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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 June 2015 (11.06.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/084969 A1

(51) International Patent Classification:
C12N 15/29 (2006.01)

(21) International Application Number:
PCT/US2014/068392

(22) International Filing Date:
3 December 2014 (03.12.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/911,275 3 December 2013 (03.12.2013) US
61/911,826 4 December 2013 (04.12.2013) US

(71) Applicants: **IOWA STATE UNIVERSITY RESEARCH FOUNDATION, INC.** [US/US]; 2121 State Avenue, Ames, IA 50014 (US). **UNIVERSITY OF BONN** [DE/DE]; Regina-pacis-weg 3, 53113 Bonn (DE).

(72) Inventors: **SCHNABLE, Patrick, S.**; 2035b Roy J. Carver Co-lab, Iowa State University, Ames, IA 50011-3650 (US). **LIU, Sanzhen**; 4022 Throckmorton, Department Of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502 (US). **HOCKHOLDINGER, Frank**; University Of Bonn, Friedrich-ebert-allee-144, 53113 Bonn (DE). **NESTLER, Josefine**; University Of Bonn, Friedrich-ebert-alle 144, 53113 Bonn (DE).

(74) Agents: **MERCHANT, Mary Anthony** et al.; Ballard Spahr LLP, 999 Peachtree Street, Suite 1000, Atlanta, GA 30309 (US).

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

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

Declarations under Rule 4.17:

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

[Continued on next page]

(54) Title: PLANTS WITH IMPROVED DROUGHT TOLERANCE

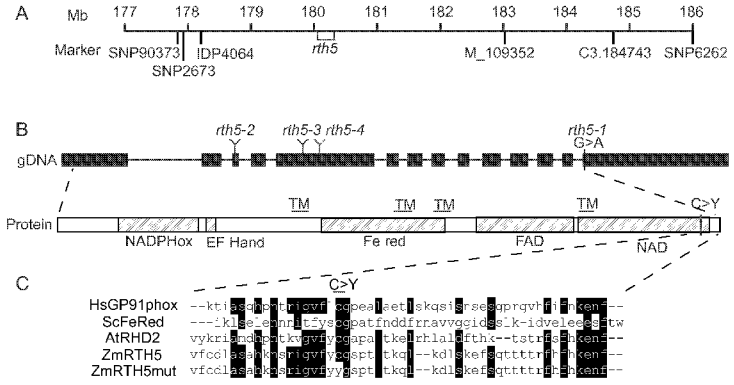

FIG. 1

(57) Abstract: Disclosed herein is a method for increasing drought tolerance in a plant. The method includes introducing into the plant a recombinant DNA construct that includes a heterologous promoter operably linked to a polynucleotide that encodes a RTH5 family member. The method can be performed in monocots and dicots. Also, disclosed herein are plants and expression cassettes that include recombinant DNA that contains a heterologous promoter operably linked to polynucleotides that encodes an RTH5 family member.

WO 2015/084969 A1

Published:

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

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

PLANTS WITH IMPROVED DROUGHT TOLERANCE

CROSS-REFERENCE TO RELATED APPLICATIONS

This Application claims the benefit of U.S. Provisional Application No. 61/911,275, filed on December 3, 2013 and the benefit of U.S. provisional Application No. 61/911,826, filed on December 4, 2013. Each application is hereby incorporated by reference in its entirety.

REFERENCE TO SEQUENCE LISTING

The Sequence Listing submitted on December 3, 2014 as a text file named "36446_0084P1_Sequence_Listing.txt," created on November 25, 2014, and having a size of 92,427 bytes is hereby incorporated by reference.

FIELD

The field of disclosure relates to plant breeding and genetics and, in particular, relates to recombinant DNA constructs useful in plants for conferring improved agronomic traits.

BACKGROUND

Improving agronomic traits in crop plants is beneficial to farmers. Several factors influence crop yield. Abiotic stress is the primary cause of crop loss worldwide, causing average yield losses of more than 50% for major crops. Among the various abiotic stresses, drought is a major factor that limits crop productivity worldwide. Exposure of plants to a water-limiting environment during various developmental stages appears to activate various physiological and developmental changes. Molecular mechanisms of abiotic stress responses and the genetic regulatory networks of drought stress tolerance have been studied.

Natural responses to abiotic stress vary among plant species and among varieties and cultivars within a plant species. Certain species, varieties or cultivars are more tolerant to abiotic stress such as drought than others. Transgenic approaches are needed for improving drought tolerance in crop plants.

SUMMARY

Compositions and methods to increase drought tolerance and to increase nutrient uptake in plants are disclosed. Compositions comprise recombinant RTH5 family polynucleotides, RTH5 polypeptides, expression cassettes, plants and seeds. The methods of

the invention comprise increasing the expression of an RTH5 polypeptide in a plant of interest. Any method for increasing the expression of the RTH5 polypeptide is encompassed. That is plants can be transformed with a DNA construct comprising an RTH5 polynucleotide operably linked with a heterolgous promoter that drives expression in plant roots. Alternatively, expression levels of the endogenous RTH5 polypeptide in the plant can be increased by methods available in the art to enhance the expression of endogenous genes. Expression constructs comprising an RTH5 polynucleotide as well as plants and seed having increased levels of an RTH5 polypeptide are provided.

Expression of RTH5 polypeptides increases root hair formation and growth. Increased expression of RTH5 polypeptides in the root of a plant thus results in increased drought tolerance and increased nutrient uptake.

The following embodiments are encompassed by the present invention:

1. A method for increasing root hair formation and growth in a plant, said method comprising increasing the expression of an RTH5 polypeptide in roots of said plant.

2. The method of embodiment 1, wherein said RTH5 polypeptide is an endogenous RTH5 polypeptide.

3. The method of embodiment 2, wherein said expression is increased by alterations to the genome of the plant said alterations comprising additions, deletions, and/or substitutions of nucleotides into the genome.

4. The method of embodiment 1, wherein said expression is increased by transforming said plant with a recombinant DNA construct comprising a polynucleotide encoding an RTH5 polypeptide operably linked to a heterologous promoter that drives expression in a plant root.

5. A method for increasing drought tolerance in a plant, said method comprising introducing into said plant a recombinant DNA construct, said construct comprising a heterologous promoter that drives expression in a plant root operably linked to a polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

6. A method for enhancing nutrient uptake in a plant, said method comprising introducing into said plant a recombinant DNA construct, said construct comprising a heterologous promoter that drives expression in a plant root operably linked to a polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

7. The method of embodiment 5 or 6, wherein said polynucleotide is stably integrated into the genome of the plant.

8. The method of any one of embodiments 5-7, wherein said plant is a plant cell.

9. The method of any one of embodiments 5-8, wherein said plant is a dicot.

10. The method of embodiment 9, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

11. The method of any one of embodiments 5-8, wherein said plant is a monocot.

12. The method of embodiment 11, wherein said monocot is maize, sugarcane, wheat, rice, barley, sorghum, millet, switchgrass, or rye.

13. A seed of the plant produced by the method of any one of embodiments 1-12.

14. A plant comprising a recombinant DNA construct comprising a heterologous promoter that drives expression in a plant root operably linked to a polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

15. The plant of embodiment 14, wherein said plant is a cell.

16. The plant of embodiment 14 or 15, wherein said plant is a monocot.

17. The plant of embodiment 16, wherein said monocot is maize, sugarcane, wheat, rice, barley, sorghum, millet, switchgrass, or rye.

18. The plant of embodiment 14 or 15, wherein said plant is a dicot.

19. The plant of embodiment 18, wherein the dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

20. The plant of any one of embodiments 14-19, wherein said polynucleotide is stably incorporated into the genome of the plant.

21. A seed of the plant of embodiment 20.

22. An expression cassette comprising a polynucleotide operably linked to a heterologous promoter that drives expression in a plant root, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

23. The expression cassette of embodiment 22, wherein said plant is a monocot.

24. The expression cassette of embodiment 23, wherein said monocot is maize, sugarcane, wheat, rice, barley, sorghum, millet, switchgrass, or rye.

25. The expression cassette of embodiment 22, wherein said plant is a dicot.

26. The expression cassette of embodiment 25, wherein the dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

27. A method of selecting or identifying an allelic variant of rth5 in a maize plant, the method comprising the steps of:

a. obtaining a population of maize plants, wherein said maize plants exhibit an alteration of at least one agronomic characteristic; wherein the at least one agronomic characteristic is selected from the group consisting of increased root hair formation and growth, increased drought tolerance, and enhanced nutrient uptake;

b. evaluating allelic variations with respect to the polynucleotide sequence encoding a protein comprising SEQ ID NO:3, or in the genomic region that regulates the expression of the polynucleotide encoding the protein;

c. associating allelic variations with said alteration of at least one agronomic characteristic; and

d. selecting or identifying an allelic variant that is associated with said alteration of at least one agronomic characteristic.

28. A method of selecting or identifying a first maize plant or a first maize germplasm that has one or more beneficial alleles of rth5, the method comprising:

(a) screening a plurality of maize plants or a plurality of maize germplasm for at least one polymorphism within a marker locus, wherein the marker locus is:

(a) a first polynucleotide having at least 90% and less than 100% nucleotide sequence identity with SEQ ID NO:1 or 2; or

(b) a second polynucleotide encoding a polypeptide comprising an amino acid sequence having at least 90% and less than 100% sequence identity to SEQ ID NO:3, wherein expression of the first or second polynucleotide in a maize plant results in a phenotype comprising an alteration of at least one agronomic characteristic when compared to a control maize plant; wherein the at least one agronomic characteristic is selected from the group consisting of increased root hair formation and growth, increased drought tolerance, and enhanced nutrient uptake; and wherein the control maize plant comprises:

(c) a polynucleotide having the nucleotide sequence of SEQ ID NO:1 or 2; or

(d) a polynucleotide encoding the amino acid sequence of SEQ ID NO:3;

(b) identifying a first maize plant or a first maize germplasm comprising the at least one polymorphism of the marker locus; and

(c) selecting the first maize plant or first maize germplasm of step (b).

29. A method of reducing soil erosion in a crop field, said method comprising introducing into a plant a recombinant DNA construct, said construct comprising a heterologous promoter that drives expression in a plant root operably linked to a

polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

30. A method of increasing infection of a plant with a desirable microorganism, said method comprising introducing into said plant a recombinant DNA construct, said construct comprising a heterologous promoter that drives expression in a plant root operably linked to a polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

31. The method of embodiment 30, wherein the beneficial microorganism is rhizobium or *Pseudomonas putida* KT2440.

32. A method of increasing the resistance to root pathogens in a plant, said method comprising reducing the expression of an RTH5 family member, or inhibiting the function of the RTH5 family member.

33. The method of embodiment 32, wherein the maize root pathogen is *Fusarium* species, like *F. verticillioides* or *F. graminearu.*

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTING

The disclosure can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing, which form a part of this application.

Figure 1 shows the mapping of *rth5*. (A) Map-based cloning experiments positioned the *rth5* gene within a 280 kb interval of chromosome 3. (B) Sequencing revealed a G to A (G>A) substitution in the last exon of candidate gene GRMZM2G426953 resulting in a cysteine by tyrosine (C>Y) substitution in the protein sequence for the *rth5-1* allele at position 821. The positions of *Mu* transposon insertions in three additional alleles, *rth5-2*, *rth5-3*, and *rth5-4*, identified in transposon tagged populations, are shown. (C) Alignment of the C-terminal 50 amino acids of human, yeast and Arabidopsis NADPH oxidases and wild-type and mutant RTH5. TM: trans-membrane, NADPHox: NADPH oxidase domain, EF hand: Ca^{2+} binding site, Fe red: Ferric reduction domain, FAD / NAD: FAD / NAD binding site, Hs: *Homo sapiens*, Sc: *Saccharomyces cerevisiae*, At: *Arabidopsis thaliana*, Zm: *Zea mays*.

Figure 2 shows a phylogenetic reconstruction of NADPH oxidases of maize, rice, Arabidopsis, and soybean. The phylogenetic relations of RTH5 were reconstructed by blasting the RTH5 sequence against the maize (*Zea mays*, maizesequence.org), rice (*Oryza sativa*, rice.plantbiology.msu.edu), soybean (*Glycine max*, soybean.org) and *Arabidopsis thaliana* (arabidopsis.org) databases. MEGA4 was used for alignment and tree generation via the Neighbor Joining algorithm. Two subfamilies were discriminated by presence or absence of several conserved domains in group I and II . Monocot specific clades / sub clades are highlighted with gray boxes.

Figure 3 shows an alignment of RTH5 and its homologs from maize, soybean, rice and Arabidopsis. Proteins are listed according to their order in the phylogenetic tree (Figure 2). Sequences were aggregated in units of ten amino acids (aa). Five and more aa per unit are depicted as boxes, >5 as lines. Color code: red - conserved in both groups, blue - conserved in group I, green - conserved in group II, dark gray - annotated functional domain, light gray - no functional annotation.

Figure 4 shows tissue-specific expression of *rth5* transcripts and accumulation of RTH5 proteins. (A) Expression of the *rth5* gene family members in root hairs. (B) *rth5* expression in different maize tissues. All expression levels are relative to the *Myosin* gene (Genbank AC: 486090G09.x1). Error bars indicate standard deviation; fold changes are indicated relative to root hairs. *P*-value obtained via Student's t-test, *p*-value * ≤ 0.05, n = 3.

RNA *in situ* hybridization of *rth5* antisense (C) and sense (D) probes on paraffin- embedded cross-sections. (E-J) Immunohistochemistry was performed on paraffin-embedded cross (E, F) and longitudinal sections of the differentiation (G, H) or elongation zone (I, J). (E, C, F) Sections incubated with anti-RTH5 antibody. (B, D, F) Control sections incubated with buffer. Arrows indicate root hairs. Scale bars in C, D: 50 µm, E-J: 100 µm. For all experiments three-days-old B73 seedling were used.

Figure 5 shows the expression of the *rth5* gene in various tissues in the normalized expression value of FPKM (fragments per kilobase of exon per million of fragments mapped). The data was downloaded from qteller.com as of 4/1/2013.

Figure 6 shows an expression analysis of maize *rboh* genes. Quantitative RT-PCR analysis of different root tissues of all maize *rboh* gene family members in three-day-old B73 seedlings is shown. Coleoptilar nodes and leaves were harvested from seven-day-old seedlings. Expression was calculated (n = 3) relative to reference gene *myosin*. Error bars represent standard deviation. Please note the different relative expression levels on the Y-axes of the different genes have been adjusted to highlight tissue-specific expression differences for individual genes.

Figure 7 is a schematic model for the predicted function of RTH5 in root hair initiation and growth. (A) Shows the suggested function of RTH5 in root hair initiation. A trichoblast-specific signal can activate the RTH5 protein through a small GTPase or calcium-dependent protein kinase (CDPK) to facilitate trichoblast differentiation. (B) Illustration of the role of RTH5 in tip growth by producing apoplastic superoxide, which is rapidly converted into hydrogen peroxide. Apoplastic peroxidases (PER) generate hydroxyl radicals, which are cleaving celluloses and hemicelluloses leading to cell wall loosening. The acidic cell wall (CW) allows oriented growth at the softened site. PM: plasma membrane.

Summary of SEQ ID NOS

Description and Abbreviation	SEQ ID NO
Maize (*rth5* genomic sequence) (Maize Genetics and Genomics Database Gene ID: RMZM2G426953)	1
Maize (*rth5* cDNA sequence) (GenBank Accession No: DQ855284)	2
Maize (RTH5 amino acid sequence) (GenBank Accession No: DQ855284)	3
GRMZM2G401179_P01_variant	4
GRMZM2G323731_P01 (amino acid)	5
GRMZM2G401179_P01	6

At3g45810.1	7
At5g60010.1	8
LOC_Os01g61880.1	9
Glyma07g15690.1	10
Glyma18g39500.1	11
LOC_Os05g38980.1	12
Glyma18g39500.1_variant	13

DETAILED DESCRIPTION

Methods for increasing the expression of an RTH5 polypeptide in a plant are provided. The methods comprise transforming the plant with a DNA construct comprising an RTH5 polynucleotide operably linked to a heterologous promoter that drives expression in a plant root. Alternatively the plant can be altered to increase the levels of expression of the endogenous RTH5 polypeptide. Such methods for altering the expression of the endogenous gene are known in the art and are encompassed by the present invention.

Using the methods and compositions of the present invention, drought tolerance and/or the uptake of nutrients in a plant can be increased. In particular, polynucleotides that encode polypeptides from the Roothairless5 (RTH5) family are provided. Polypeptides from the RTH5 family promote root hair formation and growth.

Root hairs are important for the uptake of water and nutrients in plants. As shown herein, the ROOTHAIRLESS5 (RTH5) family of polynucleotides and polypeptides are important for the formation of root hairs. Expression of RTH5 polypeptides in plant roots promotes the formation and growth of root hairs, thus improving water and nutrient uptake by the plant.

As referred to herein, the "RTH5 family" refers to a family of polynucleotides and polypeptides that have high sequence identity with and include the *Zea mays* L. (Maize) *roothairless5* (*rth5*) gene, cDNA, or polypeptide described herein as SEQ ID NOs: 1, 2, and 3, respectively, and promote the formation and growth of root hairs in a plant. An "RTH5 family member" refers to a specific polynucleotide or polypeptide included in the RTH5 family (an RTH5 polynucleotide or RTH5 polypeptide, respectively). As an example, maize *roothairless5* is an RTH5 family member. The maize "*roothairless5*" and "*rth5*" genomic sequence, cDNA, and polypeptide are provided in SEQ ID NOs: 1, 2, and 3, respectively.

Polynucleotides of the invention are polynucleotides encoding RTH5 family members and include without limitation, the polynucleotides corresponding to SEQ ID NOs: 1 and 2, as well as active fragments and variants thereof. Polypeptides of the invention include

members of the RTH5 family. Such RTH5 polypeptides include those set forth in SEQ ID NOs: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13, and active fragments and variants thereof. RTH5 family members include polynucleotides having sequences with about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, 99.5% or more sequence identity to SEQ ID NO: 2 and which encode RTH 5 polypeptides that have the ability to promote root hair formation and growth in a plant. RTH5 family members also include polypeptides having sequences with about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, 99.5% or more sequence identity to SEQ ID NO: 3 and which promote root hair formation and growth.

The RTH5 family demonstrates nicotinamide adenine dinucleotide phosphate-oxidase (NADPH oxidase or NOX) activity. As referred to herein, "NADPH oxidase activity" or "NOX activity" refers to the enzymatic activity of the polypeptide that produces the reactive oxygen species (ROS) superoxide (O_2^-) using an electron donor. Without wishing to be limited to a particular mechanism of action, ROS and ROS-related proteins are thought to play an important role in root hair tip growth. The methods of the invention utilize RTH5 family polynucleotides to increase the expression of RTH5 family polypeptides, which advantageously increase root hair formation and growth.

In general, concentration and/or activity of the RTH5 family member is increased by at least 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell which did not have the sequence of the invention introduced. The increased expression in the present invention can occur during and/or subsequent to growth of the plant to the desired stage of development. In specific embodiments, the levels and/or activity of RTH5 polypeptides of the present invention are increased in monocots, particularly maize.

The expression level of the RTH5 family polypeptide may be measured directly, for example, by assaying for the level of the RTH5 family polypeptide in the plant, or indirectly, for example, by measuring the NAPDH oxidase activity of the RTH5 family polypeptide in the plant. Methods of assaying for NOX activity are known in the art. For example, NADPH oxidase activity can be measured by such methods as detecting superoxide and/or hydrogen peroxide in the roots, as disclosed in the experimental section and as described in the art. See, for example, Foreman *et al.* (2003) *Nature* 422:442-446, herein incorporated by reference.

Root hair cells include long tubular projections referred to herein as “root hairs.” Root hairs are thought to aid plants in nutrient uptake, anchorage, and microbial interactions. Root hair growth is divided into three phases: first, defined swelling to form a bulge; second, transition to tip growth; and finally, tip growth by oriented exocytosis. Root hairs increase the surface area on a plant root, thereby increasing the ability of the root to take up water and nutrients.

Root hair formation and growth can be measured by phenotypic analysis. For example, root hair formation and growth can be measured using microscopy techniques described in the experimental section and as described by Foreman *et al.* (2003) *Nature* 422:442-446, herein incorporated by reference.

The formation and growth of root hairs leads to increased drought tolerance. As referred to herein, “drought" refers to a decrease in water availability to a plant that, especially when prolonged, can cause damage to the plant or prevent its successful growth (e.g., limiting plant growth or seed yield). Accordingly, “drought tolerance” is a trait of a plant to survive under drought conditions over prolonged periods of time without exhibiting substantial physiological or physical deterioration.

Embodiments of the present methods and compositions promote "increased drought tolerance" of a plant. Increased drought tolerance is measured relative to a reference or control plant, and is a trait of the plant to survive under drought conditions over prolonged periods of time, without exhibiting the same degree of physiological or physical deterioration relative to the reference or control plant grown under similar drought conditions. Typically, when a transgenic plant comprising a recombinant DNA construct in its genome exhibits increased drought tolerance relative to a reference or control plant due to the presence of the construct, the reference or control plant does not comprise in its genome the recombinant DNA construct.

One of ordinary skill in the art is familiar with protocols for simulating drought conditions and for evaluating drought tolerance of plants that have been subjected to simulated or naturally occurring drought conditions. For example, one can simulate drought conditions by giving plants less water than normally required or no water over a period of time, and one can evaluate drought tolerance by looking for differences in physiological and/or physical condition, including (but not limited to) vigor, growth, size, or root length, or in particular, leaf color or leaf area size. Other techniques for evaluating drought tolerance include measuring chlorophyll fluorescence, photosynthetic rates and gas exchange rates. See, for example, WO 2013/006345 herein incorporated by reference in its entirety.

One can also evaluate drought tolerance by the ability of a plant to maintain sufficient yield (at least about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% yield) in field testing under simulated or naturally occurring drought conditions (e.g., by measuring for substantially equivalent yield under drought conditions compared to non-drought conditions, or by measuring for less yield loss under drought conditions compared to a control or reference plant).

In addition, increased formation and growth of root hairs leads to increased nutrient uptake. As referred to herein, "nutrient uptake" refers to a plant's ability it to remove nutrients from a soil or growth medium. "Increased nutrient uptake" refers to a plant's ability to remove nutrients from a soil or growth medium relative to a reference or control plant, and is a trait of the plant wherein the plant demonstrates an improved agronomic characteristic compared to the reference or control plant. Typically, when a transgenic plant comprising a recombinant DNA construct in its genome exhibits increased nutrient uptake relative to a reference or control plant due to the presence of the construct, the reference or control plant does not comprise in its genome the recombinant DNA construct. Nonlimiting examples of nutrients taken up by plant roots include nitrogen, phosphorus, potassium, and carbon.

"Agronomic characteristic" or "agronomic parameter" is a measurable trait including but not limited to, abiotic stress tolerance, greenness, yield, growth rate, biomass, fresh weight at maturation, dry weight at maturation, fruit yield, seed yield, total plant nitrogen content, fruit nitrogen content, seed nitrogen content, nitrogen content in a vegetative tissue, total plant free amino acid content, fruit free amino acid content, seed free amino acid content, free amino acid content in a vegetative tissue, total plant protein content, fruit protein content, seed protein content, protein content in a vegetative tissue, drought tolerance, nitrogen uptake, root lodging, harvest index, stalk lodging, plant height, ear height, ear length, salt tolerance, early seedling vigor and seedling emergence under low temperature stress.

It is understood and herein contemplated that the plants and germplasms disclosed herein can be identified as having an alternation in at least one agronomic characteristic through the identification of a polymorphism in a polynucleotide having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, or 99.5% identity to SEQ ID NO: 1 or 2 or the identification of a polymorphism in a polynucloeotide that encodes a polypeptide having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, or 99.5% identity to SEQ ID NOs: 3, 4, 5, 6, 7, 8, 910, 11, 12, or 13. Thus, in one aspect, disclosed

herein are methods of selecting or identifying a first maize plant or a first maize germplasm that has one or more beneficial alleles of rth5, the method comprising:

(a) screening a plurality of maize plants or a plurality of maize germplasm for at least one polymorphism within a marker locus, wherein the marker locus is:

(a) a first polynucleotide having at least 90% and less than 100% nucleotide sequence identity with SEQ ID NO:1 or 2; or

(b) a second polynucleotide encoding a polypeptide comprising an amino acid sequence having at least 90% and less than 100% sequence identity to SEQ ID NO:3, wherein expression of the first or second polynucleotide in a maize plant results in a phenotype comprising an alteration of at least one agronomic characteristic when compared to a control maize plant; wherein the at least one agronomic characteristic is selected from the group consisting of increased root hair formation and growth, increased drought tolerance, and enhanced nutrient uptake; and wherein the control plant comprises:

(c) a polynucleotide having the nucleotide sequence of SEQ ID NO:1 or 2; or

(d) a polynucleotide encoding the amino acid sequence of SEQ ID NO:3

(b) identifying a first maize plant or a first maize germplasm comprising the at least one polymorphism of the marker locus; and

(c) selecting the first maize plant or first maize germplasm of step (b).

Abiotic stress can be at least one condition selected from the group consisting of: drought, water deprivation, flood, high light intensity, high temperature, low temperature, salinity, etiolation, defoliation, heavy metal toxicity, anaerobiosis, nutrient deficiency, nutrient excess, UV irradiation, atmospheric pollution (e.g., ozone) and exposure to chemicals (e.g., paraquat) that induce production of reactive oxygen species (ROS).

"Increased stress tolerance" of a plant is measured relative to a reference or control plant, and is a trait of the plant to survive under stress conditions over prolonged periods of time, without exhibiting the same degree of physiological or physical deterioration relative to the reference or control plant grown under similar stress conditions.

A plant with "increased stress tolerance" can exhibit increased tolerance to one or more different stress conditions.

"Stress tolerance activity" of a polypeptide indicates that over-expression of the polypeptide in a transgenic plant confers increased stress tolerance to the transgenic plant relative to a reference or control plant.

Methods of assaying increased nutrient uptake are known in the art. For example, one can grow a transgenic plant comprising a recombinant DNA construct and evaluate nutrient

uptake by looking for differences in physiological and/or physical condition, including (but not limited to) vigor, growth, size, or root length, leaf color, leaf area size, or crop yield. Other methods include measurement of the kinetics of plant root nutrient uptake. See, for example, the methods reviewed by Bassirirad (2000) *New Phytol.* 147:155-169, incorporated by reference herein.

In another embodiment, increased root hair formation and growth can reduce soil erosion in crop fields. It is recognized that root hairs help cling to soil particles and are therefore important to preventing soil erosion.

In addition, increased root hair formation and growth is expected to promote the infection of the plant with beneficial microorganisms, such as rhizobium and *Pseudomonas putida* KT2440. Rhizobia enter the plant via root hairs, which results in root nodule formation and increased nitrogen fixation.

In addition, plants are known to interact with a wide range of rhizosphere-colonizing bacteria. These are attracted to root surfaces by chemical components in root exudates, which are rapidly assimilated into microbial biomass (Rangel-Castro JI, *et al.* (2005) *Environ Microbiol* 7: 828–838). This so-called rhizosphere effect supports bacterial cell densities in the root vicinity up to 100-fold greater than in surrounding soil (Whipps JM *et al.* (2001) *J Exp Bot* 52: 487–511).

In another embodiment, it is recognized that the reduction of RTH5 family expression or mutation in RTH5 family members can result in reduced susceptibility of the plant to plant pathogens which infect the plant roots. For example, *Barssica* are susceptible to *Plasmodiophora brassicae*, which is the pathogen responsible for clubroot. Clubroot is caused when *P. brassicae* enter the plant via the root hairs. Therefore, it is thought that reducing the number of root hairs can limit the mode of entry of plant pathogens.

The sequences disclosed herein as SEQ ID NOs: 2 and 3 have been previously identified as *respiratory burst oxidase homolog* A (*rboh*A). See, Lin *et al.* (2009) *J. Exp. Bot.* 60:3221-3238. While regulation of the *rboh*A encoded NADPH oxidase (NOX) by a mitogen-activated protein kinase cascade signaling was demonstrated in maize leaves, rth5/rbohA has not been previously identified to be involved in root hair formation and growth.

Thus, in one aspect, disclosed herein are methods of increasing the resistance to root pathogens in a plant, said method comprising reducing the expression of an RTH5 family member, or inhibiting the function of the RTH5 family member. Maize root pathogens can be any organzims that cause a dileterious effecto on a plant and can infect the plant throught

the root or root hairs. On non-limiting list of maize root pathogens include, but are not limited to, *Fusarium* species, like *F. verticillioides* and *F. graminearu.*

The RTH5 polynucleotides disclosed herein can be provided in expression cassettes for expression in a plant of interest. An "expression cassette" comprises a recombinant nucleic acid that is operably linked to a heterologous promoter that drives expression of the nucleic acid. A "recombinant" nucleic acid is made by a combination of two otherwise separated segments of nucleic acid sequence, for example, by chemical synthesis or by the manipulation of isolated segments of polynucleic acids by genetic engineering techniques. The term "recombinant DNA construct" refers to any agent such as a plasmid, cosmid, virus, autonomously replicating sequence, phage, or linear or circular single-stranded or double-stranded DNA or RNA nucleotide sequence, derived from any source, capable of genomic integration or autonomous replication, comprising a DNA molecule where one or more DNA sequences have been linked in a functionally operative manner. Such recombinant DNA constructs are capable of introducing a 5′ regulatory sequence or promoter region and a DNA sequence for a selected gene product into a cell in such a manner that the DNA sequence is transcribed into a functional mRNA that is translated and therefore expressed. A recombinant DNA construct (and an expression cassette) may comprise a polynucleotide of interest linked to a heterologous polynucleotide such as, for example, a heterologous promoter.

The cassette can include 5' and 3' regulatory sequences operably linked to an RTH5 polynucleotide. "Operably linked" is intended to mean a functional linkage between two or more elements. For example, an operable linkage between a polynucleotide of interest and a regulatory sequence (i.e., a promoter) is functional link that allows for expression of the polynucleotide of interest. Operably linked elements may be contiguous or non-contiguous. When used to refer to the joining of two protein coding regions, by operably linked is intended that the coding regions are in the same reading frame. The cassette may additionally contain at least one additional gene to be cotransformed into the organism. Alternatively, the additional gene(s) can be provided on multiple expression cassettes. Such an expression cassette is provided with a plurality of restriction sites and/or recombination sites for insertion of the RTH5 polynucleotide to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

In preferred embodiments, the expression cassette includes a polynucleotide operably linked to a root-preferred heterologous promoter. The polynucleotide can include a

nucleotide sequence encompassed by SEQ ID NO: 2. In some embodiments the nucleotide sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, or 99.5% sequence identity to SEQ ID NO: 2, and the polynucleotide encodes a polypeptide having NADPH oxidase activity. In some embodiments the nucleotide sequence is the sequence of SEQ ID NO: 2.

In other embodiments, the expression cassette includes a nucleotide sequence that encodes the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13. In some embodiments, the nucleotide sequence encodes an amino acid sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, or 99.5% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide that has NADPH oxidase activity. In some embodiments the polynucleotide encodes a polypeptide having the amino acid sequence of SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13.

The expression cassette can include in the 5'-3' direction of transcription, a transcriptional and translational initiation region (i.e., a promoter), an RTH5 family polynucleotide as disclosed herein, and a transcriptional and translational termination region (i.e., termination region) functional in plants. The regulatory regions (i.e., promoters, transcriptional regulatory regions, and translational termination regions) and/or the RTH5 family polynucleotide of the invention may be native/analogous to the host cell or to each other. Alternatively, the regulatory regions and/or the RTH5 family polynucleotide of the invention can be heterologous to the host cell or to each other. As used herein, "heterologous" in reference to a sequence is a sequence that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous polynucleotide is from a species different from the species from which the polynucleotide was derived, or, if from the same/analogous species, one or both are substantially modified from their original form and/or genomic locus, or the promoter is not the native promoter for the operably linked polynucleotide. As used herein, a "chimeric gene" comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence.

While it may be optimal to express the sequences using heterologous promoters, the native promoter sequences can be used. Such constructs can change expression levels of RTH5 family proteins in the plant or plant cell. Thus, the phenotype of the plant or plant cell can be altered.

The termination region can be native with the transcriptional initiation region, can be native with the operably linked RTH5 family polynucleotide of interest, can be native with the plant host, or can be derived from another source (i.e., foreign or heterologous) to the promoter, the RTH5 polynucleotide of interest, the plant host, or any combination thereof. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acids Res.* 15:9627-9639.

Where appropriate, the polynucleotides may be optimized for increased expression in the transformed plant. That is, the polynucleotides can be synthesized using plant-preferred codons for improved expression. See, for example, Campbell and Gowri (1990) *Plant Physiol.* 92:1-11 for a discussion of host-preferred codon usage. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

Additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression. The G-C content of the sequence can be adjusted to levels average for a given cellular host, as calculated by reference to known genes expressed in the host cell. When possible, the sequence is modified to avoid predicted hairpin secondary mRNA structures.

The expression cassettes may additionally contain 5' leader sequences. Such leader sequences can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Gallie *et al.* (1995) *Gene* 165(2):233-238), MDMV leader (Maize Dwarf Mosaic Virus) (*Virology* 154:9-20), and human immunoglobulin heavy-chain binding protein (BiP) (Macejak *et al.* (1991) *Nature* 353:90-94); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling *et al.* (1987) *Nature* 325:622-625); tobacco mosaic virus leader (TMV) (Gallie *et al.* (1989) in *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256);

and maize chlorotic mottle virus leader (MCMV) (Lommel *et al.* (1991) *Virology* 81:382-385). See also, Della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968.

In preparing the expression cassette, the various DNA fragments can be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers can be employed to join the DNA fragments or other manipulations can be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, can be involved.

Promoters useful in the methods of the invention include those promoters that drive expression of a polypeptide in roots of a plant. Constitutive promoters can be used such as the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO 99/43838 and U.S. Patent No. 6,072,050; the core CaMV 35S promoter (Odell *et al.* (1985) *Nature* 313:810-812); rice actin (McElroy *et al.* (1990) *Plant Cell* 2:163-171); ubiquitin (Christensen *et al.* (1989) *Plant Mol. Biol.* 12:619-632 and Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689); pEMU (Last *et al.* (1991) *Theor. Appl. Genet.* 81:581-588); MAS (Velten *et al.* (1984) *EMBO J.* 3:2723-2730); ALS promoter (U.S. Patent No. 5,659,026), and the like. Other constitutive promoters include, for example, U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; and 6,177,611.

A "root-preferred" promoter, as referred to herein, is a promoter which favors spatial expression of a polynucleotide of interest in the root of a plant compared to expression in other plant tissue within the same plant. Root-preferred promoters can be heterologous or native to the plant receiving the polynucleotide of interest. Root-preferred promoters are known and can be selected from the many available from the literature or isolated de novo from various compatible species. See, for example, Hire *et al.* (1992) *Plant Mol. Biol.* 20(2):207-218 (soybean root-specific glutamine synthetase gene); Keller and Baumgartner (1991) *Plant Cell* 3(10):1051-1061 (root-specific control element in the GRP 1.8 gene of French bean); Sanger *et al.* (1990) *Plant Mol. Biol.* 14(3):433-443 (root-specific promoter of the mannopine synthase (MAS) gene of *Agrobacterium tumefaciens*); and Miao *et al.* (1991) *Plant Cell* 3(1):11-22 (full-length cDNA clone encoding cytosolic glutamine synthetase (GS), which is expressed in roots and root nodules of soybean). See also Bogusz *et al.* (1990) *Plant Cell* 2(7):633-641, where two root-specific promoters isolated from hemoglobin genes from the nitrogen-fixing nonlegume *Parasponia andersonii* and the related non-nitrogen-fixing

nonlegume *Trema tomentosa* are described. Leach and Aoyagi (1991) describe their analysis of the promoters of the highly expressed rolC and rolD root-inducing genes of *Agrobacterium rhizogenes* (see *Plant Science* (Limerick) 79(1):69-76). They concluded that enhancer and tissue-preferred DNA determinants are dissociated in those promoters. Teeri *et al.* (1989) used gene fusion to lacZ to show that the *Agrobacterium* T-DNA gene encoding octopine synthase is especially active in the epidermis of the root tip and that the TR2' gene is root specific in the intact plant and stimulated by wounding in leaf tissue, an especially desirable combination of characteristics for use with an insecticidal or larvicidal gene (see *EMBO J.* 8(2):343-350). The TR1' gene, fused to *nptII* (neomycin phosphotransferase II) showed similar characteristics. Additional root-preferred promoters include the VfENOD-GRP3 gene promoter (Kuster *et al.* (1995) *Plant Mol. Biol.* 29(4):759-772); and rolB promoter (Capana *et al.* (1994) *Plant Mol. Biol.* 25(4):681-691. See also U.S. Patent Nos. 5,837,876; 5,837,848; 5,750,386; 5,633,363; 5,459,252; 5,401,836; 5,110,732; 5,023,179; and 7,554,005.

Other root preferred promoters include the following: the maize NAS2 promoter, the maize Cyclo promoter (US 2006/0156439, published July 13, 2006), the maize ROOTMET2 promoter (WO05063998, published July 14, 2005), the CR1BIO promoter (WO06055487, published May 26, 2006), the CRWAQ81 promoter (WO05035770, published April 21, 2005) and the maize ZRP2.47 promoter (NCBI accession number: U38790; GI No. 1063664),

The above list of promoters is not meant to be limiting. Any root-preferred promoter or promoter that drives expression in the root of a plant can be used with the polynucleotides disclosed herein.

The expression cassette can also comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Many selectable markers are known in the art and any can be used in the practice of the invention.

Embodiments found herein encompass isolated, substantially purified, or recombinant polynucleotide or protein compositions. An "isolated" or "purified" polynucleotide or protein, or biologically active portion thereof, is substantially or essentially free from components that normally accompany or interact with the polynucleotide or protein as found in its naturally occurring environment. Thus, an isolated or purified polynucleotide or protein is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Optimally, an "isolated" polynucleotide is free of sequences

(optimally protein encoding sequences) that naturally flank the polynucleotide (i.e., sequences located at the 5' and 3' ends of the polynucleotide) in the genomic DNA of the organism from which the polynucleotide is derived. A protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, optimally culture medium represents less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of chemical precursors or non-protein-of-interest chemicals.

Fragments and variants of the disclosed polynucleotides and proteins encoded thereby are also encompassed by the present invention. By "fragment" is intended a portion of the polynucleotide or a portion of the amino acid sequence and hence protein encoded thereby. Fragments of a polynucleotide can encode protein fragments that retain the biological activity of the native protein and hence promote root hair formation and growth. Alternatively, fragments of a polynucleotide that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, fragments of a nucleotide sequence may range from at least about 20 nucleotides, about 50 nucleotides, about 100 nucleotides, and up to the full-length polynucleotide encoding the proteins disclosed herein.

A fragment of an RTH5 family polynucleotide that encodes a biologically active portion of an RTH5 family protein disclosed herein can encode at least 15, 25, 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, or 850 contiguous amino acids, or up to the total number of amino acids present in a full-length RTH5 family protein, for example, 852, 789, 850, 774, 822, 842, 912, 886, 843, 799, 860, 819, and 820 amino acids for SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13, respectively. Fragments of an RTH5 family polynucleotide that are useful as hybridization probes or PCR primers generally need not encode a biologically active portion of an RTH5 family protein.

Thus, a fragment of an RTH5 family polynucleotide can encode a biologically active portion of an RTH5 family protein, or it can be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed below. A biologically active portion of an RTH5 family protein can be prepared by isolating a portion of one of the RTH5 family polynucleotide disclosed herein, expressing the encoded portion of the RTH5 family protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the RTH5 family protein. Polynucleotides that are fragments of an RTH5 family nucleotide sequence comprise at least 16, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1,000, 1,100, 1,200, 1,300, 1,400, 1,500, 1,600, 1,700, 1,800, 1,900,

2,000, 2,100, 2,200, 2,300, 2,400, 2,500, 2,600, 2,700, 2,800, 2,900, 3,000, 3,100, 3,200, 3,300, 3,400, 3,500, or 3,600 contiguous nucleotides, or up to the number of nucleotides present in a full-length RTH5 family polynucleotide disclosed herein, for example, 3,592, 3,691, 3,009, 3,674, and 2,974 nucleotides for SEQ ID NO: 2.

"Variants" is intended to mean substantially similar sequences. For polynucleotides, a variant comprises a deletion and/or addition of one or more nucleotides at one or more internal sites within the native polynucleotide and/or a substitution of one or more nucleotides at one or more sites in the native polynucleotide. As used herein, a "native" polynucleotide or polypeptide comprises a naturally occurring nucleotide sequence or amino acid sequence, respectively. For polynucleotides, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of one of the RTH5 family polypeptides as disclosed herein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant polynucleotides also include synthetically derived polynucleotide, such as those generated, for example, by using site-directed mutagenesis but which still encode a RTH5 family protein. Generally, variants of a particular polynucleotide of the invention will have at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, 99.5% or more sequence identity to that particular polynucleotide as determined by sequence alignment programs and parameters described elsewhere herein.

Variants of a particular polynucleotide of the invention (i.e., the reference polynucleotide) can also be evaluated by comparison of the percent sequence identity between the polypeptide encoded by a variant polynucleotide and the polypeptide encoded by the reference polynucleotide. Thus, for example, isolated polynucleotides that encodes a polypeptide with a given percent sequence identity to the polypeptide of SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13, respectively, are disclosed. Percent sequence identity between any two polypeptides can be calculated using sequence alignment programs and parameters described elsewhere herein. Where any given pair of polynucleotides of the invention is evaluated by comparison of the percent sequence identity shared by the two polypeptides they encode, the percent sequence identity between the two encoded polypeptides is at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, 99.5% or more sequence identity.

"Variant" protein is intended to mean a protein derived from the native protein by deletion or addition of one or more amino acids at one or more internal sites in the native protein and/or substitution of one or more amino acids at one or more sites in the native protein. Variant proteins encompassed by the present invention are biologically active, that is they continue to possess the desired biological activity of the native protein, that is, root hair formation and growth and/or NADPH oxidase activity as described herein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Biologically active variants of a native RTH5 family protein will have at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.3%, 99.5% or more sequence identity to the amino acid sequence for the native protein as determined by sequence alignment programs and parameters described elsewhere herein. A biologically active variant of a protein of the invention can differ from that protein by as few as 1-15 amino acid residues, as few as 1-10, as few as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue.

In one aspect, disclosed herein are methods of selecting or identifying an allelic variant of rth5 in a maize plant, the method comprising the steps of a) obtaining a population of maize plants, wherein said maize plants exhibit an alteration of at least one agronomic characteristic; wherein the at least one agronomic characteristic is selected from the group consisting of increased root hair formation and growth, increased drought tolerance, and enhanced nutrient uptake; b) evaluating allelic variations with respect to the polynucleotide sequence (for example by performing the sequence identity comparisons described above) encoding a protein comprising SEQ ID NO:3, or in the genomic region that regulates the expression of the polynucleotide encoding the protein; c) associating allelic variations with said alteration of at least one agronomic characteristic; and d) selecting or identifying an allelic variant that is associated with said alteration of at least one agronomic characteristic.

The proteins disclosed herein may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants and fragments of the RTH5 family proteins can be prepared by mutations in the DNA. Methods for mutagenesis and polynucleotide alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of

the protein of interest may be found in the model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be optimal.

Thus, the genes and polynucleotides of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the proteins of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants continue to possess the desired root hair formation and growth activity and/or NADPH oxidase activity. The mutations that are made in the DNA encoding the variant must not place the sequence out of reading frame and optimally do not create complementary regions that produce secondary mRNA structure. See, EP Patent Application Publication No. 75,444.

The deletions, insertions, and substitutions of the protein sequences encompassed herein are not expected to produce radical changes in the characteristics of the protein. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect can be evaluated by routine screening assays. That is, the activity can be evaluated by, for example, phenotypic analysis of root hair formation and growth and/or by assaying for NADPH oxidase activity.

Variant polynucleotides and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different RTH5 family members coding sequences can be manipulated to create a new RTH5 family member possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. For example, using this approach, sequence motifs encoding a domain of interest can be shuffled between the *rth5* gene disclosed herein and other *rboh* and *rth* genes to obtain a new gene coding for a protein with an improved property of interest, such as an increased K_m in the case of an enzyme. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Crameri *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al. (1997) Proc. Natl. Acad. Sci. USA* 94:4504-4509; Crameri *et al.* (1998) *Nature* 391:288-291; and U.S. Patent Nos. 5,605,793 and 5,837,458.

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

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

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

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent sequence identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988) *CABIOS* 4:11-17; the local alignment algorithm of Smith *et al.* (1981) *Adv. Appl. Math.* 2:482; the global alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453; the search-for-local alignment method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448; the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 872264, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the GCG Wisconsin Genetics Software Package, Version 10 (available from Accelrys Inc., 9685 Scranton Road, San Diego, California, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (1988) *Gene* 73:237-244 (1988); Higgins *et al.* (1989) *CABIOS* 5:151-153; Corpet *et al.* (1988) *Nucleic Acids Res.* 16:10881-

90; Huang *et al.* (1992) *CABIOS* 8:155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol.* 24:307-331. The ALIGN program is based on the algorithm of Myers and Miller (1988) *supra*. A PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used with the ALIGN program when comparing amino acid sequences. The BLAST programs of Altschul *et al* (1990) *J. Mol. Biol.* 215:403 are based on the algorithm of Karlin and Altschul (1990) *supra*. BLAST nucleotide searches can be performed with the BLASTN program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the invention. BLAST protein searches can be performed with the BLASTX program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to a protein or polypeptide of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.* (1997) *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used. Alignment may also be performed manually by inspection.

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 using the following parameters: % identity and % similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight of 2, and the BLOSUM62 scoring matrix; or any equivalent program thereof. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

GAP uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty.

Default gap creation penalty values and gap extension penalty values in Version 10 of the GCG Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

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

(c) As used herein, "sequence identity" or "identity" in the context of two polynucleotides or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is

calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

Sequence alignments and percent identity calculations may be determined using the MEGALIGN® program of the LASERGENE® bioinformatics computing suite (DNASTAR® Inc., Madison, WI). Multiple alignment of the sequences provided herein may be performed using the Clustal V method of alignment (Higgins and Sharp (1989) CABIOS. 5:151 153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal V method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences, using the Clustal V program, it is possible to obtain “percent identity” and “divergence” values by viewing the “sequence distances” table on the same program; unless stated otherwise, percent identities and divergences provided and claimed herein were calculated in this manner.

Alternatively, the Clustal W method of alignment may be used. The Clustal W method of alignment (described by Higgins and Sharp, CABIOS. 5:151-153 (1989); Higgins, D. G. et al., Comput. Appl. Biosci. 8:189-191 (1992)) can be found in the MEGALIGN® v6.1 program of the LASERGENE® bioinformatics computing suite (DNASTAR® Inc., Madison, Wis.). Default parameters for multiple alignment correspond to GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergent Sequences=30%, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB. For pairwise alignments the default parameters are Alignment=Slow-Accurate, Gap Penalty=10.0, Gap Length=0.10, Protein Weight Matrix=Gonnet 250 and DNA Weight Matrix=IUB. After alignment of the sequences using the Clustal W program, it is possible to

obtain "percent identity" and "divergence" values by viewing the "sequence distances" table in the same program.

The use of the term "polynucleotide" is not intended to limit the present invention to polynucleotides comprising DNA. Those of ordinary skill in the art will recognize that polynucleotides can comprise ribonucleotides and combinations of ribonucleotides and deoxyribonucleotides. Such deoxyribonucleotides and ribonucleotides include both naturally occurring molecules and synthetic analogues. The polynucleotides of the invention also encompass all forms of sequences including, but not limited to, single-stranded forms, double-stranded forms, hairpins, stem-and-loop structures, and the like.

The methods of the invention involve introducing a polypeptide or polynucleotide into a plant. "Introducing" is intended to mean presenting to the plant the polynucleotide or polypeptide in such a manner that the sequence gains access to the interior of a cell of the plant. The methods of the invention do not depend on a particular method for introducing a sequence into a plant, only that the polynucleotide or polypeptides gains access to the interior of at least one cell of the plant. Methods for introducing polynucleotide or polypeptides into plants are known in the art including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

"Stable transformation" is intended to mean that the nucleotide construct introduced into a plant integrates into the genome of the plant and is capable of being inherited by the progeny thereof. "Transient transformation" is intended to mean that a polynucleotide is introduced into the plant and does not integrate into the genome of the plant or a polypeptide is introduced into a plant.

Transformation protocols as well as protocols for introducing polypeptides or polynucleotide sequences into plants may vary depending on the type of plant or plant cell, i.e., monocot or dicot, targeted for transformation. Suitable methods of introducing polypeptides and polynucleotides into plant cells include microinjection (Crossway *et al.* (1986) *Biotechniques* 4:320-334), electroporation (Riggs *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, *Agrobacterium*-mediated transformation (U.S. Patent No. 5,563,055 and U.S. Patent No. 5,981,840), direct gene transfer (Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, U.S. Patent Nos. 4,945,050; U.S. Patent No. 5,879,918; U.S. Patent No. 5,886,244; and, 5,932,782; Tomes *et al.* (1995) in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); McCabe *et al.* (1988) *Biotechnology* 6:923-926); and Lec1 transformation (WO 00/28058). Also see Weissinger *et al.* (1988) *Ann. Rev. Genet.*

22:421-477; Sanford *et al.* (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.* (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe *et al.* (1988) *Bio/Technology* 6:923-926 (soybean); Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh *et al.* (1998) *Theor. Appl. Genet.* 96:319-324 (soybean); Datta *et al.* (1990) *Biotechnology* 8:736-740 (rice); Klein *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein *et al.* (1988) *Biotechnology* 6:559-563 (maize); U.S. Patent Nos. 5,240,855; 5,322,783; and, 5,324,646; Klein *et al.* (1988) *Plant Physiol.* 91:440-444 (maize); Fromm *et al.* (1990) *Biotechnology* 8:833-839 (maize); Hooykaas-Van Slogteren *et al.* (1984) *Nature (London)* 311:763-764; U.S. Patent No. 5,736,369 (cereals); Bytebier *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet *et al.* (1985) in *The Experimental Manipulation of Ovule Tissues,* ed. Chapman *et al.* (Longman, New York), pp. 197-209 (pollen); Kaeppler *et al.* (1990) *Plant Cell Reports* 9:415-418 and Kaeppler *et al.* (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505 (electroporation); Li *et al.* (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford (1995) *Annals of Botany* 75:407-413 (rice); Osjoda *et al.* (1996) *Nature Biotechnology* 14:745-750 (maize via *Agrobacterium tumefaciens*); all of which are herein incorporated by reference.

Methods are known in the art for the targeted insertion of a polynucleotide at a specific location in the plant genome. In one embodiment, the insertion of the polynucleotide at a desired genomic location is achieved using a site-specific recombination system. See, for example, WO99/25821, WO99/25854, WO99/25840, WO99/25855, and WO99/25853, all of which are herein incorporated by reference. Briefly, the expression cassette disclosed herein can be contained in transfer cassette flanked by two non-recombinogenic recombination sites. The transfer cassette is introduced into a plant having stably incorporated into its genome a target site which is flanked by two non-recombinogenic recombination sites that correspond to the sites of the transfer cassette. An appropriate recombinase is provided and the transfer cassette is integrated at the target site. The polynucleotide of interest is thereby integrated at a specific chromosomal position in the plant genome.

It is therefore recognized that methods of the present invention do not depend on the incorporation of the entire polynucleotide into the genome, only that the plant or cell thereof is altered as a result of the introduction of the polynucleotide into a cell. In one embodiment of the invention, the genome may be altered following the introduction of the polynucleotide into a cell. For example, the polynucleotide, or any part thereof, may incorporate into the genome of the plant. Alterations to the genome of the present invention include, but are not

limited to, additions, deletions, and substitutions of nucleotides into the genome. While the methods of the present invention do not depend on additions, deletions, and substitutions of any particular number of nucleotides, it is recognized that such additions, deletions, or substitutions comprises at least one nucleotide.

As indicated, the invention also encompasses plants where the endogenous RTH5 expression is enhanced. Thus, the level and/or activity of an RTH5 polypeptide can be increased by altering the gene encoding the RTH5 family polypeptide or its promoter. See, e.g., Kmiec, U.S. Patent 5,565,350; and Zarling *et al.*, PCT/US93/03868. Therefore mutagenized plants that carry mutations in RTH5 family genes, where the mutations increase expression of the RTH5 family gene or increase the root hair formation and growth activity of the encoded RTH5 family polypeptide are provided.

Levels of endogenous RTH5 can be increased by methods known in the art. For example, endogenous RTH5 can be increased by the introduction of sequences of interest that upregulate endogenous expression. Such sequences of interest can include heterologous promoters, siRNA targeted to genomic regulatory elements, enhancer elements, and/or booster sequences. See, e.g., U. S. Patent 5,939,541; U.S. Patent 6,576,442; US 2012/0036594; and WO 1991/009955. Other methods of inserting sequences of interest to upregulate endogenous expression can include the use of various meganucleases to target polynucleotides. Such meganucleases are set forth in WO 2009/114321 (herein incorporated by reference), which describes “custom" meganucleases. See, also, Gao *et al.* (2010) *Plant Journal* 1:176-187. Additional methods of inserting sequences of interest to upregulate endogenous expression of a sequence of interest that can be employed, include but are not limited to the use of ZnFingers, meganucleases, and, TAL nucleases. See, for example, WO2010079430, WO2011072246, and US20110201118, each of which is herein incorporated by reference in their entirety.

Other embodiments include methods of detection and use of polymorphisms in the maize Rth5 promoter that are associated with increased expression of the RTH5 polypeptide.

The cells that have been transformed can be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. These plants can then be grown, and either pollinated with the same transformed strain or different strains, and the resulting progeny having constitutive expression of the desired phenotypic characteristic identified. Two or more generations can be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotypic characteristic has been

achieved. In this manner, the present invention provides transformed seed (also referred to as "transgenic seed") having a polynucleotide of the invention, for example, an expression cassette of the invention, stably incorporated into their genome.

As used herein, the term plant also includes plant cells, plant protoplasts, plant cell tissue cultures from which plants can be regenerated, plant calli, plant clumps, and plant cells that are intact in plants or parts of plants such as embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, and the like. Grain is intended to mean the mature seed produced by commercial growers for purposes other than growing or reproducing the species. Progeny, variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced polynucleotides.

In some embodiments, the composition is a seed of a plant comprising a recombinant DNA construct that includes a promoter that drives expression in roots of a plant operably linked to a polynucleotide that encodes an RTH5 family member as described herein.

The recombinant DNA constructs disclosed herein can be used for transformation of any plant species, including, but not limited to, monocots and dicots. Examples of plant species of interest include, but are not limited to, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus, B. rapa, B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor, Sorghum vulgare*), millet (e.g., 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*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), 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 americana*), 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, barley, vegetables, ornamentals, and conifers.

Vegetables include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea

(*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum.

Conifers that can be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). In specific embodiments, plants of the present invention are crop plants (for example, corn, alfalfa, sunflower, *Brassica*, soybean, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, etc.). In other embodiments, corn and soybean and sugarcane plants are optimal, and in yet other embodiments corn plants are optimal.

Another embodiment is a method for transforming a cell (or microorganism) comprising transforming a cell (or microorganism) with any of the isolated polynucleotides or recombinant DNA constructs of the present invention. The cell (or microorganism) transformed by this method is also included. In particular embodiments, the cell is eukaryotic cell, e.g., a yeast, insect or plant cell, or prokaryotic, e.g., a bacterial cell. The microorganism can be Agrobacterium, e.g. Agrobacterium tumefaciens or Agrobacterium rhizogenes.

Another embodiment is a recombinant DNA construct (and corresponding cells, plants, seeds and methods) comprising a promoter (heterologous or native) that drives expression in a plant root operably linked to a polynucleotide, wherein said polynucleotide encodes a polypeptide that comprises the following regions: four trans-membrane (TM) domains, two EF hand motifs, FAD and NAD cofactor binding sites, a ferric reductase domain, and the NADPH oxidase domain, wherein each of these regions have at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% amino acid sequence identity to the corresponding regions from SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polypeptide has NADPH oxidase activity. The polypeptide can be further characterized by having nine additional amino acids in the FAD cofactor domain, and by lacking 15 amino acids in the NADPH binding region, as compared to the dicot group I proteins (AT3G45810; AT5G60010; Glyma07g15690; Glyma18g39500) of the subclade that includes RTH5.

EXAMPLES

Example 1: Cloning and characterization of the maize *roothairless5* gene, which controls root hair initiation and outgrowth and encodes a monocot-specific NADPH oxidase involved in superoxide production

The roothairless5 gene controls both root hair length and density

An EMS mutagenesis screen yielded a recessive mutant (Schnable Lab Accession # 1350) that affects root hair development but has no other obvious effects on plant growth and development. Genetic crosses demonstrated that this mutant is not allelic to the previously described *rth1*, *rth2* and *rth3* mutants (Wen *et al.* (1994) *Am. J. Bot.* 81:833-842). The affected gene and reference allele were designated *rth5* and *rth5-1*, respectively. Compared to wild-type, the *rth5* mutant displays significantly shorter root hairs. Root hair length was determined by WinRhizo software (n=12) and density of wild-type and *rth5* primary roots was counted and calculated for per mm^2 (n=3), and then analyzed via eSEM (environmental scanning electron microscopy). *P*-values were obtained via Student's t-test. The length of root hairs on primary roots of 4-day-old (4 d) *rth5* mutants was significantly decreased to 4% of wild-type length (*p*-value <0.001). Moreover, in these mutant seedlings root hair density was significantly reduced to 64% of wild-type density (*p*-value <0.001). In 10 d seedlings it is possible to also observe root hairs on seminal and shoot-borne roots. The lengths and numbers of the root hairs on these roots including primary, seminal, crown and lateral were also significantly reduced in the mutant (*p*-value <0.001, n=12).

Map-based cloning of rth5 identified an NADPH oxidase

Using Sequenom-based bulked segregant analysis (BSA) (Liu *et al.* (2010) *Genetics* 184:19-26) the *rth5* locus was mapped to the long arm of chromosome 3. By analyzing F_2 and F_1BC populations (Methods) the *rth5* gene was mapped to the interval 180.1-180.4 Mb of chromosome 3 (Refgen1), flanked by IDP (insertion deletion polymorphism) markers IDP4064 (SEQ ID NO: 16) and C3.184743 (SEQ ID NO: 18) (Table 1). This interval harbors only five gene models (4a53), including GRMZM2G426953, an NADPH oxidase (NOX), which contains a G to A transition at position 2462, resulting in a Cysteine to Tyrosine conversion at amino acid position 821 close to the C-terminus of the protein (Figure 1B). This amino acid exchange is the result of a G-to-A transition relative to the B73 allele,

which is characteristic of EMS-induced mutations. This cysteine is conserved among all 26 inbred parents of the maize NAM population.

The affected cysteine residue is also conserved among divergent NADPH oxidases ranging from plants to yeast and human (Figure 1C). Hence, an alteration of this conserved residue in the predicted NAD substrate binding region can lead to a functional deficiency in the *rth5* mutant.

Table 1. Detailed Information of Markers Used in the Fine-Mapping Experiment

Marker	Chr_B73Ref1	Target_position_B73Ref1	Polymorphic_site_B73Ref1	Chr_B73Ref2	Target_position_B73Ref	Polymorphic_site_B73Ref2	Target_sequence	Polymorphism
SNP90373	chr3	177701752	177701852	chr3	179374007	179374117	ATCCATATTATATAGGAATGTTGGTCCAAGTGAAATTACCTTGCCACCAATAAAAGATGTTTGT	A/G
							CGCAGTGAAGAGGAGAAAATTGAGATGGAGGAGATAJA/G]AACTGACGCCGTATTATATAGG	
							AATACAAAGGTTGACGTCGAGATCTATGACGCCAAGGTCGCGACATAGACCTTGGCGTCTAT	
							CCCCTACCACATAGGC	
SNP2673	chr3	177872101	177872270	chr3	179637328	179637497	ACCCACACCACATGCCTGTTGCCAGTAGTGCTTAAGATAANCATGTGCCTTTTCCCCCCTTGG	T/C
							AAGACAAAANNATATATCCCTACNTTTTNGGCGTAAATTTCAACNCTGGTTACANATTGATTT	
							CGTACCGAAAATGANCTTCATTTTAGATCTCACCATCCCAGAG[C/T]TTCGTCGATTGTTTTGTA	
							AATACTCTGGGGGTGCACAAACGCGTCATTTCTGATTTGCTTAACCTGTTCCCCCCACCAAAAG	
							GAGTACTAGTTCNNATATGGATAGAGTGCTAGAGACATNATTGCAA	
IDP4064	chr3	178222219	NA	chr3	179987446	NA	CAAAAATATCTATAACATTTCCCTCTTCCCAACAAACAGCTCGTCCGTGAGCTGTACGTCGTC	InDel
							GTTTGAACTCATCAAGCTGCTCCCAAGATGCATCATAAGGCCACCTGTCCAACCCACTGTACCA	
							GCAACTCCCACACACCCCTGTTTAAGTGTGACTTCACGACCTTGGCTGGAGTAGGAAGTACTT	
							TGCCATGTAGAATGTCTGGCAAGGGAACCACCTCTGCTGGGAGCATGCCTTCAAACTTCTTGA	
							GGAGTGCGACATGAAATACGTCATGGATCTGTGCCTTGGCCGGAAGCTTCAACTGTAGGAG	
							ATCTCCCTAATGCATTGCTGAATTTGATAAGGATCGTAAAACTTTGGTCCGAGTTTGGAATGA	
							GTTGCAGCTGTTACTCCCACCACCGAACGTTGCTGGAGGCGCAACCAAACCCATTCCCCTACT	
							GCATATTGGGACTGCCCTACGGTTCTTGTCTTGTGCATGCTTCATCGTAACATGGGCCTAGAGC	
							GTTCGAACTCATCAAGCTGCTCCCAAGATGCATCATAAGGCCACCTGTCCAACCCACTGTACCA	
							AGACGGTTCTGGACCTCCTTCAAAAATCCATCCCGAGCCAATAGCTGTTGATCTACTGCCACG	
							ACACAAGCTGTTCCCGGCTGATAGGACAACTGGGTTGAGGGTCACACCCATAAACAATTAT	
							GAAAGGAGTACACTTAACAGCCGTCTGGTAAGAAGAATTGTAACAAAACTCGGCCCAAAGGA	
							GCCATTGTAACCAAGTCTTGGGTCTGTCGCCTGCCAAGGAGCGAAGATACATGAACAATGATCC	
							TATTTACCACCTCTGACTGACCATCCGTCTGTGGAATGGCTGAGCTCAGGAGAAGACTTGA	
							CGCCAGCCAAGCAGAACAGTTCTATCCAAAAAAGGCTAGTAAAGACTTTATCCCGGTCGCTTA	
							CTATGGAGCATGGAATACCCATGAAGTTGCACAATCCCCCTGAAGAATGCCTTGGCTACCGAC	
							GCAGCCGTGTATGGGTGACCAAGCGGAATAAAATGTGCAAACTTAGAGAATATGTCCACCAC	
							AGTGAGAACTACTGATTTGTCGCCAACTTTGGGGAAACCTTCCACAAAGTCCATTGCAATGTC	
							ACTCCAAACCTCTGATAGTACTGGTAAAGGCTGAAGCAGTCCCGCGAGATGTAAATGTTTCAT	
							TTTGTTCTTCTAGCAAGTGGTACAAACCCTTCACAAATTCCCCGGACCTGACGTGTTGCCTGGGG	
							GCTGTAGAATGATGCTCGCCACTAGTGTTAATGTCTTCTCAGTGCCTTCATGACCACTGTTGTA	
							GGCATTTAGCCAAGTTATTAGCCAAACCGATGAAGTTGTAGGCACAAACACTTTTCCGTGAAA	
							TTGTAGCAGACCATCCCTGATCGCCCACCCCTCCTTGGTTCCCGCTGTGAGTTTGCTGCG	
							CAGTTCAGATACTTGAGGGTCACTATAGGCTTTGGCTCTCAGAGTATTCAAACAACTGAAATTG	
							TGGAGTGGAAATGGTGTTGAGAGTTGCAGCCTCCTCATCAGCGCGAGATAATGCATCTGCTG	
							CACCATTCAATTTGCCAAGTCTGTACTTGCAGAGATATCATAGCCAAACAACTTACTAACCCA	
							TGTATGTTGAGGGATGGTGGACAAGCGCCGATCAAGAAGAAACTTGAGACTATAGTGACCAG	
M_109352	chr3	182687479	182687539	chr3	184435019	184435079	ACTGGTATTTTAAGCTATATATCGAGGTGTGCCTGTCTCTAGTCATTTATTATGGTGTTTA[G/C]	G/C
							AATGGTCTACACAGTTCGTTAACGGTGTATCGTGGATGCATGTTGCCGCGAGCAGATACG	
C3.184743	chr3	184774364	NA	chr3	186489224	NA	AACAAAGGGACCGTTTGTACAACTCAAGCCCATATCTGCTAGAGTGTGAGGANAAATTTCATTA	InDel
							TTTGGGTCTTAATTCGGCTTCCTGTGATCTTTAGGGCCTGTTCGTTTTTCACGGATTCAAGCCC	
							TGGAATAGTTCCTACGCGGATTACTTCTCTAATTTATATAAATTTTGATGAGCTGGAATGAATC	
							CGGGTTCATTCCTGAAGAAACGAACAAGCCCTTAGGGGGACTCAATTCATCTGCCTTCCAGTA	
							TTTGACACCCAGGCTCACCAACTCNTTAAACCGGGTGTTTCTTGGATGTTGGCTTACATTTTAT	
							TGATTTNCTGGT	
SNP6262	chr3	186876354	186876191	chr3	188616163	188616000	TGTTTATTTAGAAATAGATATACGTGAGGTTATTTTCACCTTTCTATATAATGGCATCCCATGG	C/G
							TTATATTGTTTTCAGAAAAAATATAGTTATCCTTGAAAGAAACACCGGGAGATACTACTAACGAG	
							CTATATAAGATTCTCTTAATTTTGGTTTCTCTTTTC[G/C]TTGCTCAAATAAGCCATGNATTAAA	
							CCTATGCTTAGTGGAGTTAGTAAGTTCCATTGGAAGCCCTATAGCTGTTTTCNATGGACTCCATC	
							GAATGTGTACTACCAGTGCCGGTTGGCATAAACCAGAAACTTCA	

Confirmation of rth5 identity via the generation of independent alleles

To confirm that the mutation in candidate gene GRMZM2G426953 indeed confers the phenotype of the *rth5* mutant, several novel putative *Mutator* alleles were isolated from Pioneer´s trait utility system of corn (TUSC, Bensen *et al.* (1995) *The Plant Cell* 7:75-84). Among these putative *Mu* insertion alleles, three displayed the roothairless phenotype and each of these alleles contained a *Mutator* insertion in the candidate gene as demonstrated by sequencing. Allele *rth5-3* contains a *Mu1* insertion in exon 5 (Figure 1B) while *rth5-2* and *rth5-4* contain *MuDR* insertions in exons 3 and 5, respectively (Figure 1B).

The rth5 gene structure

The B73 allele of the *rth5* gene consists of 14 exons and 13 introns (Figure 1B) encoding a 3,792 bp mRNA (including 5' and 3' UTR). This mRNA encodes an 852 amino acid protein with a predicted molecular weight of ~96 kDa.

The RTH5 protein is predicted to contain four trans-membrane (TM) domains, two EF hand motifs, FAD and NAD cofactor binding sites, a ferric reductase domain, and the NADPH oxidase domain characteristic for NADPH oxidases (Figure 1B).

The NADPH oxidase family in maize

Based on sequence similarity with the RTH5 protein 17 respiratory burst oxidases (Rbohs) were identified in the maize filtered gene set (FGS, 4a53), a set of high confidence genes. Phylogenetic reconstructions based on the full-length protein sequences of all known members of the maize, Arabidopsis, soybean and rice NADPH oxidase gene families (Figure 2) were performed. Two major groups of NADPH oxidases were observed. All the NADPH oxidases exhibit diverse sequences in their N-terminus. The smaller group II is characterized by less conservation or the absence of the NADPH oxidase domain and a different conserved sequence in the NAD cofactor binding region (Figure 3). Interestingly, no soybean and two rice and only one maize proteins were included in group II, whereas eight Arabidopsis RBOH proteins belong to that subgroup. Along with two other maize and two rice RBOHs RTH5 is a member of a monocot-specific sub-clade of group I (Figure 2). No homeolog of RTH5 resulting from an ancient genome duplication was found in the maize genome. *In silico* searches for syntenic homologs (genomeevolution.org/CoGe) identified the rice gene Os01g61880, which was also the closest homolog of RTH5 in the phylogenetic tree (Figure 2). RTH5 and its two closest maize homologs RBOH11 and RBOH13 are located in a subgroup of the phylogenetic tree characterized by the presence of a few additional amino acids at the end of the FAD cofactor binding site (Figure 3).

The rth5 gene is preferentially expressed in root hairs

Tissue-specific expression of *rth5* and all 17 *Rboh* family members was examined via qRT-PCR in six different seedling tissues including the cap of the primary root, elongation zone, differentiation zone without root hairs, isolated root hairs, and the coleoptilar node and the first leaf. *Rth5* transcripts accumulated in all tested tissues, but the significantly highest signal was detected in root hairs (Figure 4A). Moreover, the qTeller tool was used to compare the accumulation of *rth5* transcripts in RNA-Seq data from a wide range of tissues and organs. The highest expression values were observed in seedling roots and shoots (Figure 5). Expression data of the remaining *Rboh* family members in these tissues are summarized in Figure 6. The relative expression levels of all 17 *Rboh* genes in root hairs were determined via quantitative real time PCR (qRT-PCR). The *rth5* gene displayed the highest expression, accounting for 52% of the total *Rboh* expression in maize root hairs (Figure 4A). *Rboh11* and *Rboh13,* the closest homologs of *rth5* were only very weakly expressed in root hairs.

RNA *in situ* hybridization was used to study root tissue-specific expression patterns. Cross sections hybridized with an *in vitro* transcribed *rth5* RNA antisense probe resulted in a signal in root hairs (Figure 4C), whereas the sense probe yielded no signal (Figure 4D).

The RTH5 protein is localized in epidermal cells of the primary root

Immunohistochemical experiments were performed to localize the RTH5 protein *in situ*. Cross-sections of wild-type primary roots with emerging root hairs were incubated with the RTH5 antibody (Figure 4E) or buffer solution (Figure 4F) followed by the secondary antibody. A signal of the RTH5 antibody was detected in all epidermis cells and in root hairs; whereas the control sections displayed no signal.

In longitudinal sections an RTH5 signal was detected in root hairs and all epidermis cells of the root hair zone (Figure 4G). Interestingly, an RTH5 signal was also detected in longitudinal sections of the elongation zone, which does not form root hairs (Figure 4I). The control sections showed no detectable color reaction (Figure 4H & 4J).

Both, cross- and longitudinal sections were performed for wild-type and *rth5* mutant primary roots, but displayed no difference in signal intensity or tissue organization apart from the root hair defect.

The rth5 mutant accumulates less ROS during root hair tip growth

Superoxide is a very short-lived radical that is rapidly converted into hydrogen peroxide (i.e., NADPH oxidase generates superoxide which in turn permutates into hydrogen peroxide). Staining for both molecules was performed in wild-type and *rth5* mutant seedlings and presence or absence of high ROS signals in root hair tips were quantified. While 74% of wild-type root hairs displayed the superoxide signal at the tip, the frequency was significantly reduced to 24% in *rth5* root hairs. Similarly, 77% and 71% of wild-type root hairs exhibited a high hydrogen peroxide signal in the tips of root hairs as detected using DAB or H_2DCF-DA, respectively, while only 27% and 21% of *rth5* root hairs displayed this signal using the same dyes. These differences were significant in Fisher´s exact test at $p \leq 0.001$.

RNA-Seq of rth5 vs. wild-type

To understand the influence of *rth5* on global patterns of transcription, RNA-Seq was performed on 6-day-old roots from the *rth5* mutant and wild-type seedlings (Methods). Two biological replicates were analyzed for each genotype. Approximately 98% of raw reads from each sample passed the quality check and trimming procedure. 3.7-4.7 million reads per replicate (~88% of the post-trimmed reads) were uniquely and confidently mapped to the B73 reference genome (Table 2) (Methods). Consistent with the results from the qRT-PCR experiment, the *rth5* gene was preferentially expressed in wild-type vs *rth5* root hairs with a relatively small fold change (1.4). Among 21,007 genes that have ≥20 RNA-Seq reads across all four biological replicates, 1,257 genes differentially expressed genes (DEGs) were identified at a 10% false discovery rate (FDR). 634 and 623 DEGs were significantly up- and down-regulated, respectively. A gene ontology (GO) enrichment analysis of the 1,257 DEGs revealed statistically significant over-representation of many functional categories, including monooxygenase activity, peroxidase activity, oxidoreductase activity, response to oxidative stress, heme binding, phospholipid-translocating ATPase activity, cellulose synthase activity, and lipid biosynthetic process (Table 3).

Separate gene ontology (GO) enrichment analyses of the 634 and 623 up and down-regulated DEGs indicate that DEGs are significantly over-represented in some functional categories in both DEG groups, such as monooxygenase activity and heme binding (Table 3). Genes associated with response to oxidative stress, response to biotic stimulus, as well as sucrose metabolic process are significantly enriched in the up-regulated DEG group; whereas genes involved in cellulose biosynthetic process, phospholipid-translocating ATPase activity, and myosin complex are significantly enriched in the down-regulated DEG group. Overall,

these RNA-Seq results are consistent with the view that *rth5/rbohA* functions in producing ROS required for elongation of root hairs.

Of 17 maize *rboh* homologous genes, 14 are at the detectable level in the RNA-Seq data. Six *rboh* genes are DEGs (Table 4). Statistically, the *rboh* genes are 7.2x over-represented in the DEGs (χ^2 test, p-value=$1.5x10^{-7}$). Four and two *rboh* genes were up- and down-regulated in the mutant relative to the wild-type, respectively. One DEG, *rboh7* (GRMZM2G065144) was up-regulated ≥8 fold in the mutant.

Table 2. Summary of Reads from the RNA-Seq Experiment

Sample	Number of raw reads	Number of remaining trimmed reads (% raw reads)	Number of mapped reads (% trimmed reads)	Number of uniquely & confidently mapped reads (% trimmed reads)
mut_rep1	5,301,418	5,197,927 (98.0%)	5,086,480 (97.9%)	4,569,445 (87.9%)
mut_rep1	4,436,707	4,352,149 (98.1%)	4,260,690 (97.9%)	3,833,694 (88.1%)
wt_rep1	4,232,619	4,152,065 (98.1%)	4,065,642 (97.9%)	3,666,583 (88.3%)
wt_rep2	5,437,462	5,329,121 (98.0%)	5,217,366 (97.9%)	4,694,561 (88.1%)

Table 3. Significantly Over-Represented GOs in Differentially Expressed Genes (DEGs)

GO	pvalue	qvalue	Term	Def
GO:0004497	1.00E-04	0.02	monooxygenase activity	Catalysis of the incorporation of one atom from molecular oxygen into a compound and the reduction of the other atom of oxygen to water.
GO:0004601	1.00E-04	0.02	peroxidase activity	Catalysis of the reaction: donor + hydrogen peroxide = oxidized donor + 2 H2O.
GO:0009055	1.00E-04	0.02	electron carrier activity	Any molecular entity that serves as an electron acceptor and electron donor in an electron transport system.
GO:0016491	1.00E-04	0.02	oxidoreductase activity	Catalysis of an oxidation-reduction (redox) reaction, a reversible chemical reaction in which the oxidation state of an atom or atoms within a molecule is altered. One substrate acts as a hydrogen or electron donor and becomes oxidized, while the other acts as hydrogen or electron acceptor and becomes reduced.
GO:0016760	1.00E-04	0.02	cellulose synthase (UDP-forming) activity	Catalysis of the reaction: UDP-glucose + ((1,4)-beta-D-glucosyl)(n) = UDP + ((1,4)-beta-D-glucosyl)(n+1).
GO:0020037	1.00E-04	0.02	heme binding	Interacting selectively and non-covalently with heme, any compound of iron complexed in a porphyrin (tetrapyrrole) ring.
GO:0045449	1.00E-04	0.02	NA	NA
GO:0055114	1.00E-04	0.02	oxidation-reduction process	A metabolic process that results in the removal or addition of one or more electrons to or from a substance, with or without the concomitant removal or addition of a proton or protons.
GO:0005506	0.00020	0.03	iron ion binding	Interacting selectively and non-covalently with iron (Fe) ions.
GO:0006979	0.00020	0.03	response to oxidative stress	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of oxidative stress, a state often resulting from exposure to high levels of reactive oxygen species, e.g. superoxide anions, hydrogen peroxide (H2O2), and hydroxyl radicals.
GO:0030244	0.00020	0.03	cellulose biosynthetic process	The chemical reactions and pathways resulting in the formation of cellulose, a linear beta1-4 glucan of molecular mass 50-400 kDa with the pyranose units in the -4C1 conformation.
GO:0006519	0.00040	0.06	NA	NA
GO:0004012	0.00060	0.07	phospholipid-translocating ATPase activity	Catalysis of the movement of phospholipids from one membrane bilayer leaflet to the other (phospolipid 'flippase' activity), driven by the hydrolysis of ATP.
GO:0015914	0.00060	0.07	phospholipid transport	The directed movement of phospholipids into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore. Phospholipids are any lipids containing phosphoric acid as a mono- or diester.
GO:0003993	0.00070	0.08	acid phosphatase activity	Catalysis of the reaction: an orthophosphoric monoester + H2O = an alcohol + phosphate, with an acid pH optimum.
GO:0016747	0.00080	0.08	transferase activity, transferring acyl groups other than amino-acyl groups	Catalysis of the transfer of an acyl group, other than amino-acyl, from one compound (donor) to another (acceptor).
GO:0008610	0.00090	0.08	lipid biosynthetic process	The chemical reactions and pathways resulting in the formation of lipids, compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent.
GO:0015662	0.00090	0.08	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	Catalysis of the transfer of a solute or solutes from one side of a membrane to the other according to the reaction: ATP + H2O = ADP + phosphate, to directly drive the transport of ions across a membrane. The reaction is characterized by the transient formation of a high-energy aspartyl-phosphoryl-enzyme intermediate.
GO:0050664	0.00100	0.09	oxidoreductase activity, acting on NADH or NADPH, oxygen as acceptor	Catalysis of an oxidation-reduction (redox) reaction in which NADH or NADPH acts as a hydrogen or electron donor and reduces an oxygen molecule.

Table GO2. Significantly over-represented GOs in the mutant-up-regulated DEGs

GO	pvalue	qvalue	Term	Def
GO:0004601	1.00E-04	0.04	peroxidase activity	Catalysis of the reaction: donor + hydrogen peroxide = oxidized donor + 2 H2O.
GO:0016491	1.00E-04	0.04	oxidoreductase activity	Catalysis of an oxidation-reduction (redox) reaction, a reversible chemical reaction in which the oxidation state of an atom or atoms within a molecule is altered. One substrate acts as a hydrogen or electron donor and becomes oxidized, while the other acts as hydrogen or electron acceptor and becomes reduced.
GO:0020037	1.00E-04	0.04	heme binding	Interacting selectively and non-covalently with heme, any compound of iron complexed in a porphyrin (tetrapyrrole) ring.
GO:0055114	1.00E-04	0.04	oxidation-reduction process	A metabolic process that results in the removal or addition of one or more electrons to or from a substance, with or without the concomitant removal or addition of a proton or protons.
GO:0004497	0.00020	0.07	monooxygenase activity	Catalysis of the incorporation of one atom from molecular oxygen into a compound and the reduction of the other atom of oxygen to water.
GO:0009055	0.00030	0.07	electron carrier activity	Any molecular entity that serves as an electron acceptor and electron donor in an electron transport system.
GO:0009058	0.00030	0.07	biosynthetic process	The chemical reactions and pathways resulting in the formation of substances; typically the energy-requiring part of metabolism in which simpler substances are transformed into more complex ones.
GO:0004867	0.00040	0.07	serine-type endopeptidase inhibitor activity	Stops, prevents or reduces the activity of serine-type endopeptidases, enzymes that catalyze the hydrolysis of nonterminal peptide bonds in a polypeptide chain; a serine residue (and a histidine residue) are at the active center of the enzyme.
GO:0010181	0.00040	0.07	FMN binding	Interacting selectively and non-covalently with FMN, flavin mononucleotide, the coenzyme or the prosthetic group of various flavoprotein oxidoreductase enzymes.
GO:0006979	0.00060	0.09	response to oxidative stress	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of oxidative stress, a state often resulting from exposure to high levels of reactive oxygen species, e.g. superoxide anions, hydrogen peroxide (H2O2), and hydroxyl radicals.
GO:0009607	0.00060	0.09	response to biotic stimulus	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a biotic stimulus, a stimulus caused or produced by a living organism.
GO:0005985	0.00070	0.09	sucrose metabolic process	The chemical reactions and pathways involving sucrose, the disaccharide fructofuranosyl-glucopyranoside.
GO:0016157	0.00070	0.09	sucrose synthase activity	Catalysis of the reaction: UDP-glucose + D-fructose = UDP + sucrose.
GO:0003993	0.00080	0.10	acid phosphatase activity	Catalysis of the reaction: an orthophosphoric monoester + H2O = an alcohol + phosphate, with an acid pH optimum.

Table GO3. Significantly over-represented GOs in the mutant-down-regulated DEGs

GO	pvalue	qvalue	Term	Def
GO:0016760	1.00E-04	0.05	cellulose synthase (UDP-forming) activity	Catalysis of the reaction: UDP-glucose + {(1,4)-beta-D-glucosyl)(n) = UDP + {(1,4)-beta-D-glucosyl)(n+1).
GO:0030244	1.00E-04	0.05	cellulose biosynthetic process	The chemical reactions and pathways resulting in the formation of cellulose, a linear beta1-4 glucan of molecular mass 50-400 kDa with the pyranose units in the -4C1 conformation.
GO:0045449	1.00E-04	0.05	NA	NA
GO:0004012	0.00020	0.05	phospholipid-translocating ATPase activity	Catalysis of the movement of phospholipids from one membrane bilayer leaflet to the other (phospolipid 'flippase' activity), driven by the hydrolysis of ATP.
GO:0015662	0.00020	0.05	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	Catalysis of the transfer of a solute or solutes from one side of a membrane to the other according to the reaction: ATP + H2O = ADP + phosphate, to directly drive the transport of ions across a membrane. The reaction is characterized by the transient formation of a high-energy aspartyl-phosphoryl-enzyme intermediate.
GO:0015914	0.00020	0.05	phospholipid transport	The directed movement of phospholipids into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore. Phospholipids are any lipids containing phosphoric acid as a mono- or diester.
GO:0020037	0.00020	0.05	heme binding	Interacting selectively and non-covalently with heme, any compound of iron complexed in a porphyrin (tetrapyrrole) ring.
GO:0006754	0.00030	0.06	ATP biosynthetic process	The chemical reactions and pathways resulting in the formation of ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator.
GO:0004497	0.00040	0.06	monooxygenase activity	Catalysis of the incorporation of one atom from molecular oxygen into a compound and the reduction of the other atom of oxygen to water.
GO:0016459	0.00040	0.06	myosin complex	A protein complex, formed of one or more myosin heavy chains plus associated light chains and other proteins, that functions as a molecular motor; uses the energy of ATP hydrolysis to move actin filaments or to move vesicles or other cargo on fixed actin filaments; has magnesium-ATPase activity and binds actin. Myosin classes are distinguished based on sequence features of the motor, or head, domain, but also have distinct tail regions that are believed to bind specific cargoes.
GO:0030528	0.00040	0.06	NA	NA

5

Table 4. Expression of Maize *Rboh* Homologous Genes

GeneID	Symbol	wt_rep1	wt_rep2	mut_rep1	mut_rep2	mut-wt_log2FC	mut-wt_pvalue	mut-wt_qvalue	mut-wt_sig
GRMZM2G065144	Rboh7	10	24	217	192	3.61	0.000	0.000	yes
GRMZM2G426953	RTH5	409	540	365	286	-0.53	0.000	0.000	yes
GRMZM2G034896	Rboh6	243	328	398	322	0.35	0.001	0.029	yes
GRMZM2G037993	RbohAlike	45	96	107	107	0.63	0.002	0.046	yes
GRMZM2G441541	RbohD	72	84	60	40	-0.63	0.004	0.070	yes
GRMZM2G358619	Rboh12	3	6	13	16	1.71	0.004	0.072	yes
GRMZM2G043435	RbohC	525	723	537	542	-0.19	0.038	0.294	no
GRMZM2G300965	Rboh10	14	17	18	28	0.59	0.122	0.504	no
GRMZM2G089291	Rboh8	31	35	29	20	-0.42	0.162	0.566	no
GRMZM2G138152	RbohB	294	378	345	274	-0.1	0.293	0.702	no
GRMZM2G448185	Rboh14	146	152	120	151	-0.12	0.446	0.808	no
GRMZM2G022547	Rboh5	414	632	610	466	0.06	0.494	0.833	no
GRMZM2G147966	Rboh9	42	52	47	40	-0.1	0.679	0.913	no
GRMZM2G169201	Rboh15	131	153	141	139	0	0.973	0.996	no
GRMZM2G068557	RbohFR	0	0	0	0	NA	NA	NA	NA
GRMZM2G323731	Rboh11	0	0	0	0	NA	NA	NA	NA
GRMZM2G401179	Rboh13	0	0	0	0	NA	NA	NA	NA

Table 5. Oligonucleotide Primers Used For qRT-PCR Experiments.

Primer name	Sequence 5' to 3'	Accession number	Tm[a] [°C]	SEQ ID NO:	Product [bp]
RT RTH5 FP	CACAAGAACTCTCGCATAGG		60		204
RT RTH5 RP	CTGCTATGCCAATGCTATCG		60	21	
RT RbohAlike FP	ATATGCACCGACGACTGTAG		64		225
RT RbohAlike RP	CAAGCACATTGCGGAGAACC		61	23	
RT RbohB FP	ATACTTGGTGCGCAGAAATG		60		318
RT RbohB RP	ACCGACAATCCCAACTACTC		61	25	
RT RbohC FP	ATGTCACAAGGCTACCTAAC		58		145
RT RbohC RP	TACCAACCTACGCTACTTAC		58	27	
RT RbohD FP	CAAGCAGAGTTGAGGGTTTG		61		155
RT RbohD RP	CCTAGAGCACCTCAGCTATG		61	29	
RT RbohFR FP	GTCGGTTCAACGATCATCAG		62		216
RT RbohFR RP	CATTAGACGCGAAACATGAC		59	31	
RT Rboh5 FP	CACCACCACTACATTGAAAG		58		240
RT Rboh5 RP	CAGAACACCTACCCATTCTG		59	33	
RT Rboh6 FP	CTAGTGTCAAGACGCACTTC		60		217
RT Rboh6 RP	CCAGAGCTTCTCCAGTTATG		59	35	
RT Rboh7 FP	ACAAGAAGGGCGTGATCGAG	GRMZM2G065144	64	36	294
RT Rboh7 RP	TGGTTTCCCGCGAGAATTCC	GRMZM2G065144	64	37	
RT Rboh8 FP	ATTCACGCGATTTGTGTTCC		61		199
RT Rboh8 RP	AGCTCCTCAAGTAAGATCAG		58	39	
RT Rboh9 FP	TGCTGATGACCGTCTCCTTC		65		294
RT Rboh9 RP	CGCTGCTACAATCTCGAGTG		62	41	
RT Rboh10 FP	CTGGCGGGACGTCTTCAAGC		67		181
RT Rboh10 RP	GCTTCGTGGTGGTCCACATC		65	43	
RT Rboh11 FP	CATGGTCCAGTCACTTCAAC		60		208
RT Rboh11 RP	GGTACTGCTGAATTCTATCG		57	45	
RT Rboh12 FP	GGCCGAGACCGACAAGAATG		65		246

RT Rboh12 RP	CCGCCACAGTAGAACACTCC		64	47	
RT Rboh13 FP	ATCGCGACGATGTAATCGAG		62		217
RT Rboh13 RP	CCGATGCGTGAGTTCTTGTG		63	49	
RT Rboh14 FP	GAACCCGATTTGATTTCCAC		59		209
RT Rboh14 RP	CACCCTCCCATTTGGTTAAG		60	51	
RT Rboh15 FP	AAGAAGCACCACAGGGTTAG		61		223
RT Rboh15 RP	TGGCTCCTCGTATATTCATC		58	53	

[a]: melting temperature

The maize *rth5* mutant displays significantly shortened root hairs and reduced root hair density while aboveground development remains unaffected. Hence, *rth5* specifically controls both the elongation of root hairs and the specification of epidermis cells or the initiation of root hairs. More specifically, root hair elongation in the mutant *rth5* is affected during the transition from bulge to tip growth. In contrast to *rth5*, root hairs of the *rth3* mutant are characterized by disrupted bulges while, *rth1* and *rth2* root hairs fail to elongate after transition to tip growth (Wen and Schnable (1994) *Am. J. Bot.* 81:833-842). The *rth5* mutant is like *rth2* and *rth3* (Wen and Schnable (1994) *Am. J. Bot.* 81:833-842), which are specifically affected in root hair formation. In contrast, the maize mutants *rth1* (Wen and Schnable (1994) *Am. J. Bot.* 81:833-842), *dil1* and *dil2* (Lid *et al.*, (2004) *Planta* 218:370-378) which are defective in root hair formation also display pleiotropic effects during development. The *rth1* mutation displays general growth abnormalities during development (Wen and Schnable (1994) *Am. J. Bot.* 81:833-842) while *dil1* and *dil2* control cell division orientation in the aleurone, leaf and root epidermis. The *rth3* mutant shows no difference in root hair density as compared to wild-type, while the *rth1* and *rth2* mutants form twice as many hairs as wild-type plants. In contrast, the maize mutant *rth5* forms fewer root hairs as compared to wild-type primary roots indicating that the *rth5* gene is involved in not only root hair elongation but also in epidermis specification and/or root hair initiation.

In *rth5* all root types including primary, seminal, lateral, and shoot-borne roots display defects in root hair formation, indicating that the RTH5 protein is critical for root hair formation in all root types. A general defect in root hair formation affecting all root types has also been observed for the *rth1*, *rth2* and *rth3* mutants (Hochholdinger *et al.* (2008) *Plant J.* 54:888-898, (Wen and Schnable (1994) *Am. J. Bot.* 81:833-842). In contrast to this general mechanism of root hair formation in all root types, the control of lateral root formation is root-type specific. For instance, *rum1* (Woll *et al.* (2005) Plant Physiol. 139:1255-1267) and *lrt1* (Hochholdinger and Feix (1998) *Plant J.* 16:247-255) display root-type specific defects,

that are confined to embryonic roots. In these mutants postembryonic shoot-borne roots do not display any lateral root defects.

The RTH5 protein is characterized by several functional domains (Figure 1B) including four trans-membrane domains, responsible for the membrane embedding, and two EF hand motifs, which are required for Ca^{2+} regulation. Moreover, the FAD and NAD domains confer binding to the cofactor FAD and the substrate NADPH, respectively. Regulation of the RBOHA/RTH5 NADPH oxidase by a mitogen-activated protein kinase cascade controlled by abscisic acid signaling was demonstrated in maize leaves (Lin *et al.* (2009) *J. Exp. Bot.* 60:3221-3238).

Consistent with this co-factor binding site and functional assignment, the *rth5* gene encodes a predicted NADPH oxidase (NOX). The NOX protein family is present in most eukaryotic species (Bedard *et al.* (2007) *Biochimie* 89:1107-1112). The *rth5* mutant was induced by EMS, which resulted in a Guanine to Adenine substitution in the genomic sequence changing the cysteine (C) residue in amino acid position 821 to tyrosine (Y) (Figure 1B). Cysteine residues often form intra- or inter-protein disulfide bridges to enable secondary structures. As this cysteine residue is highly conserved among a very diverse set of species (Figure 1C), it is likely that the C821Y mutation is disabling a functionally important protein domain. A point mutation of this conserved C in the human NADPH oxidase 2 (NOX2) gene results in the genetic disorder chronic granulomatous disease (Kawahara *et al.* (2007) *BMC Evol. Biol.* 7:109). This disease affects cells of the immune system, which have difficulties in forming reactive oxygen compounds, including superoxide used to kill certain ingested pathogens.

Phylogenetic reconstruction of RBOH (RESPIRATORY BURST OXIDASE HOMOLOG) proteins of rice, maize, soybean, and Arabidopsis revealed the existence of two subgroups. These subgroups are distinguished by inter-group differences in the sequences of the NAD binding region, which are conserved within each group (Figure 3). Ten Arabidopsis RBOH proteins are classified to group I, which is consistent with previous annotations of superoxide-producing NOX proteins (Kawahara *et al.* (2007) *BMC Evol. Biol.* 7:109). Interestingly, all soybean proteins are assigned to group I, while eight of 18 members of the Arabidopsis RBOH protein family are classified in group II. According to their annotation, group II proteins are often connected to iron deficiency and can function as ferric reductases which are NOX proteins that do not produce superoxide (Ivanov *et al.* (2012) *Mol. Plant* 5:27-42).

The nine members of the group I subclade that includes RTH5 are characterized by nine additional amino acids in the FAD cofactor domain as compared to the remaining group I proteins. Within this subclade, all five proteins in the monocot-specific group of this subclade lack 15 amino acids in their NADPH binding region. Both, the additional amino acids in the FAD, and the missing amino acids in the NAD region can confer altered binding to cofactors or substrates leading to specialized functions.

While several types of NOX proteins are found in mammals ranging from ancestral types to peroxidase-containing DUOXs, in plants only homologs of EF-hand-containing NOX5-types have been identified (Bedard *et al.* (2007) *Biochimie* 89:1107-1112). Human NOX proteins consist of two transmembrane heterodimers ($gp91^{phox}$ and $p22^{phox}$) and four regulatory subunits ($p40^{phox}$, $p47^{phox}$, $p67^{phox}$, and Rac2) in the cytoplasm (Lam *et al.* 2010). Thus far, no homologs for the regulatory subunits $p47^{phox}$, $p67^{phox}$, or $p22^{phox}$ have been found in plants (Bedard *et al.* (2007) *Biochimie* 89:1107-1112).

The human NOX2 proteins have been shown to play a role in the production of superoxide as a first defense response to microorganisms invading neutrophils and macrophages (Lam *et al.* (2010) *Seminars in Immunopathology* 32:415-430, Nauseef (2008) *J. Biol. Chem.* 283:16961-16965). Plant NOX proteins have been shown to function in plant immunity and several developmental processes. Pepper and tobacco NOX proteins were demonstrated to function in plant immunity (Wi *et al.* (2012) *Plant Physiol.* 159:251-265, Yi *et al.* (2010) *New Phytol.* 185:701-715). Aerchenyma formation after waterlogging was shown to be conferred by the maize NOX protein GRMZM2G300965 *(Yamauchi et al. (2011) Plant Signal Behav.* 6:759-761), which showed specific expression in the root cap and the elongation zone (Figure 5, RBOH10). Finally, the Arabidopsis RBOHD and RBOHF are involved in ABA-dependent stomatal closure (Kwak *et al.* (2003) *EMBO J* 22:2623-2633) and lead to an oxidative burst following infection (Galletti *et al.* (2008) *Plant Physiol* 148:1695-1706).

The Arabidopsis RBOH protein RBOHC (RHD2), which is distantly related to RTH5 also controls root hair growth. Arabidopsis plants defective in *rhd2* gene function form very short root hairs that do initiate bulges, but that do not elongate (Schiefelbein and Somerville (1990) *Plant Cell* 2:235-243). Moreover, such plants display stunned root growth (Foreman *et al.* (2003) *Nature* 422:422-446), an effect that was not observed for *rth5* mutant roots. *Rhd2* is expressed in root epidermal cells, but also in the cortical cells of the elongation- and differentiation zone (Foreman *et al.* (2003) *Nature* 422:422-446).

Consistent with its mutation in the C-terminus of the protein, the expression of the *rth5-1* mutant reference allele was only slightly reduced as compared to wild-type primary roots relative to the *Myosin* gene (Genbank AC: 48090G09.x1). Similarly, protein levels detected by immunohistochemistry and Western blot analyses using three-days-old seedlings did not show any obvious change between wild-type and *rth5-1* mutants (for example, see Figure 4). Nevertheless, although expression and protein abundance are not dramatically affected, this mutation significantly affects the function of the protein as illustrated by the mutant phenotype.

Expression of *rth5* was detected in several shoot and root tissues at low levels with expression maxima in root hairs (Figure 5 and Figure 5). Similarly, the rice *rboh* genes are also expressed in multiple plant organs and tissues (Wong *et al.* (2007) *Plant Cell* 19:4022-4034). In maize root hairs *the rth5* gene accounts for 52% of the expression of all 17 *rboh* genes and hence the majority of NADPH oxidase transcripts in root hairs. The RTH5 protein was detected preferentially in root hairs and epidermis cells, but weaker signals were also detected in root cortex cells (Figure 4). In two studies mainly focusing on the influence of hormones on NADPH oxidase activity in maize leaves it was demonstrated that expression of *rth5* (*rbohA*) is induced after abscisic acid and brassinosteroid treatment as well as hydrogen peroxide application (Lin *et al.*(2009) *J. Exp. Bot.* 60:3221-3238, Zhang *et al.* (2010) *J. Exp. Bot.* 61:4399-4411). These experiments suggest that RTH5 may play a role in hormone-mediated oxidative defense in leaves.

Arabidopsis root hairs display high Ca^{2+} and superoxide concentrations in their growing tips (Foreman *et al.* (2003) *Nature* 422:422-446). The tips of root hairs of the mutant *rth5* displayed significantly less superoxide staining and accumulate less hydrogen peroxide (p-value <0.001). These results indicate that as a consequence of a functional defect of the RTH5 protein, the *rth5* mutant is unable to establish high levels of ROS in the tips of root hairs. Not all growing tips of wild-type root hairs displayed staining for superoxide and hydrogen peroxide. Some root hairs may have died during the staining procedure; some may not have taken up the indicator chemicals. Consistent with expression data the presence of RTH5 homologous RBOH proteins may account for the residual ROS content detected in the mutant (Figure 6).

A model for the function of RTH5 in root hair elongation is summarized in Figure 7B. Superoxide molecules, localized in the root hair tips produced by RTH5 are transformed rapidly into hydrogen peroxide, which in turn is converted to hydroxyl radicals by apoplastic peroxidases (Liszkay *et al.* (2003) *Planta* 217:658-667). These hydroxyl radicals are cleaving

polysaccharides (Fry (1998) *Biochem. J.* 332:507-515) and therefore cause cell wall loosening, thereby enabling polarized cell growth (Bibikova *et al.* (1998) *Development* 125:2925-2934, Liszkay *et al.* (2004) *Plant Physiol.* 136:3114-3123).

In Arabidopsis the molecular interactions resulting in position-dependent root hair initiation are well characterized (reviewed in: Hochholdinger and Nestler (2012) *Encyclopedia of Genetics*, Elsevier: New York, vol. 5, p. 349-352). In contrast, the molecular network underlying random root hair patterning in maize remains enigmatic (Clowes (2000) *New Phytologist* 146: 83-94). The reduced root hair density observed in mutant seedlings indicates that RTH5 is involved not only in root hair tip growth, but also in root hair initiation or epidermis specification (Figure 7). The presence of RTH5 in all epidermal cells is consistent with this hypothesis that superoxide is only produced upon activation of RTH5, which thus acts as a signal converter, similar to the human NOX2 *(Lam et al. (2010) Seminars in Immunopathology* 32:415-430), which subsequently promotes root hair initiation. This is supported by the finding of selective root hair cell superoxide staining in the maize differentiation zone. In contrast, all epidermis cells display superoxide staining in the elongation zone (Liszkay *et al.* (2004) *Plant Physiol.* 136:3114-3123). Residual root hair initiation in the *rth5* mutant might be controlled by RTH5-independent pathways. Due to the lack of functional RTH5 proteins these root hairs fail the transition to tip growth.

Experimental Procedures

Plant material and growth conditions

The mutation was initially induced by EMS (Ethyl methanesulfonate) mutagenesis of the Pioneer inbred line 114748/AD. Mutants used in the present study were progeny of plants backcrossed 10 times to the inbred line B73 (Schnable laboratory stock 00g-1542-1).

Kernels were surface sterilized as described previously *(Nestler et al. (2011) J. Proteome Res.* 10:2525-2537) and germinated in moist paper rolls in a 16 h light, 8 h dark cycle at 26 °C in distilled water.

Binocular imaging and environmental scanning electron microscopy

A Zeiss Stemi SV8 Binocular coupled with a Powershot G2 camera was used to document root hairs at a 20x magnification. Root hair length was determined via WinRhizo Software. A Philips XL30 FEI environmental scanning electron microscope (eSEM) was used for surface illustration of fresh four-day-old primary roots without fixation or tissue drying.

Bulked segregant analysis (BSA)

Sequenom-based BSA (bulked segregant analysis) (Liu *et al.* (2010) *Genetics* 184:19-26) was used to map the *rth5* gene. 89 individuals from an F_2-family were phenotypically divided into two bulks: mutant and non-mutant. Equal quantities of root tissue were pools from each individual in each bulk. DNA was extracted from the two tissue bulks and then subjected to Sequenom-based SNPtyping using ~1,000 SNP markers.

Fine-scale mapping of rth5

Multiple markers located on the long arm of chromosome 3 were used to genotype 80 *rth5* mutant individuals from an F_2-population. An *rth5* mutant individual in a B73 genetic background was crossed with the Mo17 inbred line and the resulting F_1-seeds were backcrossed to Mo17 to create an F_1BC population. Two SNP (single nucleotide polymorphism) markers, SNP2673 (SEQ ID NO: 15) and SNP6262 (SEQ ID NO: 19), were used to genotype 512 F_1BC individuals. In total, 65 recombinants were identified between SNP2673 and SNP6262, each of which was self-pollinated. The root hair phenotypes of self-pollinated progeny were scored to infer whether or not specific F_1BC_1 recombinants carried the *rth5* mutant allele. In parallel, the recombinants were genotyped with a series of

molecular markers (Figure 1A). Based on an analysis of the resulting genotyping and phenotyping data, the *rth5* mutant was mapped to the 179.9-182.3 Mb interval of chromosome 3 in AGPv1. To further narrow down the *rth5* interval, a larger F_1BC population (N=~5,000) was genotyped with SNP markers SNP90373 (SEQ ID NO: 14) and M_109352 (SEQ ID NO: 17). In total, 440 recombinants between these two marker loci were identified and genotyped for *rth5* as described above. After genotyping all available recombinants (65 + 440) with 35 SNP markers (15 of which are co-dominant in this population) the *rth5* gene was mapped to the interval 180.1-180.4 Mb.

Identification of novel rth5 mutant alleles

Three novel alleles, *rth5-2, rth5-3, and rth5-4*, were identified by a reverse genetic screen of *Mutator* stocks at Pioneer Hi-Bred. Confirmation of the candidate seedlings displaying a roothairless phenotype was performed by PCR mapping of the *Mutator* insertion with a general *Mutator* (Mu) oligonucleotide primer (Dietrich *et al.* 2002) and an *rth5*-specific oligonucleotide primer (5'-GCACATCTCCCGGATAAATTG-3') (SEQ ID NO: 54). The amplification product contained 39 bp beyond the Mu TIR sequence sufficient to identify the *Mutator* element (Dietrich *et al.* (2002) *Genetics* 160:697-716). *Mu* insertions were found in positions 1300, 1931, and 1968 bp of the genomic *rth5* sequence counting from the ATG start codon in the alleles, *rth5-2*, *rth5-3*, and *rth5-4*, respectively.

RTH5 protein domain prediction

Several bioinformatics tools were used to predict protein domains: InterProScan, MyHits (EMBL- myhits.isb-sib.ch), SMART (simple modular architecture research tool-smart.embl-heidelberg.de) and TMHMM. Domains predicted by at least two search tools were summarized in the protein domain structure (Figure 1B).

Phylogenetic analyses

The protein sequences of RTH5 homologs were obtained by blasting the RTH5 sequence against species-specific databases at maizesequence.org, rice.plantbiology.msu.edu, soybean.org, and TAIR.org. The DNA Baser software was used to combine all sequences into a Fasta file which was used for alignment and calculation of a phylogenetic tree by MEGA4.0 (Tamura *et al.* (2007) *Mol. Biol. Evol.* 24:1596-1599) using the Neighbor joining algorithm (Bootstrap, 1000 replications).

RNA in situ hybridization

Primary roots of three-day-old seedlings were fixed and embedded in paraffin as described previously (Hochholdinger *et al.* (2008) *Plant J.* 54:888-898). The RNA probe for hybridization was generated by amplification of 282 bp from the 3′-end of the *rth5* gene with the oligonucleotide primers forward 5'-GTGTACCCGAAGATCCGATG-3'(SEQ ID NO: 55), and reverse 5'-GACAGCTCGGGCAGAAAGAC-3'(SEQ ID NO: 56). The amplicon was cloned into the pGEM T-easy vector (Promega,) in sense and antisense directions. *In vitro* transcription of the probes and Digoxigenin-labeling was performed using the SP6 polymerase (NEB) according to the manufacturer's protocol. RNA *in situ* hybridization was performed according to (Jackson (1992) *Mol. Plant Pathol.* Oxford 163-174). A Zeiss Axioplan 2 microscope in combination with a Photometrics Cool Snapcamera (Roper Scientific GmbH) was used for image acquisition.

Quantitative RT-PCR

For qRT-PCR, RNA was extracted from approximately 100 mg of different maize tissues with the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's protocol, including on-column DNase I digestion (Fermentas). Isolation and harvesting of root hairs was performed as described previously (Nestler, *et al.* (2011) *J. Proteome Res* 10:2525-2537). cDNA was prepared from 500 ng total RNA by using the qScript cDNA Super Mix (Quanta BioSciences). qRT-PCR experiments were performed using MESA Green or Blue qPCR Mastermix Plus for SYBR Assay no ROX kit (Eurogentec). All experiments were conducted in three biological replicates and three technical replicates per biological replicate in a CFX384 Real-Time PCR Detection System (Bio-Rad). The efficiency for every oligonucleotide primer pair (summarized in Table 5) was determined by a dilution series. Expression was calculated relative to the reference gene *myosin* (Genbank AC486090G09.x1), which has previously been used as expression standard in maize root assays (Dembinsky *et al.* (2007) *Plant Physiol.* 145: 575-588).

Antibody preparation and immune histochemistry

A polyclonal antibody was produced by incubating rabbits with a specific RTH5 peptide with the amino acid sequence VAGMRPGRMTRMQSSAQM (SEQ ID NO: 57). Sample preparation was slightly modified from the RNA *in situ* hybridization protocol. Fixation was performed using 4% formamide solution. After sectioning, the samples were

deparaffined using RotiClear (Roth) and rehydrated using decreasing ethanol concentrations (90%, 50%, 25%) in Microtubules stabilizing buffer (MtSB) (Albertini *et al.* (1984) *Eur. J. Cell. Biol.* 33: 134-143). Unspecific binding was blocked by 3% BSA in 1x MtSB, followed by incubation with the RTH5 antibody (1:100 in 1x MtSB) overnight at 4 °C. After six wash steps in 1x MtSB the samples were incubated for 2 h with the secondary antibody ZytoChem Plus AP Polymer anti-Rabbit (ZytoMed) according to the manufacturer's protocol. After washing with a buffer containing 100 mM Tris, 100 mM NaCl, and 50 mM $MgCl_2$, the signal was detected using the Roche NBT/BCIP stock solution by incubation in the dark for 5-10 min. The reaction was stopped by three water incubations. Images were obtained as described for RNA *in situ* hybridization experiments.

Detection of superoxide and hydrogen peroxide

Superoxide was detected by incubating root samples in 0.5 mM Nitro blue tetrazolium chloride (NBT) in 0.1 M KCl/ 0.1 M NaCl solution which forms an insoluble blue formazan precipitate upon reaction with superoxide (Bielski *et al.* (1980) *J. Phys. Chem-US* 84:830-833). To detect hydrogen peroxide two dyes were used: DAB (3,3-diaminobenzidine) and 2,7-dichlorodihydrofluorescein diacetate (H_2DCF-DA). DAB forms a brown precipitate when oxidized by peroxidase activity (Thordal-Christensen *et al.* (1997) Plant J. 11:1187-1194). Seedlings were incubated overnight in 1mg/ml DAB dissolved in water. H_2DCF-DA emits a green fluorescence that corresponds to cytoplasmic H_2O_2 levels (Keston and Brandt (1965) Anal. Biochem. 11:1-5). The stock solution was prepared by dissolving 1 mg DCHF-DA in 1 ml DMSO, and mixing with 1 ml H_2O. To obtain the working solution a dilution in 200 ml H_2O was prepared.

Three-day-old seedlings were incubated for 45 min in the dark in petri dishes, each of which contained a detection solution. The signal was detected by mounting primary roots in water on a microscopic slide covered by a 60 mm cover slip. Excitation of DCHF-DA was performed at 488 nm. Emission was detected at 525 nm using a Zeiss Axioplan 2 microscope.

RNA-Seq experiment

Seeds from an F_2-family segregating for *rth5* after nine generations of backcrossing to the inbred B73 were grown in water-soaked germination paper in an incubator at 25 °C. 3 cm root tips were collected from 6-day-old *rth5* mutants and wild-type siblings. Two mutant and two wild-type pools were collected. Each pool consisted of root tissues from six individuals.

RNA was extracted using RNeasy mini kits (Qiagen) with DNaseI treatment per the manufacturer's protocol. RNA quality was analyzed using a Bioanalyer 2100 RNA Nano chip. RNA-Seq libraries were constructed using an Illumina RNA-Seq sample preparation kit following the manufacturer's protocol. In total four libraries were prepared (two mutant and two wild-type) and sequenced on an Illumina HiSeq2000 at the Iowa State University DNA facility, generating 99 bp single-end reads (GenBank accession no. SRP020528).

Trimming and mapping of RNA-Seq reads

Raw reads were subjected to quality checking and trimming to remove low quality bases using a custom trimming pipeline (Liu *et al.* (2012) PLoS ONE 7:e36406). Trimmed reads were aligned to the B73 reference genome (B73Ref2) using GSNAP (Wu and Nacu (2010) *Bioinformatics* 26:873-881) , allowing at most 2 mismatches every 50 bp and 2 bp tails per 50 bp. Only uniquely mapped reads were used for subsequent analyses. The read depth of each gene in the filtered gene set (ZmB73_5b_FGS) was computed based on the coordinates of mapped reads and the annotated locations of genes in the B73 reference genome.

Differential expression analysis of RNA-Seq

Genes with an average of at least five uniquely mapped reads across the four samples were tested for differential expression between the *rth5* mutant and wild-type using the R package *QuasiSeq* (http://cran.r-project.org/web/packages/QuasiSeq). The negative binomial *QLSpline* method implemented in the *QuasiSeq* package was used to compute a p-value for each gene. The 0.75 quantile of reads from each sample was used as the normalization factor (Bullard *et al.* (2010) *BMC Bioinformatics* 11:94). An approach for controlling for multiple testing (Benjamini and Hochberg (1995) *J. Roy. Statistical Society, Series B* 57:289-300) was used to convert p-values to q-values. To approximately control the false discovery rate (FDR) at 10%, only genes with q-values ≤ 0.1 were declared to be differentially expressed.

Gene ontology (GO) enrichment analysis of significantly differentially expressed genes

The software GOseq (Young *et al.* (2010) *Genome Biol.* 11:R14) was used to perform the GO analysis for differentially expressed genes (DEGs), up-regulated DEGs, and down-regulated DEGs in the mutants. p-values for over-representation of DEGs for each GO category were generated by randomly sampling the same number of differentially expressed

genes 10,000 times. Genes were weighted with their read counts from all samples to account for the potential bias in power to detect differential expression in genes with different levels of read counts.

Example 2: Preparation of cDNA Libraries and Isolation and Sequencing of cDNA Clones

mRNAs can be isolated using the Qiagen® RNA isolation kit for total RNA isolation, followed by mRNA isolation via attachment to oligo(dT) Dynabeads from Invitrogen (Life Technologies, Carlsbad, CA), and sequencing libraries can be prepared using the standard mRNA-Seq kit and protocol from Illumina, Inc. (San Diego, CA). In this method, mRNAs are fragmented using a ZnCl2 solution, reverse transcribed into cDNA using random primers, end repaired to create blunt end fragments, 3' A-tailed, and ligated with Illumina paired-end library adaptors. Ligated cDNA fragments can then be PCR amplified using Illumina paired-end library primers, and purified PCR products can be checked for quality and quantity on the Agilent Bioanalyzer DNA 1000 chip prior to sequencing on the Genome Analyzer II equipped with a paired end module.

Reads from the sequencing runs can be soft-trimmed prior to assembly such that the first base pair of each read with an observed FASTQ quality score lower than 15 and all subsequent bases are clipped using a Python script. The Velvet assembler (Zerbino *et al.* (2008) *Genome Research* 18:821-9) can be run under varying kmer and coverage cutoff parameters to produce several putative assemblies along a range of stringency. The contiguous sequences (contigs) within those assemblies can be combined into clusters using Vmatch software (available on the Vmatch website) such that contigs which are identified as substrings of longer contigs are grouped and eliminated, leaving a non-redundant set of longest "sentinel" contigs. These non-redundant sets can be used in alignments to homologous sequences from known model plant species.

In cases where the sequence assemblies are in fragments, the percent identity to other homologous genes can be used to infer which fragments represent a single gene. The fragments that appear to belong together can be computationally assembled such that a translation of the resulting nucleotide sequence returns the amino acid sequence of the homologous protein in a single open-reading frame. These computer-generated assemblies can then be aligned with other polypeptides of the invention.

The article "a" and "an" are used herein to refer to one or more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one or more element.

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

SEQUENCES

SEQ ID NO: 1: Maize (*rth5* genomic sequence) (Maize Genetics and Genomics Database Gene ID: RMZM2G426953)

gcttcctctt gccccccccac cccaccccac cgccaccacc cacccctcgc ccgcaggtgg 60
gcccagcctc catcctcctc ctctgagagc gagccggcgg cggcgagcgg cattgcgggt 120
tatctggtgg agcgggacat ggcggggtac ggggaccggc gatcgccgcc gctggagggc 180
atcaccgtcg acggggccag cagggcccaa tcgccccggg cgggtgcggg ggcgggggcg 240
gggaggctgc cgccgccgcc aggaggcttc gcgcgcgggc tcatgaagca gccgtcgcgg 300
ctggcgtccg gggtgcggca gttcgcttcg cgtgtgtcga tgaaggtgcc cgagggcgtg 360
gccgggatgc ggccgggacg gatgacgcgg atgcagtcta gcgcgcagat ggggctccgt 420
ggcttgcgct tcctcgacaa gacctccggc ggcaaggagg gatggaaggc cgtcgagcgc 480
cgcttcgacg acatggccaa gggcagcgga cgactccaga aggagagctt cggcaagtgc 540
atcggtaagc aaagtgctag tgtgctgctt gctcttgctc catcgagatt gcctcttctg 600
tcggttggcc acatgctgcc taagtgcttg gtgtgatcca gactggaaat aaaccccaaa 660
taatcgatag ttgagacgct gataagaacc gctaggcgta ggcactcata tgacagcaat 720
tgttaatatg tgcatcgaat ctctttcctt gcttttagga attcttgttt gctttacttt 780
ttttttgag tgtttcaaag agctctcatg gagtcatggt actgcaatga actgtgcctg 840
tttccaccaa gcctacctaa ctgagatgag atccaccgtc attcgctccc caagggtctt 900
cccaaatcat gctgaaagaa aagactcttc ccaaatttta atgaagtttc tccccaaagt 960
ggtcctttag cactaatcgg tcagctcgga caaacaagat tctagctagt aacgccatgg 1020
gccaccacca tgtaattcta tcccatgctt tcatctgcca gcaagcgtaa aatcatttgc 1080
tggcgagctg ttcgtgttgc tggtgcccgt ggcgtgatca atggatgatc cgtggctgac 1140
tgactgatct cattcccaag gcatggggga ctccaaggag tttgccggcg agctgttcgt 1200
ggcgctggcg cggaggcgga gcctggagcc ggaggatggc atcaccaaag agcagctcaa 1260
ggagttctgg gaggagatga ccgaccagaa cttcgactcg cggctgcgca ttttctttga 1320
catgtaaagc tccaccgcgc cctcgttccc ggcccaacgg acatgttcat cgttaggcta 1380
ctgataaaac gaattcccct ctgcgtctga ctcaggtgcg acaagaacgg cgatgggatg 1440
ctcacggaag atgaggtcaa agaggtgagc gacggcaact tttcgcgatt tggatgacca 1500
aatatttagg agtggacgaa cttattaaat tatcaacaaa ttaatttgtt ttttaatttg 1560
ttccccccgg caggttatta tactgagcgc gtcagcgaac aagctagcca agctgaaggg 1620
tcacgccgcg acctacgcgt ccctgatcat ggaagagctc gacccggacg accgcggcta 1680
cattgaggtc gtgcttgcag aaacaacatt ttgctttctg ttcgtatcca acagcgataa 1740
gtggtacgcg ttgcttatac aaccaaacgg tgtgatgcag atttggcaac tggagacgtt 1800
gctccggggc atggtgagcg cgcaggctcc cgagaagctg aagcggacga cgtcgagcct 1860
ggcgcggacg atgatcccgt cgcggtaccg gaacccgctg aagcggcacc tgtccaagac 1920
ggtggacttc atccacgaga actggaagcg gatatggctc gtgacgctgt ggctggtcgt 1980
caacgtcgcc ctgttcgtgt acaagttcga gcagtacaag cgccgaaccg cgttccaggt 2040
gatgggctac tgcgtctgcg tcgccaaggg tgccgctgag atcctcaagc tcaacatggc 2100
tctcatcctg ctgcccgtct gccggaatac gctgacgacg ctcaggtcca ccgcgctcag 2160
ccatgtcata cccttcgatg acaacatcaa cttccacaag gtcatcgcgc tgtccatcgc 2220
gatcgccaca gcgatccaca cgctcgcaca cgtgacctgc gacttcccaa ggctgatcag 2280
ctgcccgacg gacaagttca tggccacctt ggggtccaac ttccactaca agcagccgac 2340
ttacctgggc ttgctggaga gcacacccgg ggttaccgga atcctcatga tcatcataat 2400
gtccttctcc ttcacgctgg caacacattc cttcaggcgg agtgtggtga agctgccatc 2460
gccgctacac caccttgccg gtttcaatgc cttctggtac gctcaccacc tgctggtcct 2520
tgcgtatgtc ctgctggtgg tgcactccta cttcatattc ctcaccaggg agtggtacaa 2580
gaagacggta aacctcggac accttgcttt gagatgaaac tcattctccc ctctgcatgt 2640
tcagtcagtg ttgggagcag caaaataaca tatggtctca attgtttatt cagacatgga 2700
tgtacctgat tgtccctgtc ctcttctatg cctgtgaaag agtcatcagg aaatttcgtg 2760
agaacaacta ccatgcggga attgtgaggg taagctattg aactaaattg atttggttca 2820
ttctgagttt ctggctggta attgttaatc aaattctgtt atgtatatgc aggcagcaat 2880
ttatccggga gatgtgctct ctattcacat gaagaagcca cagggtttca agtacaagag 2940
tgggatgtat ctgtttgtta aatgcccaga agtctcgccc ttcgagtggt accatttcat 3000
taattgttgc atacatctcc aagaaataat ttatctgtgg ttcttgctta ccgttctctc 3060
atatttgcag gcacccctttc tctataactt cggcaccagg cgatgactac ttgagtgtgc 3120
atatccgtac gctgggtgac tggacatccg aactgcggat gctttttggg aaggttagtt 3180
tgaggcagca aaagtgttaa gagttggaag agtaacaata ctgtttttttg aatggcagaa 3240
taatgtggac tatcagaaat gtataatgtg gattattcgc acaggcttgc caggcacaag 3300
taacttccaa gaaggctacc cttacaagac ttgaaactac agttgtggca gacgcccaga 3360

```
cagaggacac taggtgctta cctttctttc atttttttcac accctatagt ttgattggtt 3420
ttagagaaac tactgcaact ttaagatttt attaacctgg tcagctttgg tgtgtttgct 3480
ttaggtttcc caaggtctac atagacgggc catacggtgc accagcacaa aattacagga 3540
aatatgacat tcttctgctt attggccttg gaataggagc aactcctttc atcagcatac 3600
tgaaggatat gttgaacaac ctaaaatcca acgaagtaat ttccctttt aatgcattat 3660
ttcgatactg ttaagaaaat caaggatagc tgataaccgt actgagttga ttaaacagga 3720
ggtggaaagc atccacggct ctgagatagg cagcttcaag aacaatggtc caggaagggc 3780
ttacttctac tgggtcacca gagagcaagg atctttcgaa tggttcaaag gagtcatgaa 3840
tgaggttgca gggagcgatc acagcgtact gtctcactct catttctctt aaatcttaat 3900
tacagcgtat gttcataaac taagttgaaa tctggctgct gcagaatgtt atagagatgc 3960
acaattacct gaccagcgtg tatgaagaag gtgacgcaag gtcagctctg attgccatgg 4020
tacagtcact tcagcgtgct aaaaacggcg tggatatcgt ctccggcagc aaggtatacg 4080
ataatttccc ttttgctcaa ctgtaaagtg ttttgcatat gctatctgtt ttggttaacg 4140
gagcactctg ttgcagattc gaacacattt tgcaagacca aactggagga aggtattctg 4200
tgatttggcc agcgcacaca agaactctcg cataggtaag ccacttgtca gtgaacatcg 4260
agtgacacga ttttcagatt cagtacaaca gttaatgttt tctctaaaaa aaaattcagg 4320
agttttctat tgtggatctc cgacgctcac aaaacaactg aaggatcttt cgaaagaatt 4380
cagccagaca accacaaccc ggttccattt ccacaaagag aacttctgag acgtgtaccc 4440
gaagatccga tggactggaa acataattgt atagggaaaa aatacgatag cattggcata 4500
gcagatttag ttttacaagt tttgatgtat gcggggttgt acaaaatatg tgtagaaagc 4560
tagatgtcac catcatacat agattctgaa atgcttgcag ctatatatat tgcattgcat 4620
aagtgaaacc acttgcttcc taggataacc gagttctaga tctcctgctt tgaacacgca 4680
aattttctct tttcagtctt tctgcccgag ctgtcattgc attgcgaata agggtgctat 4740
tctgtttcct agaagattct tgtgtatcat cttgcactgg cacaagtgcc acagaacctg 4800
gtgtgtcttc agcgccctga gaatttccag gtgcattcct gttctttgga ggcctgcctc 4860
ttttttagg ctgcttatca gcagataaag atgtatcttt tcgaggcctg cccctcttcc 4920
taggttgatt actagcctgc aagccgtcca ctctaactat cgccatctta ttgtttgtat 4980
catttccctg ggatagtgta actatatgtg atgttttctc catactgtca tccactgtag 5040
catatatttt ccttggtcta cccacctttg gtttggagtc agaaactgat cccttaagtt 5100
tgcacttgtc gataacaagt gttttaagtg gaacagttat gtccaatggc agtagaactt 5160
tctcaagttc acgccattca agatcttcct ggtctccatc atcataaatg acggtgtacc 5220
agttgctttc aatgtcatac ttggccactt tcccgacgaa atatgtgtct ccaaaaagct 5280
tacgcacttt cctcccttct aaccattcac caaaacgagg atcaacatga gattctgcgt 5340
catttttac tgcagattga gaaccagttg tgcaatcaaa aggaaaatct ggaacgccat 5400
tgataatgat ggcaggctct ttaggaactt tgattataga tgtagtttct gactcttgct 5460
tgccatcagc cccctccatg gttttctata tggaaaatat ggaagaatca ggaaacagat 5520
tcat                                                              5524
```

SEQ ID NO: 2: Maize (*rth5* cDNA sequence) (GenBank Accession No: DQ855284)

```
gtgggcccag cctccatcct cctcctctga gagcgagccg gcggcggcga gcggcattgc 60
gggttatctg gtggagcggg acatggcggg gtacggggac cggcgatcgc cgccgctgga 120
gggcatcacc gtcgacgggg ccagcagggc ccaatcgccc ggggcgggtg cgggggcggg 180
ggcggggagg ctgccgccgc cgccaggagg cttcgcgcgc gggctcatga agcagccgtc 240
gcggctggcg tccggggtgc ggcagttcgc ttcgcgtgtg tcgatgaagg tgcccgaggg 300
cgtggccggg atgcggccgg gacggatgac gcggatgcag tctagcgcgc agatggggct 360
ccgtggcttg cgcttcctcg acaagacctc cggcggcaag gagggatgga aggccgtcga 420
gcgccgcttc gacgacatgg ccaagggcag cggacgactc cagaaggaga gcttcggcaa 480
gtgcatcggc atgggggact ccaaggagtt tgccggcgag ctgttcgtgg cgctggcgcg 540
gaggcggagc ctggagccgg aggatggcat caccaaagag cagctcaagg agttctggga 600
ggagatgacc gaccagaact tcgactcgcg gctgcgcatt ttctttgaca tgtgcgacaa 660
gaacggcgat gggatgctca cggaagatga ggtcaaagag gttattatac tgagcgcgtc 720
agcgaacaag ctagccaagc tgaagggtca cgccgcgacc tacgcgtccc tgatcatgga 780
agagctcgac ccggacgacc gcggctacat tgagatttgg caactggaga cgttgctccg 840
gggcatggtg agcgcgcagg ctcccgagaa gctgaagcgg acgacgtcga gcctggcgcg 900
gacgatgatc ccgtcgcggt accggaaccc gctgaagcgg cacctgtcca agacggtgga 960
cttcatccac gagaactgga agcggatatg gctcgtgacg ctgtggctgg tcgtcaacgt 1020
cgccctgttc gtgtacaagt tcgagcagta caagcgccga accgcgttcc aggtgatggg 1080
ctactgcgtc tgcgtcgcca agggtgccgc tgagatcctc aagctcaaca tggctctcat 1140
cctgctgccc gtctgccgga atacgctgac gacgctcagg tccaccgcgc tcagccatgt 1200
```

```
cataccettc gatgacaaca tcaacttcca caaggtcatc gcgctgtcca tcgcgatcgc 1260
cacagcgatc cacacgctcg cacacgtgac ctgcgacttc ccaaggctga tcagctgccc 1320
gacggacaag ttcatggcca ccttggggtc caacttccac tacaagcagc cgacttacct 1380
gggcttgctg gagagcacac ccggggttaa cggaaacctc atgatcatca taatgtcctt 1440
ctccttcaag ctggcaacac attccttcag gcggagtgtg gtgaagctgc catcgccgct 1500
acaccacctt gccggtttca atgccttctg gtacgctcac cacctgctgg tccttgcgta 1560
tgtcctgctg gtggtgcact cctacttcat attcctcacc agggagtggt acaagaagac 1620
gacatggatg tacctgattg tccctgtcct cttctatgcc tgtgaaagag tcatcaggaa 1680
atttcgtgag aacaactacc atgcgggaat tgtgagggca gcaatttatc cgggagatgt 1740
gctctctatt cacatgaaga agccacaggg tttcaagtac aagagtggga tgtatctgtt 1800
tgttaaatgc ccagaagtct cgcccttcga gtggcacccc ttctctataa cttcggcacc 1860
aggcgatgac tacttgagtg tgcatatccg tacgctgggt gactggacat ccgaactgcg 1920
gatgcttttt gggaaggctt gccaggcaca agtaacttcc aagaaggcta cccttacaag 1980
acttgaaact acagttgtgg cagacgccca gacagaggac actaggtttc ccaaggtcta 2040
catagacggg ccatacggtg caccagcaca aaattacagg aaatatgaca ttcttctgct 2100
tattggcctt ggaataggag caactccttt catcagcata ctgaaggata tgttgaacaa 2160
cctaaaatcc aacgaagagg tgggaagcat ccacggctct gagataggca gcttcaagaa 2220
caatggtcca ggaagggctt acttctactg ggtcaccaga gagcaaggat ctttcgaatg 2280
gttcaaagga gtcatgaatg aggttgcagg gagcgatcac agcaatgtta tagagatgca 2340
caattacctg accagcgtgt atgaagaagg tgacgcaagg tcagctctga ttgccatggt 2400
acagtcactt cagcgtgcta aaaacggcgt ggatatcgtc tccggcagca agattcgaac 2460
acattttgca agaccaaact ggaggaaggt attctgtgat ttggccagcg cacacaagaa 2520
ctctcgcata ggagttttct attgtggatc tccgacgctc acaaaacaac tgaaggatct 2580
ttcgaaagaa ttcagccaga caaccacaac ccggttccat ttccacaagg aaaacttctg 2640
agacgacgtg tacccgaaga tccgatggac tggaaacata attgtatagg gaaaaaatac 2700
gatagcattg gcatagcaga tttagtttta caagttttga tgtatgcggg gttgtacaaa 2760
atatgtgtag aaagctagat gtcaccatca tacatagatt ctgaaatgct tgcagatata 2820
tatattgcat tgcataagtg aaaccacttg cttcctagga taaccgagtt ctagatctcc 2880
tgctttgaac acgcaaattt tctcttttca gtctttctgc ccgagctgtc attgcattgc 2940
gaataagggt gctattctgt ttcctagaag attcttgtgt atcatcttgc actggcacaa 3000
gtgccacaga acctggtgtg tcttcagcgc cctgagaatt tccaggtgca ttcctgttct 3060
ttggaggcct gcctcttttt ttaggctgct tatcagcaga taaagatgta tcttttcgag 3120
gcctgcccct cttcctaggt tgattactag cctgcaagcc gtccactcta actatcgcca 3180
tcttattgtt tgtatcattt ccctgggata gtgtaactat atgtgatgtt ttctccatac 3240
tgtcatccac tgtagcatat attttccttg gtctacccac ctttggtttg gagtcagaaa 3300
ctgatccctt aagtttgcac ttgtcgataa caagtgtttt aagtggaaca gctatgtcca 3360
atggcagtag aactttctca agttcacgcc attcaagatc ttcctggtct ccatcatcat 3420
aaatgacggt gtaccagttg ctttcaatgt catacttggc cactttcccg acgaaatatg 3480
tgtctccaaa aagcttacgc actttcctcc cttctaacca ttcaccaaaa cgaggatcaa 3540
catgagattc tgcgtcattt tttactgcag attgagaacc agttgtgcaa tc         3592
```

SEQ ID NO: 3: Maize (RTH5 amino acid sequence) (GenBank Accession No: DQ855284)

```
Met Ala Gly Tyr Gly Asp Arg Arg Ser Pro Pro Leu Glu Gly Ile Thr
 1               5                  10                  15
Val Asp Gly Ala Ser Arg Ala Gln Ser Pro Gly Ala Gly Ala Gly Ala
            20                  25                  30
Gly Ala Gly Arg Leu Pro Pro Pro Pro Gly Gly Phe Ala Arg Gly Leu
        35                  40                  45
Met Lys Gln Pro Ser Arg Leu Ala Ser Gly Val Arg Gln Phe Ala Ser
    50                  55                  60
Arg Val Ser Met Lys Val Pro Glu Gly Val Ala Gly Met Arg Pro Gly
65                  70                  75                  80
Arg Met Thr Arg Met Gln Ser Ser Ala Gln Met Gly Leu Arg Gly Leu
                85                  90                  95
Arg Phe Leu Asp Lys Thr Ser Gly Gly Lys Glu Gly Trp Lys Ala Val
            100                 105                 110
Glu Arg Arg Phe Asp Asp Met Ala Lys Gly Ser Gly Arg Leu Gln Lys
        115                 120                 125
Glu Ser Phe Gly Lys Cys Ile Gly Met Gly Asp Ser Lys Glu Phe Ala
```

```
    130                 135                 140
Gly Glu Leu Phe Val Ala Leu Ala Arg Arg Arg Ser Leu Glu Pro Glu
145                 150                 155                 160
Asp Gly Ile Thr Lys Glu Gln Leu Lys Glu Phe Trp Glu Glu Met Thr
                165                 170                 175
Asp Gln Asn Phe Asp Ser Arg Leu Arg Ile Phe Phe Asp Met Cys Asp
            180                 185                 190
Lys Asn Gly Asp Gly Met Leu Thr Glu Asp Glu Val Lys Glu Val Ile
        195                 200                 205
Ile Leu Ser Ala Ser Ala Asn Lys Leu Ala Lys Leu Lys Gly His Ala
    210                 215                 220
Ala Thr Tyr Ala Ser Leu Ile Met Glu Glu Leu Asp Pro Asp Asp Arg
225                 230                 235                 240
Gly Tyr Ile Glu Ile Trp Gln Leu Glu Thr Leu Leu Arg Gly Met Val
                245                 250                 255
Ser Ala Gln Ala Pro Glu Lys Leu Lys Arg Thr Thr Ser Ser Leu Ala
            260                 265                 270
Arg Thr Met Ile Pro Ser Arg Tyr Arg Asn Pro Leu Lys Arg His Leu
        275                 280                 285
Ser Lys Thr Val Asp Phe Ile His Glu Asn Trp Lys Arg Ile Trp Leu
    290                 295                 300
Val Thr Leu Trp Leu Val Val Asn Val Ala Leu Phe Val Tyr Lys Phe
305                 310                 315                 320
Glu Gln Tyr Lys Arg Arg Thr Ala Phe Gln Val Met Gly Tyr Cys Val
                325                 330                 335
Cys Val Ala Lys Gly Ala Ala Glu Ile Leu Lys Leu Asn Met Ala Leu
            340                 345                 350
Ile Leu Leu Pro Val Cys Arg Asn Thr Leu Thr Thr Leu Arg Ser Thr
        355                 360                 365
Ala Leu Ser His Val Ile Pro Phe Asp Asp Asn Ile Asn Phe His Lys
    370                 375                 380
Val Ile Ala Leu Ser Ile Ala Ile Ala Thr Ala Ile His Thr Leu Ala
385                 390                 395                 400
His Val Thr Cys Asp Phe Pro Arg Leu Ile Ser Cys Pro Thr Asp Lys
                405                 410                 415
Phe Met Ala Thr Leu Gly Ser Asn Phe His Tyr Lys Gln Pro Thr Tyr
            420                 425                 430
Leu Gly Leu Leu Glu Ser Thr Pro Gly Val Thr Gly Ile Leu Met Ile
        435                 440                 445
Ile Ile Met Ser Phe Ser Phe Thr Leu Ala Thr His Ser Phe Arg Arg
    450                 455                 460
Ser Val Val Lys Leu Pro Ser Pro Leu His His Leu Ala Gly Phe Asn
465                 470                 475                 480
Ala Phe Trp Tyr Ala His His Leu Leu Val Leu Ala Tyr Val Leu Leu
                485                 490                 495
Val Val His Ser Tyr Phe Ile Phe Leu Thr Arg Glu Trp Tyr Lys Lys
            500                 505                 510
Thr Thr Trp Met Tyr Leu Ile Val Pro Val Leu Phe Tyr Ala Cys Glu
        515                 520                 525
Arg Val Ile Arg Lys Phe Arg Glu Asn Asn Tyr His Ala Gly Ile Val
    530                 535                 540
Arg Ala Ala Ile Tyr Pro Gly Asp Val Leu Ser Ile His Met Lys Lys
545                 550                 555                 560
Pro Gln Gly Phe Lys Tyr Lys Ser Gly Met Tyr Leu Phe Val Lys Cys
                565                 570                 575
Pro Glu Val Ser Pro Phe Glu Trp His Pro Phe Ser Ile Thr Ser Ala
            580                 585                 590
Pro Gly Asp Asp Tyr Leu Ser Val His Ile Arg Thr Leu Gly Asp Trp
        595                 600                 605
Thr Ser Glu Leu Arg Met Leu Phe Gly Lys Ala Cys Gln Ala Gln Val
```

```
    610                 615                 620
Thr Ser Lys Lys Ala Thr Leu Thr Arg Leu Glu Thr Thr Val Val Ala
625                 630                 635                 640
Asp Ala Gln Thr Glu Asp Thr Arg Phe Pro Lys Val Tyr Ile Asp Gly
                645                 650                 655
Pro Tyr Gly Ala Pro Ala Gln Asn Tyr Arg Lys Tyr Asp Ile Leu Leu
            660                 665                 670
Leu Ile Gly Leu Gly Ile Gly Ala Thr Pro Phe Ile Ser Ile Leu Lys
        675                 680                 685
Asp Met Leu Asn Asn Leu Lys Ser Asn Glu Glu Val Glu Ser Ile His
    690                 695                 700
Gly Ser Glu Ile Gly Ser Phe Lys Asn Asn Gly Pro Gly Arg Ala Tyr
705                 710                 715                 720
Phe Tyr Trp Val Thr Arg Glu Gln Gly Ser Phe Glu Trp Phe Lys Gly
                725                 730                 735
Val Met Asn Glu Val Ala Gly Ser Asp His Ser Asn Val Ile Glu Met
            740                 745                 750
His Asn Tyr Leu Thr Ser Val Tyr Glu Glu Gly Asp Ala Arg Ser Ala
        755                 760                 765
Leu Ile Ala Met Val Gln Ser Leu Gln Arg Ala Lys Asn Gly Val Asp
    770                 775                 780
Ile Val Ser Gly Ser Lys Ile Arg Thr His Phe Ala Arg Pro Asn Trp
785                 790                 795                 800
Arg Lys Val Phe Cys Asp Leu Ala Ser Ala His Lys Asn Ser Arg Ile
                805                 810                 815
Gly Val Phe Tyr Cys Gly Ser Pro Thr Leu Thr Lys Gln Leu Lys Asp
            820                 825                 830
Leu Ser Lys Glu Phe Ser Gln Thr Thr Thr Thr Arg Phe His Phe His
        835                 840                 845
Lys Glu Asn Phe
    850
```

SEQ ID NO: 4: GRMZM2G401179_P01_variant; corn; 21-aa at position 215 removed to get better alignment

```
MTSSAGYADGGTGGLGDGDPPVPPLRKQPSRIASGMRRLASKVSAAVPEMRGLKRTHSGTQSGLRGLRFLDKTSA
GKDGWKSVEKRFDEMNTEGRLQRENFAKCIGMADSNEFASEVFVALARRTHINPDDGVTKEQLKQFWEEMTDQNF
DSRLRIFFDMCDKNGDGKLTEDEVKEVIVLSASANKLAKLKKHAATYASLIMEELDPDHRGYIEIWQLETLLRGM
VTASAPPEKMNMASASLARTMVPSSHRSPLQRRINKAVDFVHENWKRIWVLSLWGVLCIALFVFKFIQYRRRAVF
EVMGYCVCIAKGAAETLKLNMALILLPVCRNTLTRLRSTALSKVVPFDDNINFHKVIALAIAIGSATHTLAHVLC
DFPRLVACPKDKFMEKLGPFFNYVQPTWPTLLASIPGWTGILLILIMSFSFTLATHSFRRSVVKLPSPLHHLAGF
NAFWYAHHLLVIAYVLLVMHSYFIFLTKQWYKRTTWMYLAVPVVFYASERSIRKIREQSYRVSIIKAAIYPGNVL
SLYMTKPPGFKYKSGMYMFVKCPDVSPFEWHPFSITSAPGDDYLSVHIRTLGDWTSELRNLFGKACEAEVTSKKA
TLARLETTVVAHGLAEDTRFPKVFIDGPYGAPAQNYRKYDILLLIGLGIGATPFISILKDLLNNIKSNEEMHDAE
LGCSLKTNGPGRAYFYWVTREQGSFEWFKGVMNDVAESDRDDVIEMHNYLTSVYEEGDARSALIAMVQSLQHAKN
GVDIVSGSKIRTHFARPNWRKVFSDLANAHKNSRIGVFYCGSPTLTKTLRDLSIEFSSTTTTRFHFHKENF
```

SEQ ID NO: 5: GRMZM2G323731_P01_ Maize (amino acid)

```
Met Ala Ser Ser Ser Gly Tyr Ala Asp Gly Gly Thr Gly Gly Leu Gly
 1               5                  10                  15
Asp Gly Asp Pro Pro Val Pro Ser Met Arg Lys Gln Pro Ser Arg Ile
            20                  25                  30
Ala Ser Gly Met Arg Arg Leu Ala Ser Lys Val Ser Ala Ala Val Pro
        35                  40                  45
Glu Met Arg Gly Leu Lys Arg Thr Gln Ser Gly Ala Gln Ser Gly Leu
    50                  55                  60
```

```
Arg Gly Leu Arg Phe Leu Asp Lys Thr Ser Ala Gly Lys Asp Gly Trp
65                  70                  75                  80
Lys Thr Val Glu Lys Arg Phe Asp Glu Met Ser Thr Asp Gly Arg Leu
                85                  90                  95
Gln Arg Glu Asn Phe Ala Lys Cys Ile Gly Met Ala Asp Ser Lys Glu
            100                 105                 110
Phe Ala Ser Glu Val Phe Val Ala Leu Ala Arg Arg Arg His Ile Asn
        115                 120                 125
Pro Asp Asp Gly Val Thr Lys Glu Gln Leu Lys Glu Phe Trp Glu Glu
    130                 135                 140
Met Thr Asp Gln Asn Phe Asp Ser Arg Leu Arg Ile Phe Phe Asp Met
145                 150                 155                 160
Cys Asp Lys Asn Gly Asp Gly Lys Leu Thr Glu Asp Glu Val Lys Glu
                165                 170                 175
Val Ile Val Leu Ser Ala Ser Ala Asn Lys Leu Ala Lys Leu Lys Lys
            180                 185                 190
His Ala Ala Thr Tyr Ala Ser Leu Ile Met Glu Glu Leu Asp Pro Asp
        195                 200                 205
His Arg Gly Tyr Ile Glu Ile Trp Gln Leu Glu Thr Leu Leu Arg Gly
    210                 215                 220
Met Val Thr Ala Ser Gly Pro Thr Asn Met Gly Ala Ser Ala Ser Leu
225                 230                 235                 240
Ala Arg Thr Met Val Pro Ser Ser His Arg Thr Pro Leu Gln Arg Arg
                245                 250                 255
Met Asn Lys Ala Val Asp Leu Val His Glu Asn Trp Lys Arg Ile Trp
            260                 265                 270
Val Leu Ser Leu Trp Gly Val Leu Asn Met Ala Leu Phe Val Phe Lys
        275                 280                 285
Phe Thr Gln Tyr Arg Arg Arg Ala Val Phe Glu Val Met Gly Tyr Cys
    290                 295                 300
Val Cys Ile Ala Lys Gly Ala Ala Glu Thr Leu Lys Leu Asn Met Ala
305                 310                 315                 320
Leu Ile Leu Leu Pro Val Cys Arg Asn Thr Leu Thr Arg Leu Arg Ser
                325                 330                 335
Thr Ala Leu Ser Lys Val Val Pro Phe Asp Asp Asn Ile Asn Phe His
            340                 345                 350
Lys Val Ile Ala Leu Ala Ile Ala Ile Gly Ser Ala Thr His Thr Leu
        355                 360                 365
Ala His Val Leu Cys Asp Phe Pro Arg Leu Val Ala Cys Pro Lys Asp
    370                 375                 380
Lys Phe Met Glu Lys Leu Gly Pro Phe Phe Asn Tyr Ala Gln Pro Thr
385                 390                 395                 400
Trp Ala Thr Leu Leu Ser Ser Ile Pro Gly Trp Thr Gly Ile Leu Leu
                405                 410                 415
Ile Leu Ile Met Ser Phe Ser Phe Thr Leu Ala Thr His Ser Phe Arg
            420                 425                 430
Arg Ser Val Val Lys Leu Pro Ser Pro Leu His His Leu Ala Gly Phe
        435                 440                 445
Asn Ala Phe Trp Tyr Ala His His Leu Leu Val Ile Ala Tyr Val Leu
    450                 455                 460
Leu Val Met His Ser Tyr Phe Ile Phe Leu Thr Lys Gln Trp Tyr Lys
465                 470                 475                 480
Arg Thr Thr Trp Met Tyr Leu Ala Val Pro Val Val Phe Tyr Ala Ser
                485                 490                 495
Glu Arg Ser Ile Arg Arg Ile Arg Glu Lys Ser Tyr Arg Val Ser Ile
            500                 505                 510
Ile Lys Ala Ala Ile Tyr Pro Gly Asn Val Leu Ser Leu Tyr Met Lys
        515                 520                 525
Lys Pro Thr Ser Phe Lys Tyr Lys Ser Gly Met Tyr Met Phe Val Lys
    530                 535                 540
```

```
Cys Pro Asp Val Ser Pro Phe Glu Trp His Pro Phe Ser Ile Thr Ser
545                 550                 555                 560
Ala Pro Gly Asp Asp Tyr Leu Ser Val His Ile Arg Thr Leu Gly Asp
                565                 570                 575
Trp Thr Ser Glu Leu Arg Asn Leu Phe Gly Lys Ala Cys Glu Ala Glu
            580                 585                 590
Val Thr Ser Lys Lys Ala Thr Leu Ala Arg Leu Glu Thr Thr Val Val
        595                 600                 605
Ala His Gly Leu Ala Glu Asp Thr Arg Phe Pro Lys Val Phe Ile Asp
    610                 615                 620
Gly Pro Tyr Gly Ala Pro Ala Gln Asn Tyr Arg Lys Tyr Asp Ile Leu
625                 630                 635                 640
Leu Leu Ile Gly Leu Gly Ile Gly Ala Thr Pro Phe Ile Ser Ile Leu
                645                 650                 655
Lys Asp Leu Leu Asn Asn Ile Lys Ser Asn Glu Glu Met Gln Ser Met
            660                 665                 670
His Asp Thr Glu Leu Gly Cys Ser Phe Lys Thr Asn Gly Pro Gly Arg
        675                 680                 685
Ala Tyr Phe Tyr Trp Val Thr Arg Glu Gln Gly Ser Phe Glu Trp Phe
    690                 695                 700
Lys Gly Val Met Asn Asp Val Ala Glu Ser Asp His Asp Asp Val Ile
705                 710                 715                 720
Glu Met His Asn Tyr Leu Thr Ser Val Tyr Glu Glu Gly Asp Ala Arg
                725                 730                 735
Ser Ala Leu Ile Ala Met Val Gln Ser Leu Gln His Ala Lys Asp Gly
            740                 745                 750
Val Asp Ile Val Ser Gly Ser Lys Ile Arg Thr His Phe Ala Arg Pro
        755                 760                 765
Asn Trp Arg Lys Val Phe Ser Asp Leu Ala Asn Ala His Lys Asn Ser
    770                 775                 780
Arg Ile Gly Val Phe Tyr Cys Gly Ser Pro Thr Leu Thr Lys Thr Leu
785                 790                 795                 800
Arg Asp Leu Ser Ile Glu Phe Ser Ser Thr Thr Thr Thr Arg Phe His
                805                 810                 815
Phe His Lys Glu Asn Phe
            820
```

SEQ ID NO: 6: GRMZM2G401179_P01_Maize

```
Met Thr Ser Ser Ala Gly Tyr Ala Asp Gly Gly Thr Gly Gly Leu Gly
 1               5                  10                  15
Asp Gly Asp Pro Pro Val Pro Pro Leu Arg Lys Gln Pro Ser Arg Ile
            20                  25                  30
Ala Ser Gly Met Arg Arg Leu Ala Ser Lys Val Ser Ala Ala Val Pro
        35                  40                  45
Glu Met Arg Gly Leu Lys Arg Thr His Ser Gly Thr Gln Ser Gly Leu
    50                  55                  60
Arg Gly Leu Arg Phe Leu Asp Lys Thr Ser Ala Gly Lys Asp Gly Trp
65                  70                  75                  80
Lys Ser Val Glu Lys Arg Phe Asp Glu Met Asn Thr Glu Gly Arg Leu
                85                  90                  95
Gln Arg Glu Asn Phe Ala Lys Cys Ile Gly Met Ala Asp Ser Asn Glu
            100                 105                 110
Phe Ala Ser Glu Val Phe Val Ala Leu Ala Arg Arg Thr His Ile Asn
        115                 120                 125
Pro Asp Asp Gly Val Thr Lys Glu Gln Leu Lys Gln Phe Trp Glu Glu
    130                 135                 140
Met Thr Asp Gln Asn Phe Asp Ser Arg Leu Arg Ile Phe Phe Asp Met
145                 150                 155                 160
Cys Asp Lys Asn Gly Asp Gly Lys Leu Thr Glu Asp Glu Val Lys Glu
                165                 170                 175
```

```
Val Ile Val Leu Ser Ala Ser Ala Asn Lys Leu Ala Lys Leu Lys Lys
            180                 185                 190
His Ala Ala Thr Tyr Ala Ser Leu Ile Met Glu Glu Leu Asp Pro Asp
        195                 200                 205
His Arg Gly Tyr Ile Glu Val His Leu Ala Leu Phe Ala Cys Leu Ser
    210                 215                 220
Arg Ser Ser Ile Leu Thr Glu Asp Arg Lys Gln Ile Trp Gln Leu Glu
225                 230                 235                 240
Thr Leu Leu Arg Gly Met Val Thr Ala Ser Ala Pro Pro Glu Lys Met
                245                 250                 255
Asn Met Ala Ser Ala Ser Leu Ala Arg Thr Met Val Pro Ser Ser His
            260                 265                 270
Arg Ser Pro Leu Gln Arg Arg Ile Asn Lys Ala Val Asp Phe Val His
        275                 280                 285
Glu Asn Trp Lys Arg Ile Trp Val Leu Ser Leu Trp Gly Val Leu Cys
    290                 295                 300
Ile Ala Leu Phe Val Phe Lys Phe Ile Gln Tyr Arg Arg Arg Ala Val
305                 310                 315                 320
Phe Glu Val Met Gly Tyr Cys Val Cys Ile Ala Lys Gly Ala Ala Glu
                325                 330                 335
Thr Leu Lys Leu Asn Met Ala Leu Ile Leu Leu Pro Val Cys Arg Asn
            340                 345                 350
Thr Leu Thr Arg Leu Arg Ser Thr Ala Leu Ser Lys Val Val Pro Phe
        355                 360                 365
Asp Asp Asn Ile Asn Phe His Lys Val Ile Ala Leu Ala Ile Ala Ile
    370                 375                 380
Gly Ser Ala Thr His Thr Leu Ala His Val Leu Cys Asp Phe Pro Arg
385                 390                 395                 400
Leu Val Ala Cys Pro Lys Asp Lys Phe Met Glu Lys Leu Gly Pro Phe
                405                 410                 415
Phe Asn Tyr Val Gln Pro Thr Trp Pro Thr Leu Leu Ala Ser Ile Pro
            420                 425                 430
Gly Trp Thr Gly Ile Leu Leu Ile Leu Ile Met Ser Phe Ser Phe Thr
        435                 440                 445
Leu Ala Thr His Ser Phe Arg Arg Ser Val Val Lys Leu Pro Ser Pro
    450                 455                 460
Leu His His Leu Ala Gly Phe Asn Ala Phe Trp Tyr Ala His His Leu
465                 470                 475                 480
Leu Val Ile Ala Tyr Val Leu Leu Val Met His Ser Tyr Phe Ile Phe
                485                 490                 495
Leu Thr Lys Gln Trp Tyr Lys Arg Thr Thr Trp Met Tyr Leu Ala Val
            500                 505                 510
Pro Val Val Phe Tyr Ala Ser Glu Arg Ser Ile Arg Lys Ile Arg Glu
        515                 520                 525
Gln Ser Tyr Arg Val Ser Ile Ile Lys Ala Ala Ile Tyr Pro Gly Asn
    530                 535                 540
Val Leu Ser Leu Tyr Met Thr Lys Pro Pro Gly Phe Lys Tyr Lys Ser
545                 550                 555                 560
Gly Met Tyr Met Phe Val Lys Cys Pro Asp Val Ser Pro Phe Glu Trp
                565                 570                 575
His Pro Phe Ser Ile Thr Ser Ala Pro Gly Asp Asp Tyr Leu Ser Val
            580                 585                 590
His Ile Arg Thr Leu Gly Asp Trp Thr Ser Glu Leu Arg Asn Leu Phe
        595                 600                 605
Gly Lys Ala Cys Glu Ala Glu Val Thr Ser Lys Lys Ala Thr Leu Ala
    610                 615                 620
Arg Leu Glu Thr Thr Val Val Ala His Gly Leu Ala Glu Asp Thr Arg
625                 630                 635                 640
Phe Pro Lys Val Phe Ile Asp Gly Pro Tyr Gly Ala Pro Ala Gln Asn
                645                 650                 655
```

```
Tyr Arg Lys Tyr Asp Ile Leu Leu Leu Ile Gly Leu Gly Ile Gly Ala
            660                 665                 670
Thr Pro Phe Ile Ser Ile Leu Lys Asp Leu Leu Asn Asn Ile Lys Ser
        675                 680                 685
Asn Glu Glu Met His Asp Ala Glu Leu Gly Cys Ser Leu Lys Thr Asn
    690                 695                 700
Gly Pro Gly Arg Ala Tyr Phe Tyr Trp Val Thr Arg Glu Gln Gly Ser
705                 710                 715                 720
Phe Glu Trp Phe Lys Gly Val Met Asn Asp Val Ala Glu Ser Asp Arg
                725                 730                 735
Asp Asp Val Ile Glu Met His Asn Tyr Leu Thr Ser Val Tyr Glu Glu
            740                 745                 750
Gly Asp Ala Arg Ser Ala Leu Ile Ala Met Val Gln Ser Leu Gln His
        755                 760                 765
Ala Lys Asn Gly Val Asp Ile Val Ser Gly Ser Lys Ile Arg Thr His
    770                 775                 780
Phe Ala Arg Pro Asn Trp Arg Lys Val Phe Ser Asp Leu Ala Asn Ala
785                 790                 795                 800
His Lys Asn Ser Arg Ile Gly Val Phe Tyr Cys Gly Ser Pro Thr Leu
                805                 810                 815
Thr Lys Thr Leu Arg Asp Leu Ser Ile Glu Phe Ser Ser Thr Thr Thr
            820                 825                 830
Thr Arg Phe His Phe His Lys Glu Asn Phe
        835                 840
```

SEQ ID NO: 7: At3g45810.1_ Arabidopsis thaliana

```
Met Lys Asn Asn Lys Lys Val Gly Thr Glu Asp Ser Thr Lys Trp Met
 1               5                  10                  15
Leu Glu Ser Val Glu Ile Asp Pro Lys Gly Asp Ser Ser Val Lys Gln
            20                  25                  30
Pro Glu Ser Thr Ile Asn Ser Asn Asn Pro Glu Ser Ser Gly Ala Gly
        35                  40                  45
Gly Gly Ile Leu Lys Asn Val Ser Lys Asn Leu Ala Val Gly Ser Ile
    50                  55                  60
Ile Arg Ser Met Ser Val Asn Lys Trp Arg Lys Ser Gly Asn Leu Gly
65                  70                  75                  80
Ser Pro Ser Thr Arg Lys Ser Gly Asn Leu Gly Pro Pro Leu Pro Val
                85                  90                  95
Ser Gln Val Lys Arg Pro Gly Pro Gln Arg Val Glu Arg Thr Thr Ser
            100                 105                 110
Ser Ala Ala Arg Gly Leu Gln Ser Leu Arg Phe Leu Asp Arg Thr Val
        115                 120                 125
Thr Gly Arg Glu Arg Asp Ser Trp Arg Ser Ile Glu Asn Arg Phe Asn
    130                 135                 140
Gln Phe Ala Val Asp Gly Arg Leu Pro Lys Asp Lys Phe Gly Val Cys
145                 150                 155                 160
Ile Gly Met Gly Asp Thr Leu Glu Phe Ala Ala Lys Val Tyr Glu Ala
                165                 170                 175
Leu Gly Arg Arg Arg Gln Ile Lys Thr Glu Asn Gly Ile Asp Lys Glu
            180                 185                 190
Gln Leu Lys Leu Phe Trp Glu Asp Met Ile Lys Lys Asp Leu Asp Cys
        195                 200                 205
Arg Leu Gln Ile Phe Phe Asp Met Cys Asp Lys Asp Gly Asp Gly Lys
    210                 215                 220
Leu Thr Glu Glu Glu Val Lys Glu Val Ile Val Leu Ser Ala Ser Ala
225                 230                 235                 240
Asn Arg Leu Val Asn Leu Lys Lys Asn Ala Ala Ser Tyr Ala Ser Leu
                245                 250                 255
Ile Met Glu Glu Leu Asp Pro Asn Glu Gln Gly Tyr Ile Glu Met Trp
```

```
                260                 265                 270
Gln Leu Glu Val Leu Leu Thr Gly Ile Val Ser Asn Ala Asp Ser His
        275                 280                 285
Lys Val Val Arg Lys Ser Gln Gln Leu Thr Arg Ala Met Ile Pro Lys
    290                 295                 300
Arg Tyr Arg Thr Pro Thr Ser Lys Tyr Val Val Val Thr Ala Glu Leu
305                 310                 315                 320
Met Tyr Glu His Trp Lys Lys Ile Trp Val Val Thr Leu Trp Leu Ala
                325                 330                 335
Val Asn Val Val Leu Phe Met Trp Lys Tyr Glu Glu Phe Thr Thr Ser
            340                 345                 350
Pro Leu Tyr Asn Ile Thr Gly Arg Cys Leu Cys Ala Ala Lys Gly Thr
        355                 360                 365
Ala Glu Ile Leu Lys Leu Asn Met Ala Leu Ile Leu Val Pro Val Leu
    370                 375                 380
Arg Arg Thr Leu Thr Phe Leu Arg Ser Thr Phe Leu Asn His Leu Ile
385                 390                 395                 400
Pro Phe Asp Asp Asn Ile Asn Phe His Lys Leu Ile Ala Val Ala Ile
                405                 410                 415
Ala Val Ile Ser Leu Leu His Thr Ala Leu His Met Leu Cys Asn Tyr
            420                 425                 430
Pro Arg Leu Ser Ser Cys Pro Tyr Asn Phe Tyr Ser Asp Tyr Ala Gly
        435                 440                 445
Asn Leu Leu Gly Ala Lys Gln Pro Thr Tyr Leu Gly Leu Met Leu Thr
    450                 455                 460
Pro Val Ser Val Thr Gly Val Leu Met Ile Ile Phe Met Gly Ile Ser
465                 470                 475                 480
Phe Thr Leu Ala Met His Tyr Phe Arg Arg Asn Ile Val Lys Leu Pro
                485                 490                 495
Ile Pro Phe Asn Arg Leu Ala Gly Phe Asn Ser Phe Trp Tyr Ala His
            500                 505                 510
His Leu Leu Val Ile Ala Tyr Ala Leu Leu Ile Ile His Gly Tyr Ile
        515                 520                 525
Leu Ile Ile Glu Lys Pro Trp Tyr Gln Lys Thr Thr Trp Met Tyr Val
    530                 535                 540
Ala Ile Pro Met Val Leu Tyr Ala Ser Glu Arg Leu Phe Ser Arg Val
545                 550                 555                 560
Gln Glu His Asn His Arg Val His Ile Ile Lys Ala Ile Val Tyr Ser
                565                 570                 575
Gly Asn Val Leu Ala Leu Tyr Met Thr Lys Pro Gln Gly Phe Lys Tyr
            580                 585                 590
Lys Ser Gly Met Tyr Met Phe Val Lys Cys Pro Asp Ile Ser Lys Phe
        595                 600                 605
Glu Trp His Pro Phe Ser Ile Thr Ser Ala Pro Gly Asp Glu Tyr Leu
    610                 615                 620
Ser Val His Ile Arg Ala Leu Gly Asp Trp Thr Ser Glu Leu Arg Asn
625                 630                 635                 640
Arg Phe Ala Glu Thr Cys Glu Pro His Gln Lys Ser Lys Pro Ser Pro
                645                 650                 655
Asn Asp Leu Ile Arg Met Glu Thr Arg Ala Arg Gly Ala Asn Pro His
            660                 665                 670
Val Glu Glu Ser Gln Ala Leu Phe Pro Arg Ile Phe Ile Lys Gly Pro
        675                 680                 685
Tyr Gly Ala Pro Ala Gln Ser Tyr Gln Lys Phe Asp Ile Leu Leu Leu
    690                 695                 700
Ile Gly Leu Gly Ile Gly Ala Thr Pro Phe Ile Ser Ile Leu Lys Asp
705                 710                 715                 720
Met Leu Asn Asn Leu Lys Pro Gly Ile Pro Lys Thr Gly Gln Lys Tyr
                725                 730                 735
Glu Gly Ser Val Gly Gly Glu Ser Leu Gly Gly Ser Ser Val Tyr Gly
```

```
                740                 745                 750
Gly Ser Ser Val Asn Gly Gly Gly Ser Val Asn Gly Gly Gly Ser Val
        755                 760                 765
Ser Gly Gly Gly Arg Lys Phe Pro Gln Arg Ala Tyr Phe Tyr Trp Val
    770                 775                 780
Thr Arg Glu Gln Ala Ser Phe Glu Trp Phe Lys Gly Val Met Asp Asp
785                 790                 795                 800
Ile Ala Val Tyr Asp Lys Thr Asn Val Ile Glu Met His Asn Tyr Leu
                805                 810                 815
Thr Ser Met Tyr Glu Ala Gly Asp Ala Arg Ser Ala Leu Ile Ala Met
            820                 825                 830
Val Gln Lys Leu Gln His Ala Lys Asn Gly Val Asp Ile Val Ser Glu
        835                 840                 845
Ser Arg Ile Arg Thr His Phe Ala Arg Pro Asn Trp Arg Lys Val Phe
    850                 855                 860
Ser Glu Leu Ser Asn Lys His Glu Thr Ser Arg Ile Gly Val Phe Tyr
865                 870                 875                 880
Cys Gly Ser Pro Thr Leu Val Arg Pro Leu Lys Ser Leu Cys Gln Glu
                885                 890                 895
Phe Ser Leu Glu Ser Ser Thr Arg Phe Thr Phe His Lys Glu Asn Phe
            900                 905                 910
```

SEQ ID NO: 8: At5g60010.1_ Arabidopsis thaliana

```
Met Lys Ser Asn Thr Pro Thr Glu Asp Ser Thr Lys Trp Met Leu Glu
 1               5                  10                  15
Ser Val Glu Ile Asp Ser Met Gly Glu Ser Ser Ser Lys Glu Pro Glu
            20                  25                  30
Ile Asn Leu Asn Lys Asn Glu Gly Gly Leu Lys Lys Asn Ala Ser Arg
        35                  40                  45
Asn Leu Gly Val Gly Ser Ile Ile Arg Thr Leu Ser Val Ser Asn Trp
    50                  55                  60
Arg Lys Ser Gly Asn Leu Gly Ser Pro Ser Thr Arg Lys Ser Gly Asn
65                  70                  75                  80
Leu Gly Pro Pro Thr Asn Ala Val Pro Lys Lys Thr Gly Pro Gln Arg
                85                  90                  95
Val Glu Arg Thr Thr Ser Ser Ala Ala Arg Gly Leu Gln Ser Leu Arg
            100                 105                 110
Phe Leu Asp Arg Thr Val Thr Gly Arg Glu Arg Asp Ala Trp Arg Ser
        115                 120                 125
Ile Glu Asn Arg Phe Asn Gln Phe Ser Val Asp Gly Lys Leu Pro Lys
    130                 135                 140
Glu Lys Phe Gly Val Cys Ile Gly Met Gly Asp Thr Met Glu Phe Ala
145                 150                 155                 160
Ala Glu Val Tyr Glu Ala Leu Gly Arg Arg Arg Gln Ile Glu Thr Glu
                165                 170                 175
Asn Gly Ile Asp Lys Glu Gln Leu Lys Leu Phe Trp Glu Asp Met Ile
            180                 185                 190
Lys Lys Asp Leu Asp Cys Arg Leu Gln Ile Phe Phe Asp Met Cys Asp
        195                 200                 205
Lys Asn Gly Asp Gly Lys Leu Thr Glu Glu Glu Val Lys Glu Val Ile
    210                 215                 220
Val Leu Ser Ala Ser Ala Asn Arg Leu Gly Asn Leu Lys Lys Asn Ala
225                 230                 235                 240
Ala Ala Tyr Ala Ser Leu Ile Met Glu Glu Leu Asp Pro Asp His Lys
                245                 250                 255
Gly Tyr Ile Glu Met Trp Gln Leu Glu Ile Leu Leu Thr Gly Met Val
            260                 265                 270

Thr Asn Ala Asp Thr Glu Lys Met Lys Lys Ser Gln Thr Leu Thr Arg
```

```
         275                 280                 285
Ala Met Ile Pro Glu Arg Tyr Arg Thr Pro Met Ser Lys Tyr Val Ser
    290                 295                 300
Val Thr Ala Glu Leu Met His Glu Asn Trp Lys Lys Leu Trp Val Leu
305                 310                 315                 320
Ala Leu Trp Ala Ile Ile Asn Val Tyr Leu Phe Met Trp Lys Tyr Glu
                325                 330                 335
Glu Phe Met Arg Asn Pro Leu Tyr Asn Ile Thr Gly Arg Cys Val Cys
            340                 345                 350
Ala Ala Lys Gly Ala Ala Glu Thr Leu Lys Leu Asn Met Ala Leu Ile
        355                 360                 365
Leu Val Pro Val Cys Arg Lys Thr Leu Thr Ile Leu Arg Ser Thr Phe
    370                 375                 380
Leu Asn Arg Val Val Pro Phe Asp Asp Asn Ile Asn Phe His Lys Val
385                 390                 395                 400
Ile Ala Tyr Met Ile Ala Phe Gln Ala Leu Leu His Thr Ala Leu His
                405                 410                 415
Ile Phe Cys Asn Tyr Pro Arg Leu Ser Ser Cys Ser Tyr Asp Val Phe
            420                 425                 430
Leu Thr Tyr Ala Gly Ala Ala Leu Gly Asn Thr Gln Pro Ser Tyr Leu
        435                 440                 445
Gly Leu Met Leu Thr Ser Val Ser Ile Thr Gly Val Leu Met Ile Phe
    450                 455                 460
Phe Met Gly Phe Ser Phe Thr Leu Ala Met His Tyr Phe Arg Arg Asn
465                 470                 475                 480
Ile Val Lys Leu Pro Lys Pro Phe Asn Val Leu Ala Gly Phe Asn Ala
                485                 490                 495
Phe Trp Tyr Ala His His Leu Leu Val Leu Ala Tyr Ile Leu Leu Ile
            500                 505                 510
Ile His Gly Tyr Tyr Leu Ile Ile Glu Lys Pro Trp Tyr Gln Lys Thr
        515                 520                 525
Thr Trp Met Tyr Leu Ala Val Pro Met Leu Phe Tyr Ala Ser Glu Arg
    530                 535                 540
Leu Phe Ser Arg Leu Leu Gln Glu His Ser His Arg Val Asn Val Ile
545                 550                 555                 560
Lys Ala Ile Val Tyr Ser Gly Asn Val Leu Ala Leu Tyr Val Thr Lys
                565                 570                 575
Pro Pro Gly Phe Lys Tyr Lys Ser Gly Met Tyr Met Phe Val Lys Cys
            580                 585                 590
Pro Asp Leu Ser Lys Phe Glu Trp His Pro Phe Ser Ile Thr Ser Ala
        595                 600                 605
Pro Gly Asp Asp Tyr Leu Ser Val His Ile Arg Ala Leu Gly Asp Trp
    610                 615                 620
Thr Thr Glu Leu Arg Ser Arg Phe Ala Lys Thr Cys Glu Pro Thr Gln
625                 630                 635                 640
Ala Ala Ala Lys Pro Lys Pro Asn Ser Leu Met Arg Met Glu Thr Arg
                645                 650                 655
Ala Ala Gly Val Asn Pro His Ile Glu Glu Ser Gln Val Leu Phe Pro
            660                 665                 670
Lys Ile Phe Ile Lys Gly Pro Tyr Gly Ala Pro Ala Gln Asn Tyr Gln
        675                 680                 685
Lys Phe Asp Ile Leu Leu Leu Val Gly Leu Gly Ile Gly Ala Thr Pro
    690                 695                 700

Phe Ile Ser Ile Leu Lys Asp Met Leu Asn His Leu Lys Pro Gly Ile
705                 710                 715                 720
Pro Arg Ser Gly Gln Lys Tyr Glu Gly Ser Val Gly Gly Glu Ser Ile
                725                 730                 735
Gly Gly Asp Ser Val Ser Gly Gly Gly Gly Lys Lys Phe Pro Gln Arg
            740                 745                 750
```

```
Ala Tyr Phe Phe Trp Val Thr Arg Glu Gln Ala Ser Phe Asp Trp Phe
        755                 760                 765
Lys Gly Val Met Asp Asp Ile Ala Glu Tyr Asp Lys Thr His Val Ile
    770                 775                 780
Glu Met His Asn Tyr Leu Thr Ser Met Tyr Glu Ala Gly Asp Ala Arg
785                 790                 795                 800
Ser Ala Leu Ile Ala Met Val Gln Lys Leu Gln His Ala Lys Asn Gly
                805                 810                 815
Val Asp Ile Val Ser Glu Ser Arg Ile Arg Thr His Phe Ala Arg Pro
            820                 825                 830
Asn Trp Arg Lys Val Phe Ser Glu Leu Ser Ser Lys His Glu Ala Cys
        835                 840                 845
Arg Ile Gly Val Phe Tyr Cys Gly Ser Pro Thr Leu Val Arg Pro Leu
    850                 855                 860
Lys Glu Leu Cys Gln Glu Phe Ser Leu Glu Ser Ser Thr Arg Phe Thr
865                 870                 875                 880
Phe His Lys Glu Asn Phe
                885
```

SEQ ID NO: 9: LOC_Os01g61880.1_ Oryza sativa

```
Met Ala Ser Pro Tyr Asp His Gln Ser Pro His Ala Gln His Pro Ser
 1               5                  10                  15
Gly Leu Pro Arg Pro Pro Gly Ala Gly Ala Gly Ala Ala Ala Gly Gly
            20                  25                  30
Phe Ala Arg Gly Leu Met Lys Gln Pro Ser Arg Leu Ala Ser Gly Val
        35                  40                  45
Arg Gln Phe Ala Ser Arg Val Ser Met Lys Val Pro Glu Gly Val Gly
    50                  55                  60
Gly Met Arg Pro Gly Gly Gly Arg Met Thr Arg Met Gln Ser Ser Ala
65                  70                  75                  80
Gln Val Gly Leu Arg Gly Leu Arg Phe Leu Asp Lys Thr Ser Gly Gly
                85                  90                  95
Lys Glu Gly Trp Lys Ser Val Glu Arg Arg Phe Asp Glu Met Asn Arg
            100                 105                 110
Asn Gly Arg Leu Pro Lys Glu Ser Phe Gly Lys Cys Ile Gly Met Gly
        115                 120                 125
Asp Ser Lys Glu Phe Ala Gly Glu Leu Phe Val Ala Leu Ala Arg Arg
    130                 135                 140
Arg Asn Leu Glu Pro Glu Asp Gly Ile Thr Lys Glu Gln Leu Lys Glu
145                 150                 155                 160
Phe Trp Glu Glu Met Thr Asp Gln Asn Phe Asp Ser Arg Leu Arg Ile
                165                 170                 175
Phe Phe Asp Met Cys Asp Lys Asn Gly Asp Gly Met Leu Thr Glu Asp
            180                 185                 190
Glu Val Lys Glu Val Ile Ile Leu Ser Ala Ser Ala Asn Lys Leu Ala
        195                 200                 205
Lys Leu Lys Gly His Ala Ala Thr Tyr Ala Ser Leu Ile Met Glu Glu
    210                 215                 220
Leu Asp Pro Asp Asp Arg Gly Tyr Ile Glu Ile Trp Gln Leu Glu Thr
225                 230                 235                 240
Leu Leu Arg Gly Met Val Ser Ala Gln Ala Ala Pro Glu Lys Met Lys
                245                 250                 255
Arg Thr Thr Ser Ser Leu Ala Arg Thr Met Ile Pro Ser Arg Tyr Arg
            260                 265                 270
Ser Pro Leu Lys Arg His Val Ser Arg Thr Val Asp Phe Val His Glu
        275                 280                 285
Asn Trp Lys Arg Ile Trp Leu Val Ala Leu Trp Leu Ala Val Asn Val
    290                 295                 300
Gly Leu Phe Ala Tyr Lys Phe Glu Gln Tyr Glu Arg Arg Ala Ala Phe
```

```
305                 310                 315                 320
Gln Val Met Gly His Cys Val Cys Val Ala Lys Gly Ala Ala Glu Val
                325                 330                 335
Leu Lys Leu Asn Met Ala Leu Ile Leu Leu Pro Val Cys Arg Asn Thr
            340                 345                 350
Leu Thr Thr Leu Arg Ser Thr Ala Leu Ser His Val Ile Pro Phe Asp
        355                 360                 365
Asp Asn Ile Asn Phe His Lys Val Ile Ala Ala Thr Ile Ala Ala Ala
    370                 375                 380
Thr Ala Val His Thr Leu Ala His Val Thr Cys Asp Phe Pro Arg Leu
385                 390                 395                 400
Ile Asn Cys Pro Ser Asp Lys Phe Met Ala Thr Leu Gly Pro Asn Phe
                405                 410                 415
Gly Tyr Arg Gln Pro Thr Tyr Ala Asp Leu Leu Glu Ser Ala Pro Gly
            420                 425                 430
Val Thr Gly Ile Leu Met Ile Ile Ile Met Ser Phe Ser Phe Thr Leu
        435                 440                 445
Ala Thr His Ser Phe Arg Arg Ser Val Val Lys Leu Pro Ser Pro Leu
    450                 455                 460
His His Leu Ala Gly Phe Asn Ala Phe Trp Tyr Ala His His Leu Leu
465                 470                 475                 480
Val Leu Ala Tyr Val Leu Leu Val Val His Ser Tyr Phe Ile Phe Leu
                485                 490                 495
Thr Arg Glu Trp Tyr Lys Lys Thr Thr Trp Met Tyr Leu Ile Val Pro
            500                 505                 510
Val Leu Phe Tyr Ala Cys Glu Arg Thr Ile Arg Lys Val Arg Glu Asn
        515                 520                 525
Asn Tyr Arg Val Ser Ile Val Lys Ala Ala Ile Tyr Pro Gly Asn Val
    530                 535                 540
Leu Ser Leu His Met Lys Lys Pro Pro Gly Phe Lys Tyr Lys Ser Gly
545                 550                 555                 560
Met Tyr Leu Phe Val Lys Cys Pro Asp Val Ser Pro Phe Glu Trp His
                565                 570                 575
Pro Phe Ser Ile Thr Ser Ala Pro Gly Asp Asp Tyr Leu Ser Val His
            580                 585                 590
Ile Arg Thr Leu Gly Asp Trp Thr Thr Glu Leu Arg Asn Leu Phe Gly
        595                 600                 605
Lys Ala Cys Glu Ala Gln Val Thr Ser Lys Lys Ala Thr Leu Ser Arg
    610                 615                 620
Leu Glu Thr Thr Val Val Ala Asp Ala Gln Thr Glu Asp Thr Arg Phe
625                 630                 635                 640
Pro Lys Val Leu Ile Asp Gly Pro Tyr Gly Ala Pro Ala Gln Asn Tyr
                645                 650                 655

Lys Lys Tyr Asp Ile Leu Leu Leu Ile Gly Leu Gly Ile Gly Ala Thr
            660                 665                 670
Pro Phe Ile Ser Ile Leu Lys Asp Leu Leu Asn Asn Ile Lys Ser Asn
        675                 680                 685
Glu Glu Val Glu Ser Ile His Gly Ser Glu Ile Gly Ser Phe Lys Asn
    690                 695                 700
Asn Gly Pro Gly Arg Ala Tyr Phe Tyr Trp Val Thr Arg Glu Gln Gly
705                 710                 715                 720
Ser Phe Glu Trp Phe Lys Gly Val Met Asn Asp Val Ala Glu Ser Asp
                725                 730                 735
His Asn Asn Ile Ile Glu Met His Asn Tyr Leu Thr Ser Val Tyr Glu
            740                 745                 750
Glu Gly Asp Ala Arg Ser Ala Leu Ile Ala Met Val Gln Ser Leu Gln
        755                 760                 765
His Ala Lys Asn Gly Val Asp Ile Val Ser Gly Ser Arg Ile Arg Thr
    770                 775                 780
```

```
His Phe Ala Arg Pro Asn Trp Arg Lys Val Phe Ser Asp Leu Ala Asn
785                 790                 795                 800
Ala His Lys Asn Ser Arg Ile Gly Val Phe Tyr Cys Gly Ser Pro Thr
                805                 810                 815
Leu Thr Lys Gln Leu Lys Asp Leu Ser Lys Glu Phe Ser Gln Thr Thr
            820                 825                 830
Thr Thr Arg Phe His Phe His Lys Glu Asn Phe
        835                 840
```

SEQ ID NO: 10: Glyma07g15690.1_ Glycine max

```
Arg Lys Met Met Arg Thr Glu Ser Gly Ala Ala Arg Gly Ile Lys Gly
 1               5                  10                  15
Leu Arg Phe Leu Asp Arg Thr Val Thr Gly Arg Glu Thr Asp Ala Trp
            20                  25                  30
Lys Ser Ile Glu Lys Arg Phe Thr Gln Asn Ala Val Asp Gly Lys Leu
        35                  40                  45
Thr Lys Asp Lys Phe Gly Thr Cys Met Gly Met Gly Ala Glu Ser Lys
    50                  55                  60
Asp Phe Ala Gly Glu Leu Tyr Glu Ala Leu Ala Arg Arg Arg Lys Val
65                  70                  75                  80
Tyr Ala Glu Asn Gly Ile Ser Leu Asp Glu Ala Lys Val Phe Trp Glu
                85                  90                  95
Asp Met Thr Asn Lys Asp Phe Glu Ser Arg Leu Gln Val Phe Phe Asp
            100                 105                 110
Met Cys Asp Lys Asn Gly Asp Gly Lys Leu Ser Glu Asp Glu Val Lys
        115                 120                 125
Glu Val Ile Val Leu Ser Ala Ser Ala Asn Lys Leu Gly Asn Leu Lys
    130                 135                 140
Met His Ala Asp Gly Tyr Ala Ser Leu Ile Met Glu Glu Leu Asp Pro
145                 150                 155                 160
Asp His Asn Gly Tyr Ile Glu Ile Trp Gln Leu Glu Thr Leu Leu Lys
                165                 170                 175
Glu Met Val Ser Ser Glu Glu Gly Thr Lys Lys Leu Asp Gln Cys Arg
            180                 185                 190
Ala Met Thr Leu Ser Lys Ala Met Ile Pro Ser Lys Tyr Arg Thr Pro
        195                 200                 205
Val Ser Lys Phe Leu Ser Lys Thr Thr Glu Phe Ala Leu Asp Lys Trp
    210                 215                 220
Lys Lys Ile Trp Val Phe Ala Leu Trp Leu Ala Ile Asn Leu Val Leu
225                 230                 235                 240
Phe Ile Trp Lys Phe Lys Gln Tyr Arg Glu Lys Lys Ala Phe Gln Val
                245                 250                 255
Met Gly Tyr Cys Leu Cys Phe Ala Lys Gly Ala Ala Glu Thr Leu Lys
            260                 265                 270
Phe Asn Met Ala Leu Ile Val Leu Thr Met Cys Arg Arg Thr Leu Thr
        275                 280                 285
Lys Leu Arg Gly Ser Phe Leu Ser Arg Ile Ile Pro Phe Asp Asp Asn
    290                 295                 300
Ile Asn Phe His Lys Thr Ile Ala Val Ala Val Val Ile Gly Thr Phe
305                 310                 315                 320
Ile His Val Met Met His Ile Thr Cys Asp Phe Pro Arg Leu Ile Ser
                325                 330                 335
Cys Pro Glu Asn Lys Phe Phe Ser Ile Phe Gly Asp Gly Phe Asn Tyr
            340                 345                 350
Glu Gln Pro Thr Tyr Tyr Thr Leu Val Lys Ser Ile Pro Gly Leu Thr
        355                 360                 365
Gly Ile Leu Met Val Leu Ile Met Ala Phe Thr Phe Thr Leu Ala Thr
    370                 375                 380
His Tyr Phe Arg Lys Ser Val Val Lys Leu Pro Ser Pro Leu His Arg
```

385 390 395 400
Leu Ala Gly Phe Asn Ala Phe Trp Tyr Ala His His Leu Leu Ile Val
405 410 415
Val Tyr Ile Leu Leu Ile Ile His Gly Tyr Phe Leu Phe Leu Thr Lys
420 425 430
Glu Trp Asn Lys Lys Thr Thr Trp Met Tyr Leu Val Val Pro Leu Ala
435 440 445
Leu Tyr Ala Phe Glu Arg Ile His Pro Phe Phe Arg Ser Lys Asp His
450 455 460
Arg Val Ser Ile Ile Lys Ala Ile Ile Tyr Thr Gly Asn Val Leu Ala
465 470 475 480
Leu Tyr Met Thr Lys Pro Gln Gly Phe Lys Tyr Glu Ser Gly Met Tyr
485 490 495
Leu Phe Val Lys Cys Pro Asp Ile Ser Thr Phe Glu Trp His Pro Phe
500 505 510
Ser Ile Thr Ser Ala Pro Gly Asp Asp Tyr Leu Ser Val His Ile Arg
515 520 525
Thr Leu Gly Asp Trp Thr Thr Glu Leu Lys Asn Thr Phe Ala Gln Val
530 535 540
Cys Glu Pro His Asn Ala Gln Pro Arg Lys Gly Asn Leu Met Arg Met
545 550 555 560
Glu Thr Arg Ala Pro Asn Ser Thr Tyr Asn His Pro Ser Lys Ser Arg
565 570 575
Ile Arg Tyr Pro Lys Ile Leu Ile Lys Gly Pro Tyr Gly Ala Pro Ala
580 585 590
Gln Ser Tyr Lys Asn Tyr Asp Val Leu Phe Leu Ile Gly Leu Gly Ile
595 600 605
Gly Ala Thr Pro Met Ile Ser Ile Leu Lys Asp Met Leu Asn Asn Met
610 615 620
Lys Ser Glu Ser Pro Lys Glu Val Ser Gly Arg Leu Tyr Ile Leu Phe
625 630 635 640
Leu Gln Gly Thr Tyr Met Gln Asp Ser Val Pro Ser Ser Asn Ser Asp
645 650 655

Asp Gln Ile Lys Lys Gly Pro Glu Arg Ala Tyr Phe Tyr Trp Val Thr
660 665 670
Arg Glu Gln Ser Ser Phe Glu Trp Phe Lys Gly Val Met Asp Asp Ile
675 680 685
Ala Asp Tyr Asp Cys Asp Asn Ile Ile Glu Met His Asn Tyr Leu Thr
690 695 700
Ser Val Tyr Glu Glu Gly Asp Ala Arg Ser Ala Leu Ile Ala Met Ile
705 710 715 720
Gln Arg Leu Gln His Ala Lys Asn Gly Val Asp Val Val Ser Glu Ser
725 730 735
Arg Ile Arg Thr His Phe Ala Arg Pro Asn Trp Lys Lys Val Phe Thr
740 745 750
Glu Leu Ala Asn Ala His Gln Ser Ser Arg Ile Gly Val Phe Tyr Cys
755 760 765
Gly Ser Pro Thr Leu Thr Lys Thr Leu Lys Glu Leu Cys His Glu Phe
770 775 780
Ser Leu Lys Ser Ser Thr Arg Phe Gln Phe His Lys Glu Asn Phe
785 790 795

SEQ ID NO: 11: Glyma18g39500.1_ Glycine max

Met Ser Thr Glu Lys Gly Gly Lys Asp Glu Ser Ser Thr Trp Ile Leu
1 5 10 15
Glu Gly Ile Asp Ile Asp Pro Ile Val Glu Arg Pro Lys Arg Asp Asp
20 25 30
Glu Ala Pro Pro Asn Gly Ile Met Gly Asn Asn Gly Gly Arg Lys Met

```
         35                  40                  45
Met Arg Ala Glu Ser Gly Ala Ala Arg Gly Ile Lys Ser Leu Arg Phe
    50                  55                  60
Leu Asp Arg Thr Val Thr Gly Lys Glu Ala Asp Ala Trp Lys Ser Ile
65                  70                  75                  80
Glu Lys Arg Phe Thr Gln Asn Ala Val Asp Gly Lys Leu Thr Lys Asp
                85                  90                  95
Lys Phe Gly Thr Cys Met Gly Met Gly Ala Glu Ser Lys Asp Phe Ala
            100                 105                 110
Gly Glu Leu Tyr Glu Ala Leu Ala Arg Arg Arg Asn Val Cys Ala Glu
        115                 120                 125
Asn Gly Ile Thr Leu Asp Glu Val Lys Val Phe Trp Glu Asp Met Thr
    130                 135                 140
Asn Arg Asp Leu Glu Ser Arg Leu Gln Val Phe Phe Asp Met Cys Asp
145                 150                 155                 160
Lys Asn Gly Asp Gly Arg Leu Ser Glu Glu Glu Val Lys Glu Val Ile
                165                 170                 175
Val Leu Ser Ala Ser Ala Asn Lys Leu Gly Asn Leu Lys Val His Ala
            180                 185                 190
Asp Ala Tyr Ala Ser Leu Ile Met Glu Glu Leu Asp Pro Asp His Asn
        195                 200                 205
Gly Tyr Ile Glu Val Arg Ser Glu Lys Phe Leu Leu Leu Ser Asn Phe
    210                 215                 220
Ile Glu Phe Tyr Ile Asn Leu His Leu Leu Ala Met Thr Leu Ser Arg
225                 230                 235                 240
Ala Met Ile Pro Ser Lys Tyr Arg Thr Pro Val Ser Lys Phe Leu Ser
                245                 250                 255
Thr Thr Ala Glu Phe Ala Leu Asp Lys Trp Lys Lys Ile Trp Val Val
            260                 265                 270
Ala Leu Trp Leu Ala Ile Asn Leu Val Leu Phe Ile Trp Lys Phe Lys
        275                 280                 285
Gln Tyr Arg Glu Arg Glu Ala Phe Lys Val Met Gly Tyr Cys Leu Cys
    290                 295                 300
Phe Ala Lys Gly Ala Ala Glu Thr Leu Lys Phe Asn Met Ala Leu Ile
305                 310                 315                 320
Val Leu Thr Met Cys Arg Arg Thr Leu Thr Lys Leu Arg Gly Ser Phe
                325                 330                 335
Leu Asn Arg Ile Ile Pro Phe Asp Asp Asn Ile Asn Phe His Lys Thr
            340                 345                 350
Ile Ala Val Ala Val Val Ile Gly Thr Phe Ile His Val Met Met His
        355                 360                 365
Ile Thr Cys Asp Phe Pro Arg Leu Ile Ser Cys Pro Glu Asn Lys Phe
    370                 375                 380
Met Ser Ile Leu Gly Gln Asp Phe Asn Tyr Glu Gln Pro Thr Phe Tyr
385                 390                 395                 400
Thr Leu Leu Lys Ser Ile Leu Gly Val Thr Gly Ile Leu Met Val Leu
                405                 410                 415
Leu Met Ala Phe Ile Phe Thr Leu Ala Thr His Tyr Phe Arg Lys Ser
            420                 425                 430
Val Val Lys Leu Pro Leu Ser Leu His Arg Leu Ala Gly Phe Asn Ala
        435                 440                 445
Phe Trp Tyr Ala His His Leu Leu Ile Val Val Tyr Ile Leu Leu Ile
    450                 455                 460
Ile His Gly Tyr Phe Leu Phe Leu Thr Lys Glu Trp Asp Lys Lys Thr
465                 470                 475                 480
Thr Trp Met Tyr Leu Val Val Pro Leu Val Leu Tyr Ala Phe Glu Arg
                485                 490                 495
Ile His Pro Phe Phe Arg Gly Lys Asp His Arg Val Ser Ile Ile Lys
            500                 505                 510
Ala Ile Ile Tyr Thr Gly Asn Val Leu Ala Leu Tyr Met Thr Lys Pro
```

```
         515                 520                 525
Gln Gly Phe Lys Tyr Lys Ser Gly Met Tyr Ile Phe Val Lys Cys Pro
    530                 535                 540
Asp Ile Ser Ser Phe Glu Trp His Pro Phe Ser Ile Thr Ser Ala Pro
545                 550                 555                 560
Gly Asp Asp Tyr Leu Ser Val His Ile Arg Thr Leu Gly Asp Trp Thr
                565                 570                 575
Thr Glu Leu Lys Asn Lys Phe Thr Gln Val Cys Glu Pro His Ser Ala
            580                 585                 590
Gln Pro Arg Lys Gly Asn Leu Met Arg Met Glu Thr Arg Ala Pro Ser
        595                 600                 605
Ser Asn Tyr Asn His Ser Ser Asn Ser Ser Ile Arg Tyr Pro Lys Ile
    610                 615                 620
Leu Ile Lys Gly Pro Tyr Gly Ala Pro Ala Gln Ser Tyr Lys Asn Tyr
625                 630                 635                 640
Asp Val Leu Met Leu Ile Gly Leu Gly Ile Gly Ala Thr Pro Met Ile
                645                 650                 655
Ser Ile Leu Lys Asp Met Leu Asn Asn Met Lys Ser Glu Ser Pro Lys
            660                 665                 670
Glu Val Ser His Arg Leu Tyr Ile Leu Phe Trp Leu Ala Ala Tyr Val
        675                 680                 685
Tyr Leu Ser Leu Leu Val Glu Ile Ile Phe Ser Lys Thr Phe Lys Gly
    690                 695                 700

Thr Tyr Met Gln Asp Ser Asp His Ser Tyr His Leu Asp Asp Gln Ile
705                 710                 715                 720
Lys Lys Gly Pro Glu Arg Ala Tyr Phe Tyr Trp Val Thr Arg Glu Gln
                725                 730                 735
Ser Ser Phe Glu Trp Phe Lys Gly Val Met Asp Asp Ile Ala Asp Tyr
            740                 745                 750
Asp His Asp Asn Ile Ile Glu Met His Asn Tyr Leu Thr Ser Val Tyr
        755                 760                 765
Glu Glu Gly Asp Ala Arg Ser Ala Leu Ile Ala Met Ile Gln Lys Leu
    770                 775                 780
Gln His Ala Lys Asn Gly Val Asp Val Val Ser Glu Ser Arg Ile Arg
785                 790                 795                 800
Thr His Phe Ala Arg Pro Asn Trp Lys Lys Val Phe Thr Gln Leu Ala
                805                 810                 815
Asn Ala His Gln Ser Ser Arg Ile Gly Val Phe Tyr Cys Gly Ser Pro
            820                 825                 830
Thr Leu Thr Lys Thr Leu Lys Glu Leu Cys Leu Glu Phe Ser Leu Asn
        835                 840                 845
Ser Ser Thr Arg Phe Gln Phe His Lys Glu Asn Phe
    850                 855                 860
```

SEQ ID NO: 12: LOC_Os05g38980.1_ Oryza sativa

```
Met Ala Gly Asp Tyr Val Asp Val Pro Leu Gly Gly Gly Gly Gln Ser
 1               5                  10                  15
Thr Leu Pro Pro Val Ala Pro Leu Lys Lys Gln Pro Ser Arg Leu Ala
            20                  25                  30
Ser Gly Met Lys Arg Leu Ala Ser Met Val Pro Asp Thr Met Lys Leu
        35                  40                  45
Lys Arg Thr His Ser Ser Ala Gln Pro Ala Leu Arg Gly Leu Arg Phe
    50                  55                  60
Leu Asp Lys Thr Ser Ala Gly Lys Asp Gly Trp Lys Asn Val Glu Lys
65                  70                  75                  80
Arg Phe Asp Glu Met Ser Ala Asp Gly Arg Leu Pro Gln Glu Ser Phe
                85                  90                  95
Ala Lys Cys Ile Gly Met Ala Asp Ser Lys Glu Phe Ala Ser Glu Val
```

```
                100                 105                 110
Phe Val Ala Leu Ala Arg Arg Arg Ser Ile Lys Pro Glu Asp Gly Ile
        115                 120                 125
Thr Lys Glu Gln Leu Lys Glu Phe Trp Glu Glu Leu Thr Asp Gln Asn
    130                 135                 140
Phe Asp Ser Arg Leu Arg Ile Phe Phe Asp Met Cys Asp Lys Asn Gly
145                 150                 155                 160
Asp Gly Gln Leu Thr Glu Asp Glu Val Lys Glu Val Ile Val Leu Ser
                165                 170                 175
Ala Ala Ala Asn Lys Leu Ala Lys Leu Lys Ser His Ala Ala Thr Tyr
            180                 185                 190
Ala Ser Leu Ile Met Glu Glu Leu Asp Pro Asp His Arg Gly Tyr Ile
        195                 200                 205
Glu Ile Trp Gln Leu Glu Thr Leu Leu Arg Gly Met Val Thr Ala Gln
    210                 215                 220
Gly Pro Pro Glu Lys Val Lys Leu Ala Ser Ala Ser Leu Ala Arg Thr
225                 230                 235                 240
Met Val Pro Ser Ser His Arg Ser Pro Met Gln Arg Arg Phe Asn Lys
                245                 250                 255
Thr Val Asp Phe Ile His Glu Asn Trp Lys Arg Ile Trp Val Leu Ser
            260                 265                 270
Leu Trp Ala Ile Leu Asn Ile Ala Leu Phe Met Tyr Lys Phe Val Gln
        275                 280                 285
Tyr Ser Arg Arg Asp Ala Phe Gln Val Met Gly Tyr Cys Val Cys Ile
    290                 295                 300
Ala Lys Gly Ala Ala Glu Thr Leu Lys Leu Asn Met Ala Val Ile Leu
305                 310                 315                 320
Leu Pro Val Cys Arg Asn Thr Leu Thr Arg Leu Arg Ser Thr Ala Leu
                325                 330                 335
Ser Lys Val Val Pro Phe Asp Asp Asn Ile Asn Phe His Lys Val Ile
            340                 345                 350
Ala Leu Thr Ile Ala Ile Gly Ala Ala Thr His Thr Leu Ala His Val
        355                 360                 365
Thr Cys Asp Phe Pro Arg Leu Val Ser Cys Pro Arg Asp Lys Phe Glu
    370                 375                 380
Ala Thr Leu Gly Pro Tyr Phe Asn Tyr Val Gln Pro Thr Tyr Ser Ser
385                 390                 395                 400
Leu Val Ala Ser Thr Pro Gly Trp Thr Gly Ile Leu Met Ile Leu Ile
                405                 410                 415
Met Ser Phe Ser Phe Thr Leu Ala Thr His Ser Phe Arg Arg Ser Val
            420                 425                 430
Val Lys Leu Pro Ser Pro Leu His His Leu Ala Gly Phe Asn Ala Phe
        435                 440                 445
Trp Tyr Ala His His Leu Leu Val Ile Ala Tyr Ile Leu Leu Val Leu
    450                 455                 460
His Ser Tyr Phe Ile Phe Leu Thr Lys Gln Trp Tyr Asn Arg Thr Thr
465                 470                 475                 480
Trp Met Phe Leu Ala Val Pro Val Leu Phe Tyr Ser Cys Glu Arg Thr
                485                 490                 495
Ile Arg Arg Val Arg Glu Ser Ser Tyr Gly Val Thr Val Ile Lys Ala
            500                 505                 510
Ala Ile Tyr Pro Gly Asn Val Leu Ser Ile His Met Asn Lys Pro Ser
        515                 520                 525
Ser Phe Lys Tyr Lys Ser Gly Met Tyr Met Phe Val Lys Cys Pro Asp
    530                 535                 540
Val Ser Pro Phe Glu Trp His Pro Phe Ser Ile Thr Ser Ala Pro Gly
545                 550                 555                 560
Asp Asp Tyr Leu Ser Val His Ile Arg Thr Leu Gly Asp Trp Thr Thr
                565                 570                 575
Glu Leu Arg Asn Leu Phe Gly Lys Ala Cys Glu Ala Gln Val Ser Ser
```

```
            580                 585                 590
Lys Lys Ala Thr Leu Ala Arg Leu Glu Thr Thr Ile Ile Ala Asp Gly
        595                 600                 605
Leu Lys Glu Glu Thr Cys Phe Pro Lys Val Phe Ile Asp Gly Pro Phe
    610                 615                 620
Gly Ala Pro Ala Gln Asn Tyr Lys Lys Tyr Asp Ile Leu Leu Leu Ile
625                 630                 635                 640
Gly Leu Gly Ile Gly Ala Thr Pro Phe Ile Ser Ile Leu Lys Asp Leu
                645                 650                 655
Leu Asn Asn Ile Lys Ser Asn Gly Asp Val Gln Ser Thr His Asp Ala
            660                 665                 670
Glu Leu Gly Cys Thr Phe Lys Ser Asn Gly Pro Gly Arg Ala Tyr Phe
        675                 680                 685

Tyr Trp Val Thr Arg Glu Gln Gly Ser Phe Glu Trp Phe Lys Gly Val
    690                 695                 700
Met Asn Asp Val Ala Glu Ser Asp His Asp Asn Val Ile Glu Met His
705                 710                 715                 720
Asn Tyr Leu Thr Ser Val Tyr Glu Glu Gly Asp Ala Arg Ser Ala Leu
                725                 730                 735
Ile Ala Met Val Gln Ser Leu Gln His Ala Lys Asn Gly Val Asp Ile
            740                 745                 750
Val Ser Gly Ser Lys Ile Arg Thr His Phe Ala Arg Pro Asn Trp Arg
        755                 760                 765
Lys Val Phe Ser Asp Leu Ala Asn Ala His Gln Asn Ser Arg Ile Gly
    770                 775                 780
Val Phe Tyr Cys Gly Ser Pro Thr Leu Thr Lys Met Leu Arg Asp Leu
785                 790                 795                 800
Ser Leu Glu Phe Ser Gln Thr Thr Thr Thr Arg Phe His Phe His Lys
                805                 810                 815
Glu Asn Phe
```

SEQ ID NO: 13: Glyma18g39500.1_variant; soybean; 19-aa at position 683 removed to get better alignment.

```
MSTEKGGKDESSTWILEGIDIDPIVERPKRDDEAPPNGIMGNNGGRKMMRAESGAARGIKSLRFLDRTVTGKEAD
AWKSIEKRFTQNAVDGKLTKDKFGTCMGMGAESKDFAGELYEALARRRNVCAENGITLDEVKVFWEDMTNRDLES
RLQVFFDMCDKNGDGRLSEEEVKEVIVLSASANKLGNLKVHADAYASLIMEELDPDHNGYIEVRSEKFLLLSNFI
EFYINLHLLAMTLSRAMIPSKYRTPVSKFLSTTAEFALDKWKKIWVVALWLAINLVLFIWKFKQYREREAFKVMG
YCLCFAKGAAETLKFNMALIVLTMCRRTLTKLRGSFLNRIIPFDDNINFHKTIAVAVVIGTFIHVMMHITCDFPR
LISCPENKFMSILGQDFNYEQPTFYTLLKSILGVTGILMVLLMAFIFTLATHYFRKSVVKLPLSLHRLAGFNAFW
YAHHLLIVVYILLIIHGYFLFLTKEWDKKTTWMYLVVPLVLYAFERIHPFFRGKDHRVSIIKAIIYTGNVLALYM
TKPQGFKYKSGMYIFVKCPDISSFEWHPFSITSAPGDDYLSVHIRTLGDWTTELKNKFTQVCEPHSAQPRKGNLM
RMETRAPSSNYNHSSNSSIRYPKILIKGPYGAPAQSYKNYDVLMLIGLGIGATPMISILKDMLNNMKSESPKEVS
HRLYILFFKGTYMQDSDHSYHLDDQIKKGPERAYFYWVTREQSSFEWFKGVMDDIADYDHDNIIEMHNYLTSVYE
EGDARSALIAMIQKLQHAKNGVDVVSESRIRTHFARPNWKKVFTQLANAHQSSRIGVFYCGSPTLTKTLKELCLE
FSLNSSTRFQFHKENF
```

What is claimed is:

1. A method for increasing drought tolerance in a plant, said method comprising introducing into said plant a recombinant DNA construct, said construct comprising a root-preferred promoter operably linked to a polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

2. A method for enhancing nutrient uptake in a plant, said method comprising introducing into said plant a recombinant DNA construct, said construct comprising a root-preferred promoter operably linked to a polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

3. The method of claim 1 or 2, wherein said polynucleotide is stably integrated into the genome of the plant.

4. The method of claim 1, 2, or 3, wherein said plant is a plant cell.

5. The method of claim 1, 2, or 3, wherein said plant is a dicot.

6. The method of claim 5, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

7. The method of claim 1, 2 or 3, wherein said plant is a monocot.

8. The method of claim 7, wherein said monocot is maize, sugarcane, wheat, rice, barley, sorghum, or rye.

9. A seed of the plant produced by the method of any one of claims 1-8.

10. A plant comprising a recombinant DNA construct comprising a root-preferred promoter operably linked to a polynucleotide, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

11. The plant of claim 10, wherein said plant is a cell.

12. The plant of claim 10, wherein said plant is a monocot.

13. The plant of claim 12, wherein said monocot is maize, sugarcane, wheat, rice, barley, sorghum, or rye.

14. The plant of claim 10, wherein said plant is a dicot.

15. The plant of claim 14, wherein the dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

16. The plant of any one of claims 10 to 15, wherein said polynucleotide is stably incorporated into the genome of the plant.

17. A seed of the plant of claim 16.

18. An expression cassette comprising a polynucleotide operably linked to a heterologous promoter that drives expression of the polynucleotide in a root of a plant, wherein said polynucleotide comprises a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence comprising SEQ ID NO: 2;

(b) a nucleotide sequence encoding the amino acid sequence set forth in SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13;

(c) a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity; and

(d) a nucleotide sequence encoding an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13, wherein said polynucleotide encodes a polypeptide having NADPH oxidase activity.

19. The expression cassette of claim 18, wherein said plant is a monocot.

20. The expression cassette of claim 19, wherein said monocot is maize, sugarcane, wheat, rice, barley, sorghum, or rye.

21. The expression cassette of claim 18, wherein said plant is a dicot.

22. The expression cassette of claim 21, wherein the dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

23. A method of selecting an allelic variant of rth5 in a maize plant, the method comprising the steps of:

a. obtaining a population of maize plants, wherein said maize plants exhibit an alteration of at least one agronomic characteristic; wherein the at least one agronomic characteristic is selected from the group consisting of increased root hair formation and growth, increased drought tolerance, and enhanced nutrient uptake;

b. evaluating allelic variations with respect to the polynucleotide sequence encoding a protein comprising SEQ ID NO:3, or in the genomic region that regulates the expression of the polynucleotide encoding the protein;

c. associating allelic variations with said alteration of at least one agronomic characteristic; and

d. selecting an allelic variant that is associated with said alteration of at least one agronomic characteristic.

24. A method of selecting a first maize plant or a first maize germplasm that has one or more beneficial alleles of rth5, the method comprising:

(a) screening a plurality of maize plants or a plurality of maize germplasm for at least one polymorphism within a marker locus, wherein the marker locus is:

(a) a first polynucleotide having at least 90% and less than 100% nucleotide sequence identity with SEQ ID NO:1 or 2; or

(b) a second polynucleotide encoding a polypeptide comprising an amino acid sequence having at least 90% and less than 100% sequence identity to SEQ ID NO:3, wherein expression of the first or second polynucleotide in a maize plant results in a phenotype comprising an alteration of at least one agronomic characteristic when compared to a control maize plant; wherein the at least one agronomic characteristic is selected from the group consisting of increased root hair formation and growth, increased drought tolerance, and enhanced nutrient uptake; and wherein the control maize plant comprises:

(c) a polynucleotide having the nucleotide sequence of SEQ ID NO:1 or 2; or

(d) a polynucleotide encoding the amino acid sequence of SEQ ID NO:3;

(b) identifying a first maize plant or a first maize germplasm comprising the at least one polymorphism of the marker locus; and

(c) selecting the first maize plant or first maize germplasm of step (b).

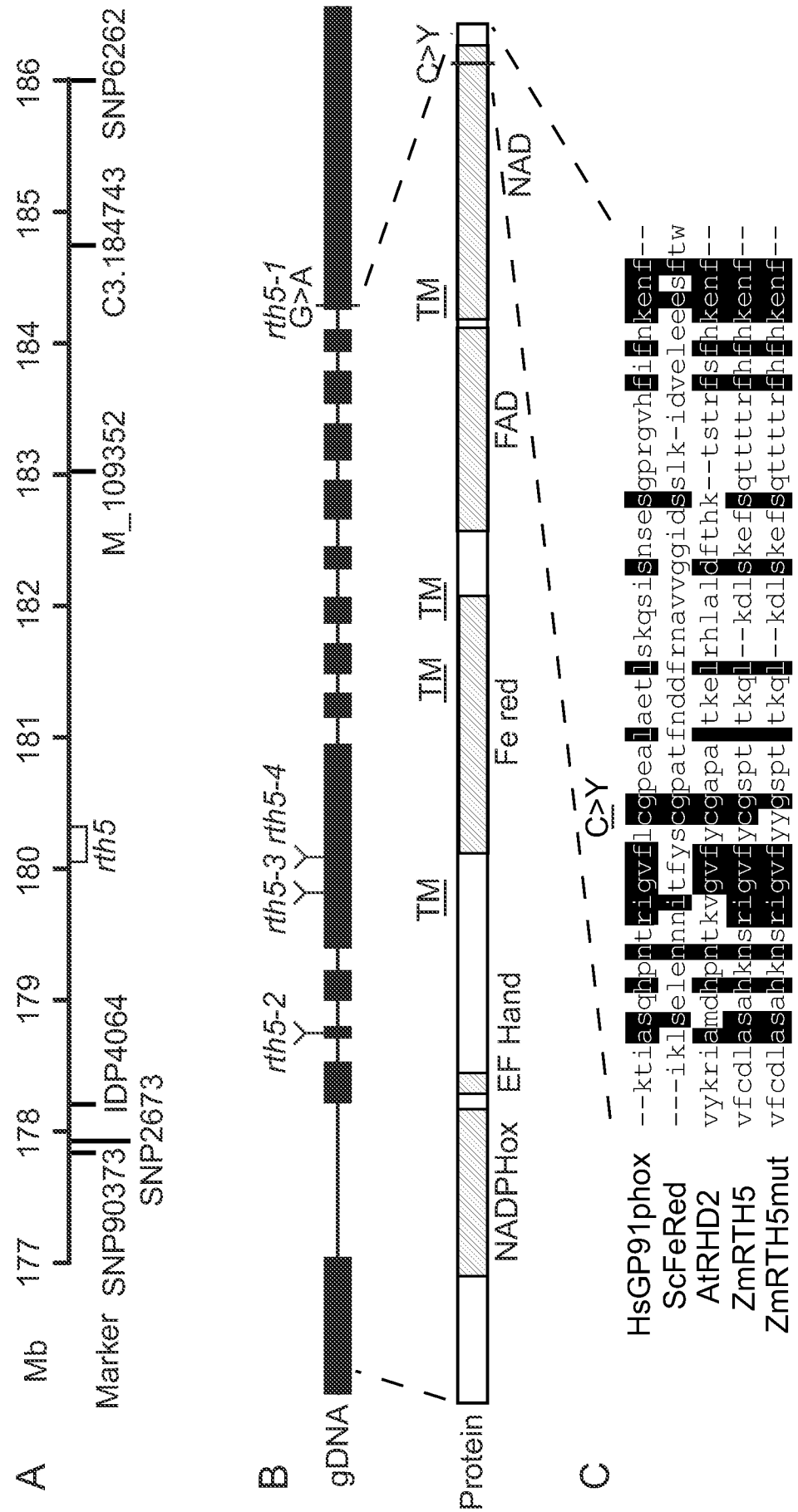
A
Mb
177
178
179
180
181
182
183
184
185
186
Marker
SNP90373
IDP4064
SNP2673
rth5
M_109352
C3.184743
SNP6262
B
gDNA
rth5-2
rth5-3 rth5-4
rth5-1
G>A
C>Y
Protein
TM
TM
TM
TM
NADPHox
EF Hand
Fe red
FAD
NAD
C
C>Y
HsGP91phox
ScFeRed
AtRHD2
ZmRTH5
ZmRTH5mut

FIG. 1

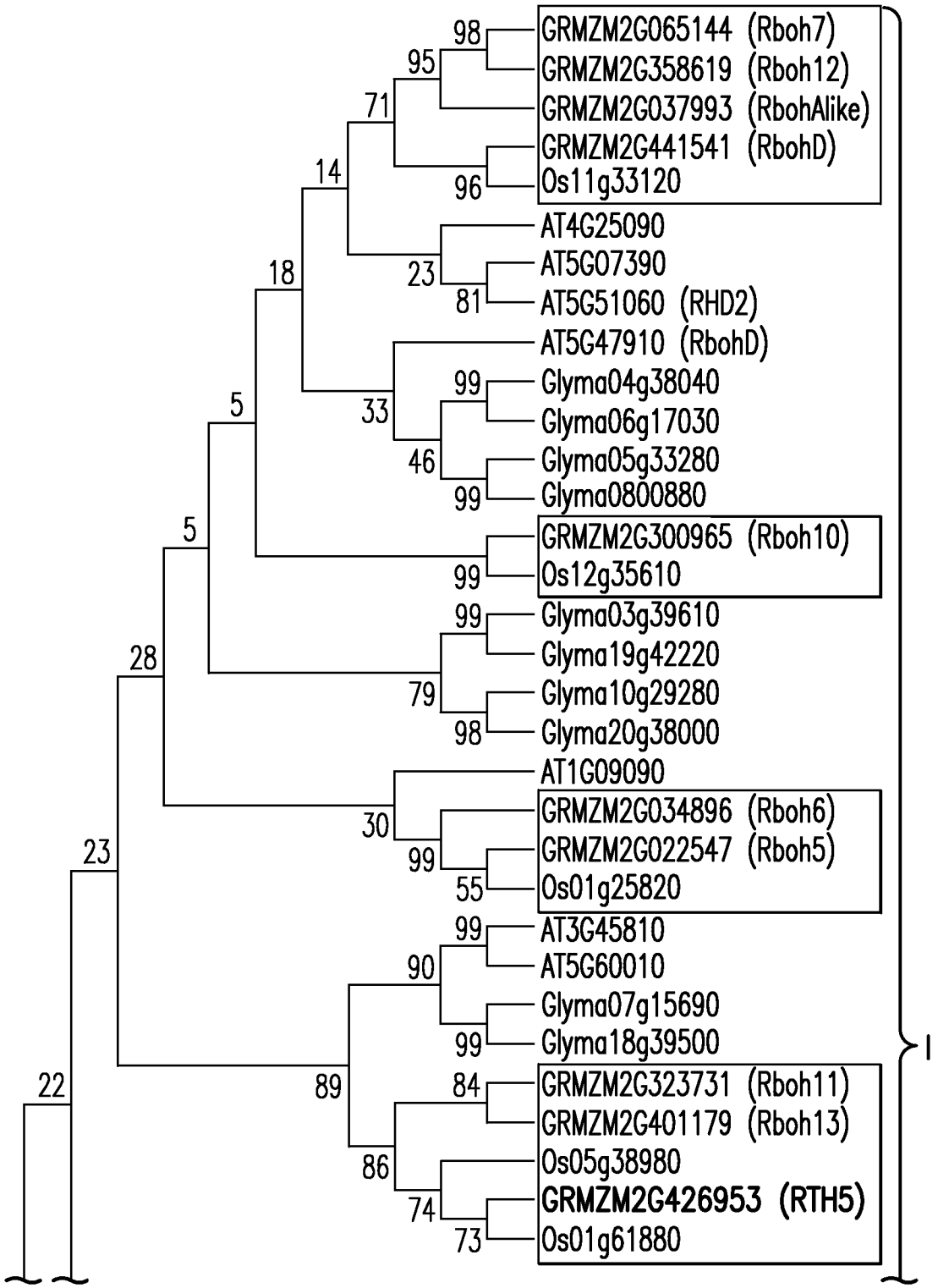
GRMZM2G065144 (Rboh7)
GRMZM2G358619 (Rboh12)
GRMZM2G037993 (RbohAlike)
GRMZM2G441541 (RbohD)
Os11g33120
AT4G25090
AT5G07390
AT5G51060 (RHD2)
AT5G47910 (RbohD)
Glyma04g38040
Glyma06g17030
Glyma05g33280
Glyma08g00880
GRMZM2G300965 (Rboh10)
Os12g35610
Glyma03g39610
Glyma19g42220
Glyma10g29280
Glyma20g38000
AT1G09090
GRMZM2G034896 (Rboh6)
GRMZM2G022547 (Rboh5)
Os01g25820
AT3G45810
AT5G60010
Glyma07g15690
Glyma18g39500
GRMZM2G323731 (Rboh11)
GRMZM2G401179 (Rboh13)
Os05g38980
GRMZM2G426953 (RTH5)
Os01g61880
I

FIG. 2

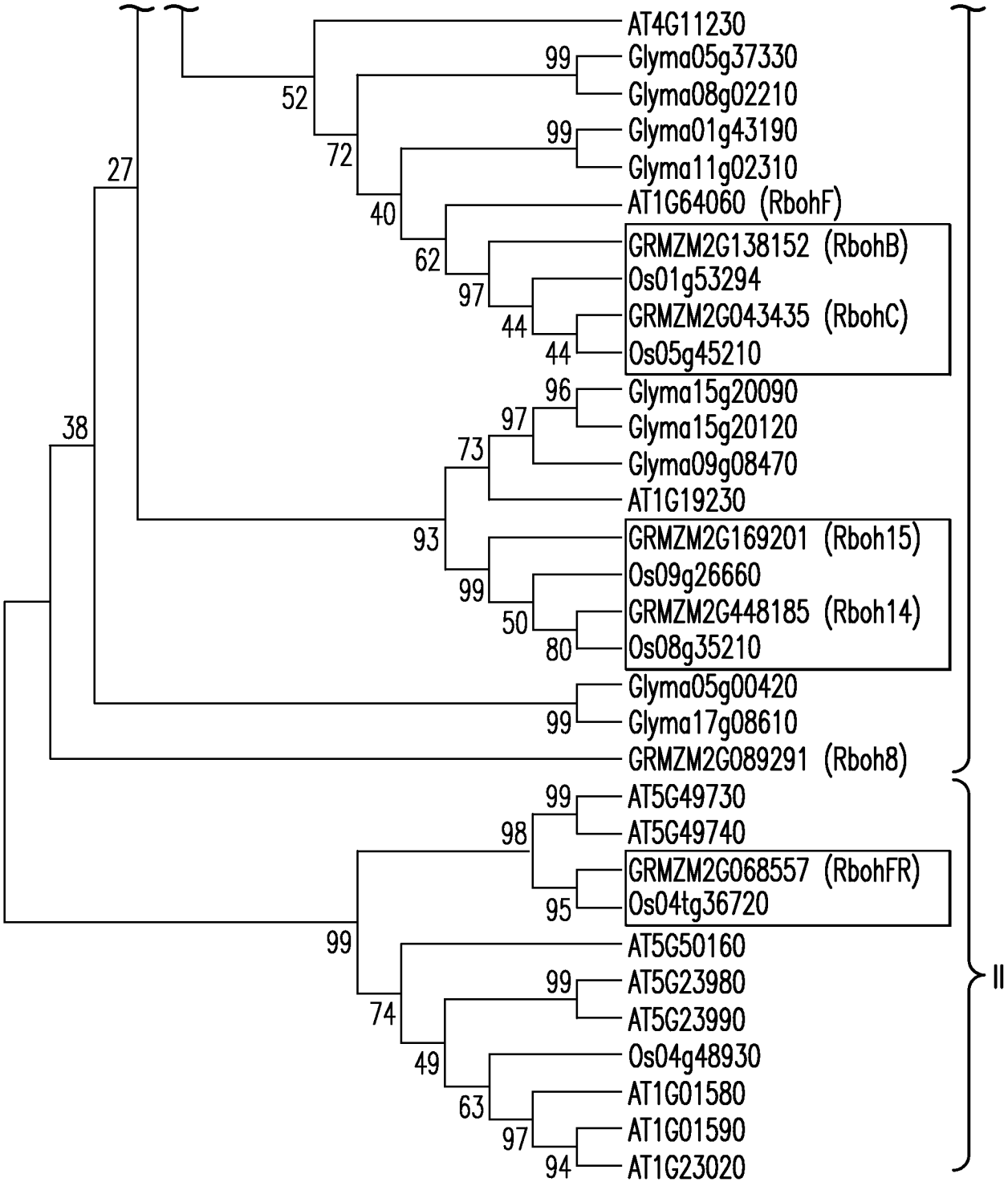

FIG. 2 (Continued)

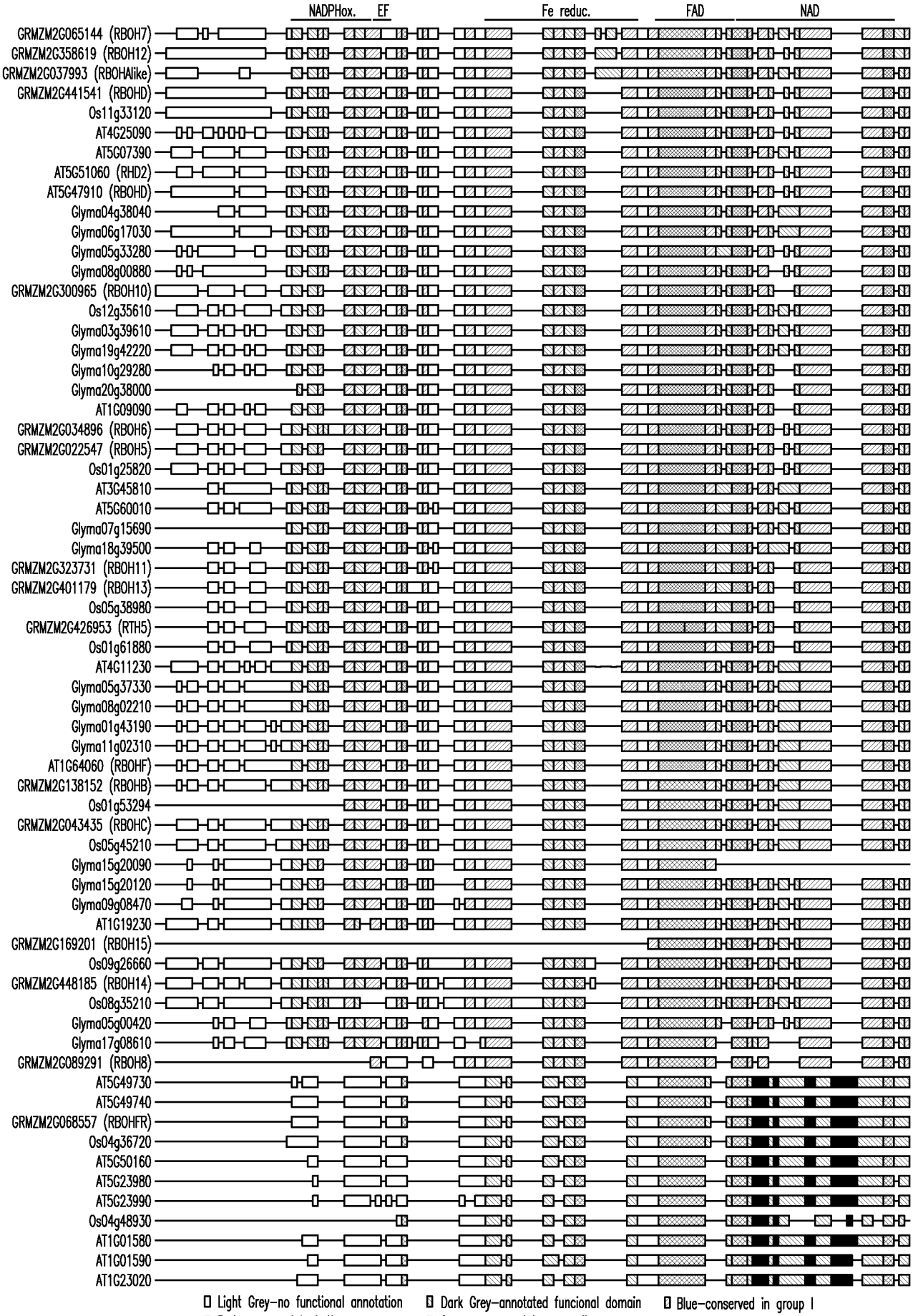
NADPHox.
EF
Fe reduc.
FAD
NAD
GRMZM2G065144 (RBOH7)
GRMZM2G358619 (RBOH12)
GRMZM2G037993 (RBOHAlike)
GRMZM2G441541 (RBOHD)
Os11g33120
AT4G25090
AT5G07390
AT5G51060 (RHD2)
AT5G47910 (RBOHD)
Glyma04g38040
Glyma06g17030
Glyma05g33280
Glyma08g00880
GRMZM2G300965 (RBOH10)
Os12g35610
Glyma03g39610
Glyma19g42220
Glyma10g29280
Glyma20g38000
AT1G09090
GRMZM2G034896 (RBOH6)
GRMZM2G022547 (RBOH5)
Os01g25820
AT3G45810
AT5G60010
Glyma07g15690
Glyma18g39500
GRMZM2G323731 (RBOH11)
GRMZM2G401179 (RBOH13)
Os05g38980
GRMZM2G426953 (RTH5)
Os01g61880
AT4G11230
Glyma05g37330
Glyma08g02210
Glyma01g43190
Glyma11g02310
AT1G64060 (RBOHF)
GRMZM2G138152 (RBOHB)
Os01g53294
GRMZM2G043435 (RBOHC)
Os05g45210
Glyma15g20090
Glyma15g20120
Glyma09g08470
AT1G19230
GRMZM2G169201 (RBOH15)
Os09g26660
GRMZM2G448185 (RBOH14)
Os08g35210
Glyma05g00420
Glyma17g08610
GRMZM2G089291 (RBOH8)
AT5G49730
AT5G49740
GRMZM2G068557 (RBOHFR)
Os04g36720
AT5G50160
AT5G23980
AT5G23990
Os04g48930
AT1G01580
AT1G01590
AT1G23020
Light Grey–no functional annotation
Dark Grey–annotated funcional domain
Blue–conserved in group I
Red–conserved in both groups
Green–conserved in group II

FIG. 3

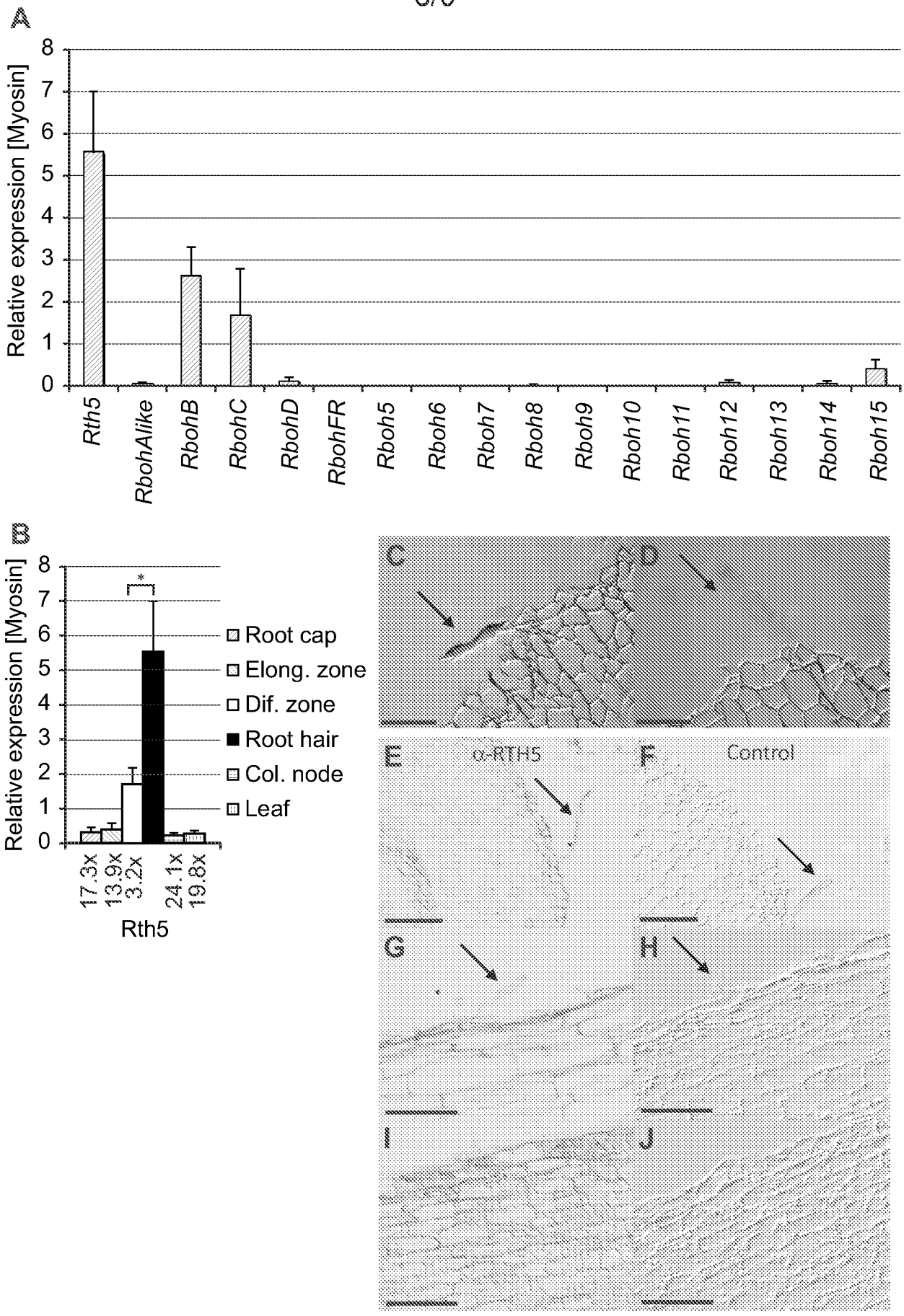
A
Relative expression [Myosin]
0
1
2
3
4
5
6
7
8
Rth5
RbohAlike
RbohB
RbohC
RbohD
RbohFR
Rboh5
Rboh6
Rboh7
Rboh8
Rboh9
Rboh10
Rboh11
Rboh12
Rboh13
Rboh14
Rboh15
B
Relative expression [Myosin]
*
17.3x
13.9x
3.2x
24.1x
19.8x
Rth5
Root cap
Elong. zone
Dif. zone
Root hair
Col. node
Leaf
C
D
E
α-RTH5
F
Control
G
H
I
J

FIG. 4

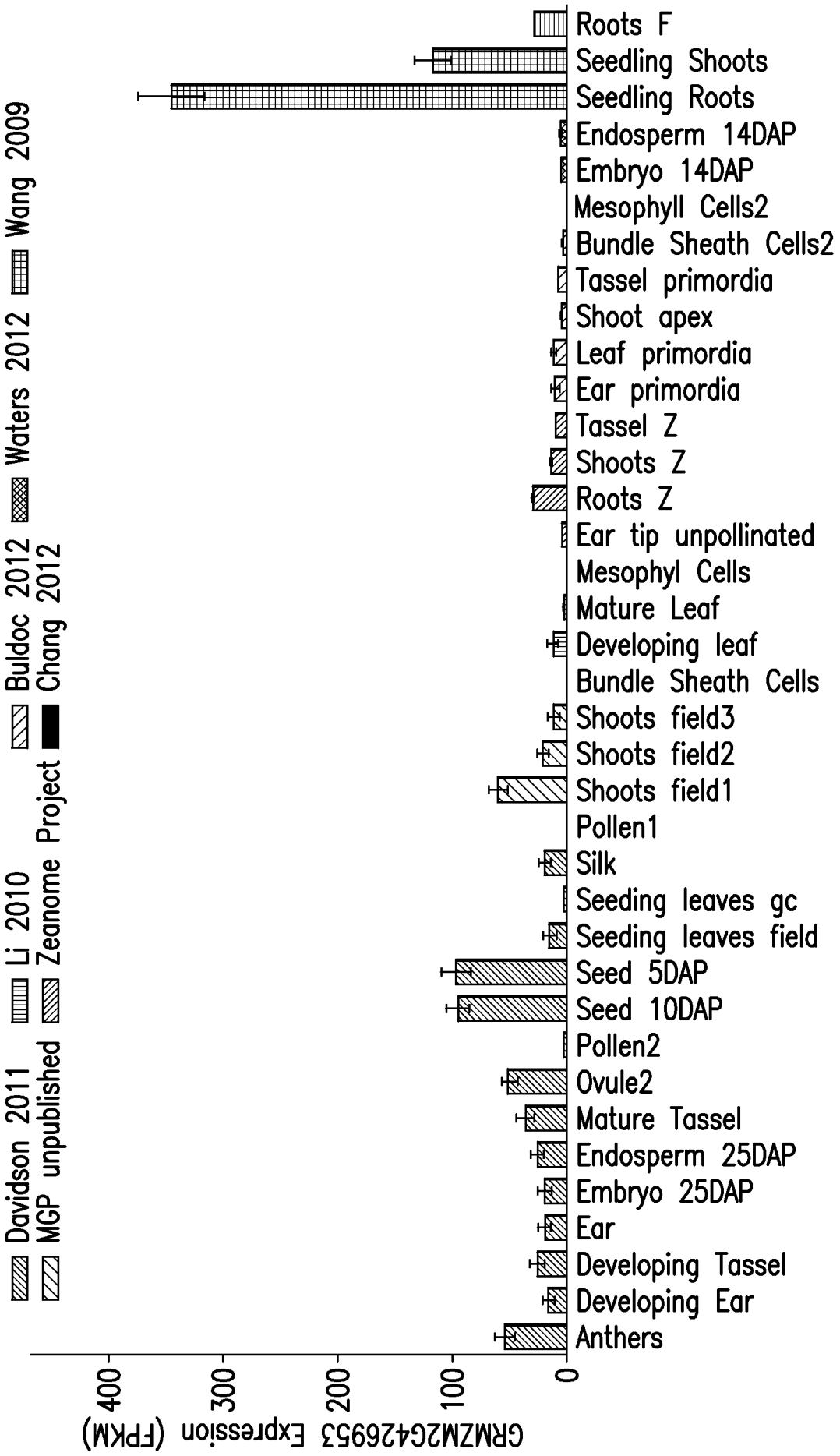
Davidson 2011
MGP unpublished
Li 2010
Zeanome Project
Buldoc 2012
Chang 2012
Waters 2012
Wang 2009
Roots F
Seedling Shoots
Seedling Roots
Endosperm 14DAP
Embryo 14DAP
Mesophyll Cells2
Bundle Sheath Cells2
Tassel primordia
Shoot apex
Leaf primordia
Ear primordia
Tassel Z
Shoots Z
Roots Z
Ear tip unpollinated
Mesophyl Cells
Mature Leaf
Developing leaf
Bundle Sheath Cells
Shoots field3
Shoots field2
Shoots field1
Pollen1
Silk
Seeding leaves gc
Seeding leaves field
Seed 5DAP
Seed 10DAP
Pollen2
Ovule2
Mature Tassel
Endosperm 25DAP
Embryo 25DAP
Ear
Developing Tassel
Developing Ear
Anthers
0
100
200
300
400
GRMZM2G426953 Expression (FPKM)

FIG. 5

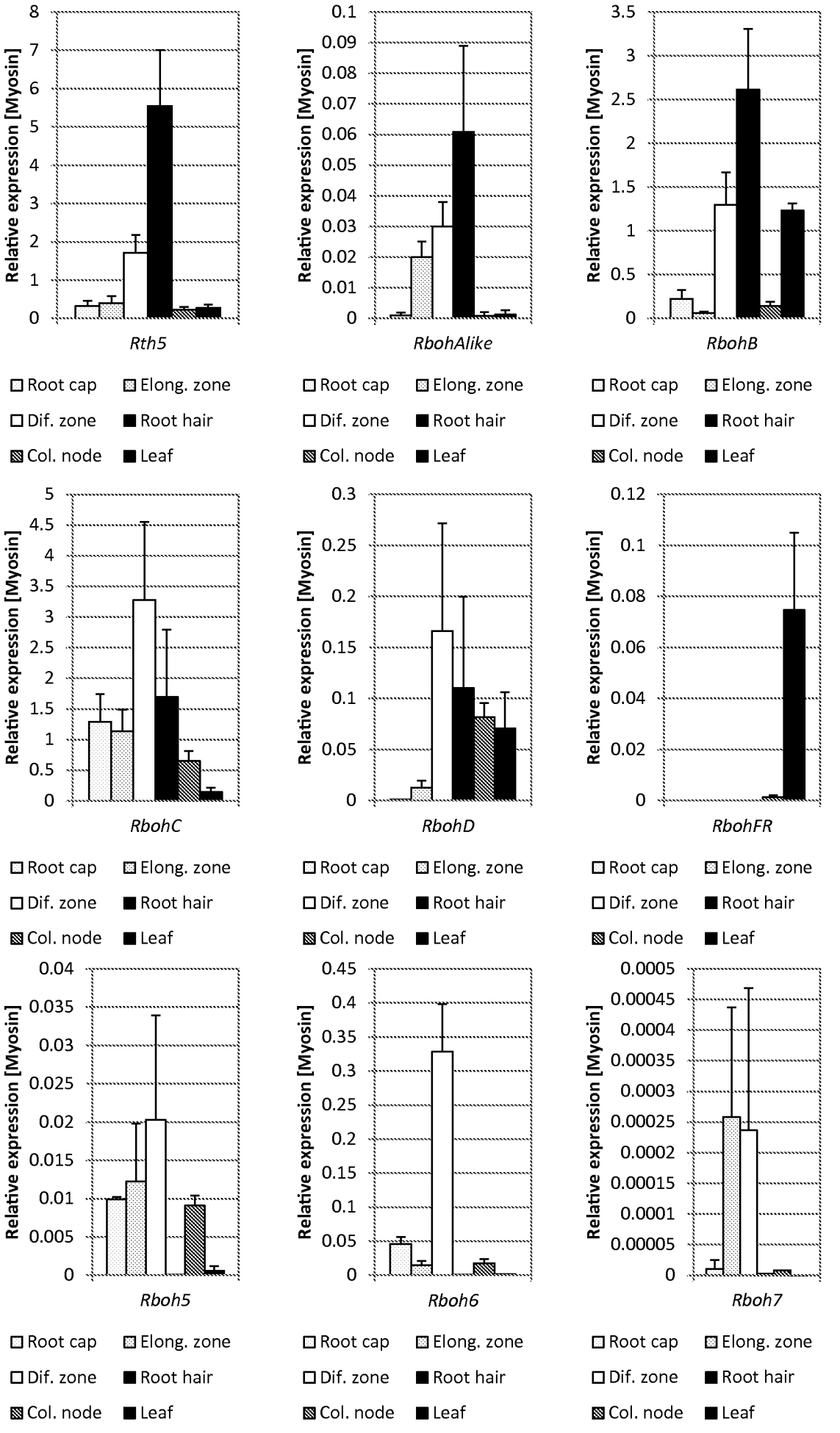
Relative expression [Myosin]
Rth5
RbohAlike
RbohB
RbohC
RbohD
RbohFR
Rboh5
Rboh6
Rboh7
Root cap
Elong. zone
Dif. zone
Root hair
Col. node
Leaf

FIG. 6

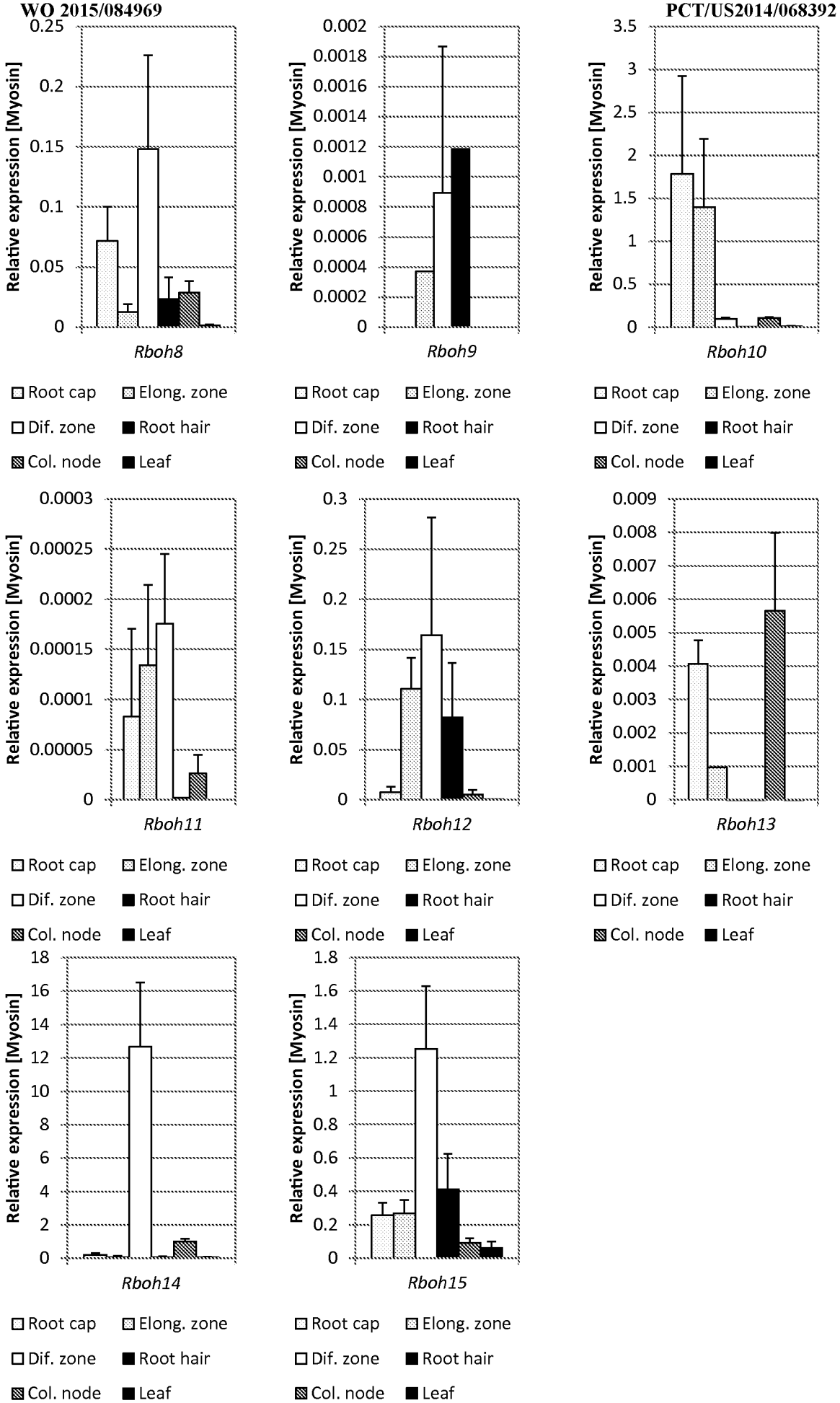

FIG. 6 (Continued)

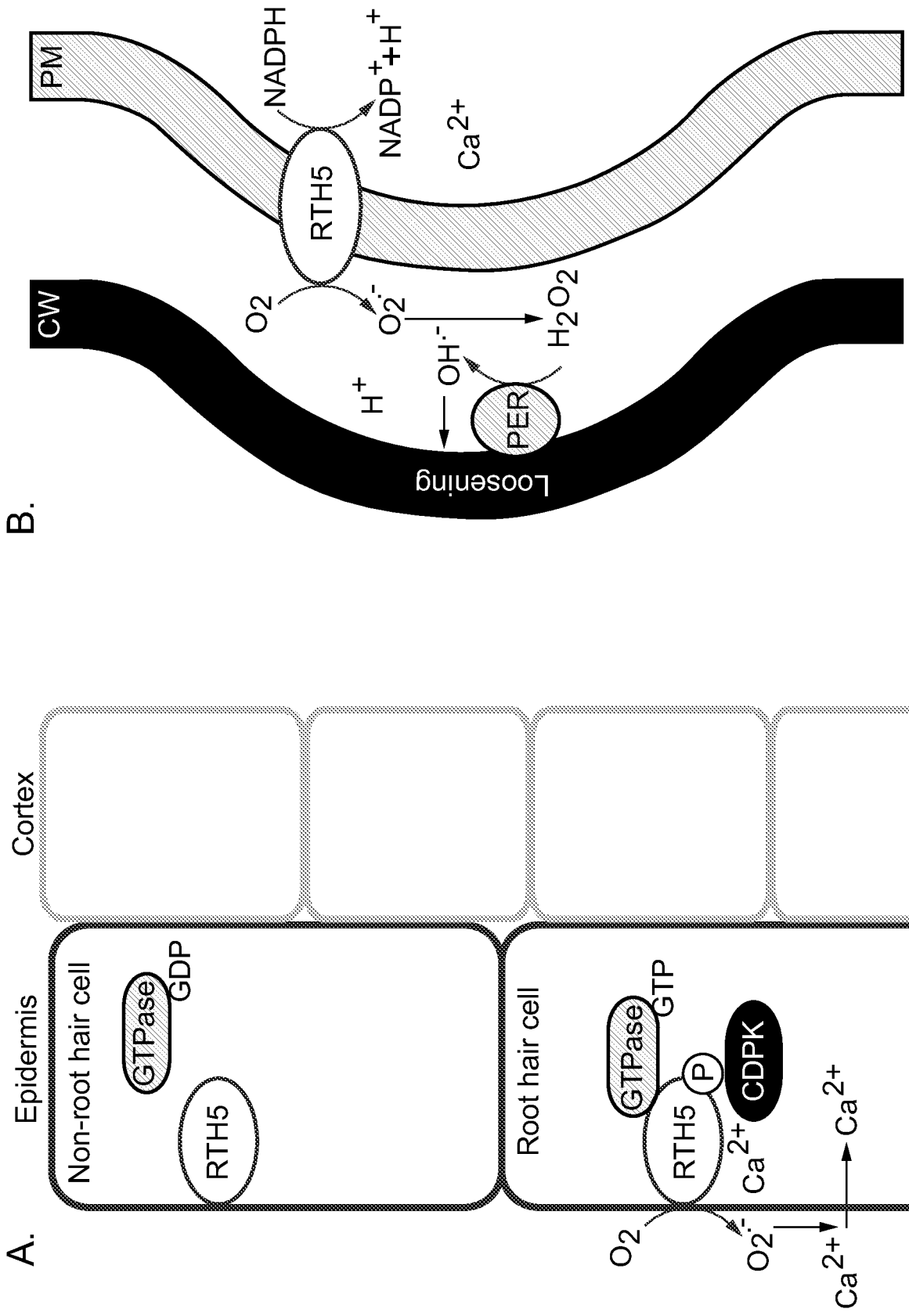
A.
Epidermis
Cortex
Non-root hair cell
GTPase
GDP
RTH5
Root hair cell
GTPase
GTP
RTH5
P
CDPK
Ca^{2+}
Ca^{2+}
O_2
$O_2^{\cdot-}$
Ca^{2+}
B.
PM
CW
NADPH
$NADP^+ + H^+$
RTH5
Ca^{2+}
O_2
$O_2^{\cdot-}$
H_2O_2
H^+
$OH^{\cdot-}$
PER
Loosening

FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/068392

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 15/29 (2015.01)
CPC - C12N 15/8261 (2015.01)
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - A01H 5/10; C12N 15/00, 15/09, 15/11, 15/29 (2015.01)
USPC - 536/1.11, 18.7, 22.1, 23.1, 23.6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - C12N 15/63, 15/79, 15/82, 15/8241, 15/8261 (2015.01)
(keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Orbit, Google Patents, PubMed, Google
Search terms used: rth5 roothairless5 plant maize corn

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012/0017292 A1 (KOVALIC et al) 19 January 2012 (19.01.2012) entire document	1-3, 10-22
---		---
Y		24
Y	LIN et al. "Positive feedback regulation of maize NADPH oxidase by mitogen-activated protein kinase cascade in abscisic acid signalling," Journal of Experimental Botany, 10 July 2009 (10.07.2009), Vol. 60, No. 11, Pgs. 3221–3238. entire document	24
A	US 8,404,929 B2 (GRUIS et al) 26 March 2013 (26.03.2013) entire document	1-3, 10-22, 24
A	US 2004/0006783 A1 (YANG et al) 08 January 2004 (08.01.2004) entire document	1-3, 10-22, 24
A	US 7,241,934 B2 (DUVICK et al) 10 July 2007 (10.07.2007) entire document	1-3, 10-22, 24

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 February 2015	25 MAR 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/068392

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-9
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/068392

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

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

   a. (means)

   ☐ on paper

   ☒ in electronic form

   b. (time)

   ☒ in the international application as filed

   ☐ together with the international application in electronic form

   ☐ subsequently to this Authority for the purposes of search

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

3. Additional comments:

SEQ ID NOs:1-13 were searched.