1) wound-responsive promoter by the inverse polymerase chain reaction and its characterization in transgenic tobacco. *Plant J* 3:191-201.
- Wasmann CC *et al.* (1986) *Mol Gen Genet* 205:446-453.
- 5 Waterhouse PM, Graham MW & Wang MB (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci U S A* 95:13959-13964.
- Weeks JT, Anderson OD & Blechl AE (1993) Rapid Production of Multiple Independent Lines of Fertile Transgenic Wheat (*Triticum aestivum*). *Plant Physiol* 102:1077-1084.
- 10 Welsh JR (1981) Fundamentals of Plant Genetics and Breeding, John Wiley & Sons, New York, New York, United States of America.
- White J, Chang SY & Bibb MJ (1990) A cassette containing the bar gene of *Streptomyces hygroscopicus*: a selectable marker for plant transformation. *Nucl Acids Res* 18:1062.
- 15 Wilcox, JR., Premachandra, GS, Young, KS, Sebastian, SA, *Plant Physiol.* 128, 650 (2002)
- Wilcox, JR, Premachandra, GS, Young, KS, Raboy, V. Isolation of High Seed Inorganic P, Low-Phytate Soybean Mutants, *Crop Sci.* 40, 1601 (2000).
- 20 Wilson, MP, Majerus, PW, Characterization of a cDNA encoding Arabidopsis thaliana inositol 1,3,4-trisphosphate 5/6-kinase, *Biochem. Biophys. Res. Commun.*, 232, 678-681 (1997).
- Wilson, MP, Majerus, PW, Isolation of Inositol 1,3,4-Trisphosphate 5/6-Kinase, cDNA Cloning, and Expression of the Recombinant Enzyme *J. Biol. Chem.* 271,11904
- 25 (1996).
- Winkler RG & Feldmann KA (1989) PCR-based identification of T-DNA insertion mutants. *Meth Mol Biol* 82:129-136.
- Wood DR (ed.) (1983) Crop Breeding, American Society of Agronomy, Madison, Wisconsin, United States of America.
- 30 Wricke G & Weber WE (1986) Quantitative Genetics and Selection Plant Breeding, Walter de Gruyter and Co., Berlin, Germany.
- Xu D, McElroy D, Thornburg RW & Wu R (1993) Systemic induction of a potato pin2

promoter by wounding, methyl jasmonate, and abscisic acid in transgenic rice plants. *Plant Mol Biol* 22:573-588.

York, JD, Odom, AR, Murphy, R, Ives, EB, Went, SR, A Phospholipase C-Dependent Inositol Polyphosphate Kinase Pathway Required for Efficient Messenger RNA Export, *Science* 285, 96 (1999).

Zhang HM, Yang H, Rech EL, Golds TJ, Davis AS, Mulligan BJ, Cocking EC & Davey MR (1988) Transgenic rice plants produced by electroporation-mediated plasmid uptake into protoplasts. *Plant Cell Reports* 7: 379-384.

Zhao ZY, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J & Pierce D (2000) Agrobacterium-mediated sorghum transformation. *Plant Mol Biol* 44:789-98.

Zhu T, Peterson DJ, Tagliani L, St Clair G, Baszczyński CL & Bowen B (1999) Targeted manipulation of maize genes in vivo using chimeric RNA/DNA oligonucleotides. *Proc Natl Acad Sci U S A* 96:8768-8773.

It will be understood that various details of the presently disclosed subject matter may be changed without departing from the scope of the presently disclosed subject matter. Furthermore, the foregoing description is for the purpose of illustration only, and not for the purpose of limitation.

CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule encoding an inositol phosphate kinase polypeptide, wherein the nucleic acid molecule is selected from the group consisting of:

- 5 (a) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of one of even numbered SEQ ID NOs: 2-10 or a polypeptide at least 40% identical to even numbered SEQ ID NOs: 2-10 and having inositol phosphate kinase activity;
- (b) a nucleic acid molecule comprising a nucleic acid sequence of one of odd
10 numbered SEQ ID NOs:1-9;
- (c) a nucleic acid molecule that has a nucleic acid sequence at least 90% identical to the nucleic acid sequence of the nucleic acid molecule of (a) or (b);
- (d) a nucleic acid molecule that hybridizes to (a) or (b) under stringent
15 hybridization conditions;
- (e) a nucleic acid molecule comprising a nucleic acid sequence complementary to (a); and
- (f) a nucleic acid molecule comprising a nucleic acid sequence that is the full reverse complement of (a).

20 2. A vector, comprising the nucleic acid of claim 1.

3. An expression cassette, comprising the nucleic acid of claim 1 operably linked to a promoter.

4. The expression cassette of claim 3, wherein the promoter is a plant promoter.

5. The expression cassette of claim 3, wherein the promoter is a constitutive
25 promoter.

6. The expression cassette of claim 3, wherein the promoter is a tissue-specific or a cell type-specific promoter.

7. The expression cassette of claim 6, wherein the tissue-specific or cell type-specific promoter directs expression of the expression cassette in a location selected
30 from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, seed, and combinations thereof.

8. The expression cassette of claim 7, wherein the tissue-specific or cell type-

specific promoter directs expression of the expression cassette in seed.

9. An isolated inositol phosphate kinase polypeptide having inositol phosphate kinase activity selected from the group consisting of:

- (j) an amino acid sequence of one of even numbered SEQ ID NOs: 2-10;
- (k) an amino acid sequence that is at least 40% identical to (a);
- (l) an amino acid sequence encoded by a nucleotide sequence substantially identical to a nucleotide sequence of one of odd numbered SEQ ID NOs: 1-9; and
- (m) an amino acid sequence encoded by a nucleic acid molecule capable of hybridizing under stringent conditions to a nucleic acid molecule of one of odd numbered SEQ ID NOs: 1-9 or to a sequence fully complementary thereto.

10. A method for producing a polypeptide of claim 9, comprising the steps of:

- (a) growing cells comprising an expression cassette under suitable growth conditions, the expression cassette comprising a nucleic acid molecule of claim 1; and
- (b) isolating the polypeptide from the cells.

11. A transgenic plant cell comprising a homozygous disruption in at least one endogenous inositol phosphate kinase gene homologous to the nucleic acid molecules of claim 1, wherein the disruption substantially inhibits the expression of a functional inositol phosphate polypeptide.

12. A transgenic plant comprising at least one of the transgenic plant cells of claim 11.

13. The transgenic plant of claim 12, wherein the plant comprises a decreased level of phytic acid when compared to a non-transformed parental plant.

14. The transgenic plant of claim 12, wherein the plant comprises an increased level of non-phytic acid phosphorous when compared to a non-transformed parental plant.

15. The transgenic plant of claim 12, wherein the homozygous disruption is tissue-specific.

16. The transgenic plant of claim 15, wherein the tissue is selected from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith,

leaf, flower, seed, and combinations thereof.

17. The transgenic plant of claim 16, wherein the tissue is seed.

18. The transgenic plant of claim 17, wherein the seed comprises a decreased level of phytic acid when compared to a seed from a non-transformed parental plant.

5 19. The transgenic plant of claim 17, wherein the seed comprises an increased level of non-phytic acid phosphorous when compared to a seed from a non-transformed parental plant.

20. The transgenic plant of claim 12, wherein the plant is selected from the group consisting of *Arabidopsis thaliana*, corn (*Zea mays*), *Brassica* sp., alfalfa (*Medicago sativa*), rice (*Oryza sativa* ssp.), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*),
10 peanut (*Arachis hypogaea*), cotton, sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea utilane*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*),
15 cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, duckweed (*Lemna*), and barley.

21. The transgenic plant of claim 20, wherein the plant is *Arabidopsis thaliana*.

22. The transgenic plant of claim 12, wherein the plant is selected from the group
25 consisting of a vegetable, an ornamental, and a conifer.

23. The transgenic plant of claim 22, wherein the vegetable is selected from the group consisting of tomatoes, lettuce, guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, green bean, lima bean, pea, and members of the genus *Cucumis*.

30 24. The transgenic plant of claim 22, wherein the ornamental is selected from the group consisting of impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia,

Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, azalea, hydrangea, hibiscus, rose, tulip, daffodil, petunia, carnation, poinsettia, and chrysanthemum.

5 25. The transgenic plant of claim 22, wherein the conifer is selected from the group consisting of loblolly pine, slash pine, ponderosa pine, lodgepole pine, Monterey pine, Douglas-fir, Western hemlock, Sitka spruce, redwood, silver fir, balsam fir, Western red cedar, and Alaska yellow-cedar.

10 26. The transgenic plant of claim 11, wherein the homozygous disruption to the at least one endogenous inositol phosphate kinase gene is to at least a gene encoding an lpk1 polypeptide and a gene encoding an lpk2 polypeptide

 27. A method of modulating production of phytate in a plant, comprising modulating the enzymatic activity of at least one inositol phosphate kinase polypeptide selected from the group consisting of lpk1 and lpk2.

15 28. The method of claim 27, wherein modulating the enzymatic activity comprises reducing the enzymatic activity of the at least one inositol phosphate kinase polypeptide.

20 29. The method of claim 28, wherein reducing the enzymatic activity of the at least one inositol phosphate kinase polypeptide comprises substantially inhibiting the expression of a functional form of the at least one inositol phosphate kinase polypeptide.

 30. The method of claim 29, wherein substantially inhibiting the expression of the functional form of the at least one inositol phosphate kinase polypeptide results in at least about a 90% reduction in the phytate content of the plant.

25 31. The method of claim 29, wherein substantially inhibiting the expression of the functional inositol phosphate polypeptide comprises homozygous disruption of an endogenous gene encoding the at least one inositol phosphate kinase polypeptide.

30 32. The method of claim 31, wherein disrupting the endogenous inositol phosphate kinase gene comprises stably inserting a foreign DNA element into the endogenous inositol phosphate kinase gene.

 33. The method of claim 31, wherein the foreign DNA element is T-DNA.

 34. The method of claim 29, wherein substantially inhibiting the expression of the

functional inositol phosphate polypeptide comprises post-transcriptionally silencing expression of the at least one inositol phosphate polypeptide.

35. The method of claim 34, wherein post-transcriptionally silencing expression of the functional inositol phosphate polypeptide comprises introducing a ribonucleic acid (RNA) into at least one cell of the plant in an amount sufficient to inhibit expression of the at least one inositol phosphate polypeptide, wherein the RNA comprises a ribonucleotide sequence which corresponds to a coding strand of a gene encoding the at least one inositol phosphate polypeptide.

36. The method of claim 34, wherein post-transcriptionally silencing expression of the functional inositol phosphate polypeptide comprises RNAi.

37. The method of claim 27, wherein the modulating the enzymatic activity of the at least one inositol phosphate polypeptide is a tissue-specific modulating.

38. The method of claim 37, wherein the tissue specifically modulated is selected from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, seed, and combinations thereof.

39. The method of claim 38, wherein the tissue specifically modulated is seed.

40. The method of claim 27, wherein the at least one inositol phosphate polypeptide is lpk-1 and 2.

41. A method of producing a plant with low levels of phytate comprising modulating in the plant the enzymatic activity of at least one inositol phosphate kinase polypeptide selected from the group consisting of lpk1 and lpk2.

42. The method of claim 41, wherein modulating the enzymatic activity comprises reducing the enzymatic activity of the at least one inositol phosphate kinase polypeptide.

43. The method of claim 42, wherein reducing the enzymatic activity of the at least one inositol phosphate kinase polypeptide comprises substantially inhibiting the expression of a functional form of the at least one inositol phosphate kinase polypeptide.

44. The method of claim 43, wherein substantially inhibiting the expression of the functional form of the at least one inositol phosphate kinase polypeptide results in at least about a 90% reduction in the phytate content of the plant.

45. The method of claim 44, wherein substantially inhibiting the expression of the

functional inositol phosphate polypeptide comprises homozygous disruption of an endogenous gene encoding the at least one inositol phosphate kinase polypeptide.

46. The method of claim 45, wherein disrupting the endogenous inositol phosphate kinase gene comprises stably inserting a foreign DNA element into the
5 endogenous inositol phosphate kinase gene.

47. The method of claim 45, wherein the foreign DNA element is T-DNA.

48. The method of claim 43, wherein substantially inhibiting the expression of the functional inositol phosphate polypeptide comprises post-transcriptionally silencing expression of the at least one inositol phosphate polypeptide.

49. The method of claim 48, wherein post-transcriptionally silencing expression of the functional inositol phosphate polypeptide comprises introducing a ribonucleic acid (RNA) into at least one cell of the plant in an amount sufficient to inhibit expression of the at least one inositol phosphate polypeptide, wherein the RNA comprises a ribonucleotide sequence which corresponds to a coding strand of a gene encoding the
15 at least one inositol phosphate polypeptide.

50. The method of claim 48, wherein post-transcriptionally silencing expression of the functional inositol phosphate polypeptide comprises RNAi.

51. The method of claim 41, wherein modulating the enzymatic activity of the at least one inositol phosphate polypeptide is a tissue-specific modulating.

52. The method of claim 51, wherein the tissue specifically modulated is selected from the group consisting of epidermis, root, vascular tissue, meristem, cambium, cortex, pith, leaf, flower, seed, and combinations thereof.

53. The method of claim 52, wherein the tissue specifically modulated is seed.

54. The method of claim 41, wherein the at least one inositol phosphate polypeptide is lpk-1 and 2.
25

55. A method of screening a plurality of compounds for a modulator of the enzymatic activity of an inositol phosphate kinase polypeptide, the method comprising:

- (k) providing a library of test compounds;
- (l) contacting an inositol phosphate kinase polypeptide selected from
30 the group consisting of lpk1 and lpk2 with each test compound;
- (m) detecting an interaction between a test compound and the inositol phosphate kinase polypeptide;

- (n) identifying a test compound that interacts with the inositol phosphate kinase polypeptide; and
- (o) isolating the test compound that interacts with the inositol phosphate kinase polypeptide, whereby a plurality of compounds is screened for a modulator of inositol phosphate kinase polypeptide enzymatic activity.

5

56. The method of claim 55, wherein the test compounds are bound to a substrate.

57. The method of claim 56, wherein the test compounds are synthesized directly

10 on the substrate.

58. The method of claim 55, wherein the modulator of the inositol phosphate kinase polypeptide enzymatic activity is an antagonist of the inositol phosphate kinase polypeptide enzymatic activity.

59. An isolated polypeptide having or comprising at least about 90% amino acid identity with a polypeptide as disclosed in Examples 6-11, over an entire sequence or functional fragment thereof, wherein the polypeptide functions as an inositol phosphate kinase.

15

60. An isolated polynucleotide encoding a polypeptide of claim 59.

1

Figure 1

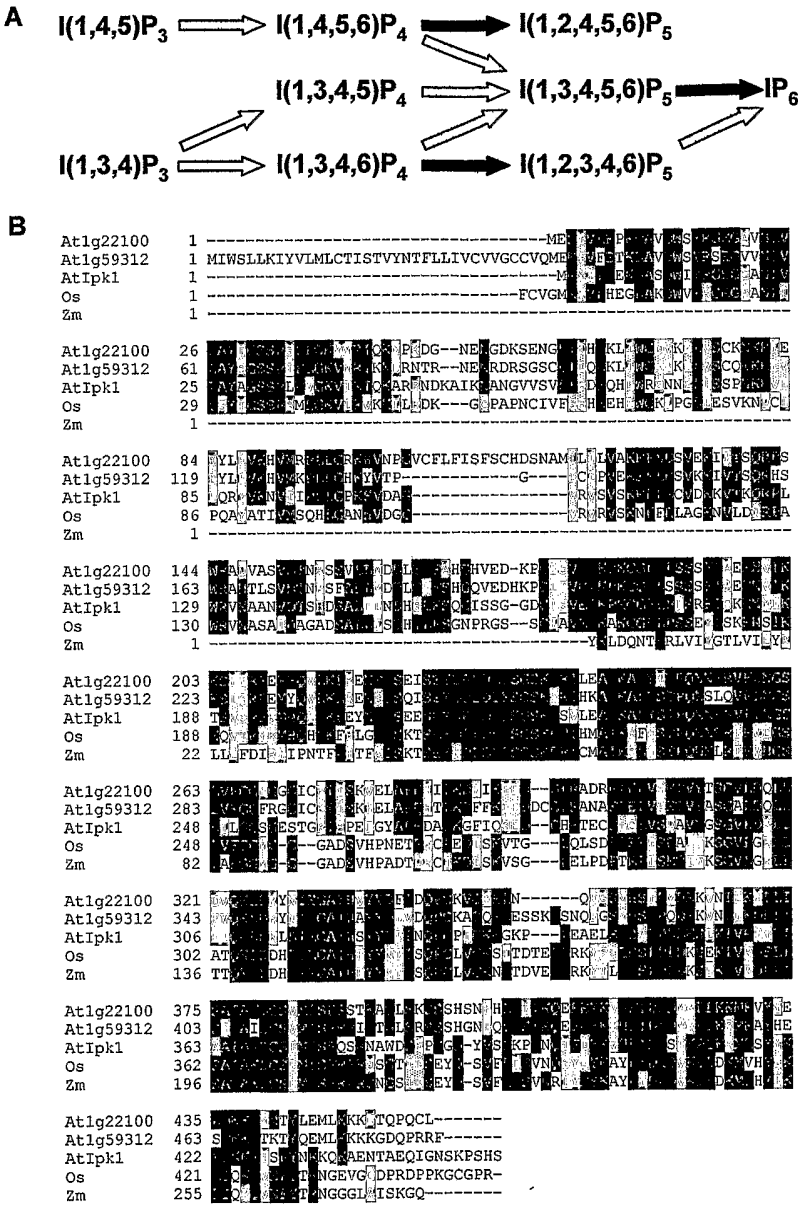


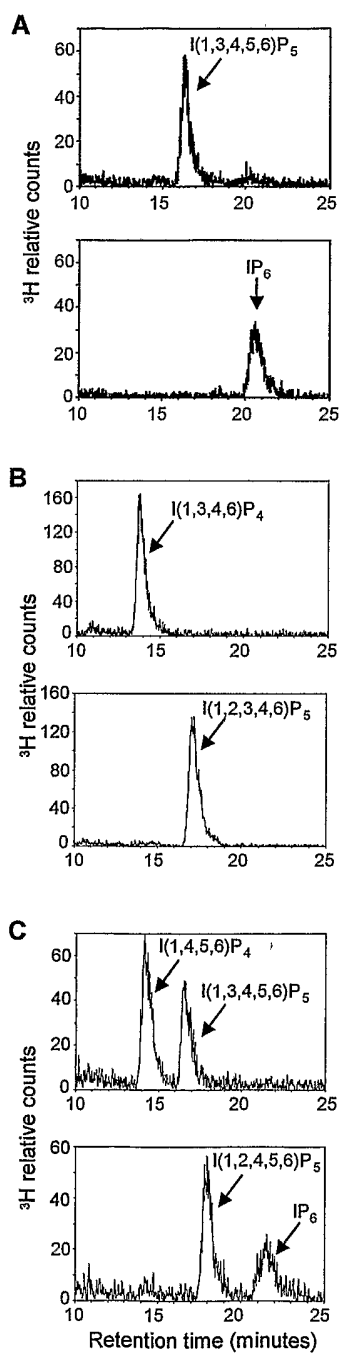
Figure 2

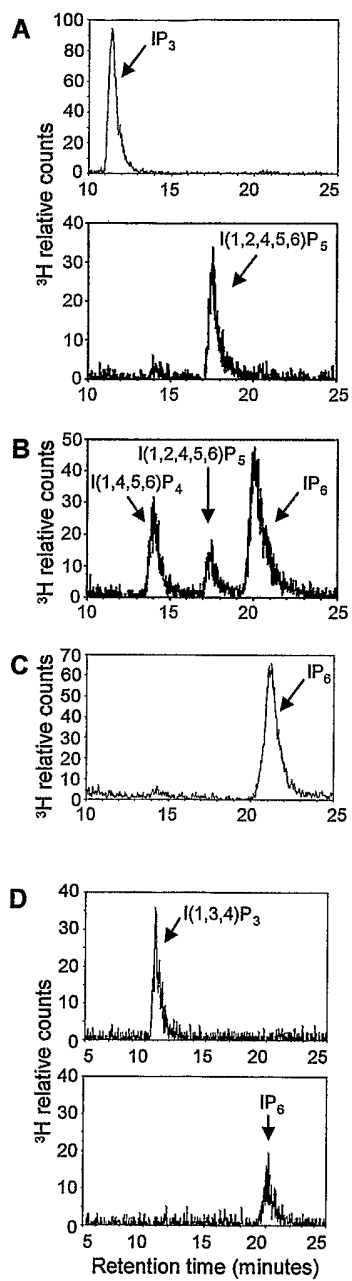
Figure 3

Figure 4



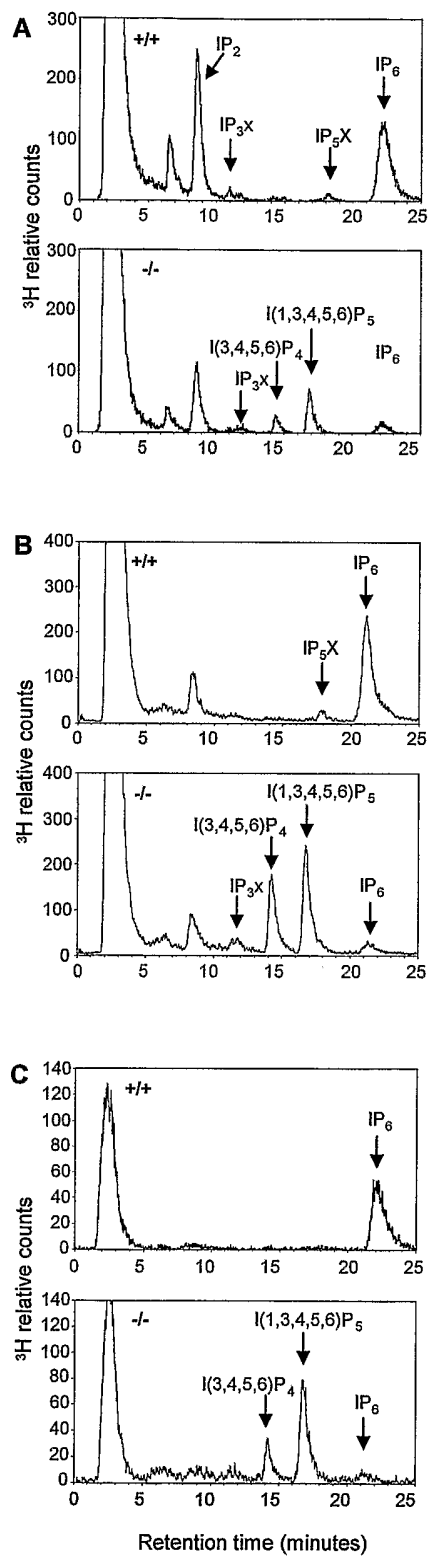
Figure 5

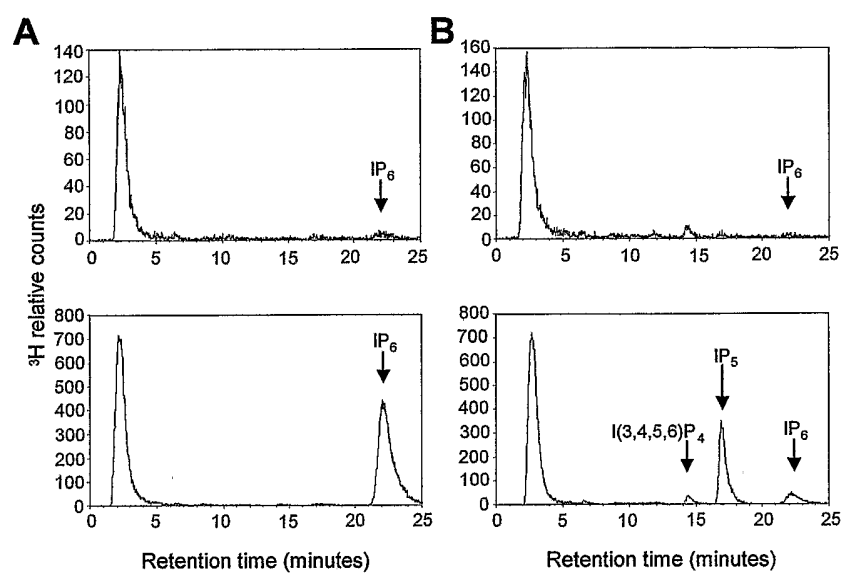
Figure 6

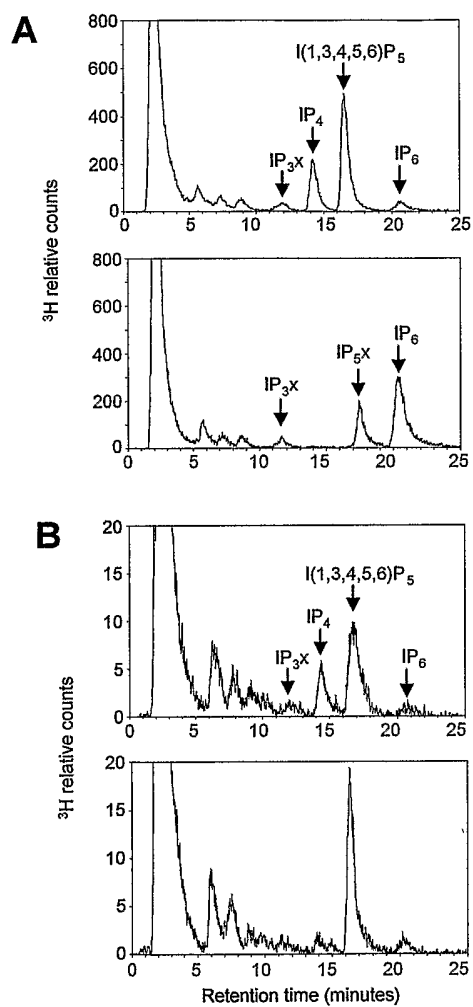
Figure 7

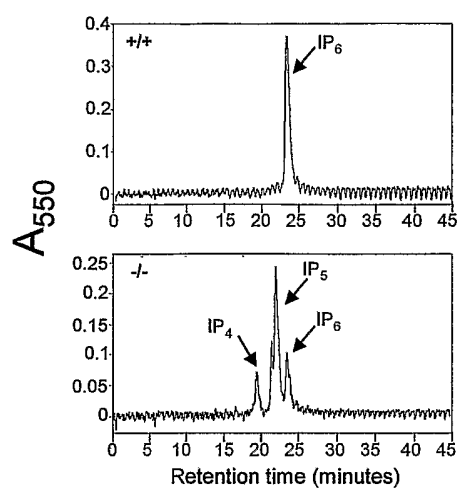
Figure 8

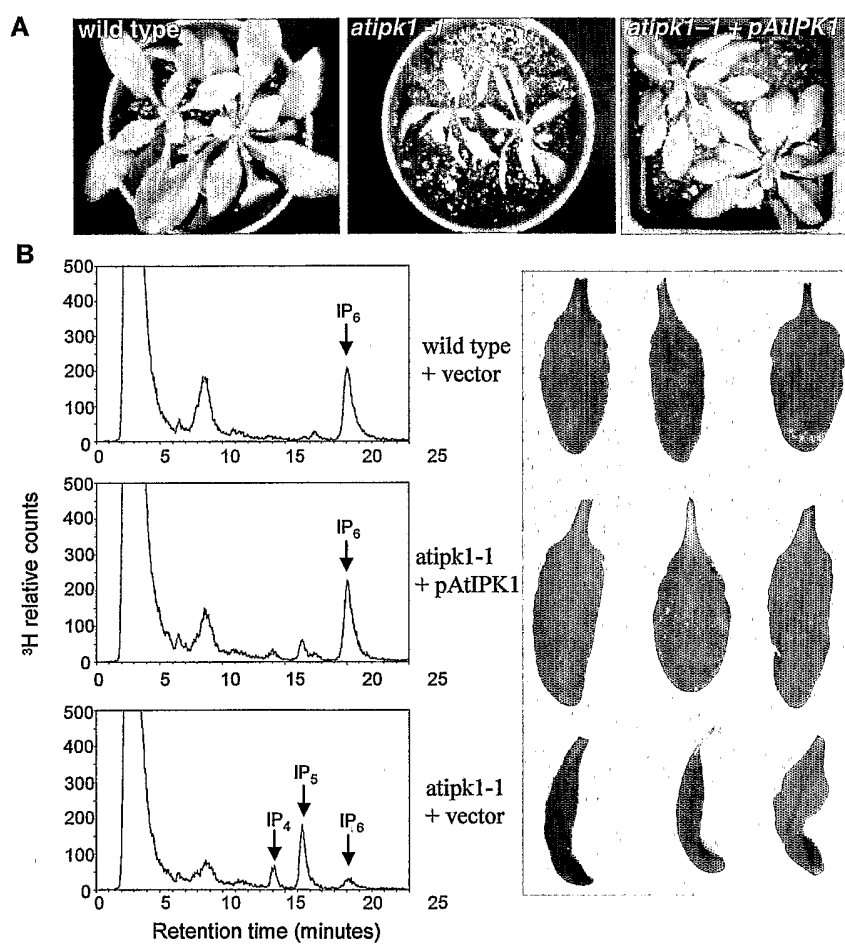
Figure 9

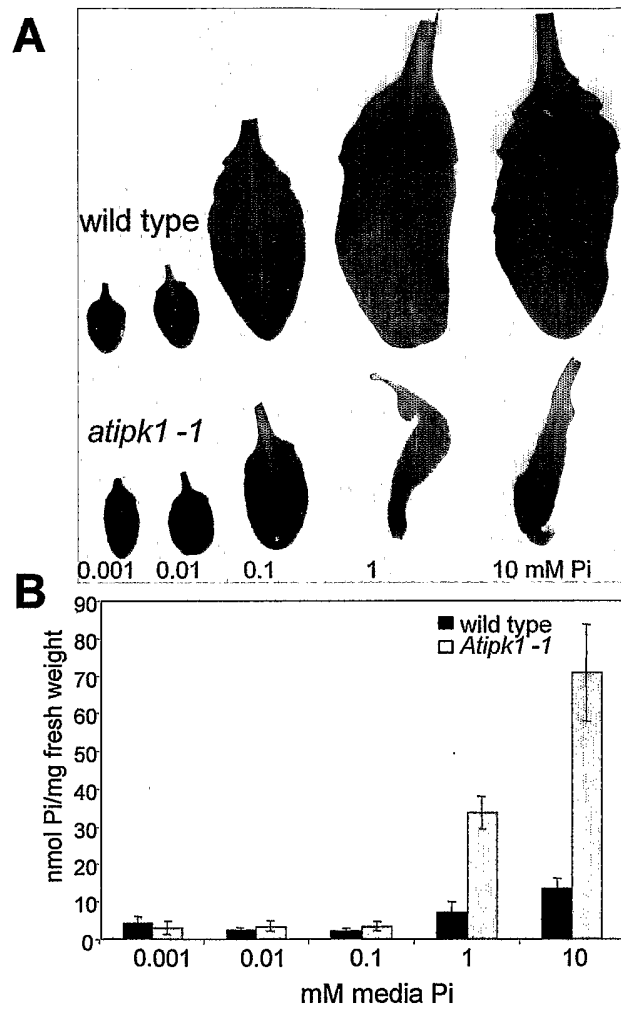
Figure 10

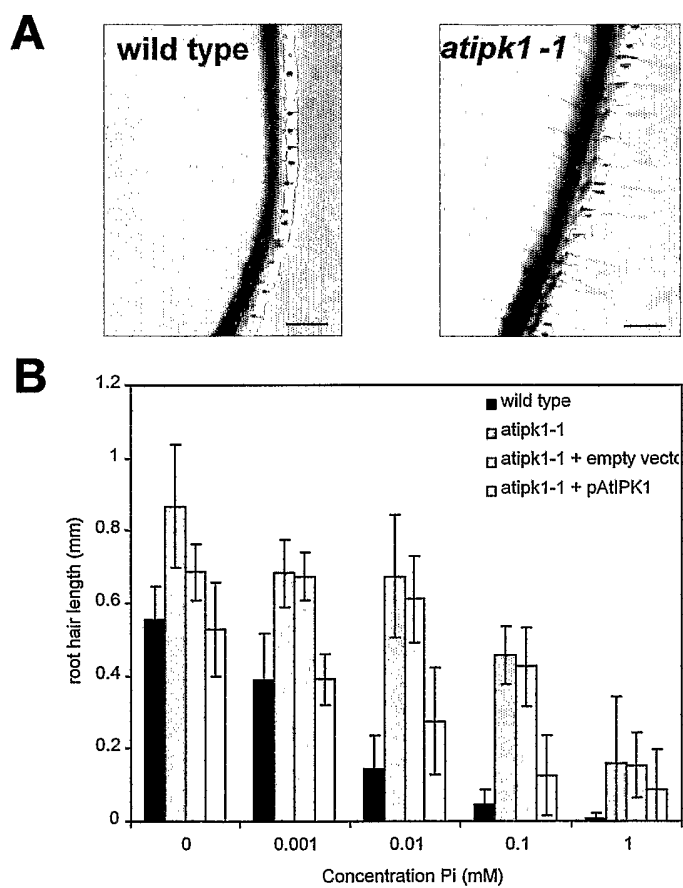
Figure 11

Figure 12

T-DNA insertional disruption of *AtIPK2 β* expression

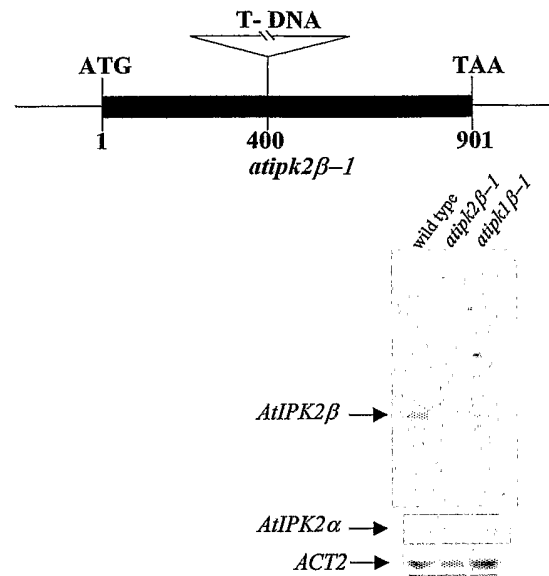
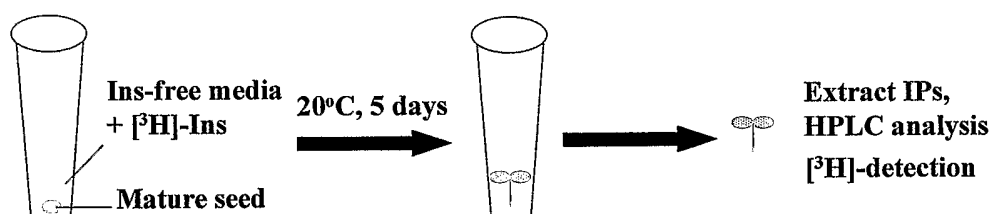


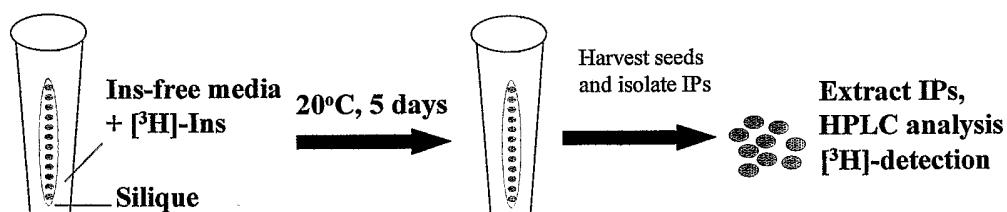
Figure 13

Inositol phosphate measurements

In vivo labeling of seedlings:



In vivo labeling of silique/seeds:



Non-radioactive mass measurements of mature seeds:

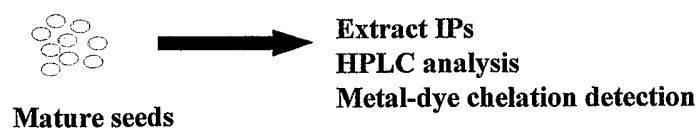


Figure 14

IP kinase mutants have low-phytate seeds

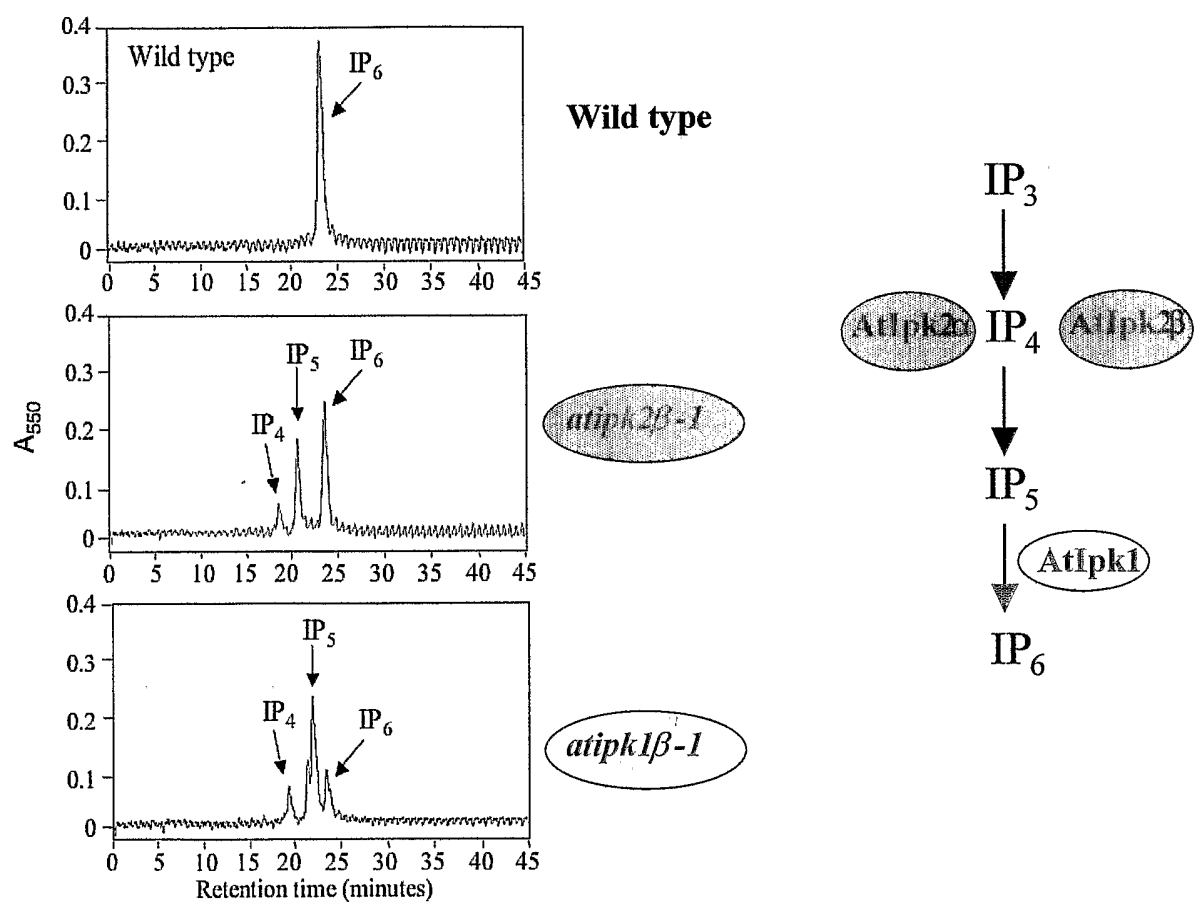


Figure 15

ipk1-1 + pIPK1 1



ipk1-1 + pIPK1 1



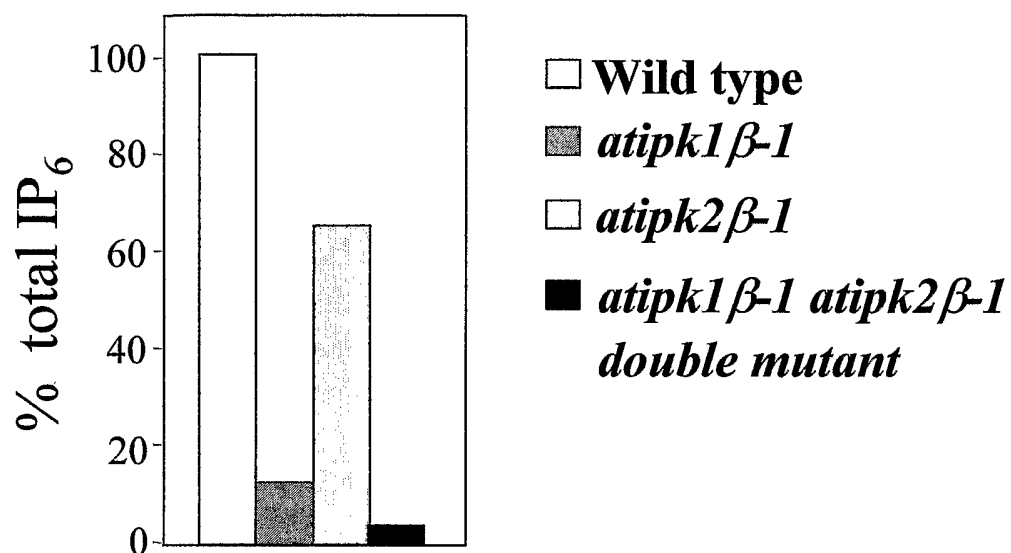
ipk1-1 ipk2β-1 (9B)



ipk1-1 ipk2β-1 (11B)



Figure 16



N=10 for each with less than 10% deviation
Mean values graphed

Figure 17

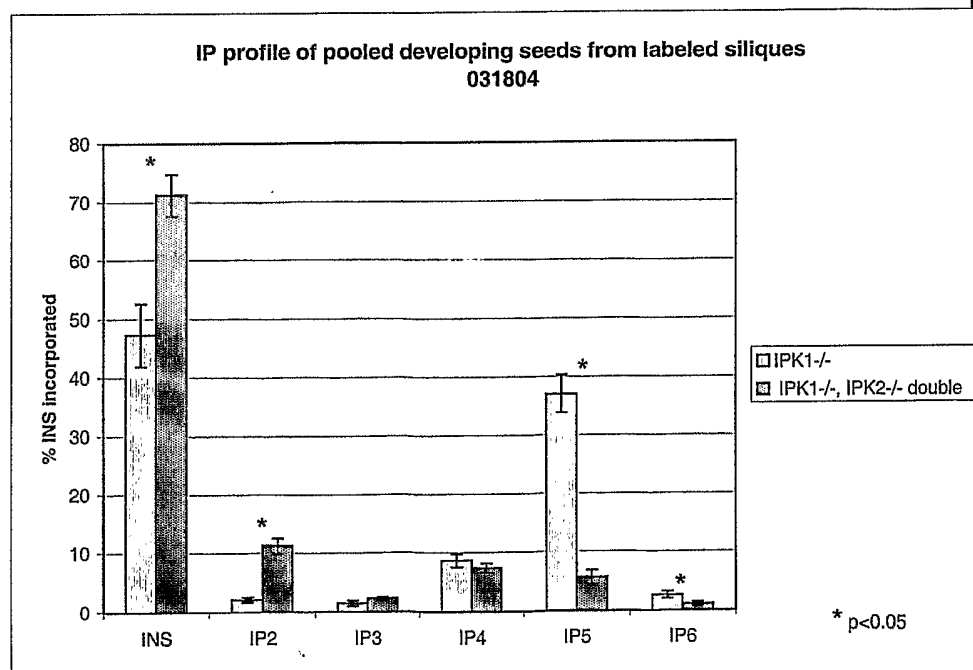
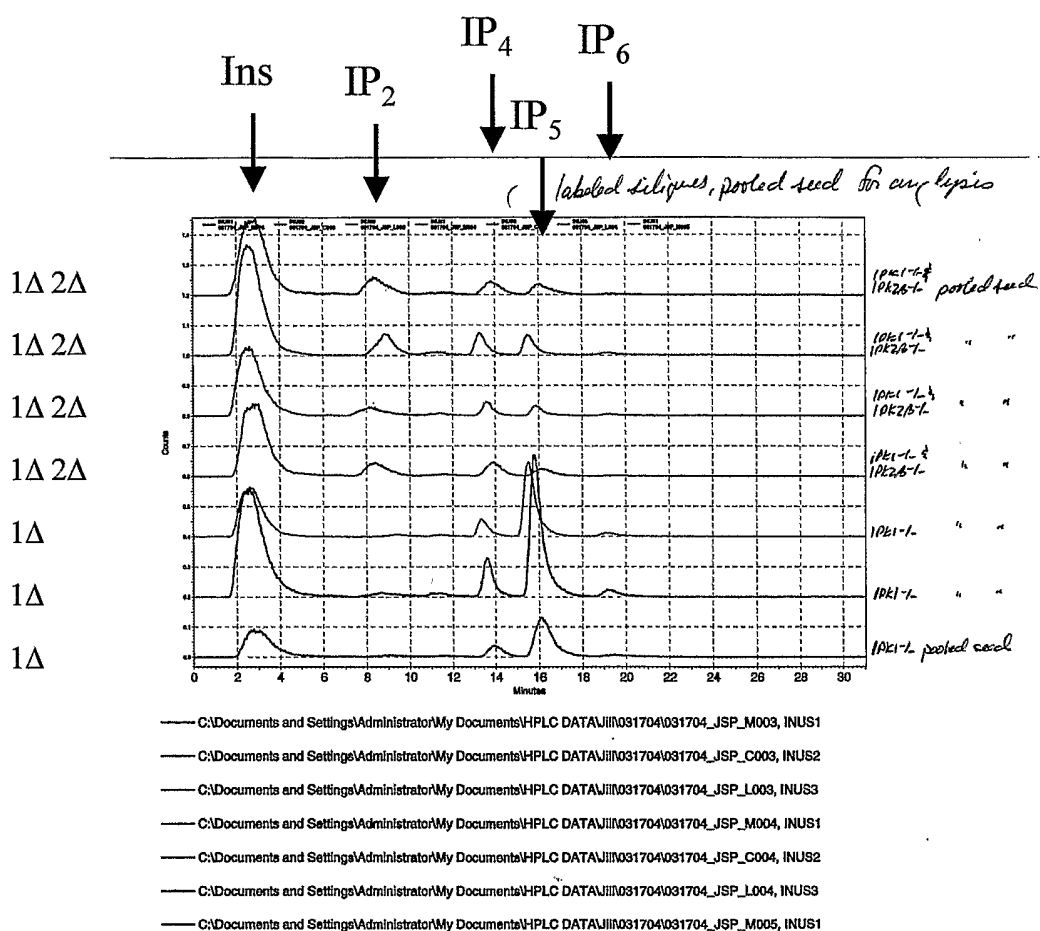


Figure 18

Query:	2	LKVP	EHQV	AGHIA	SDGK	LGPL	VDDQ	GRFF	KPLQ	GDSR	GEHE	AKFY	ESFT	SNMK	VPDHI	-	60
		:						:	:				:	:		:	
Sbjct:	13	LRP	EHQV	AGH	RAS	ADKL	GPLV	DGEG	LFYK	PLQ	AGER	GEHE	AAFY	AAFT	AHP	AVP	PRVRG 192
Query:	61	RYFP	VYHG	TQL	VEA	--	SDGS	GKLPH	LVLD	DDV	VSGY	ANPS	MDVK	IGSR	TWYP	DPV	SEEFYK 118
		:		:	:	:			:	:	:			:			:
Sbjct:	193	AF	PRFH	GTR	FLP	APAS	PGG	APY	PHI	VLDL	LAGL	PSPC	VADV	KIGACT	WPPR	SPDP	YVA 372
Query:	119	KCIK	KDRQ	TTV	SLG	FRV	SGFK	IFDH	QESS	FWRA	EKKL	VLGN	ADGA	RALR	KFFV	SSNSP 178	
		:	:	:	:	:	:	:		:	:	:		:		:	:
Sbjct:	373	KCL	AKDR	ETTS	ALL	GFRV	SGVR	VVD	ARG	GA	VWRP	DRSEL	KGIDA	AGVRR	VLRRY	VSTGG~ 549	
Query:	179	ADSN	LTPN	CAFA	SEVY	GGC	NGILA	QLLE	LKDW	FETQ	TLYH	FN	SCS	ILMI	YE-	NESIL	MQG 237
			:		:		:								:		:
Sbjct:	550	GDGL	---	DCAL	AAAV	YGEG	GVLA	QLREL	KAWFEE	QTLYH	FYSAS	ILFGY	DANAAAA	AAAAAP 720			
Query:	238	GDDA	PAPRA	QVK	LVDF	AHV	LDG	NGV	IDHN	FLG	GLCS	FIKFI	KDIL	QSV	EKH	DET	DTS 294
			:			:		:						:		:	
Sbjct:	721	GGSG	GGVR	--	VKL	VDF	AHV	DDG	DGVI	DHN	FLG	GLCS	LKFI	GDI	VAE	VT	EKASSDHS 885

A

Figure 20

Query: 1	MLKVP	EHQV	AGH	IASD	GKLG	PLVD	DDQGR	FFKPLQ	GD	SRGE	HEAK	FYES	FTSN	MKV	PD	HH	60																																						
Sbjct: 20251	MLKVP	EHQV	AGH	IASD	GKLG	PLVD	DDQGR	FFKPLQ	GD	SRGE	HEAK	FYES	FTSN	MKV	PD	HH	20072																																						
Query: 61	RYFPV	YHGT	QLV	EASD	GS	GKLP	HLVL	DDV	VSGY	ANPS	VDVK	IGSR	TWYP	DVSE	EYFK	KC	120																																						
Sbjct: 20071	RYFPV	YHGT	QLV	EASD	GS	GKLP	HLVL	DDV	VSGY	ANPS	VDVK	IGSR	TWYP	DVSE	EYFK	KC	19892																																						
Query: 121	IKKDR	QTTT	VS	LG	FR	VS	GEK	IFD	HQ	ESS	FW	RAE	KKLV	LG	YN	ADG	ARL	AR	KFV	SS	NS	PAD	180																																
Sbjct: 19891	IKKDR	QTTT	VS	LG	FR	VS	GEK	IFD	HQ	ESS	FW	RAE	KKLV	LG	YN	ADG	ARL	AR	KFV	SS	NS	PAD	19712																																
Query: 181	SNLTP	NCA	F	A	S	E	V	Y	G	C	N	G	I	L	A	Q	L	L	E	L	K	D	W	F	E	T	Q	T	L	Y	H	F	N	S	C	S	I	L	M	I	Y	E	N	E	S	I	L	M	Q	G	G	D	240		
Sbjct: 19711	SNLTP	NCA	F	A	S	E	V	Y	G	C	N	G	I	L	A	Q	L	L	E	L	K	D	W	F	E	T	Q	T	L	Y	H	F	N	S	C	S	I	L	M	I	Y	E	N	E	S	I	L	M	Q	G	G	D	19532		
Query: 241	APAPR	AQ	V	K	L	V	D	E	A	H	V	L	D	G	N	G	V	I	D	H	N	F	L	G	L	C	S	F	I	K	F	I	K	D	I	L	Q	S	V	E	K	H	D	E	T	D	T	S	L	L	E	N	G	R	300
Sbjct: 19531	APAPR	AQ	V	K	L	V	D	E	A	H	V	L	D	G	N	G	V	I	D	H	N	F	L	G	L	C	S	F	I	K	F	I	K	D	I	L	Q	S	V	E	K	H	D	E	T	D	T	S	L	L	E	N	G	R	19352

B

Figure 20 (cont'd)

Query: 2 LKVPEHQVAGHIASDGKLGPLVDDQGRFFKPLQGDSRGEHEAKFYESFTSNMKVPDHIHR 61
||||| 61
Sbjct: 46426 LKVPEHQVAGHIAKDGPGLVDDKGRFFKPLQGDSRGEIEVKFYESFSSNTEVPEHIHR 46605
61
Query: 62 YFPVYHGTQLVEASDGSGLPHLVLDVVSGYANPSVMDVKIGSRTWYPDVSEYFKKCI 121
||||| 121
Sbjct: 46606 YFPVYHGTQAVEGSDGAAMM---VLENLLAEYTKPSVMDVKMGSRTWYPDASEEYIQKCI 46776
121
Query: 122 KKDRQTTTVSLGFRVSGFKIFDHQESSFWRAEKKLVLGYNADGARIALRKFFVSSNSPADS 181
||| 181
Sbjct: 46777 KKDTGTTTVSSGFRISGFEVYDHKESSEFWKPERKLLRGLDVGARLTLRKFFVSSNSLSDT 46956
181
Query: 182 NLTPNCAFASEVYGGCNGILAQLLLELKDWFETQTLYHFNCSILMIYENESILMQGGDDA 241
|: ||||| 241
Sbjct: 46957 GSKPDSAFASSVYGGSHGILTQELLEKTWFENQTLYHFNCSILMVYENESILKGNDDDA 47136
241
Query: 242 PAPRAQVKLVDFAHVLDGNGVIDHNFGLGCSFIKFIKDILQSVK 287
| ||||| 287
Sbjct: 47137 ---RPQVKLVDFAHVLDGNGVIDHNFGLGCSFINFIREILQSPDE 47265
287

C

Query: 2 LKVPEHQVAGHIASDGKLGPLVDDQGRFFKPLQGDSRGEHEAKFY 46
|: ||||| 46
Sbjct: 67350 IRPPEHQVAGNYASTDKLGLVP---RVYKILQAGERGEHEAAFF 67225
46

A

Figure 21

Query:	3	LKVPEHQVAGHIAKDCKPGPLVDDKGRFFKPLQDGRGEIEVKFYEFSSNTEVPEHIH-	61
Sbjct:	1933	LHPPEHQVAGHRASASKLGPLIDGSGLFYKPLQAGDRGEHEVAFYEAFAHAAVPARIRD	1754
Query:	62	RYFPVYHGTQAV--EGSDGAAM--MVLENLLAEYTKPSVMDVKMGSRWTWYYPDASEEYIQK	117
Sbjct:	1753	TEFPREFHGTIRLLPTEAQPGEPPHPLVLDLLAGFQAPCVADIKIGAITWPPSSPEPYIAK	1574
Query:	118	CLKKDTGTTTVSSGFRISGFVYDVKESSEFWKPERKLLRGLYVDGARLTILRKVFVSSNSLS	177
Sbjct:	1573	CLAKDRGTTSVLLGFRVSGVRVVG-PEGAVWRTERPEVKAMDTAGVRRVLRVYVSS--VA	1403
Query:	178	DTGSKPDSAFASSVYGGSHGILTQLLELKTWFENQTLYHFNSCSILMVYENESILKGNDD	237
Sbjct:	1402	DEGM--DCALAAAVYGGKGGVLSQLRELKAWFEEQTLFHFYSASILLGYDAAAAVAAGGDG	1229
Query:	238	DARQVKLVDLAHVLDGNGVIDHNFLGGLCSFINFIREIL-QSP	280
Sbjct:	1228	GG-VTVKLVDFAHVAEGDGVIDHNFLGGLCSLIKVFVSDIVPETP	1100

Figure 21 (cont'd)

B	Query:	176	LSDTGSKPDSAFASSVYGGSHGILTQLLELKTWFENQTLYHFNSCSILMVYENESILKGN	235
	Sbjct:	920	LADEGM--DCALAAAVYGGKGGVLSQLRELKAWLEEQTLFHFYSASILLYGYDAAAVAAGG	747
	Query:	236	DDDARPQVKLVDLAHVLDGNGVIDHNFGLGCSFINFIREIL-QSP	280
	Sbjct:	746	GGGG-VTVKLVDFAHVAEGDGVIDHNFGLGELC*LIKFVSDIVPETP	612
C	Query:	3	LKVPEHQVAGHIAKDGPGLVDDKGRFFKPLQGDSRGEIEVKFYESFSSNTEVPEHIH-	61
	Sbjct:	1444	LHPPEHQVAGHRASASKLGPLIDGSGLFYKPLQAGDRGEHEVAFYEAFSAHAAVPARIRD	1265
	Query:	62	RYFPVYHGTQAV--EGSDGAAM--MVLENLLAEYTKPSVMDVKMGSRWTWYDPDASEEYIQK	117
	Sbjct:	1264	TFFPRFHGTRLLPTEAQPGEPPHPLVLDLLAGFEAPCVADIKIGAITWPPSSPEPYIAK	1085
	Query:	118	CLKKDTGTTTVSSGFRI	134
	Sbjct:	1573	CLAKDRGTTSVLLGFRV	1034
D	Query:	144	ESSFWKPERKLLRGLYVDGARLLTLRKFEVSS	173
	Sbjct:	1011	EGAVWRTERPEVKAMDTVGVRRVLRRYVSS	922

Figure 21 (cont'd)

W

```

Query:      176 LSDTGSKPDSAFASSVYGGSHGILTQLLELKTWFENQTLYHFNSCSILMVYENESILKGN 235
           |:| | | |:||||| |:|:| | | | | | | | | | | |:| :| :|
Sbjct:      907 LADEGM--DCALAAAVYGKGGLVLSQLRELKAWEFGQTLFHFYASILLGYDAAAVAAG 734

Query:      236 DDDARPQVKLVDLAHVLDCNGVIDHNFGLGCSFINFIREIL-QSP 280
           ||||| ||| |:|:| | | | | | | | | | | | | |:| :| :|
Sbjct:      733 GGGG-VTVKLVDFAHVAEGDGVIDHNFGLGC*LIKFSVDIVPETP 599

```

L

[illegible]

```
Query: 125 TTTVSSGFRISGFVEYDHKESSEFWKPERKILRLGLYVDGARLTIRKFVSS 173
      ||:| |||:| | | : | | : : | | ||:| | |
Sbjct: 1052 TTSVLLGFRVSGVRVVG-PEGAVWRTERPEVKAMDIVGVRRLRRYVSS 909
```

A

Figure 22

Query: 1	MQLKVPEHQVAGHIAKDGKPGPLVDDKGRFFKPLQGDSRGEIEVKFYESFSSNTEVPEHI	60
Sbjct: 46420	MQLKVPEHQVAGHIAKDGKPGPLVDDKGRFFKPLQGDSRGEIEVKFYESFSSNTEVPEHI	46599
Query: 61	HRYFPVYHGTQAVEGSDGAAMMVLENLLAEYTKPSVMDVKMGSRWTWYPDASEEYIQKCLK	120
Sbjct: 46600	HRYFPVYHGTQAVEGSDGAAMMVLENLLAEYTKPSVMDVKMGSRWTWYPDASEEYIQKCLK	46779
Query: 121	KDTGTTTVSSGFRISGFVYDHKESSEFWKPERKLLRGLYVDGARLTLRKFFVSSNSLSDTG	180
Sbjct: 46780	KDTGTTTVSSGFRISGFVYDHKESSEFWKPERKLLRGLDVGARLTLRKFFVSSNSLSDTG	46959
Query: 181	SKPDSAFASSVYGGSHGILTQLLELKTWFENQTLYHFNSCSILMVYENESILKGNDDDAR	240
Sbjct: 46960	SKPDSAFASSVYGGSHGILTQLLELKTWFENQTLYHFNSCSILMVYENESILKGNDDDAR	47139
Query: 241	PQVKLVDLAHVLDGNGVIDHNFGLGCSFINFIREILQSPDESADS	286
Sbjct: 47140	PQVKLVDFAHVLDGNGVIDHNFGLGCSFINFIREILQSPDESADS	47277

Figure 22 (cont'd)

Query:	3	LKVPEHQVAGHIAKDGKPGPLVDDKGRFFKPLQGD	SRGEIEVKFYESFSSNTEVPEHIHR	62			
			:	:			
Sbjct:	20248	LKVPEHQVAGHIASDGKLGPLVDDQGRFFKPLQGD	SRGEHEAKFYESTSNMKVPDHIHR	20069			
Query:	63	YFPVYHGTQAVEGSDGAAMM--VLENLLAEYTKPSV	MDVKMGSR	TWYPDASEEYIQKCL	119		
			:	:	:		
Sbjct:	20068	YFPVYHGTQLVEASDGSGLPHLVLDVVSGYANPSV	MDVKIGSR	TWYPDVSEYFYKKCI	19889		
Query:	120	KKDTGTTTVSSGFRISGFEVYDHHKESFWKPERK	LLRGLYVDGAR	LLTLRK	FVSSNLS	SDT	179
			: :	:		:	
Sbjct:	19888	KKDRQTTTVSLGFRVSGFKIFDHOESSFWRAEKK	LVLGYNADGAR	LALRK	FVSSN	SPADS	19709
Query:	180	GSKPDSAFASSVYGGSHGILTQLLELKTWFENQ	TLYHFNSCSILMVYENESILKGNDDDA	239			
		:		:			
Sbjct:	19708	NLTPNCAFASEVYGGCGILAQLLELKDWFETQ	TLYHFNSCSILMIYENESILMQGDDA	19529			
Query:	240	---RPQVKLVDLAHVLDGNGVIDHNFGLGCSFIN	FIREILQSPDE	282			
							:
Sbjct:	19528	PAPRAQVKLVDEFAHVLDGNGVIDHNFGLGCSFI	KFIKDILQSV	19391			


```
Query: 3 LKVPESHQVAGHIAKDGKPGPLVDDKGRFFKPLQGDSRGEIEVKFY 47
      |:|||||:| | |||| | :| | | |||| | | :
Sbjct: 67350 LRPPEHQVAGNYASTDKLGPLVP---RVYKILQAGERGEHEAAFF 67225
```

Figure 23

A

Exon I

Query: 1 MEMILEEKDASDWIYRGECCANLVLAYAGSSPLFV 35
||::| | || ||::||| ||::|:| |||| |
Sbjct: 123170 MEVVLHEGDAKDWYKGECAANLILSYTGSSPSMV 123066

Exon II

Query: 35 VGKVIRIQKARRNDKAIKNANGVSVLTSDSEQHLWRENNELISSPNKEVLEQRYVKHVIIPLLGSKHVDAGV 106
:|||:|::| : || | :| :||| : | : | : | : | : ||: ||| ||
Sbjct: 122224 LGKVLRVKKILK-DKGQAPAPNCIV--FSSHEEHLWGKIPGLLESVKNDCLPQAYATIVMSQHLGANHVDDGGV 122018

Exon III

Query: 106 VRVSVSKEFELECVDKKVTKQRPLWRVNAANVDTSHDSALILNDHSLFS 153
||| ||| | | | | || ||||: :| |||::||:|||
Sbjct: 121825 VRVRVSKNFFELAGKNVLDNRPAWRVNASAIADAGADSALLISDHTLFS 121682

Exon V

Query: 170 KCGFLPTSMFIGKENILKTSVSRFKMHQLLKLEYNEI 206
|||||:| :| ||| :| |::||| | | :
Sbjct: 120948 KCGFLPSSEYISKENSIIKKQVTRYKMHQHLKFHLGEV 120838

Exon VI

[illegible]

Figure 23 (cont'd)

Query:	193	FKMHQLLKLEYN--EISEESEYDPLDLFSGSKERVLEAIKALYSTPQNNFRVFLNGSLIL	250
Sbjct:	120646	FQISLVLNIFLNTFQISKTSFYDPDLDFSGSKERIHMIAKSFFSTPQNNFRIFVDGSLVF	120467
Query:	251	GGGESTGRTP-EIGYAFEDALKGFIQSEGDQRTECFQLVSDAVYGSGVLDRLLEIQK	309
Sbjct:	120466	GGMGGGADSVHPNETEKCLE-----LSKVTLGLQSDLFIELLSEAIFKSGVLGKLATQK	120302
Query:	310	LKDLDIEGAIHLYYDLINQPCPICK---EGRPLEAEELSHALPLDESLKIVKEYLIAATA	366
Sbjct:	120301	LDDHDIEGAIHLYYNIISQPCLVCKSITDTELLRKYSTLHSLPLDKSEKIVRDFLISATA	120122
Query:	367	KDCSIMISFQSRNAWDSEPSPGDYVSCLKPTNQTFDYKV-----	403
Sbjct:	120121	KDCSLMISRPRQGTTDSEYDSVFLDQSNQSYDYKVLM*YVSCLRSSISHLPVCGTIVE	119942
Query:	404	-----HFIDLSLKPLKRMEYYKLDKKIISFYNR	432
Sbjct:	119941	FFLKLTLEINVSEFPPTH*FAHLLLQAYFIDLDVKPLDKMVHYFKLDQKIVNFYTR	119780

Figure 23 (cont'd)


```
Query: 106 VRVSVSKEFLECVDDKKVKQRPPLWRVNAANVDTSKDSALILNDHSLFS 153
      ||| ||| | | | | || ||||| :| ||||::|||:|||
Sbjct: 121825 VRVRVSKNFFELAGKNVLDRPAPWRVNASAI DAGADSAALLISDHTLFS 121682
```

D

```
Query: 1 MEMILEEKDASDWIYRGEGGANLVLAYAGSSPLFV 35
      ||::| | | |::| | | |::| | | |
Sbjct: 123170 MEVVLHEGDAKDVVYKGEAANLILSYTGSSPSMV 123066
```

W

```
Query: 170 KCGFLPTSMFIGKENILKTSVSRFFKMHQLLKLEYNEI 206
        |||||:| : ||| : | |:|:|||| || :
Sbjct: 120948 KCGFLPSSEYISKENSIIKKQVTRYKMHQHLKFHLGEV 120838
```

LL

Query: 35 VGVIRIQARRNDKAIKNANGVSVLTSDQHLWRENNELISSPNKEVLQRVVKHVII 94
:| | :| | :| | :| | :| | :| | :| | :
Sbjct: 122224 LGKVLRVKKILK-DKGQPAPNCIV--FSSHEEHLWGKIPGLLESVKNDCLPQAYATIVMS 122054

Query: 95 PLLGSKHVDAGV 106
||: ||| ||
Sbjct: 122053 QHLGANHVDGGV 122018

Figure 23 (cont'd)

G

Query: 1 MEMILEEKDASDWIYRGECCANLVLAYAGSSPLFVGKVIIRIQKARRNDKAIKNANGVVS 60
||::| || ||::||| |||:|:| ||| :|||:| :| ||| | |
Sbjct: 94 MEVVLHAGDAEDWYKGEAANLILSYTGSSPSMLGKVLRAKKVL-NDKAQAPAPNCVN-- 264

Query: 61 LTSDEQHLWRENNELISSPNKEVLEQRYVKHVIIPLLGSKHVDAGVRVSVSKEFLECVDK 120
:| || :| : ||| | : : | | ||| |||:|||| | |
Sbjct: 265 FSSYEQLVWGDIPELIESVKQGSVAQAYAVHVMSSQHLGHDHVDGGVRVSVSKDFLEIVGK 444

Query: 121 KVTQRPLWRVNAANVDTSHDSALILNDHS 150
| || ||||| :||: ||||: |||
Sbjct: 445 NVLSARPAWRVNASAITDNADSAALLITDHS 534

H

Query: 389 YVSLKPTNQTFDYKVHFIIDLKPLKRMEYYKLDKKIISFYNR 432
::| :| :||| :||:| ||| :| |:||:|:| :| :|
Sbjct: 609 FIFLQSVNRAYDYKAYFVDLDDVKPLDKMTHYFKLDQKIVNFYTK 478

SEQUENCE LISTING

<110> Duke University
York, John D
Stevenson-Paulik, Jill

<120> GENERATION OF LOW PHYTATE PLANTS BY MOLECULAR DISRUPTION OF
INOSITOL POLYPHOSPHATE KINASES

<130> 180/178 PCT

<150> US 60/493,267

<151> 2003-08-07

<160> 10

<170> PatentIn version 3.2

<210> 1

<211> 1374

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (1)..(1374)

<400> 1

atg gag gag att gtt ttg gaa cca aaa gat gca gtt gat tgg tca tat	48
Met Glu Glu Ile Val Leu Glu Pro Lys Asp Ala Val Asp Trp Ser Tyr	
1 5 10 15	
aga ggc gaa gga gct gtt aat ttg gtt ctt gct tac act gga tct tct	96
Arg Gly Glu Gly Ala Val Asn Leu Val Leu Ala Tyr Thr Gly Ser Ser	
20 25 30	
ccc tct ttc ttg gga aag atg atg agg ata caa aaa atg cca aag gat	144
Pro Ser Phe Leu Gly Lys Met Met Arg Ile Gln Lys Met Pro Lys Asp	
35 40 45	
ggg aat gaa aat gga gac aaa agt gaa aat ggt ctt act act cat gaa	192
Gly Asn Glu Asn Gly Asp Lys Ser Glu Asn Gly Leu Thr Thr His Glu	
50 55 60	
aag ctt att tgg gga gac gtt aag gat ctt gta tct tgc aag aac aag	240
Lys Leu Ile Trp Gly Asp Val Lys Asp Leu Val Ser Cys Lys Asn Lys	
65 70 75 80	
gag att gaa gag tat ttg ttt gtc aaa cat gtc atg aga cct ttg ttg	288
Glu Ile Glu Glu Tyr Leu Phe Val Lys His Val Met Arg Pro Leu Leu	
85 90 95	
ggt cgt aaa cat gtc aat cct gga gta tgt ttc ctc ttc att tcc ttc	336
Gly Arg Lys His Val Asn Pro Gly Val Cys Phe Leu Phe Ile Ser Phe	
100 105 110	
tct tgt cat gat tct aac gca atg ata ctt ctt ctt gta gca aag gaa	384
Ser Cys His Asp Ser Asn Ala Met Ile Leu Leu Leu Val Ala Lys Glu	
115 120 125	

ttt ctt gaa tct gtt gag aaa att ata aca tct caa cgt cct tct tgg	432
Phe Leu Glu Ser Val Glu Lys Ile Ile Thr Ser Gln Arg Pro Ser Trp	
130 135 140	
cga gct gat gta gct agt gtg gat act aac cgc agt tct gtg ctt ctc	480
Arg Ala Asp Val Ala Ser Val Asp Thr Asn Arg Ser Ser Val Leu Leu	
145 150 155 160	
atg gat gat ttg acg ctt ttc gcc cac ggt cat gtt gag gat aaa cca	528
Met Asp Asp Leu Thr Leu Phe Ala His Gly His Val Glu Asp Lys Pro	
165 170 175	
tgc tta agt gtt gaa ata aag ccc aaa tgt gga ttc ctt cca tct tca	576
Cys Leu Ser Val Glu Ile Lys Pro Lys Cys Gly Phe Leu Pro Ser Ser	
180 185 190	
agt ttc ata gca gaa gaa aac gtt att aag aag agt ata act cgt ttc	624
Ser Phe Ile Ala Glu Glu Asn Val Ile Lys Lys Ser Ile Thr Arg Phe	
195 200 205	
gaa atg cac caa gtt ctc aag ctt agg gaa aac gag atc tca gaa atc	672
Glu Met His Gln Val Leu Lys Leu Arg Glu Asn Glu Ile Ser Glu Ile	
210 215 220	
agt gaa tat gat cct tta gac ctt ttc tct gga tcc aaa gaa aga ata	720
Ser Glu Tyr Asp Pro Leu Asp Leu Phe Ser Gly Ser Lys Glu Arg Ile	
225 230 235 240	
ctc gaa gct ata aaa gcg ctc tac acg act cct cag aac aat ttt cgt	768
Leu Glu Ala Ile Lys Ala Leu Tyr Thr Thr Pro Gln Asn Asn Phe Arg	
245 250 255	
gtg ttt ttg aat ggt tct cta gta ttt gga ggc tta ggt ggt ggc ata	816
Val Phe Leu Asn Gly Ser Leu Val Phe Gly Gly Leu Gly Gly Gly Ile	
260 265 270	
tgc aaa aca acc tca aag gtt gaa cta gct ttt gag cat ata ctc aaa	864
Cys Lys Thr Thr Ser Lys Val Glu Leu Ala Phe Glu His Ile Leu Lys	
275 280 285	
gat atc atc aag acc gat gat ggt tta cgt gca gat cgt ttc ata gag	912
Asp Ile Ile Lys Thr Asp Asp Gly Leu Arg Ala Asp Arg Phe Ile Glu	
290 295 300	
ctt gta gca gaa act gtt tac aca tcc gga gtt cta gat cag ctt cta	960
Leu Val Ala Glu Thr Val Tyr Thr Ser Gly Val Leu Asp Gln Leu Leu	
305 310 315 320	
gac gta caa aag ctc gat aga tat aac atc gag gga gcg att cac gtg	1008
Asp Val Gln Lys Leu Asp Arg Tyr Asn Ile Glu Gly Ala Ile His Val	
325 330 335	
tac tat gac ttc att gac cag cca tgt aaa gtg tgc aga gag ttg aat	1056
Tyr Tyr Asp Phe Ile Asp Gln Pro Cys Lys Val Cys Arg Glu Leu Asn	
340 345 350	
caa ttt agc tct atg cac tca att cca atg gac gaa aaa gtg aat att	1104
Gln Phe Ser Ser Met His Ser Ile Pro Met Asp Glu Lys Val Asn Ile	
355 360 365	

tgt aag gaa ttt cta ata tct gcc act gca aag gat tgt agt gta atg 1152
 Leu Lys Glu Phe Leu Ile Ser Ala Thr Ala Lys Asp Cys Ser Val Met
 370 375 380

ata agt ttt aga tct acg gaa gct ggg ctc tca aaa tcc tct tcc cat 1200
 Ile Ser Phe Arg Ser Thr Glu Ala Gly Leu Ser Lys Ser Ser Ser His
 385 390 395 400

agt aat ctt cac ctt gaa aca aca aag caa gag ttt gaa tac aag gta 1248
 Ser Asn Leu His Leu Glu Thr Thr Lys Gln Glu Phe Glu Tyr Lys Val
 405 410 415

cat ttc att gat ctt gat atg aga cct ttg aag aaa atg gaa gtc tat 1296
 His Phe Ile Asp Leu Asp Met Arg Pro Leu Lys Lys Met Glu Val Tyr
 420 425 430

tac gag tta gac aag aag atc atg aac aca tac ttg gag atg tta aag 1344
 Tyr Glu Leu Asp Lys Lys Ile Met Asn Thr Tyr Leu Glu Met Leu Lys
 435 440 445

aag aaa gga act caa cct caa tgc ttg tag 1374
 Lys Lys Gly Thr Gln Pro Gln Cys Leu
 450 455

<210> 2
 <211> 457
 <212> PRT
 <213> Arabidopsis thaliana

<400> 2

Met Glu Glu Ile Val Leu Glu Pro Lys Asp Ala Val Asp Trp Ser Tyr
 1 5 10 15

Arg Gly Glu Gly Ala Val Asn Leu Val Leu Ala Tyr Thr Gly Ser Ser
 20 25 30

Pro Ser Phe Leu Gly Lys Met Met Arg Ile Gln Lys Met Pro Lys Asp
 35 40 45

Gly Asn Glu Asn Gly Asp Lys Ser Glu Asn Gly Leu Thr Thr His Glu
 50 55 60

Lys Leu Ile Trp Gly Asp Val Lys Asp Leu Val Ser Cys Lys Asn Lys
 65 70 75 80

Glu Ile Glu Glu Tyr Leu Phe Val Lys His Val Met Arg Pro Leu Leu
 85 90 95

Gly Arg Lys His Val Asn Pro Gly Val Cys Phe Leu Phe Ile Ser Phe
 100 105 110

Ser Cys His Asp Ser Asn Ala Met Ile Leu Leu Leu Val Ala Lys Glu
 115 120 125

Phe Leu Glu Ser Val Glu Lys Ile Ile Thr Ser Gln Arg Pro Ser Trp
 130 135 140

Arg Ala Asp Val Ala Ser Val Asp Thr Asn Arg Ser Ser Val Leu Leu
 145 150 155 160

Met Asp Asp Leu Thr Leu Phe Ala His Gly His Val Glu Asp Lys Pro
 165 170 175

Cys Leu Ser Val Glu Ile Lys Pro Lys Cys Gly Phe Leu Pro Ser Ser
 180 185 190

Ser Phe Ile Ala Glu Glu Asn Val Ile Lys Lys Ser Ile Thr Arg Phe
 195 200 205

Glu Met His Gln Val Leu Lys Leu Arg Glu Asn Glu Ile Ser Glu Ile
 210 215 220

Ser Glu Tyr Asp Pro Leu Asp Leu Phe Ser Gly Ser Lys Glu Arg Ile
 225 230 235 240

Leu Glu Ala Ile Lys Ala Leu Tyr Thr Thr Pro Gln Asn Asn Phe Arg
 245 250 255

Val Phe Leu Asn Gly Ser Leu Val Phe Gly Gly Leu Gly Gly Gly Ile
 260 265 270

Cys Lys Thr Thr Ser Lys Val Glu Leu Ala Phe Glu His Ile Leu Lys
 275 280 285

Asp Ile Ile Lys Thr Asp Asp Gly Leu Arg Ala Asp Arg Phe Ile Glu
 290 295 300

Leu Val Ala Glu Thr Val Tyr Thr Ser Gly Val Leu Asp Gln Leu Leu
 305 310 315 320

Asp Val Gln Lys Leu Asp Arg Tyr Asn Ile Glu Gly Ala Ile His Val
 325 330 335

Tyr Tyr Asp Phe Ile Asp Gln Pro Cys Lys Val Cys Arg Glu Leu Asn
 340 345 350

Gln Phe Ser Ser Met His Ser Ile Pro Met Asp Glu Lys Val Asn Ile
 355 360 365

Leu Lys Glu Phe Leu Ile Ser Ala Thr Ala Lys Asp Cys Ser Val Met
 370 375 380

Ile Ser Phe Arg Ser Thr Glu Ala Gly Leu Ser Lys Ser Ser Ser His
 385 390 395 400

Ser Asn Leu His Leu Glu Thr Thr Lys Gln Glu Phe Glu Tyr Lys Val
 405 410 415

His Phe Ile Asp Leu Asp Met Arg Pro Leu Lys Lys Met Glu Val Tyr
 420 425 430

Tyr Glu Leu Asp Lys Lys Ile Met Asn Thr Tyr Leu Glu Met Leu Lys
 435 440 445

Lys Lys Gly Thr Gln Pro Gln Cys Leu
 450 455

<210> 3
 <211> 1356
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> CDS
 <222> (1)..(1356)

<400> 3
 atg gag atg att ttg gag gag aaa gat gca tca gat tgg att tac aga 48
 Met Glu Met Ile Leu Glu Glu Lys Asp Ala Ser Asp Trp Ile Tyr Arg
 1 5 10 15
 ggc gaa ggt ggt gcc aat ctc gtt ctt gct tac gct gga tct tct cca 96
 Gly Glu Gly Gly Ala Asn Leu Val Leu Ala Tyr Ala Gly Ser Ser Pro
 20 25 30
 ctt ttt gtg ggg aaa gtg att cgt ata cag aag gct cgg agg aat gat 144
 Leu Phe Val Gly Lys Val Ile Arg Ile Gln Lys Ala Arg Arg Asn Asp
 35 40 45
 aaa gca atc aag aat gcg aat ggt gtt gtt tcg gtt tta aca agc gat 192
 Lys Ala Ile Lys Asn Ala Asn Gly Val Val Ser Val Leu Thr Ser Asp
 50 55 60
 gag caa cat cta tgg agg gaa aac aat gag ctg att tca tca cca aac 240
 Glu Gln His Leu Trp Arg Glu Asn Asn Glu Leu Ile Ser Ser Pro Asn
 65 70 75 80

aag gaa gtt ctt gaa caa aga tat gtt aag cat gtc atc atc cca ctc	288
Lys Glu Val Leu Glu Gln Arg Tyr Val Lys His Val Ile Ile Pro Leu	
85 90 95	
ttg ggt tct aaa cat gtt gat gca gga gta cgt gtt tct gtg tcc aag	336
Leu Gly Ser Lys His Val Asp Ala Gly Val Arg Val Ser Val Ser Lys	
100 105 110	
gag ttt ctt gag tgt gtt gat aag aaa gta act aag cag cgt ccg cta	384
Glu Phe Leu Glu Cys Val Asp Lys Lys Val Thr Lys Gln Arg Pro Leu	
115 120 125	
tgg cgt gtt aat gca gct aat gtt gat acc agc cat gat tcc gct ctt	432
Trp Arg Val Asn Ala Ala Asn Val Asp Thr Ser His Asp Ser Ala Leu	
130 135 140	
ata ttg aat gat cat tca cta ttc tct caa ggt att act agc ggt ggt	480
Ile Leu Asn Asp His Ser Leu Phe Ser Gln Gly Ile Thr Ser Gly Gly	
145 150 155 160	
gat tgc att agt gtt gaa ata aag cct aaa tgc gga ttt ctt cca acc	528
Asp Cys Ile Ser Val Glu Ile Lys Pro Lys Cys Gly Phe Leu Pro Thr	
165 170 175	
tca atg ttc ata ggt aaa gaa aac ata ctc aaa aca agc gta agc cgt	576
Ser Met Phe Ile Gly Lys Glu Asn Ile Leu Lys Thr Ser Val Ser Arg	
180 185 190	
ttc aaa atg cat caa ctc cta aag ttg gaa tat aac gag ata tct gaa	624
Phe Lys Met His Gln Leu Leu Lys Leu Glu Tyr Asn Glu Ile Ser Glu	
195 200 205	
gaa agc gag tac gat cct ctt gat ctc ttc tct gga tcc aaa gag aga	672
Glu Ser Glu Tyr Asp Pro Leu Asp Leu Phe Ser Gly Ser Lys Glu Arg	
210 215 220	
gtt ttg gaa gct ata aaa gct tta tat tcc act cct caa aac aat ttc	720
Val Leu Glu Ala Ile Lys Ala Leu Tyr Ser Thr Pro Gln Asn Asn Phe	
225 230 235 240	
cgc gta ttc ttg aat ggt tct ctc ata tta ggg ggt tct ggt gaa agc	768
Arg Val Phe Leu Asn Gly Ser Leu Ile Leu Gly Gly Ser Gly Glu Ser	
245 250 255	
acc ggg aga acc agt ccc gaa atc ggg tat gcc ttt gag gat gct ctc	816
Thr Gly Arg Thr Ser Pro Glu Ile Gly Tyr Ala Phe Glu Asp Ala Leu	
260 265 270	
aaa ggc ttc att caa tca gaa gac ggt cag agg aca gag tgc ttt cta	864
Lys Gly Phe Ile Gln Ser Glu Asp Gly Gln Arg Thr Glu Cys Phe Leu	
275 280 285	
cag cta gta tct gac gct gtc tat ggc tca gga gtg ctt gat aga ctt	912
Gln Leu Val Ser Asp Ala Val Tyr Gly Ser Gly Val Leu Asp Arg Leu	
290 295 300	
ctt gaa att cag aag cta gac aaa tta gac att gaa gga gcg att cat	960
Leu Glu Ile Gln Lys Leu Asp Lys Leu Asp Ile Glu Gly Ala Ile His	
305 310 315 320	

ttg tat tac gat tta atc aac cag cct tgt cct ata tgt aaa gaa ggt 1008
 Leu Tyr Tyr Asp Leu Ile Asn Gln Pro Cys Pro Ile Cys Lys Glu Gly
 325 330 335

agg cca ctg gag gcg gaa ttg tct cta cat gct tta cct tta gat gaa 1056
 Arg Pro Leu Glu Ala Glu Leu Ser Leu His Ala Leu Pro Leu Asp Glu
 340 345 350

agc ttg aag atc gtg aag gag tat ttg ata gct gca act gct aaa gac 1104
 Ser Leu Lys Ile Val Lys Glu Tyr Leu Ile Ala Ala Thr Ala Lys Asp
 355 360 365

tgt agt att atg atc agt ttt caa tca agg aat gct tgg gat tca gaa 1152
 Cys Ser Ile Met Ile Ser Phe Gln Ser Arg Asn Ala Trp Asp Ser Glu
 370 375 380

ccc tcg ggt gat tac gtc tct cta aaa ccg acc aat caa acg ttt gat 1200
 Pro Ser Gly Asp Tyr Val Ser Leu Lys Pro Thr Asn Gln Thr Phe Asp
 385 390 395 400

tac aag gta cat ttc att gat cta agc ctg aaa cca ctt aag aga atg 1248
 Tyr Lys Val His Phe Ile Asp Leu Ser Leu Lys Pro Leu Lys Arg Met
 405 410 415

gag tca tac tac aaa ttg gat aag aag ata att agc ttt tac aat cgg 1296
 Glu Ser Tyr Tyr Lys Leu Asp Lys Lys Ile Ile Ser Phe Tyr Asn Arg
 420 425 430

aag cag aaa gcc gaa aac acc gca gaa caa atc ggc aac tca aaa cct 1344
 Lys Gln Lys Ala Glu Asn Thr Ala Glu Gln Ile Gly Asn Ser Lys Pro
 435 440 445

tcc cac agc taa 1356
 Ser His Ser
 450

<210> 4
 <211> 451
 <212> PRT
 <213> Arabidopsis thaliana

<400> 4

Met Glu Met Ile Leu Glu Glu Lys Asp Ala Ser Asp Trp Ile Tyr Arg
 1 5 10 15

Gly Glu Gly Gly Ala Asn Leu Val Leu Ala Tyr Ala Gly Ser Ser Pro
 20 25 30

Leu Phe Val Gly Lys Val Ile Arg Ile Gln Lys Ala Arg Arg Asn Asp
 35 40 45

Lys Ala Ile Lys Asn Ala Asn Gly Val Val Ser Val Leu Thr Ser Asp
 50 55 60

Glu Gln His Leu Trp Arg Glu Asn Asn Glu Leu Ile Ser Ser Pro Asn
 65 70 75 80

Lys Glu Val Leu Glu Gln Arg Tyr Val Lys His Val Ile Ile Pro Leu
 85 90 95

Leu Gly Ser Lys His Val Asp Ala Gly Val Arg Val Ser Val Ser Lys
 100 105 110

Glu Phe Leu Glu Cys Val Asp Lys Lys Val Thr Lys Gln Arg Pro Leu
 115 120 125

Trp Arg Val Asn Ala Ala Asn Val Asp Thr Ser His Asp Ser Ala Leu
 130 135 140

Ile Leu Asn Asp His Ser Leu Phe Ser Gln Gly Ile Thr Ser Gly Gly
 145 150 155 160

Asp Cys Ile Ser Val Glu Ile Lys Pro Lys Cys Gly Phe Leu Pro Thr
 165 170 175

Ser Met Phe Ile Gly Lys Glu Asn Ile Leu Lys Thr Ser Val Ser Arg
 180 185 190

Phe Lys Met His Gln Leu Leu Lys Leu Glu Tyr Asn Glu Ile Ser Glu
 195 200 205

Glu Ser Glu Tyr Asp Pro Leu Asp Leu Phe Ser Gly Ser Lys Glu Arg
 210 215 220

Val Leu Glu Ala Ile Lys Ala Leu Tyr Ser Thr Pro Gln Asn Asn Phe
 225 230 235 240

Arg, Val Phe Leu Asn Gly Ser Leu Ile Leu Gly Gly Ser Gly Glu Ser
 245 250 255

Thr Gly Arg Thr Ser Pro Glu Ile Gly Tyr Ala Phe Glu Asp Ala Leu
 260 265 270

Lys Gly Phe Ile Gln Ser Glu Asp Gly Gln Arg Thr Glu Cys Phe Leu
 275 280 285

Gln Leu Val Ser Asp Ala Val Tyr Gly Ser Gly Val Leu Asp Arg Leu
 290 295 300

Leu Glu Ile Gln Lys Leu Asp Lys Leu Asp Ile Glu Gly Ala Ile His
 305 310 315 320

Leu Tyr Tyr Asp Leu Ile Asn Gln Pro Cys Pro Ile Cys Lys Glu Gly
 325 330 335

Arg Pro Leu Glu Ala Glu Leu Ser Leu His Ala Leu Pro Leu Asp Glu
 340 345 350

Ser Leu Lys Ile Val Lys Glu Tyr Leu Ile Ala Ala Thr Ala Lys Asp
 355 360 365

Cys Ser Ile Met Ile Ser Phe Gln Ser Arg Asn Ala Trp Asp Ser Glu
 370 375 380

Pro Ser Gly Asp Tyr Val Ser Leu Lys Pro Thr Asn Gln Thr Phe Asp
 385 390 395 400

Tyr Lys Val His Phe Ile Asp Leu Ser Leu Lys Pro Leu Lys Arg Met
 405 410 415

Glu Ser Tyr Tyr Lys Leu Asp Lys Lys Ile Ile Ser Phe Tyr Asn Arg
 420 425 430

Lys Gln Lys Ala Glu Asn Thr Ala Glu Gln Ile Gly Asn Ser Lys Pro
 435 440 445

Ser His Ser
 450

<210> 5
 <211> 1461
 <212> DNA
 <213> Arabidopsis thaliana

<220>
 <221> CDS
 <222> (1)..(1461)

<400> 5
 atg att tgg tca ttg ctc aag att tac gtt ttg atg cta tgt aca atc 48
 Met Ile Trp Ser Leu Leu Lys Ile Tyr Val Leu Met Leu Cys Thr Ile
 1 5 10 15
 agc act gtg tac aat acg ttc ttg tta ata gtt tgt gtt gtt ggg tgt 96
 Ser Thr Val Tyr Asn Thr Phe Leu Leu Ile Val Cys Val Val Gly Cys
 20 25 30

tgt gtt cag atg gaa gag att gtt ttt gaa aca aaa gat gca gtt gat	144
Cys Val Gln Met Glu Glu Ile Val Phe Glu Thr Lys Asp Ala Val Asp	
35 40 45	
tgg tca tat aga agc gaa gga gtt gtt aat ttg gtt ctc gct tac tct	192
Trp Ser Tyr Arg Ser Glu Gly Val Val Asn Leu Val Leu Ala Tyr Ser	
50 55 60	
gga tca tct ccc act ttc ttg gga aag gtg atg agg ata aaa aaa tta	240
Gly Ser Ser Pro Thr Phe Leu Gly Lys Val Met Arg Ile Lys Lys Leu	
65 70 75 80	
cga aac act agg aat gaa aat aga gac cga agt gga agt tgt ctt acg	288
Arg Asn Thr Arg Asn Glu Asn Arg Asp Arg Ser Gly Ser Cys Leu Thr	
85 90 95	
acc cag gaa aag ctt att tgg gga gag att aag gat ctt gta tct tgc	336
Thr Gln Glu Lys Leu Ile Trp Gly Glu Ile Lys Asp Leu Val Ser Cys	
100 105 110	
caa aac aag gag att gtg gac tat ttg ttt gtc aaa cat gtc atg aaa	384
Gln Asn Lys Glu Ile Val Asp Tyr Leu Phe Val Lys His Val Met Lys	
115 120 125	
cct ttg ttg ggt cac aaa tat gtt act cct gga att tgc ctt cct gta	432
Pro Leu Leu Gly His Lys Tyr Val Thr Pro Gly Ile Cys Leu Pro Val	
130 135 140	
gaa aag gag ttt ctt gag tct gtt aag aag ata gtg acg tct cag cgt	480
Glu Lys Glu Phe Leu Glu Ser Val Lys Lys Ile Val Thr Ser Gln Arg	
145 150 155 160	
cat tct tgg cga gct aat aca ctc agt gtg gat act aac cgc agt ttc	528
His Ser Trp Arg Ala Asn Thr Leu Ser Val Asp Thr Asn Arg Ser Phe	
165 170 175	
gcg ctt ctc atg gat gat ttg acc att ttc tcc cat ggt caa gtt gag	576
Ala Leu Leu Met Asp Asp Leu Thr Ile Phe Ser His Gly Gln Val Glu	
180 185 190	
gat cat aaa cca tgc tta act gtt gaa ata aag ccc aaa tgt gga ttt	624
Asp His Lys Pro Cys Leu Thr Val Glu Ile Lys Pro Lys Cys Gly Phe	
195 200 205	
ctt tca tct tca agt ttc atagcgc gaa gaa aat gtt atc aag aag agt	672
Leu Ser Ser Ser Ser Phe Ile Ala Glu Glu Asn Val Ile Lys Lys Ser	
210 215 220	
ata agt cgc ttc gaa atg tac caa gtt cta aag ctt agg gaa aac cag	720
Ile Ser Arg Phe Glu Met Tyr Gln Val Leu Lys Leu Arg Glu Asn Gln	
225 230 235 240	
ata tca caa atc agt gaa tat gat ccc tta gac ctt ttc tct gga tcc	768
Ile Ser Gln Ile Ser Glu Tyr Asp Pro Leu Asp Leu Phe Ser Gly Ser	
245 250 255	
aaa gat aga ata cac aaa gca ata aaa gcg ctc tac tcg act cct caa	816
Lys Asp Arg Ile His Lys Ala Ile Lys Ala Leu Tyr Ser Thr Pro Gln	
260 265 270	

aac agt tta caa gtg ttc ttg aat ggt tct cta gta ttt gga ggc ttc 864
 Asn Ser Leu Gln Val Phe Leu Asn Gly Ser Leu Val Phe Gly Gly Phe
 275 280 285

cgt ggt ggt ata tgc aaa aca acc tca aag ctt gaa tta gct ttt gag 912
 Arg Gly Gly Ile Cys Lys Thr Thr Ser Lys Leu Glu Leu Ala Phe Glu
 290 295 300

cat aca ctc aaa gat ttc ttc aaa acc gaa gat gac tgc ggc cta cgt 960
 His Thr Leu Lys Asp Phe Phe Lys Thr Glu Asp Asp Cys Gly Leu Arg
 305 310 315 320

gca aat gct ttc ata gag ctt gtt gca gaa act gtt tac gca tct gga 1008
 Ala Asn Ala Phe Ile Glu Leu Val Ala Glu Thr Val Tyr Ala Ser Gly
 325 330 335

gct tta gat cag ctt ctt gag gtt caa aag ctt gac aaa tat aac atc 1056
 Ala Leu Asp Gln Leu Leu Glu Val Gln Lys Leu Asp Lys Tyr Asn Ile
 340 345 350

gag gga gcg att cac gcg tac tat gac tta gtt gac cag cca tgt aaa 1104
 Glu Gly Ala Ile His Ala Tyr Tyr Asp Leu Val Asp Gln Pro Cys Lys
 355 360 365

gcg tgc caa gag ttg gaa agt agt aaa tta tcg aat cag ttt ggt tcc 1152
 Ala Cys Gln Glu Leu Glu Ser Ser Lys Leu Ser Asn Gln Phe Gly Ser
 370 375 380

atg cat tca ttg cca cag gat gag aaa gta aat att ttg aag gat ttt 1200
 Met His Ser Leu Pro Gln Asp Glu Lys Val Asn Ile Leu Lys Asp Phe
 385 390 395 400

cta ata tct tcc act gct ata gat tgt agt gta atg ata agt ttt aga 1248
 Leu Ile Ser Ser Thr Ala Ile Asp Cys Ser Val Met Ile Ser Phe Arg
 405 410 415

cca ata gag act ggg ctc tca aga tcc tct tcc cac ggt aac att caa 1296
 Pro Ile Glu Thr Gly Leu Ser Arg Ser Ser Ser His Gly Asn Ile Gln
 420 425 430

ctt gaa tca aca aag caa gaa ttt gag tat aag ata cac ttc att gat 1344
 Leu Glu Ser Thr Lys Gln Glu Phe Glu Tyr Lys Ile His Phe Ile Asp
 435 440 445

ctt gat atg aga cct tta aag aga atg gaa gcc tat cac gaa tca gac 1392
 Leu Asp Met Arg Pro Leu Lys Arg Met Glu Ala Tyr His Glu Ser Asp
 450 455 460

aag aag atc aca aaa aca tac cag gag atg ctc aag aag aag aaa gga 1440
 Lys Lys Ile Thr Lys Thr Tyr Gln Glu Met Leu Lys Lys Lys Lys Gly
 465 470 475 480

gat caa cct cgg cgc ttc tag 1461
 Asp Gln Pro Arg Arg Phe
 485

<210> 6

<211> 486

<212> PRT

<213> Arabidopsis thaliana

<400> 6

Met Ile Trp Ser Leu Leu Lys Ile Tyr Val Leu Met Leu Cys Thr Ile
 1 5 10 15

Ser Thr Val Tyr Asn Thr Phe Leu Leu Ile Val Cys Val Val Gly Cys
 20 25 30

Cys Val Gln Met Glu Glu Ile Val Phe Glu Thr Lys Asp Ala Val Asp
 35 40 45

Trp Ser Tyr Arg Ser Glu Gly Val Val Asn Leu Val Leu Ala Tyr Ser
 50 55 60

Gly Ser Ser Pro Thr Phe Leu Gly Lys Val Met Arg Ile Lys Lys Leu
 65 70 75 80

Arg Asn Thr Arg Asn Glu Asn Arg Asp Arg Ser Gly Ser Cys Leu Thr
 85 90 95

Thr Gln Glu Lys Leu Ile Trp Gly Glu Ile Lys Asp Leu Val Ser Cys
 100 105 110

Gln Asn Lys Glu Ile Val Asp Tyr Leu Phe Val Lys His Val Met Lys
 115 120 125

Pro Leu Leu Gly His Lys Tyr Val Thr Pro Gly Ile Cys Leu Pro Val
 130 135 140

Glu Lys Glu Phe Leu Glu Ser Val Lys Lys Ile Val Thr Ser Gln Arg
 145 150 155 160

His Ser Trp Arg Ala Asn Thr Leu Ser Val Asp Thr Asn Arg Ser Phe
 165 170 175

Ala Leu Leu Met Asp Asp Leu Thr Ile Phe Ser His Gly Gln Val Glu
 180 185 190

Asp His Lys Pro Cys Leu Thr Val Glu Ile Lys Pro Lys Cys Gly Phe
 195 200 205

Leu Ser Ser Ser Ser Phe Ile Ala Glu Glu Asn Val Ile Lys Lys Ser
 210 215 220

Ile Ser Arg Phe Glu Met Tyr Gln Val Leu Lys Leu Arg Glu Asn Gln
 225 230 235 240

Ile Ser Gln Ile Ser Glu Tyr Asp Pro Leu Asp Leu Phe Ser Gly Ser
 245 250 255
 Lys Asp Arg Ile His Lys Ala Ile Lys Ala Leu Tyr Ser Thr Pro Gln
 260 265 270
 Asn Ser Leu Gln Val Phe Leu Asn Gly Ser Leu Val Phe Gly Gly Phe
 275 280 285
 Arg Gly Gly Ile Cys Lys Thr Thr Ser Lys Leu Glu Leu Ala Phe Glu
 290 295 300
 His Thr Leu Lys Asp Phe Phe Lys Thr Glu Asp Asp Cys Gly Leu Arg
 305 310 315 320
 Ala Asn Ala Phe Ile Glu Leu Val Ala Glu Thr Val Tyr Ala Ser Gly
 325 330 335
 Ala Leu Asp Gln Leu Leu Glu Val Gln Lys Leu Asp Lys Tyr Asn Ile
 340 345 350
 Glu Gly Ala Ile His Ala Tyr Tyr Asp Leu Val Asp Gln Pro Cys Lys
 355 360 365
 Ala Cys Gln Glu Leu Glu Ser Ser Lys Leu Ser Asn Gln Phe Gly Ser
 370 375 380
 Met His Ser Leu Pro Gln Asp Glu Lys Val Asn Ile Leu Lys Asp Phe
 385 390 395 400
 Leu Ile Ser Ser Thr Ala Ile Asp Cys Ser Val Met Ile Ser Phe Arg
 405 410 415
 Pro Ile Glu Thr Gly Leu Ser Arg Ser Ser Ser His Gly Asn Ile Gln
 420 425 430
 Leu Glu Ser Thr Lys Gln Glu Phe Glu Tyr Lys Ile His Phe Ile Asp
 435 440 445
 Leu Asp Met Arg Pro Leu Lys Arg Met Glu Ala Tyr His Glu Ser Asp
 450 455 460
 Lys Lys Ile Thr Lys Thr Tyr Gln Glu Met Leu Lys Lys Lys Lys Gly
 465 470 475 480

Asp Gln Pro Arg Arg Phe
485

<210> 7
<211> 861
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (1)..(861)

<400> 7
atg cag ctc aaa gtc cct gaa cat cag gtt gca gga cac att gct aaa 48
Met Gln Leu Lys Val Pro Glu His Gln Val Ala Gly His Ile Ala Lys
1 5 10 15
gac ggg aag cct ggt cct ctc gta gat gac aag ggt cgg ttc ttc aag 96
Asp Gly Lys Pro Gly Pro Leu Val Asp Asp Lys Gly Arg Phe Phe Lys
20 25 30
cca ctt cag ggc gat tct cgt ggt gaa atc gag gta aag ttc tac gaa 144
Pro Leu Gln Gly Asp Ser Arg Gly Glu Ile Glu Val Lys Phe Tyr Glu
35 40 45
tct ttc tcc tca aac aca gag gtt cca gaa cac atc cat aga tat ttc 192
Ser Phe Ser Ser Asn Thr Glu Val Pro Glu His Ile His Arg Tyr Phe
50 55 60
ccg gtg tat cac ggc act caa gca gtt gaa ggt tct gat gga gca gcc 240
Pro Val Tyr His Gly Thr Gln Ala Val Glu Gly Ser Asp Gly Ala Ala
65 70 75 80
atg atg gtg ttg gaa aat ctt ctt gca gaa tac aca aaa cca tca gta 288
Met Met Val Leu Glu Asn Leu Leu Ala Glu Tyr Thr Lys Pro Ser Val
85 90 95
atg gat gtt aag atg ggt tcg aga aca tgg tat cct gat gca tct gaa 336
Met Asp Val Lys Met Gly Ser Arg Thr Trp Tyr Pro Asp Ala Ser Glu
100 105 110
gaa tac atc caa aaa tgt ttg aag aaa gac acg ggt acc aca acc gtg 384
Glu Tyr Ile Gln Lys Cys Leu Lys Lys Asp Thr Gly Thr Thr Thr Val
115 120 125
tca tcg ggt ttc agg atc tct ggt ttc gaa gtg tat gat cac aaa gaa 432
Ser Ser Gly Phe Arg Ile Ser Gly Phe Glu Val Tyr Asp His Lys Glu
130 135 140
tcg agt ttc tgg aag ccc gag agg aag ctt ctt cgc ggg ctc tat gta 480
Ser Ser Phe Trp Lys Pro Glu Arg Lys Leu Leu Arg Gly Leu Tyr Val
145 150 155 160
gat gga gcg aga ttg act ctg agg aag ttt gta tca tct aac tca ctt 528
Asp Gly Ala Arg Leu Thr Leu Arg Lys Phe Val Ser Ser Asn Ser Leu
165 170 175

tcg gac act ggc tcg aaa cct gac tct gct ttt gcc tcg agt gtt tac 576
 Ser Asp Thr Gly Ser Lys Pro Asp Ser Ala Phe Ala Ser Ser Val Tyr
 180 185 190

ggc ggt tcc cac ggg atc tta acg cag ttg ctg gaa ctc aag acc tgg 624
 Gly Gly Ser His Gly Ile Leu Thr Gln Leu Leu Glu Leu Lys Thr Trp
 195 200 205

ttc gag aac caa acg ctc tac cat ttc aac tct tgt tcg att tta atg 672
 Phe Glu Asn Gln Thr Leu Tyr His Phe Asn Ser Cys Ser Ile Leu Met
 210 215 220

gtc tat gag aat gaa tcc atc tta aag gga aat gat gat gat gct aga 720
 Val Tyr Glu Asn Glu Ser Ile Leu Lys Gly Asn Asp Asp Asp Ala Arg
 225 230 235 240

cca caa gtc aag ctg gtg gat ctt gct cat gtt ctt gat ggt aat ggt 768
 Pro Gln Val Lys Leu Val Asp Leu Ala His Val Leu Asp Gly Asn Gly
 245 250 255

gtc att gac cat aac ttc ttg ggt ggt ctt tgc tct ttc ata aac ttc 816
 Val Ile Asp His Asn Phe Leu Gly Gly Leu Cys Ser Phe Ile Asn Phe
 260 265 270

att cgt gag att ctt cag agc cca gat gag tct gca gat tct tag 861
 Ile Arg Glu Ile Leu Gln Ser Pro Asp Glu Ser Ala Asp Ser
 275 280 285

<210> 8

<211> 286

<212> PRT

<213> Arabidopsis thaliana

<400> 8

Met Gln Leu Lys Val Pro Glu His Gln Val Ala Gly His Ile Ala Lys
 1 5 10 15

Asp Gly Lys Pro Gly Pro Leu Val Asp Asp Lys Gly Arg Phe Phe Lys
 20 25 30

Pro Leu Gln Gly Asp Ser Arg Gly Glu Ile Glu Val Lys Phe Tyr Glu
 35 40 45

Ser Phe Ser Ser Asn Thr Glu Val Pro Glu His Ile His Arg Tyr Phe
 50 55 60

Pro Val Tyr His Gly Thr Gln Ala Val Glu Gly Ser Asp Gly Ala Ala
 65 70 75 80

Met Met Val Leu Glu Asn Leu Leu Ala Glu Tyr Thr Lys Pro Ser Val
 85 90 95

Met Asp Val Lys Met Gly Ser Arg Thr Trp Tyr Pro Asp Ala Ser Glu
100 105 110

Glu Tyr Ile Gln Lys Cys Leu Lys Lys Asp Thr Gly Thr Thr Thr Val
115 120 125

Ser Ser Gly Phe Arg Ile Ser Gly Phe Glu Val Tyr Asp His Lys Glu
130 135 140

Ser Ser Phe Trp Lys Pro Glu Arg Lys Leu Leu Arg Gly Leu Tyr Val
145 150 155 160

Asp Gly Ala Arg Leu Thr Leu Arg Lys Phe Val Ser Ser Asn Ser Leu
165 170 175

Ser Asp Thr Gly Ser Lys Pro Asp Ser Ala Phe Ala Ser Ser Val Tyr
180 185 190

Gly Gly Ser His Gly Ile Leu Thr Gln Leu Leu Glu Leu Lys Thr Trp
195 200 205

Phe Glu Asn Gln Thr Leu Tyr His Phe Asn Ser Cys Ser Ile Leu Met
210 215 220

Val Tyr Glu Asn Glu Ser Ile Leu Lys Gly Asn Asp Asp Asp Ala Arg
225 230 235 240

Pro Gln Val Lys Leu Val Asp Leu Ala His Val Leu Asp Gly Asn Gly
245 250 255

Val Ile Asp His Asn Phe Leu Gly Gly Leu Cys Ser Phe Ile Asn Phe
260 265 270

Ile Arg Glu Ile Leu Gln Ser Pro Asp Glu Ser Ala Asp Ser
275 280 285

<210> 9
<211> 903
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (1)..(903)

<400> 9

atg ctc aag gtc cct gaa cac caa gtt gct ggt cac att gct agt gat	48
Met Leu Lys Val Pro Glu His Gln Val Ala Gly His Ile Ala Ser Asp	
1 5 10 15	
ggg aag ctc ggt cca ctc gta gat gac caa ggc cgg ttc ttc aag cca	96
Gly Lys Leu Gly Pro Leu Val Asp Asp Gln Gly Arg Phe Phe Lys Pro	
20 25 30	
ctt cag gga gat tct cgt ggc gaa cac gag gct aag ttc tat gag tct	144
Leu Gln Gly Asp Ser Arg Gly Glu His Glu Ala Lys Phe Tyr Glu Ser	
35 40 45	
ttc aca tcg aac atg aag gtt cca gat cac atc cat aga tac ttc ccg	192
Phe Thr Ser Asn Met Lys Val Pro Asp His Ile His Arg Tyr Phe Pro	
50 55 60	
gtg tat cac ggc act cag cta gtt gaa gca tct gat gga tct ggc aag	240
Val Tyr His Gly Thr Gln Leu Val Glu Ala Ser Asp Gly Ser Gly Lys	
65 70 75 80	
ctt cct cat ctt gtt ctt gat gat gtt gtt tca ggg tac gca aac ccg	288
Leu Pro His Leu Val Leu Asp Asp Val Val Ser Gly Tyr Ala Asn Pro	
85 90 95	
tcg gta atg gat gtt aag att gga tct agg aca tgg tac ccg gat gta	336
Ser Val Met Asp Val Lys Ile Gly Ser Arg Thr Trp Tyr Pro Asp Val	
100 105 110	
tca gaa gaa tac ttc aag aaa tgt att aag aaa gat aga cag acc acc	384
Ser Glu Glu Tyr Phe Lys Lys Cys Ile Lys Lys Asp Arg Gln Thr Thr	
115 120 125	
acg gtt tcg ttg ggg ttc agg gtt tca ggt ttt aag att ttt gat cac	432
Thr Val Ser Leu Gly Phe Arg Val Ser Gly Phe Lys Ile Phe Asp His	
130 135 140	
caa gaa tca agt ttt tgg aga gct gag aag aag ctt gtt ctt ggg tat	480
Gln Glu Ser Ser Phe Trp Arg Ala Glu Lys Lys Leu Val Leu Gly Tyr	
145 150 155 160	
aat gca gat ggt gct aga ttg gct ctg agg aag ttt gtg tca tcg aac	528
Asn Ala Asp Gly Ala Arg Leu Ala Leu Arg Lys Phe Val Ser Ser Asn	
165 170 175	
tct ccc gct gac tct aac ttg aca cca aac tgt gct ttt gca tca gag	576
Ser Pro Ala Asp Ser Asn Leu Thr Pro Asn Cys Ala Phe Ala Ser Glu	
180 185 190	
gtt tat ggc ggt tgt aac ggg atc tta gcg cag ttg ttg gag ctt aaa	624
Val Tyr Gly Gly Cys Asn Gly Ile Leu Ala Gln Leu Leu Glu Leu Lys	
195 200 205	
gat tgg ttc gaa acc caa acg ctt tac cat ttc aat tcc tgc tcg att	672
Asp Trp Phe Glu Thr Gln Thr Leu Tyr His Phe Asn Ser Cys Ser Ile	
210 215 220	
ctg atg att tac gag aat gaa tca atc ttg atg caa gga gga gat gat	720
Leu Met Ile Tyr Glu Asn Glu Ser Ile Leu Met Gln Gly Gly Asp Asp	
225 230 235 240	

gat gaa acc gat act tcc cta ctt gag aac ggg cgc tag 903
Asp Glu Thr Asp Thr Ser Leu Leu Glu Asn Gly Arg
290 295 300

<400> 10

Thr Val Ser Leu Gly Phe Arg Val Ser Gly Phe Lys Ile Phe Asp His
130 135 140

Gln Glu Ser Ser Phe Trp Arg Ala Glu Lys Lys Leu Val Leu Gly Tyr
 145 150 155 160

Asn Ala Asp Gly Ala Arg Leu Ala Leu Arg Lys Phe Val Ser Ser Asn
 165 170 175

Ser Pro Ala Asp Ser Asn Leu Thr Pro Asn Cys Ala Phe Ala Ser Glu
 180 185 190

Val Tyr Gly Gly Cys Asn Gly Ile Leu Ala Gln Leu Leu Glu Leu Lys
 195 200 205

Asp Trp Phe Glu Thr Gln Thr Leu Tyr His Phe Asn Ser Cys Ser Ile
 210 215 220

Leu Met Ile Tyr Glu Asn Glu Ser Ile Leu Met Gln Gly Gly Asp Asp
 225 230 235 240

Ala Pro Ala Pro Arg Ala Gln Val Lys Leu Val Asp Phe Ala His Val
 245 250 255

Leu Asp Gly Asn Gly Val Ile Asp His Asn Phe Leu Gly Gly Leu Cys
 260 265 270

Ser Phe Ile Lys Phe Ile Lys Asp Ile Leu Gln Ser Val Glu Lys His
 275 280 285

Asp Glu Thr Asp Thr Ser Leu Leu Glu Asn Gly Arg
 290 295 300