

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
9 March 2006 (09.03.2006)

PCT

(10) International Publication Number
WO 2006/026756 A2(51) International Patent Classification:
A01H 1/00 (2006.01)(US). LEVY, Marganit [IL/IL]; Egoz Street 4/8, 18360
Afula (IL).(21) International Application Number:
PCT/US2005/031378(74) Agents: BASTIAN, Kevin L. et al.; Townsend and
Townsend and Crew LLP, 2 Embarcadero Center, 8th
Floor, San Francisco, CA 94111 (US).(22) International Filing Date:
2 September 2005 (02.09.2005)(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ,
OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL,
SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC,
VN, YU, ZA, ZM, ZW.

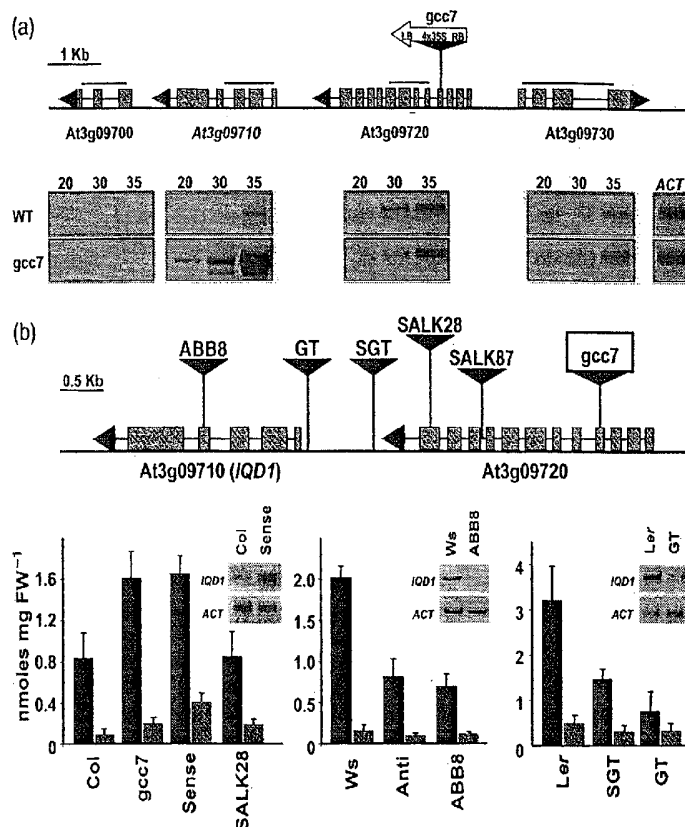
(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/607,193 3 September 2004 (03.09.2004) US(71) Applicant (*for all designated States except US*): THE
REGENTS OF THE UNIVERSITY OF CALIFOR-
NIA [US/US]; 1111 Franklin Street, 12th Floor, Oakland,
California 94607 (US).(84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,

[Continued on next page]

(54) Title: COMPOSITIONS AND METHODS FOR MODULATING GLUCOSINOLATE IN PLANTS

(57) Abstract: This invention relates to
compositions and methods for modulating
secondary metabolite levels in plants. In
particular, it relates to IQD1 nucleic acids and
proteins.



RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

Declaration under Rule 4.17:

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

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

COMPOSITIONS AND METHODS FOR MODULATING GLUCOSINOLATE IN PLANTS

CROSS-REFERENCES TO RELATED APPLICATIONS

- 5 [0001] The present application claims priority to U.S. Application No. 60/607,193, filed September 3, 2004, which is incorporated herein by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

- 10 [0002] This invention was made with Government support under Grant Nos. NRI 2001-35318-10095 and NRI 2001-35304-10919, awarded by the United States Department of Agriculture. The Government has certain rights in this invention.

FIELD OF THE INVENTION

- 15 [0003] This invention relates to compositions and methods for modulating secondary metabolite levels in plants. In particular, it relates to IQD1 nucleic acids and proteins.

BACKGROUND OF THE INVENTION

- [0004] Glucosinolates are a small but diverse class of defense-related secondary metabolites that are synthesized mainly by cruciferous species such as nutritionally important *Brassica* crops and the reference plant *Arabidopsis thaliana* (Fahey et al., *Phytochemistry*, 56:5-51 (2001); Wittstock and Halkier, *Trends Plant Sci.*, 7:263-270 (2002)). The common, glucosinolate-defining glycone structure is derived from select protein amino acids and is composed of a sulfonated oxime and a β -thioglucose residue. Extensive modification of the amino acid side chain is responsible for the chemical diversity of glucosinolates, which are stable and hydrophilic molecules and stored in vacuoles of most plant tissues. Upon tissue damage, glucosinolates are rapidly hydrolyzed by β -thioglucosidases (myrosinases) to glucose and unstable intermediates that spontaneously rearrange to various reactive products, including isothiocyanates, thiocyanates, nitriles, oxazolidine-2-thiones, or epithioalkanes (Fahey et al., *Phytochemistry*, 56:5-51 (2001); Wittstock and Halkier, *Trends Plant Sci.*, 7:263-270 (2002)). Many of these derivative compounds are biologically active and have
- 25

been implicated in plant defense against pathogens and herbivores (Kliebenstein et al., *Genetics*, 161:325-332 (2002b); Lambrix et al., *Plant Cell*, 13:2793-2807 (2001); Tierens et al., *Plant Physiol.*, 125:1688-1699 (2001)), as allelochemicals in mediating plant—insect interactions (Ratzka et al., *Proc. Natl Acad. Sci. USA*, 99:11223-11228 (2002)), or as dietary
5 inducers of detoxification enzymes that favorably modify carcinogen metabolism in mammals (Mithen et al., *J. Sci. Food Agric.*, 80:967-984 (2000); Talalay and Fahey, *J. Nutr.*, 131:3027S-3033S (2001)). The wide range of biological activities of glucosinolate breakdown products as well as the intricate intersection of indole glucosinolate metabolism and auxin homeostasis (Ljung, *Plant Mol. Biol.*, 50:309-332 (2002)) has raised interest in
10 glucosinolate biosynthesis and its regulation.

[0005] Biosynthesis of glucosinolates proceeds in three phases via (i) incremental amino acid side chain elongation; (ii) formation of the common glycone moiety to produce primary glucosinolates; and (iii) secondary modifications of the side chain to generate the known spectrum of glucosinolate compounds. More than 35 glucosinolates have been identified in
15 *Arabidopsis*, which are largely derived from methionine, tryptophan, and phenylalanine (Kliebenstein et al., *Plant Physiol.*, 126:811-825 (2001b); Reichelt et al., *Phytochemistry*, 59:663-671 (2002)). Identification of glucosinolate pathway genes in *Arabidopsis* followed by biochemical studies of recombinant enzymes confirmed the tripartite biosynthetic concept (Mikkelsen et al., *Amino Acids*, 22:279-295 (2002); Wittstock and Halkier, *Trends Plant Sci.*,
20 7:263-270 (2002)). The principle enzymes of the core pathway that converts amino acids or chain-elongated homoamino acids to primary glucosinolates have been cloned and biochemically characterized (Grubb et al., *Plant J.*, 40:893-908 (2004); Mikkelsen et al., *Plant J.*, 37:770-777 (2004); Piotrowski et al., *J. Biol. Chem.*, 279:50717-50725 (2004); Wittstock and Halkier, *Trends Plant Sci.*, 7:263-270 (2002)). Although several genes
25 involved in amino acid side chain elongation and modification have been identified (Kliebenstein et al., *Plant Cell*, 12:681-693 (2001c); Kroymann et al., *Plant Physiol.*, 127:1077-1088 (2001); de Quiros et al., *Theor. Appl. Genet.*, 101:429-437 (2000)), a number of peripheral steps in glucosinolate biosynthesis remain to be studied at the genetic and biochemical levels.

[0006] While significant progress has been made in understanding the biochemistry and genetics of glucosinolate metabolism, little is known about the regulation of glucosinolate accumulation during plant development and in response to environmental stimuli. During the life cycle of *A. thaliana*, glucosinolate content and composition change significantly among

organs and tissues. Developing roots, leaves, inflorescences, and siliques are major sites of glucosinolate synthesis or accumulation, which is consistent with a proposed function of these compounds in plant defense (Brown et al., *Phytochemistry*, 62:471-481 (2003); Petersen et al., *Planta*, 214:562-571 (2002)). A number of studies showed that a range of biotic and abiotic factors modulate leaf and seed glucosinolate profiles. These include pathogen challenge, herbivore damage and mechanical wounding, mineral nutrition, drought stress, and light conditions (Bennett and Wallsgrove, *New Phytol.*, 127:617-633 (1994); Bones and Rossiter, *Physiol. Plant.*, 97:194-208 (1996); Hirai et al., *Proc. Natl Acad. Sci. USA*, 101:10205-10210 (2004)). Treatment of *Arabidopsis* with pathogen elicitors or jasmonate (JA) induces differential accumulation of glucosinolates, and the analysis of signaling mutants indicates that JA- and salicylate (SA)-mediated defense pathways regulate the production of various glucosinolates in *Arabidopsis* (Brader et al., *Plant Physiol.*, 126:849-860 (2001); Kliebenstein et al., *Genetics*, 161:1685-1696 (2002a); Mikkelsen et al., *Plant Physiol.*, 131:298-308 (2003)). The underlying regulatory mechanisms responsible for altered glucosinolate production are largely unknown and have only recently become a target of investigation. While some of the natural variation in glucosinolate profiles among various *Arabidopsis* accessions can be explained by the allelic status of certain biosynthetic genes, other differences in glucosinolate content and composition are likely to be caused by the activity of regulatory loci (Kliebenstein et al., *Genetics*, 159:359-370 (2001a); Kliebenstein et al., *Plant Physiol.*, 126:811-825 (2001b). Kliebenstein et al., *Plant Cell*, 12:681-693 (2001c); Kliebenstein et al., *Genetics*, 161:1685-1696 (2002a); Celenza et al., *Plant Physiol.*, 137:253-262 (2005); Kim et al., *Plant Mol. Biol.*, 54:671-682 (2004)).

[0007] In response to the above needs, the present invention provides new methods for modulating levels of glucosinolate and other secondary metabolites in plants.

BRIEF SUMMARY OF THE INVENTION

[0008] This invention provides *IQD1* (also referred to as GCC7) nucleic acids and their use in modulating plant defense mechanisms, particularly levels of secondary metabolites such as glucosinolate and phytoalexins in plants. *IQD1* genes disclosed here can be used to modulate (*i.e.*, enhance or inhibit) endogenous *IQD1* expression in a variety of transgenic plants, including members of the genus *Brassica*. Typically, an expression cassette comprising a nucleic acid of the invention is introduced into the plant.

[0009] The invention provides isolated nucleic acid molecules comprising a heterologous promoter operably linked to an *IQD1* polynucleotide sequence encoding an IQD1 polypeptide which comprises an IQ76 domain as shown in SEQ ID NO: 2 and which modulates glucosinolate accumulation in a plant. The nucleic acid molecule typically encodes an IQD1 polypeptide which is at least 90% identical to SEQ ID NO: 2. In some embodiments the *IQD1* polynucleotide sequence is as shown in SEQ ID NO: 1. The promoter used in the invention is not critical and can be constitutive or inducible.

[0010] The invention also provides methods of increasing glucosinolate levels in a plant, the method comprising introducing a recombinant expression cassette comprising the *IQD1* polynucleotides of the invention. Typically, the plant is a member of the genus *Brassica*.

[0011] The invention also provides methods of decreasing glucosinolate levels in a plant, the method comprising introducing a recombinant expression cassette comprising an *IQD1* polynucleotide sequence comprising at least 10 contiguous nucleotides from SEQ ID NO: 1. Typically, the plant is a member of the genus *Brassica*.

[0012] The invention further provides transgenic plants comprising a recombinant expression cassette comprising an *IQD1* polynucleotide sequence encoding an IQD1 polypeptide which comprises an IQ76 domain as shown in SEQ ID NO: 2. Typically, the plant is a member of the genus *Brassica*.

DEFINITIONS

[0013] The term "glucosinolate", is used herein according to the commonly understood definition by those in the art. The typical glucosinolate structure is comprised of a common glycone moiety and a variable aglycone side chain.

[0014] A "IQD1 nucleic acid" or "IQD1 polynucleotide sequence" of the invention is a subsequence or full length polynucleotide sequence of a gene which encodes a polypeptide involved in control of secondary metabolites involved in plant defense. Of particular interest to the present invention are glucosinolates and phytoalexins. As shown in Figure 3, sequence similarity among IQD1-related polypeptides share a high degree of sequence conservation in their central portion, which contains three segments (I–III) of high amino acid sequence conservation. Secondary structure predictions suggest that the three segments contribute to a-helical folds. The 12 IQD1-like proteins are predicted to consist primarily of random coil structures (46–63%), a-helices (30–45%), and to a lesser extent of b strands (5–12%).

Multiple repeats of three different classes of recognition motifs for calmodulin interaction are present in segment I and separated by invariant spacing. This segment of 67 amino acid residues (between residue 107-173 of SEQ ID NO: 2) is referred to here as the IQ67 domain. The IQ67 domain contains 1-3 copies each of the IQ motif (IQxxxRGxxxR or of its more relaxed version [ILV]QxxxRxxxx[R,K]), the 1-5-10 motif ([FILVW]₃[FILV]₄[FILVW]), and the 1-8-14 motif ([FILVW]₆[FAILVW]₅[FILVW]). In addition, several conserved basic and hydrophobic amino acid residues are flanking these motifs, and the IQ67 domain is predicted to fold into a basic amphiphilic helix. The IQ motif was initially thought to mediate Ca²⁺-independent retention of calmodulin, although the motif was subsequently found to be present in some proteins that interact with calmodulin in a Ca²⁺-dependent manner (Bahler and Rhoads, 2002). The 1-5-10 and 1-8-14 motifs mediate Ca²⁺-dependent calmodulin binding and were named based on the conserved spacing of hydrophobic residues (Choi et al., 2002; Rhoads and Friedberg, 1997). Another hallmark of IQD1-related proteins is the presence of clusters of basic amino acid residues that satisfy structural characteristics of three classes of nuclear localization signals. As indicated in Figure 3, the N- and C-terminal regions of IQD1-related proteins contain several basic clusters that conform to the SV40-type, MATα2-type, and bipartite type of nuclear localization signals (Abel and Theologis, 1995). Thus, IQD1 binds to calmodulin and functions in the cell nucleus.

[0015] A *IQD1* polynucleotide is typically comprises at least 10 contiguous nucleotides from SEQ ID NO: 1, often at least about 20, and usually at least about 50 nucleotides. The polynucleotides typically comprise up to about 7000 nucleotides in length. More preferably, IQD1 polynucleotides contain a coding sequence of from about 100 to about 5500 nucleotides, often from about 500 to about 3600 nucleotides in length. A IQD1 polypeptide is typically at least 500 amino acids, typically at least 1000 amino acids, more typically at least 1500 amino acids. In some embodiments, a IQD1 polypeptide comprises fewer than 2000 amino acids, more typically fewer than 3000 amino acid and still more typically fewer than 5000 or 7500 amino acid in length.

[0016] In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "*IQD1* polynucleotide". In addition, the term specifically includes those sequences with a *IQD1* polynucleotide sequence as disclosed here and that encode polypeptides that are either wild type IQD1 polypeptides or variants that

retain the function of the IQD1 polypeptide (e.g., resulting from conservative substitutions of amino acids in the IQD1 polypeptide).

[0017] Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. 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 according to, e.g., the algorithm of Meyers & Miller, Computer Applic. Biol. Sci. 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

[0018] The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to a sequence or subsequence that has at least 40% sequence identity with a reference sequence. Alternatively, percent identity can be any integer from 40% to 100%. More preferred embodiments include at least: 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below. This definition also refers to the complement of a test sequence, when the test sequence has substantial identity to a reference sequence.

[0019] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0020] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

[0021] One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program

parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

[0022] Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0023] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

[0024] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded

sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

[0025] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

[0026] The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

[0027] The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30°C below the T_m. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 55°C, sometimes 60°C, and sometimes 65°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of

destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

[0028] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

[0029] In the present invention, genomic DNA or cDNA comprising IQD1 nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C and sometimes 65°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

[0030] A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

[0031] The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Those of skill will recognize that DNA molecule or sequence represented as a single strand will include a complementary strand, unless the context indicates otherwise (for example in the case of probes and primers).

[0032] A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental

conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. A "tissue specific" promoter is a transcriptional control element that is active in particular cells or tissues at specific times during plant development, such as in vegetative tissues or reproductive tissues. In addition, tissue-specific promoters may drive
5 expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

[0033] The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding
10 sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

[0034] The term "plant" includes whole plants, shoot vegetative organs/structures (*e.g.* leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.* bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed
15 coat) and fruit (the mature ovary), plant tissue (*e.g.* vascular tissue, ground tissue, and the like) and cells (*e.g.* guard cells, egg cells, trichomes and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, and
20 multicellular algae. It includes plants of a variety of ploidy levels, including aneuploid, polyploid, diploid, haploid and hemizygous.

[0035] A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous
25 coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally-occurring allelic variants.

[0036] A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant, or a predecessor generation of the plant, by any means other than by
30 a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like.

[0037] An “expression cassette” refers to a nucleic acid construct, which when introduced into a host cell, results in transcription and/or translation of an RNA or polypeptide, respectively. Antisense or sense constructs that are not or cannot be translated are expressly included by this definition.

5

BRIEF DESCRIPTION OF THE DRAWINGS

[0038] Figure 1 shows the identification of *IQDI*(At3gO9710) and verification of its role in glucosinolate accumulation. (a) T-DNA insertion site of line *gcc7* and mRNA expression analysis. The upper panel depicts the original T-DNA insertion in gene At3gO9710 and the organization of neighboring genes. The lower panel shows results from semiquantitative RT-PCR performed with gene-specific primers on RNA extracted from 4-week-old wild type (WT) and *gcc7* plants. ACTINI primers were used to control for cDNA input (ACT).

Numbers of PCR cycles are given above the gels. The region corresponding to the amplified cDNA is depicted as a solid line above each gene model. (b) Glucosinolate content of wild type, T-DNA insertion lines, and transgenic *IQDI* overexpression (sense) and antisense suppression lines. The T-DNA insertion sites of lines ABB8 (Ws-0), GT6935 (GT, Ler, SGT16820 (SGT, Lev), SALK_042728 (SALK28, Col-O), SALK_008887 (SALK87, Col-O), and *gcc7* (Col-0) are indicated above the gene models (upper panel). Glucosinolates were extracted from shoots of 14-day-old plants grown in agar medium and analyzed by HPLC.

Mean contents ($\pm$ SD, n = 3-4 individual plants) of methionine-derived (black bars) and tryptophan-derived (gray bars) glucosinolates are given for each line and its respective wild-type accession (lower panel). Insets show analysis of *IQDI* and ACTIN (ACT) mRNA expression of select lines by semiquantitative RT-PCR.

[0039] Figure 2 shows overall morphology of plants expressing *IOD1* gain- and loss-of-function alleles. Growth of lines *gcc7* (a), ABB8 (b), SGT16820 (c-e), and GT6935 (f) was compared with growth of wild type (plants on the right in each panel): Col (a), Ws-0 (b), and Ler (c-f). The arrow in (f) points to rosette leaves seriously damaged by herbivory. Bar = 1 cm. (a) 28 days (long-day), (b) 21 days (short-day), (c) 12 days (short-day), (d) 18 days (long-day, MS agar), and (e, f) 35 days (long-day).

[0040] Figure 3 shows amino acid sequence alignment of *IQDI* and related proteins. Aligned are closely related proteins from *Arabidopsis thaliana* (At*IQDI*-At*IQD6*), *Oryza*

sativa (OsIQD1-OsIQD5), and *Helianthus annuus* (SF16). The three segments of high amino acid sequence conservation are underlined and indicated by roman numerals. Amino acid residues conserved in at least 11 proteins are highlighted in black, and residues conserved in at least nine polypeptides are shaded (gray). Segment I contains repeats of three motifs with roles in calmodulin interaction, the IQ motif, the 1-8-14 motif (lower asterisks), and the 1-5-10 motifs (upper asterisks). Clusters of basic amino acid residues that conform to the SV40-type, MATa2-type, and bipartite type of nuclear localization signals are shaded. AtIQD1 (At3g09710), AtIQD2 (At5g03040), AtIQD3 (At3g52290), AtIQD4 (At2g26410), AtIQD5 (At3g22190), AtIQD6 (At2g26180), OSIQD1 (OS05m00240), OsIQD2 (OS01m06082), OSIQD3 (OsO3m04309), OsIQD4 (Os03m05627), OsIQD5 (OS01m00929), and SF16 (CAA52782).

[0041] Figure 4 shows Cat⁺-dependent in vitro interaction of IQD1 and calmodulin. Calmodulin-agarose beads were incubated in the presence of Cat⁺ or absence of Ca²⁺ (+EGTA) with soluble proteins prepared from induced bacterial cultures expressing a T7-tagged IQD1 protein and treated as described in Experimental procedures. Proteins of the total bacterial extract, the supernatant fraction, the entire pellet (beads) fraction, and of the last wash were resolved by SDS-PAGE, transferred to a membrane, and probed with a HRP-conjugated T7-Tag monoclonal antibody.

[0042] Figure 5 shows Subcellular localization of GFP and IQD1-GFP in root cells of transgenic Arabidopsis plants. Root tips of transgenic seedlings were incubated with propidium iodide and imaged as described in Experimental procedures. GFP-generated fluorescence (GFP), propidium iodide-generated fluorescence (PI), and merged propidium iodide/ GFP images of GFP and IQD1-GFP transgenic seedlings are shown.

[0043] Figure 6. Relative steady-state mRNA levels of select glucosinolate pathway genes. Semiquantitative and real-time RT-PCR was performed with total RNA prepared from shoots of 2-week-old plants and with gene-specific primers for glucosinolate pathway genes. Each PCR method was conducted twice with two independently grown sets of plants, and similar results were obtained in both duplicate analyses. For the semiquantitative RT-PCR, 25 cycles and 27 cycles were used to monitor expression of ACTINI and glucosinolate pathway genes, respectively. (a) Comparison of gene expression by semiquantitative RT-PCR between wild type (Col-0) and gcc7 plants. (b) Comparison of gene expression by semiquantitative RT-PCR between wild type (Col-0) and the Pr035s:1QD1 recapitulation line (sense). (c)

Analysis of transcript levels by real-time RT-PCR in line *gcc7* (gray bars) and the Pro35s:*IQD1* recapitulation line (black bars) relative to wild type. The 1092 ratio of mutant (*gcc7* or sense)/wild-type (Col) transcript levels are given. The difference between duplicates was within 15%.

- 5 [0044] Figure 7 shows Analysis of tissue-specific *IQD1* expression. (a) Analysis of steady-state *IQD1* mRNA levels in wild-type plants by semiquantitative RT-PCR using gene-specific primers for *IQD1* (27 cycles) or ACTIN (ACT, 25 cycles) and total RNA from roots (Ro), flowers (FI), stems (St), siliques (Si), inflorescence stem (In), and whole shoots (Sh). (b-v) Histochemical GUS staining of gene trap line GT6935 in seedlings at 6 days (b-d), 9
10 days (e-g), 14 days (h, i), in the inflorescence of seedlings between 21 and 32 days (j-n), and during the transition of seedlings to flowering between 12 and 18 days (o-v).

- [0045] Figure 8 shows *IQD7* overexpression reduces insect herbivory. (a) Weight gain assays of *Trichoplusia ni* larvae on 5-week-old Arabidopsis plants grown in short days (8 h light). The left panel shows the mean and SE of fresh weight gain of newly hatched cabbage
15 looper larvae feeding for 10 days on Arabidopsis lines Col-0 and *gcc7* (Col-0, $n = 39$; *gcc7*, $n = 40$; t test, $P < 0.0001$), as well as on Ws-0 and ABBI ($n = 24$ for each line; Mann-Whitney rank-sum test, $P = 0.2725$). The right panel shows representative samples of larvae that were left for 10 days on wild type (Col-0) and *gcc7* plants, and the lower panel shows a representative sample of both plant lines after 10 days of *T. ni* herbivory. (b) Dual-choice
20 assays of *Myzus persicae* nymph deposition. The graphs show the preference of green peach aphids (represented by percentage) for either wild-type Col-0 or line *gcc7* (upper graph; $n = 85$ choice pairs, $P = 0.0044$), and for either wild-type Ws-0 or line A668 (lower graph; $n = 67$ choice pairs, $P = 0.025$) in dual-choice assays (see Experimental procedures). The calculated P -values of the one-tailed binominal tests showed significant evidence for repellence by the
25 high-glucosinolate line in both choice pairs, i.e., *gcc7* (Col-0) and Ws-0 wild type.

- [0046] Figure 9 shows analysis of steady-state *IQD7* mRNA expression levels in response to hormones and mechanical stimuli. Semiquantitative and quantitative RT-PCR reactions were performed with total RNA prepared from shoots and gene-specific primers. Each PCR method was conducted at least twice with two independently grown sets of plants, and similar
30 results were obtained in both duplicate analyses. Numbers of cycles are indicated above the gels for semiquantitative RT-PCR, and the results of quantitative RT-PCR are shown only when significant differences in *IQD1* mRNA expression were measured. (a) Hormone

treatments. Soil-grown plants (3-weekold, short-days) were mock-treated (spraying) or treated for 4 and 8 h with the indicated hormones as described in Experimental procedures. Relative mRNA levels of *IQD1*, PR1(SA-inducible), PDF1.2 (JA and ethylene-inducible), and ACTIN (ACT) gene expression are shown for semiquantitative RT-PCR analyses. (b) Hormone mutants. Comparison of *IOD1* mRNA expression between wild type (Col) and various hormone-related mutants (2-week old, MS agar) by semiquantitative RT-PCR. (c) Mechanical and biotic stimuli. Comparison of *IOD1* mRNA expression between untreated, wounded (4 h), and aphid-infested (3 days), agar-grown plants (2-week old) by semiquantitative and quantitative (-fold change) RT-PCR. The difference between qRT-PCR duplicates was within 15%.

DETAILED DESCRIPTION

[0047] Compositions and methods are provided for the modification of plant defense-related gene networks, including phytoalexins and glucosinolates. Phytoalexins and glucosinolates are defense related secondary metabolites that are induced in response to various abiotic and biotic stresses, such as attack by pests (*e.g.*, insects and nematodes), as well as pathogens (*e.g.*, fungi, bacteria, and viruses) Thus, enhanced phytoalexin or glucosinolate levels can be used to enhance resistance to pathogens, pests, and other stresses.

[0048] In addition, studies suggest that consumption of members of the genus *Brassica*, such as broccoli, cauliflower, cabbage or brussel sprouts reduces the incidence of cancer in humans and other mammals (*see e.g.*, Prochaska et al., *Proc. Natl. Acad. Sci. USA*, **89**:2394-2398 (1992)). The cancer prevention activity is thought to result from the presence of metabolites of glucosinolates, which are known to be inducers of phase II enzymes, that detoxify carcinogens and mutagens in various mammalian organs (Prochaska et al., (1992) *supra* and Zhang et al., *Proc. Natl. Acad. Sci. USA*, **91**:3147-3150 (1994)).

[0049] In some contexts, glucosinolates are undesirable. For example, canola (*Brassica napus*) meal is used as a protein supplement in animal feed. The feeding value of canola meal is reduced due to the anti-nutritive effects of the breakdown products of the glucosinolates, which reduce feed intake and growth in non-ruminant animals. Methods for reducing or eliminating glucosinolates from these products has been described (*see* US Patent No. 5,866,762).

[0050] The present invention is based, at least in part, on the isolation of a T-DNA activation-tagged mutant with a high-glucosinolate chemotype. Functional characterization

of the tagged gene (IQD1) revealed it as a modifier of glucosinolate metabolism. An exemplary *IQD1* gene is shown in SEQ ID NO: 1, which is identical to GenBank Accession No. NM 111805. *IQD1* gene expression also positively correlates with camalexin (a phytoalexin) accumulation and resistance to the necrotrophic fungus *Botrytis cinerea*. Thus, *IQD1* regulates a broad range of plant defense-related gene networks.

[0051] *IQD1* encodes a novel nuclear-localized protein that binds to calmodulin in a Ca^{2+} -dependent fashion and positively regulates glucosinolate accumulation. *IQD1* overexpression leads to increased resistance against herbivory by generalist chewing and phloem-feeding insects. As *IQD1* is induced by mechanical stimuli and integrates Ca^{2+} -dependent signaling to augment and fine-tune glucosinolate accumulation in response to biotic challenge. Thus, controlling expression of *IQD1* can be conveniently used to modulate glucosinolate levels in plants.

NUCLEIC ACIDS OF THE INVENTION

[0052] The present invention involves *IQD1* polynucleotides. Polynucleotides and polypeptides are not limited to the sequences disclosed herein but include fragments and variants thereof. Those of skill in the art will recognize that conservative amino acid substitutions, as well as amino acid additions or deletions, may not result in any change in biological activity. For normal function, *IQD1* polypeptides contain a IQ67 motif described above. Thus, alterations in amino acids outside of these motifs are most likely to maintain activity.

[0053] Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described herein are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

[0054] The isolation of nucleic acids of the invention may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation,

e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as ovules, and a cDNA library which contains the *IQD1* gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which genes or homologs are expressed.

[0055] The cDNA or genomic library can then be screened using a probe based upon the sequence of a gene sequence disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against a polypeptide of the invention can be used to screen an mRNA expression library.

[0056] Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify polynucleotide sequences of the invention directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications* (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), *Academic Press*, San Diego (1990).

[0057] Appropriate primers and probes for identifying sequences of the invention from plant tissues are generated from comparisons of the sequences provided here with other related genes. Using these techniques, one of skill can identify conserved regions in the nucleic acids disclosed here to prepare the appropriate primer and probe sequences. Primers that specifically hybridize to conserved regions (e.g., the IQD167 motif) in sequences of the invention can be used to amplify sequences from widely divergent plant species.

[0058] Standard nucleic acid hybridization techniques using the conditions disclosed above can then be used to identify full-length cDNA or genomic clones.

[0059] IQD1 sequences can be tested for activity by any method known to those of skill in the art. For example, IQD1 activity can be measured by determining the ability of an encoded polypeptide to bind calmodulin in a calcium dependent manner. Suitable assays are described in the example section below. Other useful tests for IQD1 activity include assays for the ability to modulate expression of genes encoding proteins involved with glucosinolate

biosynthesis or degradation. Such assays are also described below. Yet another measure of IQD1 activity is to assay insect herbivory, as described in the example section below.

[0060] The polynucleotide sequence of the invention include polynucleotides altered to coincide with the codon usage of a particular host. For example, the codon usage of a monocot plant can be used to derive a polynucleotide that encodes a polypeptide of the invention and comprises preferred monocot codons. The frequency of preferred codon usage exhibited by a host cell can be calculated by averaging frequency of preferred codon usage in a large number of genes expressed by the host cell. This analysis is preferably limited to genes that are highly expressed by the host cell. U.S. Patent No. 5,824,864, for example, provides the frequency of codon usage by highly expressed genes exhibited by dicotyledonous plants and monocotyledonous plants. Polypeptides can also be expressed in bacteria or other microorganisms and thus, codon usage can be optimized for the particular microorganism of interest.

[0061] When synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell. The percent deviation of the frequency of preferred codon usage for a synthetic gene from that employed by a host cell is calculated first by determining the percent deviation of the frequency of usage of a single codon from that of the host cell followed by obtaining the average deviation over all codons.

MODULATING IQD1 ACTIVITY OR GENE EXPRESSION

[0062] Since *IQD1* genes are involved in glucosinolate accumulation, control of endogenous IQD1 activity or gene expression is useful in controlling glucosinolate levels in plants. For instance, an increase in *IQD1* expression can be used for increasing resistance to insects or to provide increased nutritional benefit to vegetables.

[0063] One of skill will recognize that a number of methods can be used to modulate IQD1 activity or gene expression. IQD1 activity can be modulated in the plant cell at the gene, transcriptional, posttranscriptional, translational, or posttranslational, levels. Techniques for modulating IQD1 activity at each of these levels are generally well known to one of skill and are discussed briefly below.

[0064] Isolated sequences prepared as described herein can also be used to introduce expression of a particular *IQD1* nucleic acid to enhance or increase endogenous gene

expression. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects.

[0065] One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains that perform different functions. For example, as noted above, IQD1 proteins of the invention are defined by the presence of the IQ67 domain. Thus, the gene sequences need not be full length, so long as the desired functional domain of the protein is expressed.

[0066] Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described in detail, below. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

[0067] Methods for inhibiting expression of endogenous genes are also well known. For example, methods for introducing genetic mutations into plant genes are well known. In such methods, seeds or other plant material is treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used. Plants with reduced IQD1 activity can be screened for using the assay methods described here.

[0068] Alternatively, RNA interference (RNAi) can be used. Such methods typically involve introducing an expression cassette that encodes a double stranded RNA molecule (sometimes referred to as short interfering RNA or siRNA) which prevents translation of a target mRNA. The siRNA may comprise two complementary molecules or may be constructed such that a single transcript has both the sense and complementary antisense sequences from the target gene, *e.g.*, a hairpin, which, in some embodiments, leads to production of micro RNA (miRNA). The length of the oligonucleotide is typically at least about 10 nucleotides and may be as long as the naturally-occurring target transcript. Preferably, the oligonucleotide is about 19 to about 25 nucleotides in length. Most preferably, the oligonucleotide is less than about 75, about 50, or about 25 nucleotides in length. Methods for designing double stranded RNA having the ability to inhibit gene expression in a target cell are known. (See for example, US Patent No. 6,506,559).

[0069] Other methods of inhibiting gene expression include antisense and sense suppression. For antisense suppression a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense suppression can act at all levels of gene regulation including suppression of RNA translation (*see*, Bourque *Plant Sci. (Limerick)* 105:125-149 (1995); Pantopoulos *In Progress in Nucleic Acid Research and Molecular Biology*, Vol. 48. Cohn, W. E. and K. Moldave (Ed.). Academic Press, Inc.: San Diego, California, USA; London, England, UK. p. 181-238; Heiser *et al. Plant Sci. (Shannon)* 127:61-69 (1997)) and by preventing the accumulation of mRNA which encodes the protein of interest, (*see*, Baulcombe *Plant Mol. Bio.* 32:79-88 (1996); Prins and Goldbach *Arch. Virol.* 141:2259-2276 (1996); Metzlaff *et al. Cell* 88:845-854 (1997), Sheehy *et al., Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al.*, U.S. Patent No. 4,801,340).

For an examples of the use of sense suppression to modulate expression of endogenous genes *see*, Assaad *et al. Plant Mol. Bio.* 22:1067-1085 (1993); Flavell *Proc. Natl. Acad. Sci. USA* 91:3490-3496 (1994); Stam *et al. Annals Bot.* 79:3-12 (1997); Napoli *et al., The Plant Cell* 2:279-289 (1990); Klink, *et al., J Plant Growth Regul.* 19(4):371-84 (2000); Matzke, *et al., Curr Opin Genet Dev.* 11(2):221-7 (2001); and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184.

[0070] In most of the methods described above, the nucleic acid segment to be introduced will be substantially identical to at least a portion of the endogenous *IQD1* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene. The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed.

[0071] Alternatively, homologous recombination can be used to induce targeted gene disruptions by specifically deleting or altering the *IQD1* gene *in vivo* (*see, generally*, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al., Genes Dev.* 10:2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al., Experientia*

50:277-284 (1994), Swoboda *et al.*, *EMBO J.* 13:484-489 (1994); Offringa *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); Kempin *et al.* *Nature* 389:802-803 (1997); Puchta, *Plant Mol Biol.* 48(1-2):173-82 (2002).

[0072] Gene expression can also be inactivated by transforming plant cells with constructs comprising transposons or T-DNA sequences. *IQDI* mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of *IQDI* mRNA, *e.g.*, by northern blots or reverse transcriptase PCR (RT-PCR).

PREPARATION OF RECOMBINANT VECTORS

[0073] To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

[0074] For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill.

[0075] Alternatively, the plant promoter may direct expression of the *IQDI* nucleic acid in a specific tissue or may be otherwise under more precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic (low oxygen) conditions, elevated temperature, or the presence of light. Such promoters are referred to here as "inducible". Promoters that direct expression to a specific tissue are referred to as "tissue-specific" promoters. One of skill will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

[0076] Examples of promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as fruit, seeds, flowers, roots, leaves, shoots, etc. Promoters that direct expression of nucleic acids in ovules, flowers or seeds are particularly useful in the present invention.

5 [0077] The vector comprising the sequences (*e.g.*, promoters or coding regions) from genes of the invention will typically comprise a marker gene that confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

10 PRODUCTION OF TRANSGENIC PLANTS

[0078] DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced into the genomic DNA of the plant cell using *Agrobacterium* or ballistic methods, such as DNA particle bombardment. Ballistic transformation techniques are described in,
15 *e.g.*, Klein et al. *Nature* 327:70-73 (1987).

[0079] Typically, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the
20 bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. *See*, for example Horsch et al. *Science* 233:496-498 (1984), and Fraley et al. *Proc. Natl. Acad. Sci. USA* 80:4803 (1983).

[0080] Transformed plant cells which are derived by any of the above transformation
25 techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as increased seed mass. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Regeneration can also be obtained from plant callus,
30 explants, organs, or parts thereof are well known.

[0081] The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from

the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*,
5 *Oryza*, *Panicum*, *Pennisetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, and *Zea*.

[0082] One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by
10 sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[0083] Seed obtained from plants of the present invention can be analyzed according to well known procedures to identify plants with the desired trait. For example, Northern blot analysis can be used to screen for desired plants. In addition, IQD1 activity can be assayed as
15 described above.

EXAMPLES

[0084] The following examples are offered to illustrate, but not to limit the claimed invention.

[0085] In an attempt to elucidate glucosinolate biosynthesis and its regulation in
20 *Arabidopsis*, we undertook a molecular genetic approach (Gross et al., *Plant Sci.*, 159:265-272 (2000); Grubb et al., *Plant Sci.*, 162:143-152 (2002)) and developed a novel screening procedure for T-DNA-induced mutations that cause altered glucosinolate accumulation (Wang et al., *Phytochem. Anal.*, 13:152-157 (2002)). Here, we describe the isolation of a T-DNA activation-tagged mutant with a high-glucosinolate chemotype followed by the
25 functional characterization of IQD1 as a modifier of glucosinolate metabolism. We demonstrate that IQD1 encodes a novel nuclear-localized protein that binds to calmodulin in a Ca^{2+} -dependent fashion and positively regulates glucosinolate accumulation. Interestingly, we observed that IQD1 overexpression leads to increased resistance against herbivory by generalist chewing and phloem-feeding insects. As IQD1 is induced by mechanical stimuli,
30 the prospect arises that IQD1 integrates Ca^{2+} -dependent signaling to augment and fine-tune glucosinolate accumulation in response to biotic challenge.

EXAMPLE 1:Experimental proceduresPlant material and growth conditions

5 [0086] Wild-type *A. thaliana* (L.) Heynh. accessions Columbia (Col-0), Landsberg *erecta* (*Ler*) and Wassilewskija (Ws-0), pools of T-DNA activation-tagged lines (Col-0) (Weigel et al., *Plant Physiol.*, 122:1003-1013 (2000)), T-DNA insertion lines SALK_008887 and SALK_042728 (Col-0), and mutant lines *npr1*, *ein2*, *jar1*, and *nahG* (Col-0) were obtained from the ABRC (Columbus, OH, USA). Gene trap line GT6935 (*Ler*) was provided by J.

10 Simorowski (CSHL, Cold Spring Harbor, NY, USA) (Sundaresan et al., *Genes Dev.*, 15:1797-1810 (1995)), T-DNA insertion line ABB8 (Ws) by INRA (Versailles, France), *Ds* transposon-tagged line SGT16820 (*Ler*) (Parinov et al., *Plant Cell*, 11:2263-2270 (1999)) by V. Sundaresan (University of California, Davis, CA, USA), the *fad3-2 fad7-2 fad8-1* (Col-0) triple mutant (McConn and Browse, *Plant Cell*, 8:403-416 (1996)) by J. Browse (Washington

15 State University, Pullman, WA, USA), and the *cpr1* (Col-0) mutant (Bowling et al., *Plant Cell*, 6:1845-1857 (1994)) by X. Dong (Duke University, Durham, NC, USA). All insertion lines were backcrossed once to the respective wild-type accessions prior to phenotypic analysis. Plants were grown in soil at 22°C and 70% relative humidity under illumination with fluorescent and incandescent light at a photo fluence rate of approximately 120 $\mu\text{mol m}^{-2}$

20 sec^{-1} . For plant growth in sterile conditions, seeds were surface-sterilized for 10 min in 1.5% (w/v) sodium hypochlorite and placed on solid medium containing 8 g l⁻¹ phytagar, 30 g l⁻¹ 'sucrose, 2.15 g l⁻¹ (0.5x) Murashige-Skoog (MS) salts, pH 5.6 (Murashige and Skoog, *Physiol. Plant.*, 15:473-497 (1962)), and 1x MS vitamin mix (Sigma, St Louis, MO, USA). After 2–3 days of stratification at 4°C, plants were grown at 24°C at a photon fluence rate of

25 approximately 160 $\mu\text{mol m}^{-2} \text{sec}^{-1}$ for 18 h (long days) or 8 h (short days) per day. For selection of transgenic plants, the agar medium was supplemented with 25 mg ml⁻¹ kanamycin (Sigma), or seedlings grown in soil were sprayed with 1% (v/v) Basta herbicide (Finale; Farnam Companies, Phoenix, AZ, USA). For hormone treatments, soil-grown or agar-grown plants were sprayed with 250 μM MeJA, 5 mM SA, 2.5 mM ACC (Sigma) and

30 with 0.025% ethanol as the solvent control. For external mechanical stimuli, plants were sprayed with water or wounded with forceps a few times in each leaf. For herbivory challenge, five green peach aphid nymphs (*M. persicae*) were placed onto each plant grown in MS agar medium.

Screen for glucosinolate accumulation mutants

[0087] T₄ progeny of T-DNA activation-tagged lines (Weigel et al., *Plant Physiol.*, 122:1003-1013 (2000)) were grown in sterile conditions as previously described (Grubb et al., *Plant Sci.*, 162:143-152 (2002)). After 14-15 days of growth, two single leaf disks (3 mm diameter) were excised from opposite leaves per seedling. Leaf disks were individually tested for leaf QR inducer potency using a colorimetric bioassay of QR activity in cultured murine hepatoma cells, which was optimized for direct analysis of leaf disks (Wang et al., *Phytochem. Anal.*, 13:152-157 (2002)). Fifty-five pools of 100 lines each were screened by analyzing 300 T₄ progeny per pool. Putative mutants were re-evaluated by assaying eight leaf disks obtained from four T₅ progeny. Glucosinolate profiles of T₅ lines that showed heritable changes in leaf QR inducer potency were analyzed by HPLC.

HPLC analysis of desulfoglucosinolates

[0088] For glucosinolate analysis of plant tissues, desulfoglucosinolate extracts were prepared according to Wang et al., *Phytochem. Anal.*, 13:152-157 (2002) with minor modifications. Leaves of 2-week-old plants grown on agar medium were harvested and extracted by boiling for 10 min in 1 ml H₂O supplemented with 1 µl of 100 µM benzylglucosinolate (Merck, Darmstadt, Germany) as internal standard. The supernatant was collected and the plant residue washed with 1 ml H₂O. The combined extract was applied to a DEAE Sephadex A-25 (60 mg) column that was equilibrated with 2 ml of 0.5 M pyridine acetate and washed with 1 ml H₂O. Subsequently, the column was washed with 3 ml of 0.02 M pyridine acetate and 2 ml H₂O. The glucosinolates were converted into desulfoglucosinolates by overnight treatment with 75 µl of 0.1% (1.4 U) aryl sulfatase (Sigma). Desulfoglucosinolates were eluted with 1 ml H₂O, and aliquots of 100 µl were analyzed by HPLC using a Shimadzu VP Liquid Chromatograph with a dual wavelength spectrophotometer (Shimadzu, Tokyo, Japan). Samples were separated at 45°C on a Waters Spherisorb C₁₈ column (Waters, Milford, MA, USA; 150 x 4.6 µm i.d.; 5 µm particle size) using acetonitrile and water at a flow rate of 1 ml min⁻¹. The column was developed by isocratic elution with 1.5% acetonitrile (5 min) followed by a linear gradient to 20% acetonitrile (15 min) and isocratic elution with 20% acetonitrile (10 min). Absorbance was detected at 226 and 280 nm. Desulfoglucosinolate concentrations were calculated based on published response factors (Haughn et al., *Plant Physiol.*, 97:217-226 (1991); Petersen et al., *Planta*, 214:562-571 (2002)) and the internal benzylglucosinolate standard.

Plasmid rescue and identification of the T-DNA insertion site

[0089] Genomic DNA of T₅ progeny was isolated using the Nucleon Phytopure Plant DNA Extraction Kit (Amersham Pharmacia, Piscataway, NJ, USA), restricted with *Bam*HI, ligated (T4 DNA ligase), and transformed into *E. coli* DH5 α for plasmid rescue of the T-DNA insert (Weigel et al., *Plant Physiol.*, 122:1003-1013 (2000)). Recovered plant genomic DNA fragments adjacent to the left border of the T-DNA insert were analyzed by DNA sequencing using the T-DNA left border primer JL202 (5'-CATTTTATAATAACGCTGCGGACATCTAC-3') and primer T7 (5'-TAATACGACTCACTATAGGGCGAAT-3') complementary to the T7 promoter (Weigel et al., *Plant Physiol.*, 122:1003-1013 (2000)).

mRNA detection by reverse transcriptase-mediated PCR

[0090] Total RNA was extracted from rosette leaves of 14- to 28-day-old seedlings (as indicated) using the RNeasy Kit (Qiagen, Valencia, CA, USA) followed by treatment with RNase-free DNase (Qiagen) to remove genomic DNA contamination. Five micrograms of total RNA was reverse-transcribed with the First-Strand cDNA Synthesis SSII Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The resulting cDNA was diluted fivefold, and 1 μ l was used for semiquantitative RT-PCR reactions, which were performed in a total volume of 25 μ l in PCR buffer (Eppendorf, Hamburg, Germany) containing 250 μ M dNTPs, 1.5 mM MgCl₂, 10 pmol of forward and reverse primers and 2.5 U of *Taq* DNA polymerase (Eppendorf). PCR conditions were 2 min at 94°C followed by 15-35 cycles (as indicated) of 94°C for 15 sec; 60°C for 15 sec, and 72°C for 1 min. Ten microliters of the PCR was analyzed by gel electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining. *ACTIN1*-specific primers were used as a control for the amount of template cDNA. RT-PCR reactions were performed at least twice with independent RNA preparations.

[0091] For real-time RT-PCR analyses, the resulting cDNA was diluted 500-fold and 5 μ l was used as template. Reactions were performed as previously described (Grubb *et al.*, 2004) using an ABI 7300 thermocycler and the SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. The results were analyzed with SDS software (version 2.1) and default settings were used for baseline and threshold cycle values (Applied Biosystems). All mRNA levels were calculated from threshold cycle values and as relative to wild-type controls and normalized with respect to

ACTIN 1 transcript levels (Applied Biosystems user bulletin no. 2, ABI Prism 7700 Sequence Detection System).

[0092] Gene-specific amplimers for semiquantitative and quantitative RT-PCR are listed in Table 1.

5 **Table 1** Amplimers for semi-quantitative and quantitative RT-PCR

a) Semi-quantitative RT-PCR

At3g09700-F (5'-ACCAGTTAGGCCTAGAATGCGAAAA-3')

10 At3g09700-R (5'-TGGCTTCATTGATCTTAGACGCAAG-3')

At3g09710-F (5'-CAGTCTACGTTTAGGGGCCATTTG-3')

At3g09710-R (5'-CTTGAGAACGCCTTTTCCACCTT-3')

At3g09720-F (5'-GATGACTAAACCCCTTGTCAAAAC-3')

At3g09720-R (5'-GCTCTGATGTTTTACATTTTCAGT-3')

15 At3g09730-F (5'-CCTACTCCCGAGAATTACGCTGAA-3')

At3g09730-R (5'-ATCAATCTGAACGGTCCAAGAGAGG-3')

CYP79F1-F (5'-TTTTTAGACACCATCTTGTTTTCTTCTTC-3')

CYP79F1-R (5'-AAAGCTCAATGGGTAGAAT-3')

CYP79F2-F (5'-AAAGCTCAATGCGTCGAAT-3')

20 CYP79F2-R (5'-GCGTCGAAACACATCACAGAG-3')

CYP79B2-F (5'-AACCCACCATTAAGGAGC-3')

CYP79B2-R (5'-TCATAAAATATATACGGCGTCG-3')

CYP79B3-F (5'-AAACCAACCATTAAGGAACT-3')

CYP79B3-R (5'-TCCTCGCCGTACGTCACCG-3')

25 CYP83B1-F (5'-AGCTACTCAAACTCAAGACCTCAA-3')

CYP83B1-R (5'-TTTGTAGATCTGCATCAAAAGATCA-3')

UGT74B1-F (5' -CGTTTCGTATCCGTGGCTTA-3')

UGT74B1-R (5' -GAATCAGAGCTCTTACTTCCCTAAACTCTCTATAAAC-3')

TGG1-F (5'- CTAAACTTTTCAACAGTGGCAATTT- 3')

TGG1-R (5'-CAATGCATATCCTCTAGTAGGCACT- 3')

5 PR1-F (5'-GCCCACAAGATTATCTAAGGG-3')

PR1- R (5'-ACCTCCTGCATATGATGCTCCT-3')

PDF1.2- F (5'- TCATGGCTAAGTTTGCTTCC-3')

PDF1.2 -R (5'AATACACACGATTTAGCAC-3')

ACTIN1-F (5'- TGGAAGTGGGAATGGTTAAGGCTGG-3')

10 ACTIN1-R (5'- TCTCCAGAGTCGAGCACAATACCG-3')

b) Quantitative RT-PCR

qCYP79F1-F (5'-CCCATCTTGCGCGTCAA-3')

15 qCYP79F1-R (5'-GGTGACGCTCCGGTTTGTATAC-3')

qCYP79F2-F (5'-ACCTCATGTTGCCCCGTCAAG-3')

qCYP79F2-R (5'-CGCTCCGGTTCGTATGCTA-3')

qCYP79B2-F (5'-AAGACGAACAAGGCAACCCA-3')

qCYP79B2-R (5'-TGTCGATCTCTTCCATTGCTTTAC-3')

20 qCYP79B3-F (5'-ACGAAGCTGGCCAGCCT-3')

qCYP79B3-R (5'-CTCTATCGATCTCTTCCATAGCTTTGT-3')

qCYP83B1-F (5'-TTCTCCATCAAATTCACACGAA-3')

qCYP83B1-R (5'-TCAGTTCCCGGCACAACA-3')

qTGG1-F (5'-CGTACACACTGCCTTGATGGA-3')

25 qTGG1-R (5'-CATGACCAGTTGCATTTTTAGATGT-3'),

qUGT74B1-F (5'-TTTGTCTGTGTTGCGTAAATTCTC -3')

qUGT74B1-R (5'- GGCCACCTAAACCCAATGG-3')

qACTIN1-F (5'-TCGTTGCCCCCTCCAGAGA-3')

qACTIN1-R (5'-GGTACTGAGGGAGGCCAAGAT-3')

5

Expression of recombinant IQD1 and calmodulin binding assay

[0093] A full-length cDNA fragment encoding the predicted *IQD1* protein was generated by RT-PCR using gene-specific primers clQD1-F (5'-ATGGTTAAAAAAGCGAAATG-3') and clQD1-R (5'-GACTTACACTGCACAAATGAATCAA-3'). The PCR fragment was subcloned into the TA cloning vector pGEMT (Promega, Madison, WI, USA), creating plasmid pGEMT(IQD1). After verifying the authenticity of the insert by DNA sequencing, the IQD1 cDNA fragment was mobilized into the *EcoRI* site of vector pET21 (Novagen, Madison, WI, USA), which provides an N-terminal T7-epitope tag, MASMTGGQQMG.

15 The orientation of the insert in plasmid pET21(T7-IQD1) was verified by restriction enzyme analysis. The recombinant T7-IQD1 protein was expressed in *E. coli* M15[pREP4] (Qiagen) at 37°C for 4 h by induction with 1 mM isopropyl β -D-thiogalactopyranoside (IPTG). Bacterial cells were harvested and sonicated in 1x T7-Tag washing buffer (T7-Tag Affinity Purification Kit; Novagen). After centrifugation, the supernatant was aliquoted and stored at

20 -20°C.

[0094] Aliquots of 100 μ l of calmodulin-agarose beads (phosphodiesterase -3':5'-cyclic nucleotide activator from bovine brain; Sigma), pre-equilibrated with 1x T7 washing buffer, were mixed with 500 μ l of bacterial supernatant (2 $\text{gg } \mu\text{l}^{-1}$ protein) supplemented with 1 mM CaCl_2 or 5 mM EGTA and incubated for 1 h at 4°C under gentle shaking. Calmodulin-agarose beads were sedimented by centrifugation and washed four times with 500 μ l of 1x T7 washing buffer, followed by a final wash with 100 μ l of the same solution. The bound proteins were eluted by boiling the beads for 2 min in 100 μ l of 4x SDS sample buffer. Proteins of the total extract, the initial supernatant, the last wash, and the pellet fraction were resolved on 10% (w/v) SDS-polyacrylamide gels (Laemmli, 1970). Proteins were transferred to polyvinylidene difluoride membranes (Osmonics, Westborough, MA, USA) in 15.6 mM

25

30 Tris, 120 mM glycine, 20% (v/v) methanol and 0.1% (w/v) SDS) for 60 min at constant 60 V.

Membranes were washed with 1x TTBS buffer [50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% (v/v) Tween-20] and blocked with 5% (w/v) non-fat milk in 1x TTBS buffer for 1 h at 25°C. Blots were probed with a 5000-fold dilution of T7-Tag monoclonal antibody conjugated with horse radish peroxidase (Novagen) for 16 h at 4°C. After washing the
5 membranes with 1x TTBS buffer for 3 x 15 min, signals were detected using an enhanced chemiluminescence system (SuperSignal-West Femto Maximum Kit; Pierce, Rockford, IL, USA).

Construction of transgenic Arabidopsis plants

[0095] To generate *IQD1* overexpression (recapitulation) and antisense transgenic lines, the
10 *Xba*I-*Not*I fragment of plasmid pET21(T7-IQD1) was mobilized into pATC940, a PB1101-derived binary plant trans-formation vector, to drive expression of T7-IQD1 cDNA, or of its antisense orientation, under control of the CaMV 35S promoter. For transgenic expression of translational IQD1-GFP fusions, the *Eco*RI-*Sma*I fragment of pGEMT(*IQD1*) was subcloned into the *Eco*RI-*Spe*I site of vector pEGAD (Cutler *et al.*, 2000) to create pEGAD(IQD1-GFP).
15 Plasmids pATC940(senseT7-IQD1), pATC940(antisenseT7-IQD1), pEGAD(IQD1-GFP), and pEGAD were transformed into *Agrobacterium tumefaciens* (strain GV3101) by electroporation. Transgenic Arabidopsis plants were generated by *A. tumefaciens*-mediated transformation as described previously (Clough and Bent, 1998). T₂ lines showing a segregation ratio of 3:1 for resistance to kanamycin were selected for subsequent analysis.

Histochemical analysis of transgenic plants

[0096] Gene trap line GT6935 expresses a promoter-less bacterial β -glucuronidase gene *uidA* as a transcriptional *IQD1* promoter fusion. The *uidA* insertion site is 101 bp upstream of the predicted translational start codon of *IQD1*. Histochemical staining for GUS activity was performed essentially according to Jefferson *et al.* (1987). Sample tissues were fixed in
25 80% cold acetone for 20 min and incubated for 11-16 h at 37°C in reaction buffer [50 mM Na₂HPO₄-NaH₂PO₄, pH 7.0, 0.5 mM K₃Fe(CN)₆, 0.5 mM K₄Fe(CN)₆] containing 2 mM 5-bromo-4-chloro-3-indolyl- β -D-glucuronic acid (X-Glue) as the substrate. Plant pigments were destained with ethanol, and the GUS staining patterns were recorded under a dissecting micro-scope (Nikon 5MZ800, Tokyo, Japan).

[0097] Roots of transgenic plants expressing the green fluorescence protein (GFP) reporter protein, pEGAD, or translational fusions of GFP and IQD1, pEGAD(IQD1-GFP), under control of the CaMV 35S promoter were stained with 1 μ g ml⁻¹ propidium iodide for 5 min.

After transfer to 50 mM Na₂HPO₄-NaH₂PO₄, pH 7.0, green (GFP) and red (propidium iodide) fluorescence of root tips was analyzed using a Leica TCS 4D confocal laser microscope (Leica, Wetzlar, Germany). Specimens were viewed using wavelengths specific for each stain, producing two images that were merged using Leica confocal software.

5 Aphid dual-choice assays

[0098] Green peach aphid (*M. persicae*) colonies were maintained on cabbage seedlings (*Brassica oleracea* var. *capitata*) in laboratory conditions (25 ± 5°C, 50 ± 20% relative humidity, 16 h light). Dual-choice assays were performed to study the preference of aphids for transgenic (high-glucosin late line gcc7; low-glucosin late line ABB8) or wild type
10 (Col-0; Ws-0) control Arabidopsis plants. For this purpose, aphid nymphs (late instars of alate aphids) were transferred using a fine hair brush (no. 00) and released into the center of a petri dish (100 x 25 mm; Nunc, Rochester, NY, USA) between two 15-day-old test plants of different genotypes that were grown on each half of the dish in 0.8% phytagar medium. After aphid transfer, the petri dishes were returned to the controlled growth chamber, incubated at
15 20°C (16 h light, 70 ± 5% relative humidity), and examined every 24 h for three successive days. The aphids could easily walk or fly toward the test plants, and the location of the first new-born nymphs was recorded. One-tailed binomial tests were performed to test the hypothesis that the aphids will choose the low-glucosinolate line over the high-glucosinolate line for nymph deposition (Zar, 1999).

20 Cabbage looper weight gain assay

[0099] Eggs of the cabbage looper (*T. ni*) from a highly inbred population were purchased from Enthopath, Inc. (Easton, PA, USA). One newly hatched larva was transferred with a fine hair brush (no. 00) to a 5-week-old plant of the specified genotype that was grown in soil and short-day condition (8 h light). Individual plants were confined in a thrips-screen cage
25 and returned to the environmentally controlled growth chamber (20°C, 45 ± 5% relative humidity, 8 h light). After 10 days of feeding, the fresh weights of looper larvae were individually determined. *t*-Tests were performed to compare larvae weight when justified, and Mann–Whitney rank sum tests when the assumptions for parametric tests were violated (Zar, 1999).

30 [0100] HPLC chromatograms of desulfoglucosinolates prepared from shoots of 14-day-old wild type (a) and gcc7 (b) plants. The elution of UV light-absorbing compounds in the desulfoglucosinolate fraction was monitored at 226 nm. Abbreviations of methionine and

tryptophan-derived glucosinolate side chains are: S3, 3-methylsulfinylpropyl; S4, 4-methylsulfinylbutyl; S5, 5-methylsulfinylpentyl; S7, 7-methylsulfinylheptyl; S8, 8-methylsulfinyloctyl; T4, 4-methylthiobutyl; IM, indole-3-ylmethyl; 1IM, 1-methoxyindole-3-ylmethyl; 4IM, 4-methoxyindole-3-ylmethyl. Std, internal standard.

5 Results

Screen for activation-tagged loci conferring a high-glucosinolate chemotype

[0101] We previously developed a bioassay for leaf glucosinolate content in Arabidopsis that is based on the induction of quinone reductase (QR) activity in cultured murine hepatoma cells by isothiocyanates produced from glucosinolates during plant tissue destruction (Gross et al., *Plant Sci.*, 159:265-272 (2000)). We observed a positive correlation between leaf QR inducer potency and leaf glucosinolate content for *A. thaliana* and demonstrated the feasibility to identify chemically induced mutants with altered glucosinolate profiles (Grubb et al., *Plant Sci.*, 162:143-152 (2002)). We further showed that excised leaf disks can substitute for cell-free leaf extracts as a source of glucosinolate-derived isothiocyanates, which facilitates high through-put screening of insertional mutant collections (Wang et al., *Phytochem. Anal.*, 13:152-157 (2002)). Here, we screened duplicate leaf disks of 16 500 T₄ progeny derived from 5500 independent T-DNA activation-tagged lines (Weigel et al., *Plant Physiol.*, 122:1003-1013 (2000)) for altered leaf QR inducer potency. We identified 404 putative mutants that were retested in the T₅ generation. Leaf material of T₅ progeny that inherited changes in QR inducer potency were analyzed by high-performance liquid chromatography (HPLC) for glucosinolate profiles. We identified 12 glucosinolate content and composition (gcc) lines with an altered glucosinolate chemotype that passed the three-stage screening procedure. Line gcc7, the focus of this study, accumulates about twice as much methionine- and tryptophan-derived glucosinolates as wild type (Col-0) control seedlings (Table 2).

Table 2 Glucosinolate content of wild-type accessions, T-DNA insertion lines, and transgenic *IQD1* overexpression and antisense lines

Line ^a	S3	S4	S8	T4	3-OH	IM	4IM	1IM	Total
WT (Col-0)	0.11 ± 0.02 ^b	0.62 ± 0.18	0.06 ± 0.06	0.05 ± 0.02	ND	0.05 ± 0.02	0.04 ± 0.02	0.05 ± 0.01	0.98 ± 0.33
<i>gcc7</i>	0.16 ± 0.01	1.26 ± 0.15	0.10 ± 0.02	0.11 ± 0.03	ND	0.13 ± 0.02	0.08 ± 0.02	0.07 ± 0.01	1.91 ± 0.26
Sense	0.18 ± 0.02	1.42 ± 0.15	0.06 ± 0.05	0.12 ± 0.01	ND	0.05 ± 0.01	0.14 ± 0.03	0.09 ± 0.04	2.06 ± 0.31
Salk_28	0.11 ± 0.02	0.58 ± 0.16	0.05 ± 0.03	0.05 ± 0.02	ND	0.06 ± 0.02	0.06 ± 0.02	0.06 ± 0.02	0.97 ± 0.27
WT (Ws-0)	1.25 ± 0.13	ND	0.11 ± 0.01	ND	ND	0.12 ± 0.02	0.14 ± 0.01	0.03 ± 0.01	1.65 ± 0.18
ABB8	0.64 ± 0.12	ND	0.07 ± 0.02	ND	ND	0.03 ± 0.01	0.10 ± 0.01	0.01 ± 0.00	0.85 ± 0.16
Antisense	0.76 ± 0.19	ND	0.06 ± 0.03	ND	ND	0.06 ± 0.02	0.04 ± 0.01	ND	0.92 ± 0.25
WT (<i>Ler</i>)	0.26 ± 0.08	ND	ND	ND	2.86 ± 0.60	0.32 ± 0.11	0.11 ± 0.03	0.09 ± 0.04	3.64 ± 0.86
GT6935	0.16 ± 0.07	ND	ND	ND	0.60 ± 0.37	0.12 ± 0.03	0.08 ± 0.04	0.01 ± 0.02	0.97 ± 0.53
SGT16820	0.11 ± 0.05	ND	ND	ND	1.34 ± 0.15	0.11 ± 0.02	0.09 ± 0.06	0.01 ± 0.03	1.77 ± 0.31

S3, 3-methylsulfinylpropyl; S4, 4-methylsulfinylbutyl; S8, 8-methylsulfinyloctyl; T4, 4-methylthiobutyl; 3-OH, 3-hydroxypropyl; IM, indole-3-ylmethyl; 1IM, 1-methoxyindole-3-ylmethyl; 4IM, 4-methoxyindole-3-ylmethyl; ND, not detectable.

^a The desulfoglucosinolate fraction extracted from shoots of 14-day-old seedlings was analyzed by HPLC.

^b Glucosinolate concentrations are given in nmol mg⁻¹ FW ± SD (*n* = 3-4 extractions of individual plants).

Genetic analysis of line *gcc7*

[0102] The T-DNA vector pSKI015 used for activation mutagenesis contains four repeats of the enhancer region of the constitutively active 35S promoter of *cauliflower mosaic virus* (CaMV) and the *BAR* gene for phosphinotricin (glufosinate) resistance in the T-DNA insert. Introduction of this T-DNA into the genome can cause increased gene expression near the integration site in an orientation-independent manner (Weigel et al., *Plant Physiol.*, 122:1003-1013 (2000)). Analysis of the T₅ and T₆ generation of line *gcc7* demonstrated glufosinate resistance of all progeny and a heritable high-glucosinolate chemotype. To test whether the high-glucosinolate chemotype is linked to the T-DNA insert, we backcrossed line *gcc7* to wild type and analyzed the F₁ and selfed F₂ progeny for glufosinate resistance and glucosinolate accumulation. All analyzed F₁ progeny (12 plants) were glufosinate-resistant and accumulated about 50% more total glucosinolates than wild-type plants (data not shown). The backcross F₂ generation segregated in a 3:1 ratio for both glufosinate resistance (96 resistant plants:32 susceptible plants; chi-square test, *P* = 1.0) and the high-glucosinolate chemotype (10 plants with high glucosinolate content: two plants with wild-type levels; one-tailed binominal test, *P* = 0.75, *P* = 0.39). The T-DNA was present in all mutant plants but absent in the wild-type segregants. These data are consistent with the presence of a single T-DNA insert conferring glufosinate resistance and a semi-dominant high-glucosinolate chemotype. We further performed Southern analysis on genomic DNA prepared from *gcc7* plants and digested with the restriction enzymes available for plasmid rescue in the activation-tagging vector. Hybridization of a T-DNA-specific gene probe to only one

fragment of each DNA digest revealed a single T-DNA insertion event for line gcc7; there was no indication for rearrangement of the T-DNA insert (data not shown). The dominance of the high-glucosinolate chemotype indicated that the mutation most likely resulted from increased gene expression driven by multiple 35S enhancers in the single T-DNA insert.

5 Identification and cloning of IQD1

[0103] We recovered genomic DNA fragments flanking the T-DNA in line gcc7 by plasmid rescue, which we analyzed by restriction mapping and DNA sequencing of 1.4 kb across the left T-DNA border. Figure 1(a) shows the structure of the T-DNA insertion site in the genomic region covered by BAC F11F8 of chromosome 3. The T-DNA inserted into
 10 putative exon 4 of gene At3g09720, which is predicted to encode an RNA helicase. Several gene candidates encoding proteins of unknown or hypothetical function are located adjacent to the left and right border of the T-DNA insert. Analysis of messenger RNA (mRNA) expression by semiquantitative RT-PCR demonstrated a robust overexpression of At3g09710, located 3.8 kb next to the left T-DNA border, in line gcc7 relative to the wild-type control
 15 (Figure 1a). On the contrary, steady-state mRNA expression levels of the distal genes, At3g09700 and At3g09730, were not affected in line gcc7 and indistinguishable from wild-type levels, while expression of At3g09720 was only modestly reduced. These data suggest that the high-glucosinolate chemotype of line gcc7 is caused by constitutive overexpression of At3g09710, which is hereafter referred to as *IQD1* (*IQ-DOMAIN 1*, see below).

20 Verification of IQD1 function in glucosinolate accumulation

[0104] To evaluate a proposed function of *IQD1* in glucosinolate accumulation, we analyzed five publicly available insertional mutations and generated transgenic recapitulation as well as antisense suppression lines. As expected, T-DNA insertions into At3g09720 (line SALK_042728) does not affect glucosinolate accumulation (Figure 1b; Table 2), which is
 25 also true for line SALK_008887 (data not shown). By contrast, glucosinolate concentrations are significantly reduced for *IQD1* promoter insertion lines SGT16820 and GT6935 (*Ler* accession), as well as for line ABB8 (Ws-0 accession), which harbors a T-DNA insert in the fourth exon of the predicted *IQD1* open reading frame, and for an antisense suppression line (Figure 1b; Table 2). When compared with the respective wild-type accessions, methionine-
 30 derived glucosinolates are decreased by as much as 75%, whereas tryptophan-derived glucosinolates are reduced to a lesser extent (up to 40%). The reduction in glucosinolate content correlated with reduced (GT6935) or undetectable (ABB8) levels of *IQD1* transcripts. Finally, we generated recapitulation lines (*Pro*_{35S}:*IQD1*) that accumulate 65-125% more total

glucosinolates (Figure 1b; Table 2). This observation supports genetic data indicating that the high-glucosinolate chemotype of line gcc7 is caused by overexpression of *IQD1* (At3g09710) and not by gene disruption of At3g09720. Thus, the analysis of loss- and gain-of-function alleles suggests that *IQD1* encodes a modifier of glucosinolate accumulation.

- 5 [0105] Although *IQD1*-overexpressing plants are smaller than wild type at an early growth stage (Figure 2a) they are very similar to wild type in appearance when mature (data not shown). Loss-of-function *IQD1* alleles in the *Ler* background (GT6935 and SGT16820) are slightly stunted (Figure 2c-f), which is not observed for line ABBI in the Ws-0 background (Figure 2b; data not shown). When compared with wild-type seeds, total glucosinolate
10 content of gcc7 seeds is about twofold higher, indicating that the high-glucosinolate chemotype of gcc7 is independent of the developmental stage and caused by increased glucosinolate production (data not shown).

Predicted properties of the encoded *IQD1* protein

- [0106] Conceptual translation of the *IQD1* open reading frame predicts expression of a
15 basic protein (pI of 10.4) with a molecular mass of 50.5 kDa (454 amino acid residues). We constructed a full-length *IQD1* cDNA (AY827468) and confirmed the predicted *IQD1* (At3g09710) gene model (see Experimental procedures). A BLAST search with the primary structure of *IQD1* as the query identified 11 closely related (E -value: $< e^{-20}$) hypothetical proteins: five in Arabidopsis, one in sunflower (SF16; Dudareva et al., *Plant Physiol.*,
20 106:403-404 (1994)), and five in rice. These proteins share similar molecular masses (33.6-59.6 kDa) and isoelectric points (9.2-10.9). In addition to their basic nature (Arg/Lys content of 13-20%), *IQD1*-related proteins are Ser/Thr-rich (14-21%). The amino acid and nucleotide sequence identity among the 12 genes varies from 22 to 60% and 40 to 67%, respectively, indicating that *IQD1*-like genes are quite divergent. BLAST searches with lower
25 stringency suggest the presence of 33 and at least 28 related genes in the genomes of *A. thaliana* and *Oryza sativa*, respectively (M. Levy and S. Abel, unpublished data).

- [0107] As shown in Figure 3, sequence similarity among *IQD1*-related polypeptides is primarily limited to their central portion, which contains three segments (I-III) of high amino acid sequence conservation. Secondary structure predictions suggest that the three segments
30 contribute to α -helical folds. The 12 *IQD1*-like proteins are predicted to consist primarily of random coil structures (46-63%), α -helices (30-45%), and to a lesser extent of β strands (5-12%). Multiple repeats of three different classes of recognition motifs for calmodulin

interaction are present in segment I and separated by invariant spacing. This segment of 67 amino acid residues contains two to three copies of the IQ motif, IQxxxRGxxxR, or of its more relaxed version [I,L,V]QxxxRxxxx[R,K]. The IQ motif was initially thought to mediate Ca^{2+} -independent retention of calmodulin, although the motif was subsequently found to be present in some proteins that interact with calmodulin in a Ca^{2+} -dependent manner (Bähler and Rhoads, *FEBS Lett.*, 513:107-113 (2002)). Segment I also contains two repeats of the 1-8-14 motif as well as three repeats of the 1-5-10 motif at conserved positions, which partially overlap with the IQ motifs (Figure 3). The 1-5-10 and 1-8-14 motifs mediate Ca^{2+} -dependent calmodulin binding and were named based on the conserved spacing of hydrophobic residues (Choi et al., *J. Biol. Chem.*, 277:21630-21638 (2002); Rhoads and Friedberg, *FASEB J.*, 11:331-340 (1997)). Another hallmark of IQD1-related proteins is the presence of clusters of basic amino acid residues that satisfy structural characteristics of three classes of nuclear localization signals. As indicated in Figure 3, the N- and C-terminal regions of IQD1-related proteins contain several basic clusters that conform to the SV40-type, MAT α 2-type, and bipartite type of nuclear localization signals (Abel and Theologis, *Plant J.*, 8:87-96 (1995)). Thus, the prospect arises that IQD1 binds to calmodulin and functions in the cell nucleus.

Recombinant IQD1 binds to calmodulin

[0108] To test whether recombinant IQD1 binds to calmodulin, we expressed an epitope tagged T7-IQD1 fusion protein in *Escherichia coli*. Crude extract from induced bacterial cells expressing T7-IQD1 was used in pull-down assays with bovine calmodulin-agarose beads in the presence of Ca^{2+} or in the absence (+ EGTA) of Ca^{2+} . After co-incubation of the calmodulin-agarose beads with bacterial extract, the beads were repeatedly washed, and bound proteins were eluted by suspension in sample loading buffer. Proteins of all fractions were separated by electrophoresis, transferred to a membrane, and probed with T7-Tag antibody to detect the T7-IQD1 fusion protein. As shown in Figure 4, T7-IQD1 co-sedimented with calmodulin-agarose beads only in the presence of Ca^{2+} but did not co-sediment when the incubation mix and wash buffer were supplemented with EGTA. We used recombinant T7-IAA3 as a negative control (Abel and Theologis, *Plant J.*, 8:87-96 (1995)), which did not bind to calmodulin-agarose beads (data not shown). Thus, our data suggest Ca^{2+} -dependent, but T7 epitope-independent, calmodulin binding of T7-IQD1 *in vitro*.

IQD1-GFP is targeted to the cell nucleus

[0109] To study the subcellular localization of IQD1 *in vivo*, we generated transgenic Arabidopsis lines that express GFP and an IQD1-GFP fusion protein under control of the constitutive 35S CaMV promoter. As expected, histochemical analysis revealed localization of the authentic GFP protein in the cytosol and cell nucleus, whereas the chimeric IQD1-GFP fusion protein accumulated mainly in the nucleus and was largely excluded from the cytosol (Figure 5). Thus, our data demonstrate the potential for nuclear localization of an IQD1-GFP fusion protein when stably expressed in Arabidopsis and therefore suggest the cell nucleus as a cellular compartment of IQD1 function.

IQD1 modulates expression of glucosinolate pathway genes

[0110] We used reverse transcriptase (RT)-mediated PCR to monitor steady-state mRNA levels of several genes encoding enzymes involved in glucosinolate metabolism (Figure 6). Several of these genes are members of closely related families and are expressed at low levels (e.g., CYP79s), which precluded RNA blot analysis. Both semi-quantitative and real-time RT-PCR were performed on RNA isolated from shoots of 14-day-old seedlings, and relative mRNA levels were compared between Columbia wild type, line *gcc7*, and the *Pro*_{35S}:*IQD1* recapitulation line. Relative to wild type, steady-state mRNA levels for genes with roles in Trp-derived glucosinolate biosynthesis (*CYP79B2*, *CYP79B3*, and *CYP83B1*) were elevated in line *gcc7* (2- to 16-fold), most notably for *CYP79B3* and *CYP83B1* (Figure 6a,c).

However, expression levels of genes encoding enzymes related to the biosynthesis of Met-derived glucosinolates (*CYP79F1* and *CYP79F2*) and glucosinolate degradation (myrosinase-encoding *TGG1*; Husebye et al., *Plant Physiol.*, 128:1180-1188 (2002)) were appreciably decreased (10-25% of wild-type levels), whereas expression of *UGT74B1* was only twofold reduced (Figure 6a,c). The same tendency of altered gene expression was observed in the *Pro*_{35S}:*IQD1* recapitulation line (Figure 6b,c). Thus, our data suggest that overexpression of *IQD1* deregulates at least a subset of glucosinolate pathway genes.

Tissue-specific expression of IQD1

[0111] Tissue-specific expression of *IQD1* was tested by semi-quantitative RT-PCR with RNA extracted from 21 to 28-day-old plants and was found to be present in the all major organs tested (Figure 7a). In our search for insertional loss-of-function *IQD1* alleles, we identified one gene trap line of the CSH collection, GT6935, in which the bacterial β -glucuronidase gene *uidA* under control of a minimal promoter is inserted 101 bp upstream of the predicted translational start codon of *IQD1*. Therefore, although the transcriptional start

site of *IQD1* has not been determined, line GT6935 likely reports authentic *IQD1* promoter activity. Histochemical analysis of GUS activity in up to 4-week-old GT6935 seedlings revealed reporter gene expression exclusively in the vascular bundles of hypocotyls, leaves, stems, flowers, and roots (Figure 7b-v). In 6-day-old germinating seedlings, GUS expression is detected in the shoot apical meristem, at the branching zone of the vascular tissue beneath the meristem (Figure 7b,c), and in the vasculature of the hypocotyls and roots (Figure 7d). At day 9 (Figure 7e-g) and day 14 (Figure 7h,i), GUS expression is clearly detectable in the vascular tissues of rosette leaves and of the root system; weaker expression is observed in the cotyledons. The highest level of *IQD1* promoter activity was observed in fully expanded rosette leaves and shoot meristems of non-flowering seedlings. GUS expression in roots was excluded from the elongation zone (Figure 7g). *ProIQD1::GUS* is also expressed at the receptacle of the flower and silique, in developing seeds, seed funicles, the pistil and stamen filaments (Figure 7j-n). During the transition from vegetative to reproductive development, *ProIQD1::GUS* expression gradually diminishes in rosette leaves but is clearly detectable in the developing inflorescence and flower buds; expression in the root system appears unaffected (Figure 7o-v). It is interesting to note that the observed tissue-specific pattern of *IQD1* promoter activity is strikingly similar to the expression patterns reported for several genes that encode enzymes of the glucosinolate synthesis pathway, *CYP79F1*, *CYP79F2* (Reintanz et al., *Plant Cell*, 13:351-367 (2001); Tantikanjana, *Genes Dev.*, 15:1577-1588 (2001)), *CYP79B2* (Mikkelsen et al., *J. Biol. Chem.*, 275:33712-33717 (2000)), or *UGT74B1* (Grubb et al., *Plant J.*, 40:893-908 (2004)).

IQD1 overexpression deters insect herbivory

[0112] We occasionally observed that loss-of-function *iqd1* mutants are more susceptible to insect attack when grown in soil (see Figure 2f), which prompted us to analyze the effect of altered *IQD1* gene expression on insect herbivory in more detail. Herbivory-induced tissue damage causes glucosinolate degradation and release of bioactive products that may act as attractants and oviposition stimulants for specialist or as repellents against generalist insects (Giamoustaris and Mithen, *Ann. Appl. Biol.*, 126:347-363 (1995); Kliebenstein et al., *Genetics*, 161:325-332 (2002b); Lambrix et al., *Plant Cell*, 13:2793-2807 (2001)). In a first set of experiments, we examined and compared the weight gain of newly hatched larvae of the cabbage looper (*Trichoplusia ni*), a generalist lepidopteran whose larvae feed on cruciferous and many other plant species (Shorey et al., *Ann. Entomol. Soc. Am.*, 55:591-597 (1962)). We observed that *T. ni* larvae developing on line *gcc7* were significantly smaller

and gained 25% less fresh weight than larvae on wild-type (Col-0) plants, which were appreciably more damaged by herbivory than *gcc7* plants (Figure 8a). We did not observe a significant weight difference between *T. ni* larvae grown on ABB8 and control plants (Ws-0). However, the fresh weight of larvae developing on both Ws lines was about 35% higher than the weight of larvae developing on Col-0 wild-type plants (Figure 8a), which may be explained by the different glucosinolate composition and the different types of glucosinolate hydrolysis products formed in both accessions. It has been reported that nitrile-producing *A. thaliana* accessions such as Ws-0 are more susceptible to *T. ni* herbivory than isothiocyanate-producing lines such as Col-0 (Lambrix et al., *Plant Cell*, 13:2793-2807 (2001)).

[0113] In a second experiment, we investigated the nymph deposition preference of the green peach aphid (*Myzus persicae*), a generalist phloem-feeding pest with hosts in over 40 plant families (Pollard, D.G., *Bull. Entomol. Res.*, 62:631-714 (1972)). In dual-choice assays, we placed individual late instars of alate aphids, which mature into winged adults, onto the surface of agar plates containing two seedlings of different glucosinolate chemotypes (either wild type Col-0 and *gcc7*, or wild type Ws-0 and ABB8) and scored host plant preference through an analysis of nymph deposition. Our data show that green peach aphids consistently avoided the line with the higher glucosinolate content. For each of the two choice arrangements, about two-thirds of the aphids preferred the lower glucosinolate line for reproduction, i.e., wild type Col-0 and *IQD1* loss-of-function line ABB8 (Figure 8b). In summary, our data demonstrate that overexpression of *IQD1* significantly reduces herbivory by chewing and phloem-feeding insects.

IQD1 is induced by mechanical stimuli and aphid infestation

[0114] Treatment of Arabidopsis plants with JA, SA, and ethylene leads to the accumulation of specific glucosinolates (Brader et al., *Plant Physiol.*, 126:849-860 (2001); Cipollini et al., *Mol. Ecol.*, 13:1643-1653 (2004); Mikkelsen et al., *Plant Physiol.*, 131:298-308 (2003)). As overexpression of *IQD1* causes elevated glucosinolate levels, we examined whether *IQD1* mRNA expression is responsive to JA, SA, the ethylene precursor ACC, or IAA. None of the plant hormones tested appreciably altered steady-state *IQD1* mRNA levels when externally applied (Figure 9a), which was confirmed by quantitative RT-PCR for plants treated at different developmental stages (data not shown). Induction of *PR1* by SA and of *PDF1.2* by JA and ethylene confirmed effective hormone treatment (Figure 9a). As shown in Figure 9(b), *IQD1* mRNA levels are also unaffected in various mutants defective in hormone synthesis or signaling (JA – *jar1*, *fad3-2 fad7-2 fad8*; SA – *NahG*, *npr1*, *cpr1*; and ethylene –

ein2), suggesting that *IQD1* expression is independent of these hormones. However, we noticed that mock spraying of seedlings with the solvent control or water alone caused a modest, about twofold elevation of *IQD1* mRNA levels, which was also observed after mechanical wounding of leaves, or after infestation of seedlings with green peach aphids (Figure 9c). Increased *IQD1* expression as a result of aphid herbivory is likely a consequence of both mechanical stimulation and responses to chemical cues generated during phloem feeding.

DISCUSSION

IQD1 is a modifier of glucosinolate accumulation

[0115] We used a series of constructed and publicly available loss-and gain-of-function *IQD1* alleles in three different accessions (Col-0, *Ler*, and Ws-0) to probe *IQD1* gene function in glucosinolate metabolism. Our data show that *IQD1* mRNA expression correlates positively with glucosinolate accumulation in *A. thaliana* (Table 2; Figure 1). The original T-DNA activation-tagged line, *gcc7*, and generated recapitulation lines that constitutively express a *Pro_{35S}:IQD1* transgene in the Col-0 background, accumulate about twice as much total glucosinolates as the wild-type control. In contrast, several lines with T-DNA insertions in the predicted promoter and coding region of *IQD1*/At3g09710 (*Ler*, Ws-0) as well as *Pro_{35S}:IQD1* antisense suppression lines (Ws-0) showed a significant reduction of total glucosinolate content (by 45-70%). Two insertional alleles of At3g09720, the predicted RNA helicase-encoding gene that is disrupted by the T-DNA activation tag in line *gcc7*, do not affect glucosinolate levels, which is consistent with the observed dominance of the *gcc7* glucosinolate chemotype. Thus, our genetic analysis demonstrates that *IQD1* functions as a positive regulator of glucosinolate accumulation.

[0116] Such a role for *IQD1* is further indicated by its expression patterns during plant development (Figure 7). The observed tissue-specific expression of *IQD1* conspicuously overlaps with the expression patterns of glucosinolate pathway genes for which promoter activity has been histochemically analyzed in Arabidopsis: *CYP79F1*, *CYP79F2* (Reintanz et al., *Plant Cell*, 13:351-367 (2001); Tantikanjana, *Genes Dev.*, 15:1577-1588 (2001)), *CYP79B2* (Mikkelsen et al., *J. Biol. Chem.*, 275:33712-33717 (2000)), and UGT7481 (Grubb et al., *Plant J.*, 40:893-908 (2004)). Furthermore, the temporal shift of tissue-specific *IQD1* expression during the transition to flowering and the pronounced *IQD1* promoter activity in vascular tissues and flower stalks correspond to major sites of glucosinolate biosynthesis or accumulation in Arabidopsis. Consistent with a proposed function of glucosinolates in plant

defense against herbivores and pathogens, the reproductive organs, including developing inflorescences, flowers, siliques and seeds, accumulate the highest concentrations of glucosinolates, followed by young leaves, the root system, and fully expanded leaves (Brown et al., *Phytochemistry*, 62:471-481 (2003); Du and Halkier, *Phytochemistry*, 48:1145-1150 (1998); Koroleva et al., *Plant Physiol.*, 124:599-608 (2000); Petersen et al., *Planta*, 214:562-571 (2002)). Tracer studies demonstrated *de novo* synthesis of glucosinolates in reproductive organs (Du and Halkier, *Phytochemistry*, 48:1145-1150 (1998)), phloem transport from mature leaves to inflorescences and fruits (Brudenell et al., *J. Exp. Bot.*, 50:745-756 (1999); Chen et al., *Plant Physiol.*, 127:194-201 (2001)), and glucosinolate turnover during seed germination and early seedling development (Petersen et al., *Planta*, 214:562-571 (2002)). In summary, *IQD1* gene expression correlates with glucosinolate accumulation levels and with prominent sites of glucosinolate metabolism in Arabidopsis.

IQD1 encodes a novel calmodulin-binding nuclear protein

[0117] We demonstrated Ca^{2+} -dependent binding of recombinant IQD1 to bovine calmodulin *in vitro* (Figure 4) and nuclear localization of an *IQD1*-GFP fusion protein in transgenic Arabidopsis plants (Figure 5). These experiments were prompted by the presence of putative calmodulin interaction motifs as well as nuclear localization signals in *IQD1* and closely related proteins, which are the only structural features to suggest a potential biochemical role for *IQD1* (Figure 3). Three major classes of calmodulin recruitment motifs are currently known: two related motifs, termed 1-5-10 and 1-8-14, are typified by their spacing of hydrophobic and basic amino acid residues and bind calmodulin in a Ca^{2+} -dependent fashion, whereas the IQ-motif mediates association with calmodulin in a Ca^{2+} -independent manner. However, not all characterized calmodulin-binding domains contain these features (Bähler and Rhoads, *FEBS Lett.*, 513:107-113 (2002); Choi et al., *J. Biol. Chem.*, 277:21630-21638 (2002); Hoeflich and Illura, *Cell*, 108:739-742 (2002)). Conserved segment I of 67 amino acid residues, which is shared by *IQD1* and closely related proteins, contains two or three copies of each recruitment motif at invariant positions (Figure 3). This compact arrangement of multiple calmodulin-binding motifs, referred to as the IQ67 domain, is the defining feature of more distantly related proteins of unknown biochemical functions, which are members of relatively large plant-specific families (e.g., 33 proteins in Arabidopsis; M. Levy and S. Abel, unpublished data). The majority of *IQD1*-related proteins contain putative nuclear localization signals, although not at conserved positions, which may be explained by their high frequency of basic amino acid residues. In addition to *IQD1*-like

proteins, the IQ motif occurs in several plant protein families, including myosins, cyclic NMP-gated ion channels, and a class of calmodulin-binding transcription factors involved in ethylene responses. However, the structural context of the respective IQ motif(s) is different from the IQ67 domain (Bähler and Rhoads, *FEBS Lett.*, 513:107-113 (2002); Bouche et al., *J. Biol. Chem.*, 277:21851-21861 (2002); Köhler et al., *Plant J.*, 18:97-104 (1999); Reddy and Day, *Genome Biol.*, 2:RESEARCH0024 (2001); Yang and Poovaiah, *J. Biol. Chem.*, 277:45049-45058 (2002)). Binding of recombinant IQD1 to calmodulin, possibly via multiple calmodulin recruitment motifs, suggests that the biochemical activity of IQD1 is regulated by Ca^{2+} -calmodulin *in vivo*. However, the precise Ca^{2+} sensors that interact with IQD1 remain to be identified.

[0118] Calcium signaling in plants is complex and utilizes a large repertoire of sensor and regulatory target proteins. Several classes of sensor proteins bind to Ca^{2+} via a helix-loop-helix fold known as the EF-hand motif. This motif is found in more than 250 Arabidopsis proteins (Day et al., *Genome Biol.*, 3:RESEARCH0056 (2002); Yang and Poovaiah, *Trends Plant Sci.*, 8:505-512 (2003)), including six typical calmodulins and 50 calmodulin-like proteins that differ significantly in sequence and number of EF-hand motifs (McCormack and Braam, *New Phytol.*, 159:585-598 (2003)). About 200 calmodulin-binding target proteins are currently known in plants, a number that is expected to rise (Reddy and Reddy, *Phytochemistry*, 65:1745-1776 (2004)). Plant calmodulins have been identified in different subcellular locations, including the cytosol, nucleus, peroxisome, or extracellular matrix, and a diverse set of calmodulin-binding proteins is involved in a variety of cellular processes such as cytoskeleton organization, regulation of gene expression, ion transport, disease resistance, or stress responses (Yang and Poovaiah, *Trends Plant Sci.*, 8:505-512 (2003)). For example, a subset of Arabidopsis *TCH*(touch) genes that are induced by a variety of mechanical stimuli encode calmodulin and calmodulin-related proteins, which likely mediate some plant responses to the environment (Braam et al., *Physiol. Plant.*, 98:909-916 (1996); Braam et al., *Planta*, 203:S35-S41 (1997); Sistrunk et al., *Plant Cell*, 6:1553-1565 (1994)). There is increasing evidence for the generation of nucleus-specific Ca^{2+} -signatures in plant cells (Pauly et al., *Nature*, 405:754-755 (2000); Xiong et al., *Plant J.*, 40:12-21 (2004)) and for a potential regulatory role of calmodulin and other Ca^{2+} sensor proteins in nuclear processes such as transcription or gene silencing (Anandalakshmi et al., *Science*, 290:142-144 (2000); Bouche et al., *J. Biol. Chem.*, 277:21851-21861 (2002); Perruc et al., *Plant J.*, 38:410-420 (2004); Szymanski et al., *Plant Cell*, 8:1069-1077 (1996); Yang and Poovaiah, *J. Biol.*

Chem., 277:45049-45058 (2002); Yoo et al., *J. Biol. Chem.*, 280:3697-3706 (2005)).

Although IQD1 and related proteins do not contain known DNA- or RNA-binding motifs, the high isoelectric point and Ser/Thr content of IQD1-like proteins, which are reminiscent of certain splicing factors (Chaudhary et al., *Proc. Natl Acad. Sci. USA*, 88:8189-8193 (1991)), suggest that they may associate with nucleic acids and regulate transcriptional or post-transcriptional processes of gene expression.

[0119] A regulatory role of IQD1 in glucosinolate metabolism is further supported by altered steady-state mRNA levels of multiple glucosinolate pathway genes in *IQD1* gain-of-function and loss-of-function mutants. Transcript levels of genes involved in indole glucosinolate biosynthesis, *CYP79B2*, *CYP79B3*, and *CYP83B1* are elevated in T-DNA activation-tagged and *Pro_{35S}:IQD1* recapitulation lines (Figure 6). This observation is consistent with the accumulation of indole glucosinolates to higher than wild-type levels in *IQD1*-overexpressing lines. Although overexpression of *IQD1* causes enhanced accumulation of aliphatic glucosinolates, genes with roles in methionine-derived glucosinolate biosynthesis (*CYP79F1* and *CYP79F2*) and glucosinolate degradation (*TGG1*) are significantly repressed. This unexpected result may be explained by negative feedback regulation in aliphatic glucosinolate biosynthesis. Such a feedback control is suggested by an unusually high expression of *Pro_{cyp79F1}:GUS* in homozygous versus heterozygous *cyp79F1* gene trap lines (Tantikanjana, *Genes Dev.*, 15:1577-1588 (2001)). Elevated accumulation of methionine-derived glucosinolates may also result from metabolic cross-talk between the aliphatic and indolyl branch of glucosinolate biosynthesis, which is indicated by compensatory glucosinolate production in *cyp79F1/bus1* and *cyp83A1/ref2* mutants to maintain some degree of glucosinolate homeostasis (Hemm et al., *Plant Cell*, 15:179-194 (2003); Reintanz et al., *Plant Cell*, 13:351-367 (2001)).

Role of IQD1 in plant defense

[0120] Increased *IQD1* gene expression and glucosinolate accumulation correlate with enhanced resistance to generalist chewing and phloem-feeding insects, as demonstrated by cabbage looper (*T. ni*) herbivory and green peach aphid (*M. persicae*) reproduction (Figure 8). In crucifers, glucosinolates and their breakdown products are thought to be important components of the defense arsenal against herbivores and pathogens (Rask et al., *Plant Mol. Biol.*, 42:93-113 (2000)). Recent studies showed that *T. ni* herbivory is profoundly deterred by high glucosinolate levels and that glucosinolate-derived isothiocyanates are stronger feeding repellants than the corresponding nitrile products (Jander, G. et al., *Plant Physiol.*,

126:890-898 (2001); Kliebenstein et al., *Genetics*, 161:325-332 (2002b); Lambrix et al., *Plant Cell*, 13:2793-2807 (2001)). Our data are consistent with these observations. Wild-type Col-0, an isothiocyanate-producing accession (Lambrix et al., *Plant Cell*, 13:2793-2807 (2001)), is less susceptible to *T. ni* herbivory than the Ws-0 accession, which presumably
5 produces nitriles based on our previous observation that Ws-0 leaf extracts were nearly inactive in a bioassay for leaf glucosinolate-derived isothiocyanates (Grubb et al., *Plant Sci.*, 162:143-152 (2002)). Furthermore, as evidenced by 25% reduced larvae weight gain, the high-glucosinolate line, *gcc7*, is significantly more resistant to *T. ni* herbivory than the wild type (Col-0). It is interesting to note that about a similar degree of resistance to caterpillars of
10 the generalist herbivore *Spodoptera exigua* was reported (approximately 25% reduced insect growth) when Arabidopsis plants (Col-0) were treated with JA, which led to the accumulation of twofold higher total glucosinolate levels (Cipollini et al., *Mol. Ecol.*, 13:1643-1653 (2004)). Using a second Arabidopsis-herbivore interaction system, we observed that the generalist green peach aphid avoids Arabidopsis lines with higher glucosinolate content in
15 dual-choice assays of nymph deposition. Consistent with our data, Ellis et al., *Mol. Plant Microbe Interact.*, 15:1025-1030 (2002) recently reported that JA treatment of Arabidopsis plants (Col-0) retarded growth of *M. persicae* populations, which was likely caused by elevated glucosinolate content as specific glucosinolates are known to inhibit the reproduction of green peach aphids (Fraybould and Moyes, *Heridity*, 87:383-391 (2001)).

20 **[0121]** Inducible plant defenses against chewing and phloem-feeding insects involve pathogenesis-related responses that recruit Ca^{2+} -calmodulin, JA, SA, and ethylene signaling modules (Moran and Thompson, *Plant Physiol.*, 125:1074-1085 (2001); Moran et al., *Arch. Insect Biochem. Physiol.*, 51:182-203 (2002); Reymond et al., *Plant Cell*, 16:3132-3147 (2004)). Although *IQD1* mRNA expression appears to be independent of plant hormone
25 signaling, we noticed that mechanical stimuli such as touch (mock spraying), wounding, and aphid infestation cause a moderate increase of *IQD1* transcripts (Figure 9). Enhanced *IQD1* expression due to aphid feeding likely reflects a combination of mechanical stimuli and more complex responses to salivary components (Moran et al., *Arch. Insect Biochem. Physiol.*, 51:182-203 (2002)). There is increasing evidence for integrated and temporally controlled
30 cross-talk between plant response pathways involved in wounding, herbivory, and necrotic pathogen infection (Maleck and Dietrich, *Trends Plant Sci.*, 4:215-219 (1999); Moran and Thompson, *Plant Physiol.*, 125:1074-1085 (2001); Reymond et al., *Plant Cell*, 12:707-719 (2000); Reymond et al., *Plant Cell*, 16:3132-3147 (2004)). The IQ67 domain and its

arrangement of multiple calmodulin-interacting motifs are shared by members of relatively large protein families in *Arabidopsis*, rice and likely all vascular plants (M. Levy and S. Abel, unpublished data). In view of the ubiquitous functions of IQD1-like proteins and the restriction of glucosinolate biosynthesis to select species, the prospect arises that *Arabidopsis* IQD1 has broader roles in plant defense. IQD1 and possibly related calmodulin-interacting proteins participate in the decoding of Ca^{2+} -signatures elicited by biotic and abiotic challenges. IQD1 may integrate early wound- and pathogen/elicitor-induced changes in cytoplasmic Ca^{2+} concentrations (Blume et al., *Plant Cell*, 12:1425-1440 (2000); Grant et al., *Plant J.*, 23:441-450 (2000)) to stimulate and fine-tune a wide array of coordinated defense responses, including the upregulation of glucosinolate biosynthesis (Brader et al., *Plant Physiol.*, 126:849-860 (2001); Mikkelsen et al., *Plant Physiol.*, 131:298-308 (2003); Tierens et al., *Plant Physiol.*, 125:1688-1699 (2001)). Elucidation of the biochemical activities of IQD1 may provide an important impetus for understanding the roles of Ca^{2+} -calmodulin signaling in plant defense.

EXAMPLE 2

[0122] This example presents a comparative genome-wide analysis of the entire IQD gene families in *Arabidopsis thaliana* (33 loci) and *Oryza sativa* (29 loci), which are predicted to encode proteins sharing the IQ67 domain. This genomics analysis provides the framework for future studies to dissect the function of this emerging family of novel calmodulin target proteins.

[0123] To uncover the entire family of genes coding for IQD proteins in the *Arabidopsis* genome, we searched available *Arabidopsis* databases with multiple BLAST algorithms using full-length IQD1 (454 amino acids) and its IQ67 domain as the query sequences, followed by additional searches with related sequences. In addition, we performed a pattern search with the IQ motif and its degenerate versions as the query sequences and inspected each hit for the presence of an IQ67 domain. We subsequently performed pair-wise sequence comparisons to exclude redundant entries from the initial data set, which is frequently caused by multiple identification numbers of the same DNA or protein sequence in the databases. A total of 33 non-redundant putative IQD genes were extracted from these sources (Table 3). Full-length cDNA or EST sequences were available for 26 of those genes, and we attempted to clone by reverse transcriptase-mediated PCR cDNA sequences for the remaining seven genes. The average size of IQD genes in *Arabidopsis* is 2.4 kb (Table 3).

Predicted Primary Structure and Properties of *Arabidopsis* IQD Proteins

[0124] Having identified non-redundant and verified potential IQD protein coding sequences, we developed a set of criteria for the presence of the IQ67 domain in the 33 predicted *Arabidopsis* proteins. The IQ67 domain is characterized by the precise spacing of three copies of the 11-amino acid IQ motif, which are separated by short sequences of 11 and 15 amino acid residues. The first IQ motif is best conserved (present in 32 proteins), followed by the second (26 proteins) and third (12 proteins) IQ repeat. Although the third IQ motif shows the highest degree of sequence degeneration, its initial hydrophobic amino acid and following glutamine residue are present in 31 proteins. Each IQ motif is congruent with a 1-5-10 motif of hydrophobic amino acids, which again is least conserved for the last IQ motif. A fourth 1-5-10 motif overlaps the first spacer sequence and second IQ motif. Each IQ motif also partially overlaps with a 1-8-14 motif. Besides these repetitive motifs, the IQ67 domain is characterized by the presence of additional conserved hydrophobic and basic amino acid residues flanking each IQ motif. A hallmark of *IQD* genes is the presence of a phase-0 intron at an invariant position within the coding region of the IQ67 domain that disrupts codon 16 and 17 (equivalent to codon 9 and 10 of the first IQ motif). At5g03960 is the only exception to this rule, which encodes the entire IQ67 domain on its second and central exon. Given these criteria, 32 proteins contain at least two or three discernible IQ motifs with the accompanying 1-5-10 and 1-8-14 motifs in their IQ67 domain, which we therefore consider bona fide IQD proteins. The protein encoded by At5g35670 does not meet these criteria because it only contains the first, albeit truncated IQ motif provided by the N-terminal exon of the IQ67 domain (exon 2 of At5g35670). The exon coding for the remainder of the IQ67 domain (residues 17-67) is missing and replaced by an unrelated exon in At5g35670. However, the At5g35670 protein shares five common amino acid sequence motifs outside the IQ67 domain with a large set of IQD proteins as detected by comparative MEME (Multiple Expectation Maximization for Motif Elicitation) analysis (Bailey TL et al., *Proc Int Conf Intell Syst Mol Biol*, 3:21-29 (1995)) of the complete amino acid sequences of the 33 *Arabidopsis* proteins. As most of these motifs are unique to IQD proteins, we consider At5g35670 a member of the *IQD* gene family in *Arabidopsis*. Since amino acids 17-67 of the IQ67 domain are encoded by the second or third exon of *IQD* genes, the IQ67 domain contributes to the core region of most IQD proteins. An interesting exception is At3g51380, which is the smallest member of the IQD protein family in *Arabidopsis* and essentially consists only of the IQ67 domain and a short N-terminal extension of 35 amino acid residues.

[0125] Although the predicted IQD proteins are quite diverse with respect to size (103–794 residues) and computed molecular mass (11.8–86.8 kD), they appear to be remarkably uniform in terms of their relatively high theoretical isoelectric point (10.3 ± 0.6), the only exception being At1g19870 (pI of 5.2), and with respect to the abundance of Ala (8.6 ± 2.2), Ser ($12.2\% \pm 2.2\%$), and basic amino acid residues (Arg/Lys, $17.6\% \pm 2.2\%$). To uncover the possible subcellular localization of IQD proteins in *Arabidopsis*, we searched for different signature motifs specific to cellular compartments. Because of their high content of basic residues, and as suggested by PSORT, at least half of the IQD protein family (16 members) may be localized in the cell nucleus (Table 5). This conjecture is supported by the presence of several basic clusters in IQD proteins that conform to the SV40-type, MAT α 2-type, and bipartite type of nuclear localization signals (Abel S. et al., *Plant J*, 8(1):87–96 (1995)), and by the nuclear localization of an IQD1-GFP fusion protein (Levy M et al., *Plant J*, 43(1):79–96 (2005)). The remaining IQD proteins are predicted to be localized in the mitochondria (7), chloroplasts (5), or unknown compartments (Table 5).

15 Identification and Predicted Properties of the *IQD* Protein Complement in *Oryza sativa*

[0126] We next explored the occurrence and size of the *IQD* gene family in the extensively sequenced genome of rice (Goff SA et al., *Science*, 296(5565):92–100 (2002); Yu J et al., *Science*, 296(5565):79–92 (2002)). BLAST searches in several databases of *O. sativa* ssp. *japonica* and *indica* using several *Arabidopsis* full-length *IQD* protein sequences as the queries identified 29 different loci that encode non-redundant putative *IQD* proteins in rice. The general features of rice *IQD* genes and proteins are summarized in Table 4 and Table 5. Full-length cDNA sequences are available for 16 genes and generally support the respective gene model, with the exception of two loci (OS01m05259, OS03m04309) that are incorrectly annotated (see Table 4). The putative full-length cDNA sequences of two additional genes (OS01m06663, OS06m3925) are likely truncated in their coding region when compared with the conceptual translation products of each corresponding locus. A gene model could not be derived for the OS01m06368 locus in either *O. sativa* subspecies that covers the open reading frame of a corresponding partial cDNA sequence. To date, independent evidence for gene expression has been obtained for six of the remaining ten *IQD* family members for which a full-length cDNA is currently not available, suggesting that most *IQD* genes are functional in rice (Table 4). As for *Arabidopsis*, rice *IQD* genes encode 2–6 translated exons; however, less than half of the rice family members (13 genes) contain more than four exons. Furthermore, all introns in most *OsIQD* genes are in phase-0; only six genes contain a phase-

1 intron in their 3'-region and one gene (Os04m04570) is characterized by the presence of two phase-2 and one phase-1 intron in its 5'-region. Rice *IQD* genes are slightly larger than *Arabidopsis* *IQD* genes, which is a result of increased intron length (Table 5).

[0127] Conceptual translation of full-length cDNA or predicted mRNA sequences and computation of theoretical physico-chemical protein parameters reveal that the *IQD* protein complement in rice is remarkably similar to the *IQD* protein family in *Arabidopsis* (Table 4 and Table 5). Comparative MEME analysis of the complete amino acid sequences of the 28 rice and *IQD* proteins identified a similar set of conserved sequence motifs and their distribution along the polypeptide chain as found for members of the *Arabidopsis* *IQD* protein family (Table 6). The IQ67 domain is positioned close to the core region of *IQD* polypeptides and is characterized by the same hallmarks as described for the *Arabidopsis* family, including the location and spacing of the three calmodulin-binding motifs, and the position of an invariant phase-0 intron that separates codon 16 and 17 of the IQ67 domain. It is interesting to note that the rice *IQD* gene family contains members with similar deviations from consensus properties as observed for the *IQD* gene family in *Arabidopsis*. These exceptions include loss of the phase-0 intron between the IQ67 domain-coding exons (Os01m06663, OS08m00125), replacement of the second exon coding for amino acids 17-67 of the IQ67 domain (Os06m03925), C-terminal location of the IQ67 domain (OS03m00334, Os04m04570), and an unusually large and acidic protein (Os04m05532). Since the rice *IQD* proteins display a similar range of structural and physico-chemical characteristics as the *IQD* family in *Arabidopsis*, it is very likely that we have identified most of the *IQD* family members in rice. Again, the majority of the family members (16 proteins) may be targeted to the cell nucleus; the remaining *IQD* proteins are predicted to be localized in the mitochondria (4), chloroplasts (1), or unknown compartments (Table 4).

[0128] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

Table 3
The IQD gene family in *Arabidopsis thaliana*

Gene Identifier	REFSEQ Accession	Protein ID	cDNA Accession/ Protein ID	Evidence for Expression ^b	Protein Name ^c	Size (aa)	Mass (kD)	IP	Predicted Location ^d PSORT Target ^e
At1g01110	NM_099993	NP_563618	AY085363 ^a	A, C, D	IQD18	527	59.2	10.3	N ?
At1g14380	NM_101305	NP_563950	BT005935/AA064870	A-D	IQD28	664	72.8	9.7	N ?
At1g17480	NM_101610	NP_173191	AY702665	A, C, D	IQD7	370	41.0	10.5	? ?
At1g18840	NM_101741	NP_173318	AY702666	A-D	IQD30	572	62.7	9.2	N ?
At1g19870	NM_101842	NP_564097	BT001081/AA046862	A-D	IQD32	794	86.8	5.2	N C (0.65/4)
At1g51960	NM_104077	NP_175608	-	-	IQD27	351	39.3	10.1	? ?
At1g72670	NM_105926	NP_177411	BT010652/AAR07516	A, C, D	IQD8	414	45.9	10.3	N ?
At1g74690	NM_106127	NP_177607	AY128860/AA091260	A, C, D	IQD31	587	65.2	9.6	? ?
At2g02790	NM_126334	NP_178382	-	A, C	IQD29	636	69.8	9.6	N C (0.71/4)
At2g26180	NM_128176	NP_180187	BX818988	C, D	IQD6	416	46.9	10.5	N ?
At2g26410	NM_128198	NP_180209	BX840898	A	IQD4	527	58.3	10.3	? ?
At2g33990	NM_128950	NP_180946	AU237877, AV557487	A, D	IQD9	249	28.5	10.8	N ?
At2g43680	NM_180068	NP_850399	BT008408/AA037767	A, B	IQD14	668	74.3	11.3	? ?
At3g09710	NM_111805	NP_187582	AY827468	A, C, D	IQD1	454	50.5	10.4	N ?
At3g15050	NM_112367	NP_188123	BX825987	B-D	IQD10	259	29.6	10.3	? C (0.91/1)
At3g16490	NM_112520	NP_188270	BX824788	A, D	IQD6	398	48.7	10.1	? ?
At3g22190	NM_113116	NP_188858	-	A	IQD5	400	44.5	10.1	N ?
At3g29260	NM_114785	NP_566917	BT000602/AA018171	A, B, D	IQD21	471	52.1	10.0	N ?
At3g29380	NM_114798	NP_190507	-	-	IQD15	352	40.8	10.2	N ?
At3g51380	NM_114997	NP_190706	BX838271 (FL-EST)	A, D	IQD20	103	11.8	12.4	M (0.8/2)
At3g52290	NM_115089	NP_190797	BT005639/AA064059	A-D	IQD3	430	48.1	10.6	? ?
At3g59690	NM_115831	NP_191528	BT001176/AA065063	A, D	IQD13	517	58.5	10.9	? ?
At4g00820	NM_116308	NP_567191	BX826435	A, C, D	IQD17	534	60.0	10.3	? M (0.38/5)
At4g10640	NM_117132	NP_192802	BT010145/AA022614	A, D	IQD16	423	48.7	10.1	N ?
At4g14750	NM_117560	NP_193211	BX827601	A, C, D	IQD19	387	43.9	9.7	? ?
At4g23060	NM_118435	NP_194037	AY702664	A-D	IQD22	543	60.3	10.2	? M (0.50/4)
At4g29150	NM_119059	NP_194644	BT003896/AA041944	A, D	IQD25	383	41.4	10.7	? M (0.78/3)
At5g03040	NM_120382	NP_568110	AY143972/AA028911	A-D	IQD2	461	50.5	10.6	N C (0.55/3)
At5g03960	NM_120478	NP_196016	BX829656	-	IQD12	403	46.0	10.6	? M (0.76/2)
At5g07240	NM_120806	NP_196341	BT006056/AA040401	A, C, D	IQD24	401	45.3	10.3	? M (0.54/4)
At5g13460	NM_121349	NP_196850	AY128736/AA091136	C, D	IQD11	443	50.8	10.0	N ?
At5g35670	NM_122958	NP_568529	AK128736/BAD43467	C-D	IQD33	442	49.5	8.5	? M (0.47/5)
At5g62070	NM_125600	NP_201013	AY143917/AA028856	A, C, D	IQD23	403	44.3	10.5	N C (0.51/5)

^aFull-length cDNAs (asterisk denotes a cDNA clone that is likely 5'-truncated).

^b Additional evidence for IQD gene expression provided by (A) whole-genome array [Yamada K et al., *Science*, 302(5646):842-846 (2003)], (B) community microarray data (<http://www.arabidopsis.org/>), (C) Massively Parallel Signature Sequencing (MPSS, [Meyers BC et al., *Genome Res*, 14(8):1641-1653 (2004)]; <http://mpss.udel.edu/java.html>), (D) EST clones.

^cNomenclature of IQD genes is arbitrary. Levy et al. [Levy M et al., *Plant J*, 43(1):79-96 (2005)] cloned *IQD1* and reported closely related genes *IQD2-IQD6*. The designation of *IQD7-IQD33* is based on the phylogenetic analysis presented in Figure 1a.

^dN (nucleus), C (chloroplast), M (mitochondrion). Values in parenthesis indicate score/reliability class (1-5, 1 best for TargetP predictions).

Table 4
The IQD gene family in *Oryza sativa*

Gene Identifier ^a	Clone ID	(Position) ^b	Clone ID ^c	Protein ID Coded	cdNA Accession ^e	Evidence for Expression ^f	Protein Name ^g	Size (aa)	Mass (kD)	IP	Predicted Location PSORT Target ^h	
Os01m00895	A2002743	(70239-72382)	02000445	NP_914546	AK119868	B	OsIQD22	465	49.7	10.5	N	M (0.69/3)
Os01m00929	AP002746	(152586-155207)	02000453	NP_914588	AK073282	B	OsIQD5	442	48.9	10.3	?	?
Os01m04963	AP002901	(7612-9986)	02003727	NP_916574	AK102451	A	OsIQD9	441	48.2	11.0	?	M (0.55/5)
Os01m05025	AP003288	(38561-44222)	02003743	9629.m05025 ⁱ	AK062106	A, B	OsIQD24	574	63.1	9.8	N	C (0.44/5)
Os01m05259	AP003768	(95943-99625)	02003803	NP_916047 ^j	-	A, B	OsIQD20	378	42.4	10.7	?	?
Os01m06082	AP004366	(106290-110032)	02004199	BAD73780	AK072219	A, B	OsIQD2	500	56.1	10.2	N	?
Os01m06368	AP003611	(27187-28795)	02004332	BAB63799 ^k	AK120019-	-	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Os01m06663	AP003349	(15479-17371)	02004466	NP_915152	AK105622-	A	OsIQD11	563	61.7	11.5	?	?
Os02m01875	AP005534	(59894-65564)	02005830	XP_465098	AK105486	B	OsIQD13	485	52.0	10.4	?	?
Os03m00334	AC099399	(57690-58691)	02007792	XP_470188	-	B	OsIQD26	303	32.3	11.2	N	M (0.47/4)
Os03m00584	AC105729	(135566-136967)	02029613	AAN06867	-	B	OsIQD17	417	44.3	10.2	?	M (0.61/4)
Os03m04199	AC120505	(144176-145684)	02010452	XP_468989	-	-	OsIQD16	447	48.2	10.4	?	M (0.83/3)
Os03m04309	AL731878	(118442-126461)	02014260	AK067192 ^l	AAU89191	A, B	OsIQD3	440	48.7	9.6	N	?
Os03m05627	AC084296	(48853-52578)	02011159	AAT75259	AK103438	A, B	OsIQD4	422	47.0	9.8	?	?
Os04m04570	Chr.4	(27592940-27594955) ^m	02014535	9632.m04570	-	-	OsIQD27	368	41.6	11.4	N	?
Os04m04664	AL607001	(151253-153796)	02017716	XP_473550	AK100392	A, B	OsIQD15	464	50.1	10.4	N	M (0.42/5)
Os04m05332	AL606999	(85710-91604)	02015015	XP_474230	AK066310	A, B	OsIQD25	893	98.5	4.8	N	?
Os05m00240	AC093089	(81361-85365)	02015233	AAV33309	AK065809	A, B	OsIQD1	474	52.0	10.2	?	?
Os05m00863	AC093954	(45436-47015)	02015642	XP_476075	-	-	OsIQD21	497	52.6	10.4	?	M (0.57/4)
Os05m03604	AC108500	(23568-26042)	02017442	AAU90174	-	-	OsIQD6	538	57.8	9.6	N	?
Os05m04170	Chr.5	(24971333-24973284) ⁿ	02017671	-	-	-	OsIQD19	367	40.5	10.8	N	?
Os05m04307	AC097112	(54441-58756)	02017716	XP_475770	AK101555	A, B	OsIQD23	574	63.8	9.8	N	C (0.38/5)
Os05m04352	AC104713	(37570-40338)	02017731	XP_475808	AK107193	B	OsIQD10	408	44.6	10.6	?	C (0.36/5)
Os06m00539	AP004844	(90441-94332)	02018243	BAD69297	AK099462	A, B	OsIQD7	353	39.4	10.4	N	?
Os06m02303	AP003572	(18921-22322)	02019223	BAD61625	-	-	OsIQD12	470	50.0	10.6	N	?
Os06m03925	AP003944	(329-2234)	02020217	9634.m03925	AK109238*	-	OsIQD28	432	46.1	8.3	N	M (0.55/4)
Os08m00125	AP005657	(88345-90355)	02022817	XP_479772	AK100461	A, B	OsIQD14	543	59.0	11.0	?	?
Os1Om02409	ACO27662	(17834-19903)	02029613	NP_921513	AK110922	B	OsIQD18	485	52.2	10.3	N	M (0.73/3)
Os12m04168	AL732532	(133867-139274)	02035326	9640.m04168	AK102525	A	OsIQD8	442	48.2	10.1	N	?

5

^a TIGR V2 pseudo-molecules annotation.

^b Nucleotide accession of a BAC clone coding for a rice IQD gene (<http://www.ncbi.nlm.nih.gov>). The position of the I (in parenthesis).

^c Alternative rice BAC clone from *O. sativa ssp. indica* (the prefix, AAAA, is omitted); all other accessions indicate *O. sativa ssp. japonica*.

^d For three IQD genes, protein identification numbers are only available from the TIGR Rice Genome Project database (<http://www.tigr.org/tdb/e2k/osal/>).

^e cDNAs clones are full-length if not otherwise indicated. Asterisks denote cDNA sequences that are likely 5'-truncated by comparison with predicted mRNAs and encoded OsIQD proteins.

^f Additional evidence for expression provided by (A) EST clones and (B) Massively Parallel Signature Sequencing (MPSS) [Meyers BC et al., *Genome Res*, 14(8):1641-1653 (2004)], <http://mpss.udel.edu/rice/mpss.html>).

10

^g Nomenclature of OsIQD genes is arbitrary. Levy et al. [Levy M et al., *Plant J*, 43(1):79-96 (2005)] cloned *AtIQD1* and reported closely related rice genes *OsIQD1-OsIQD5*. The designation of *OsIQD6-OsIQD29* is based on the phylogenetic analysis presented in Figure 1c.

^h N (nucleus), C (chloroplast), M (mitochondrion). Values in parenthesis indicate score/reliability class (1-5, 1 best) as predicted by TargetP.

ⁱ Region of OS01m05025 is not annotated on BAC clone AP003288 as indicated for AK062106 full-length cDNA sequence on KOME website (<http://cdna01.dna.affrc.go.jp/cDNA/>).

Therefore, no PID is available and the TIGR gene model accession is given Instead.

15

^j Predicted gene model shows an N-terminal extension by 11 amino acids (possibly incorrect start codon), which was removed to meet consensus of IQD (protein N-termini for computational analysis of protein properties.

^k Predicted protein of this gene locus is shorter for both *O. sativa* subspecies than the predicted polypeptide encode by the partial cDNA clone that is truncated in the coding region N-terminal to the predicted IQ67 domain. Therefore, theoretical physico-chemical parameters of the predicted full-length protein could not be determined (n.d.).

^l Protein coding region is misannotated when compared with the predicted protein encoded by the full-length cDNA sequence.

^m BAC clone OJ1087C03 cannot be retrieved from GenBank.

ⁿ Incorrect hyperlink from gene locus to BAC clone on RiceGB website.

Table 5**Average parameters of IQD genes and proteins from *A. thaliana* and *O. sativa***

	Arabidopsis	Rice
No. of genes	33	≥29
Gene length (kb)	2.4±0.9	3.0±1.6
No. of translated exons	4.5±1.2	4.4±1.2
Protein length (residues)	454±132	471±106
Molecular mass (kD)	50.8±14.3	51.4±11.8
Isoelectric point ^a	10.3±0.6	10.4±0.6
Frequency of Arg (%) ^a	9.3±2.4	10.6±2.5
Frequency of Lys (%) ^a	8.3±2.3	5.9±2.5
Frequency of Ser (%)	12.2±2.2	10.2±1.9
Frequency of Ala (%)	8.6±2.2	12.8±3.4

^a Computation does not include Atlg19870 (pI of 5.2) and Os04m05532 (pI of 4.8).

Table 6**Major motifs in predicted *Arabidopsis* and rice IQD proteins**

Motif ^a	Multilevel consensus sequence ^b
1	EEWAAIKIQTAFRGYLARRALRALKGLVRLQALVRGHLVRKQAA MTLRCMQALVRVQAQVRR
2	MGKKGKWFKSLFGGF
3	SWFTAVKRIFISPTK
4	NKKWKLWRTSSED
5	EKRRWSFRKSS
6	PPCPPPPPHH
7	KHAIATAIATAAAEA AVAAA
8	QAAAEVVRLTS
9	SEENQALQKQLHQKHHE
10	GEDWDDSILSK
11	EEIEAKLQMRQEAAIKRERAMAYAFSHQW
12	WKNSSKTGNPTFMDP
13	DNPNWGWNWLERWMA
14	ARPWENRLMDD
15	YEENPKIVEMDTGKPYY
16	GSMNDDESFTSCPFD
17	PNYMANTESAKAKVRCQSAPR
18	SAKKRLSFPN
19	DHVKEIEEGWCDSIG
20	WMEKLTNNAFADKLLASSPTTLPLH

^a Numbers correspond to the motifs schematically presented in the reference bars of Figure 3. Motif I corresponds to the IQ67 consensus sequence. The remaining motifs are listed in the order as they occur in the primary structures of IQD proteins, continuing with motifs most N-terminal.

^b Sequences were obtained from the MEME analysis of the 61 *Arabidopsis* and rice IQD full length proteins. Only consensus sequences that are shared by at least four *Arabidopsis* IQD proteins are listed.

WHAT IS CLAIMED IS:

1 1. An isolated nucleic acid molecule comprising a heterologous promoter
2 operably linked to an *IQD1* polynucleotide sequence encoding an IQD1 polypeptide which
3 comprises an IQ76 domain as shown in SEQ ID NO: 2 and which modulates glucosinolate
4 accumulation in a plant.

1 2. The nucleic acid molecule of claim 1, wherein the IQD1 polypeptide
2 sequence is at least 90% identical to SEQ ID NO: 2.

1 3. The nucleic acid molecule of claim 1, wherein the IQD1 polypeptide
2 sequence is as shown in SEQ ID NO: 2.

1 4. The nucleic acid molecule of claim 1, wherein the *IQD1*
2 polynucleotide sequence is as shown in SEQ ID NO: 1.

1 5. The nucleic acid molecule of claim 1, wherein the promoter is a
2 constitutive promoter.

1 6. A method of increasing glucosinolate levels in a plant, the method
2 comprising introducing a recombinant expression cassette comprising an *IQD1*
3 polynucleotide sequence encoding an IQD1 polypeptide which comprises an IQ76 domain as
4 shown in SEQ ID NO: 2.

1 7. The method of claim 6, wherein the plant is a member of the genus
2 *Brassica*.

1 8. The method of claim 6, wherein the IQD1 polypeptide sequence is at
2 least 90% identical to SEQ ID NO: 2.

1 9. The method of claim 6, wherein the IQD1 polypeptide sequence is as
2 shown in SEQ ID NO: 2.

1 10. The method of claim 6, wherein the *IQD1* polynucleotide sequence is
2 as shown in SEQ ID NO: 1.

1 11. A method of decreasing glucosinolate levels in a plant, the method
2 comprising introducing a recombinant expression cassette comprising an *IQD1*
3 polynucleotide sequence comprising at least 10 contiguous nucleotides from SEQ ID NO: 1.

1 12. The method of claim 11, wherein the plant is a member of the genus
2 *Brassica*.

1 13. A transgenic plant comprising a recombinant expression cassette
2 comprising an *IQD1* polynucleotide sequence encoding an IQD1 polypeptide which
3 comprises an IQ76 domain as shown in SEQ ID NO: 2.

1 14. The transgenic plant of claim 13, wherein the plant is a member of the
2 genus *Brassica*.

1 15. The transgenic plant of claim 13, wherein the IQD1 polypeptide
2 sequence is at least 90% identical to SEQ ID NO: 2.

1 16. The transgenic plant of claim 13, wherein the IQD1 polypeptide
2 sequence is as shown in SEQ ID NO: 2.

1 17. The transgenic plant of claim 13, wherein the *IQD1* polynucleotide
2 sequence is as shown in SEQ ID NO: 1.

1 18. Seed from the transgenic plant of claim 13.

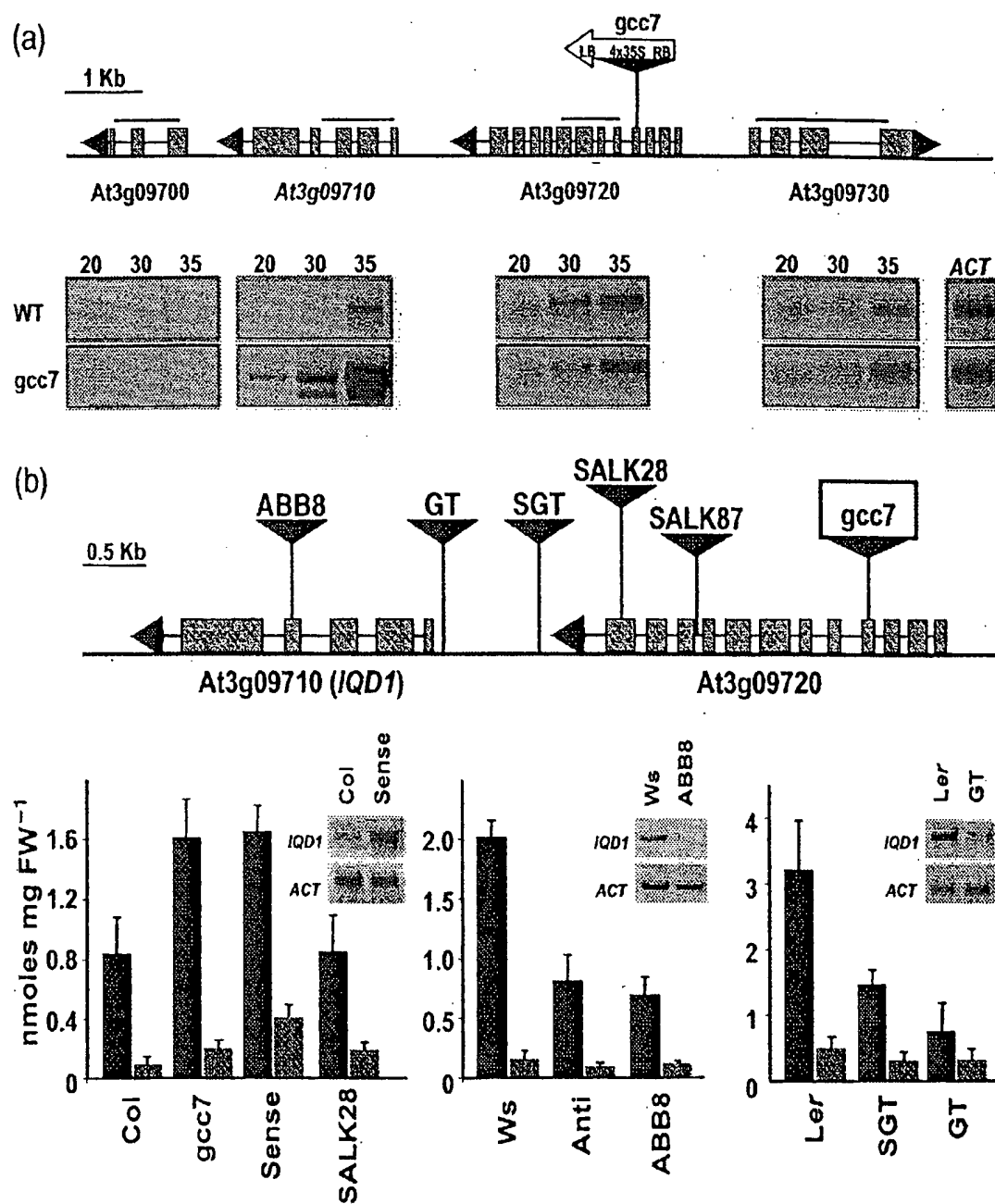
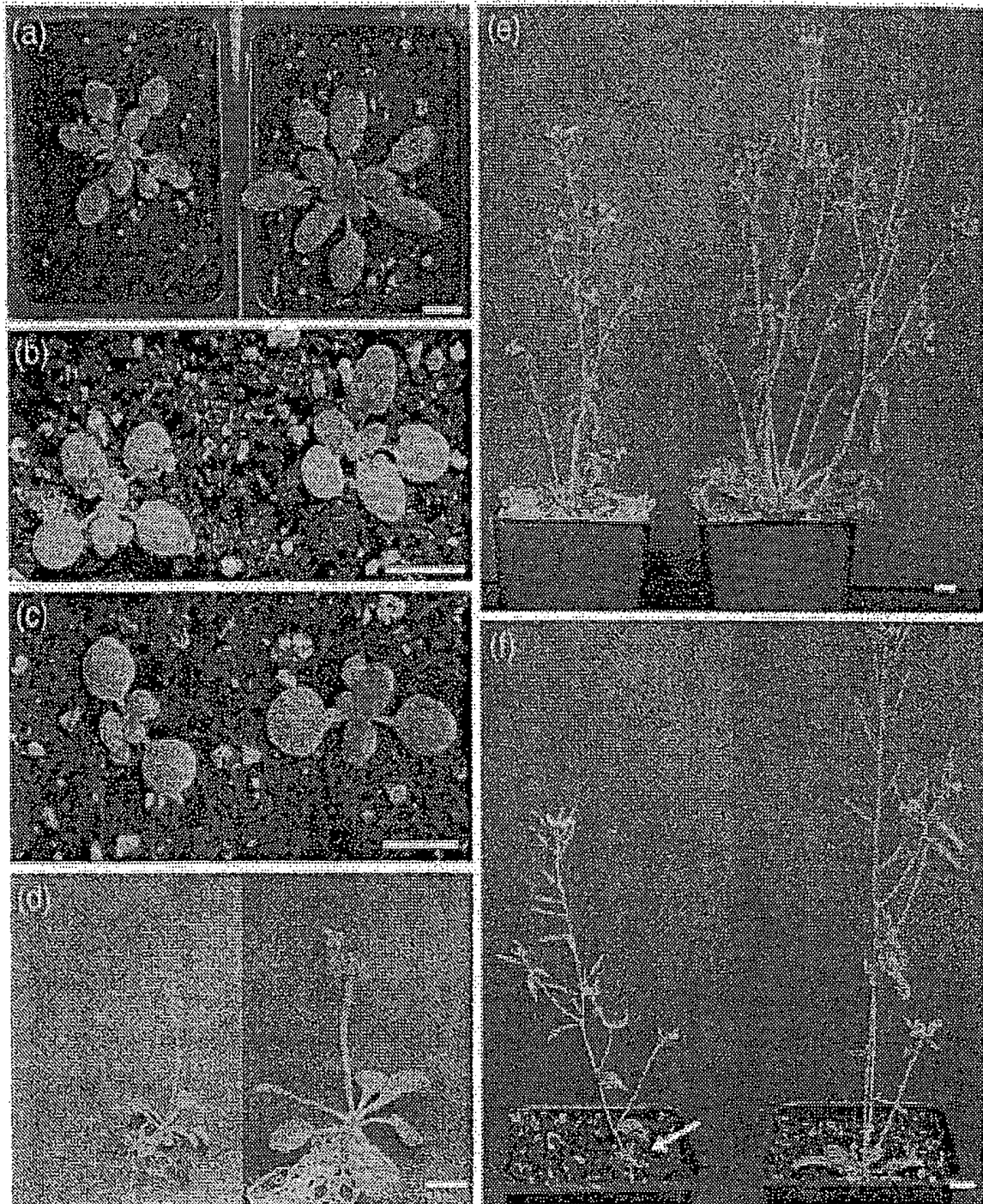


Figure 1

**Figure 2**

AtIQD1 MCKKARFSSVKA...FS...PDSK...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 60
AtIQD2 MCKKARFSSVKA...FS...PDSK...KSKQLAEGQ...NGVISNPPVDVNRQSSSSPPAL...APREVRVAEIVE... 67
AtIQD3 MCKS...FSAVKA...LSPEPKKKEQKPH...KSK...FW...FGKSKK...LDVTNSGAAYSPTVK...DAKLEIEEQ... 63
AtIQD4 MCKN...LTCSVAC...LSPG...KDKKQKPE...KPKKWSFGKQKRSRE...SFDFPLEETPP...VDPSPSSVHR...PYPPPPPLPDAFQPLLPPSPPPPPPAYINTIRYGES 101
OsIQD1 MCKKARFSAVKV...FS...PESKKEKE...RLRKLAAASNPNPPD...LTPSASLENNVVP...PPPP...PVQ...QIEEVKPEVEQE... 76
OsIQD2 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93
AtIQD5 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93
AtIQD6 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93
AtIQD6 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93
OsIQD3 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93
OsIQD4 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93
OsIQD5 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93
SF16 MCKKARFSAVKV...FSSSDPDGREAKIEKADSR...RRKWPFGKSKSDPTSTVAVPTSTAPPQPPPPFPHPIQPOPEIKOVKAETDSE... 93

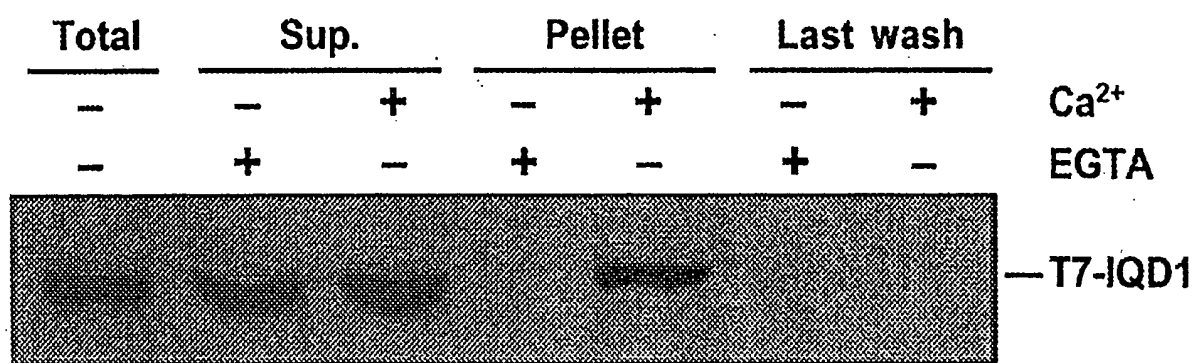
AtIQD1 .QKKNLPPS.SDSVTATVAHVLDSPSPSPESVHQAVVNRFAKSK...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 176
AtIQD2 .PNRDLSPPTADAVNVITADVPVPPSSAPGVVRRATPT.RFAGKSN...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 183
AtIQD3 .QSRHAYSAIAT...AAAAEAAVAAQAAAEVRLSALSR.FPGKSM...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 183
AtIQD4 .KESKNQALALAS...AVAAEAAVAAQAAAEVRLTTPSTPQIESK...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 216
OsIQD1 .QSKH...VTVEAPVAVPVAQTSSLP.PGVSR...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 176
AtIQD2 .QKHAVALASAVAAEAAVAAQAAAEVRLTATTAVKSPVSSKD...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 210
AtIQD5 .EDNSTRS...MVDVT.GVSTISLQSYG...GVAY...DEQS.RENRAT...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 158
AtIQD6 .KHRSUS...DGVDSST...VYSAAVATVLRAPPDFAVR...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 158
OsIQD3 .AVQSEPP...TDDNTQTVSDSISDSTLLGVHISQTEHKT...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 173
OsIQD4 .PDNGPS...EGDSNALSCSEPAFSSP...NVPVPTTEEL...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 168
AtIQD5 .HRSASS...DASDDASSAADPFTAATVATVARAPAKDMAR...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 177
SF16 .QIKH...VDVDTITTYTTVAEDVVSQAATS...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 173

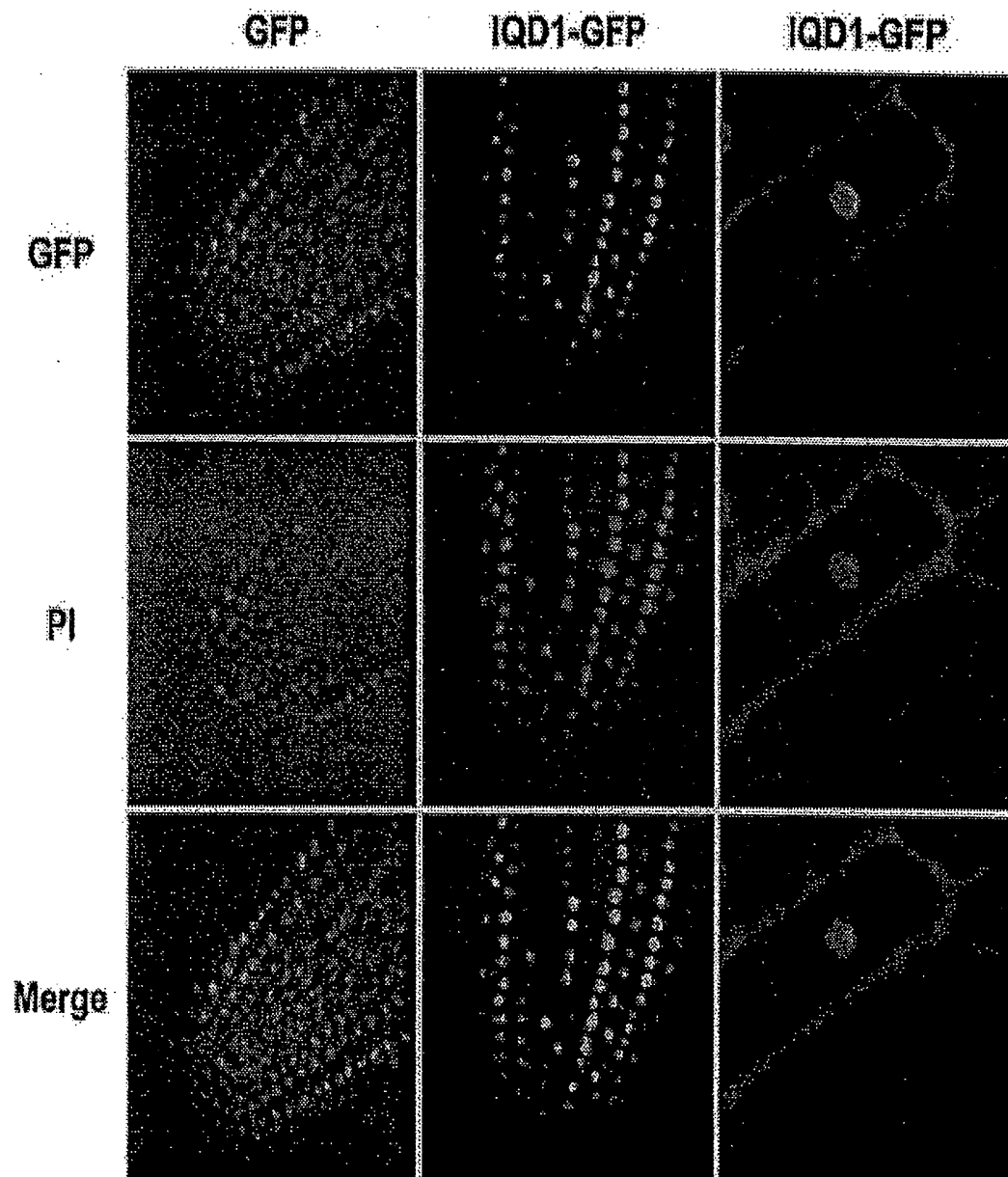
AtIQD1 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 278
AtIQD2 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 285
AtIQD3 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 271
AtIQD4 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 320
OsIQD1 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 273
OsIQD2 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 310
AtIQD5 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 259
AtIQD6 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 261
OsIQD3 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 276
OsIQD4 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 267
OsIQD5 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 293
SF16 .NARHKQLLOK.HAKELGG.LKNGG...KLEHSEVECO...DSVISYFVLATSR...SSSPQ...FEVRVDEVNYE... 267

AtIQD1 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 362
AtIQD2 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 373
AtIQD3 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 349
AtIQD4 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 425
OsIQD1 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 374
OsIQD2 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 404
AtIQD5 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 338
AtIQD6 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 335
OsIQD3 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 356
OsIQD4 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 344
OsIQD5 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 385
SF16 .QN...TTNNDS.SVKNSTNRNSQG...GETAKSSNRKLNSSTKPNTPS...ASSTA...TRNPK...KPIPPSIK.S.KSS...DDEAKS...SENRPS.IAR... 303

AtIQD1 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 454
AtIQD2 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 461
AtIQD3 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 430
AtIQD4 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 527
OsIQD1 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 474
OsIQD2 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 500
AtIQD5 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 400
AtIQD6 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 416
OsIQD3 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 440
OsIQD4 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 422
OsIQD5 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 442
SF16 .PEVS...DD.ETLSSTARRS.SNLIPT...KSAARF...KPSQ...SSRVAVTTSTTEESSILPEKAPAKRLSTASPA.PKP...RFSAPPKVEKGVKAERTP... 331

Figure 3

**Figure 4**

**Figure 5**

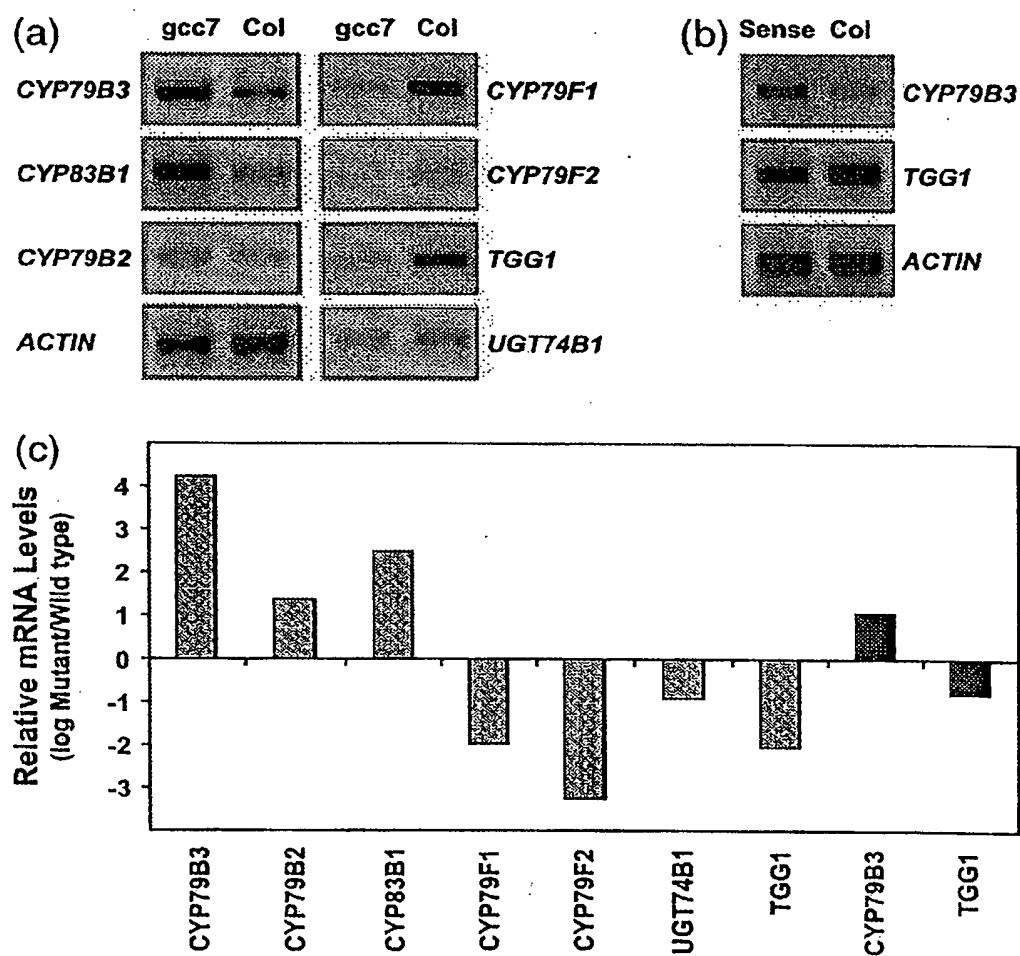
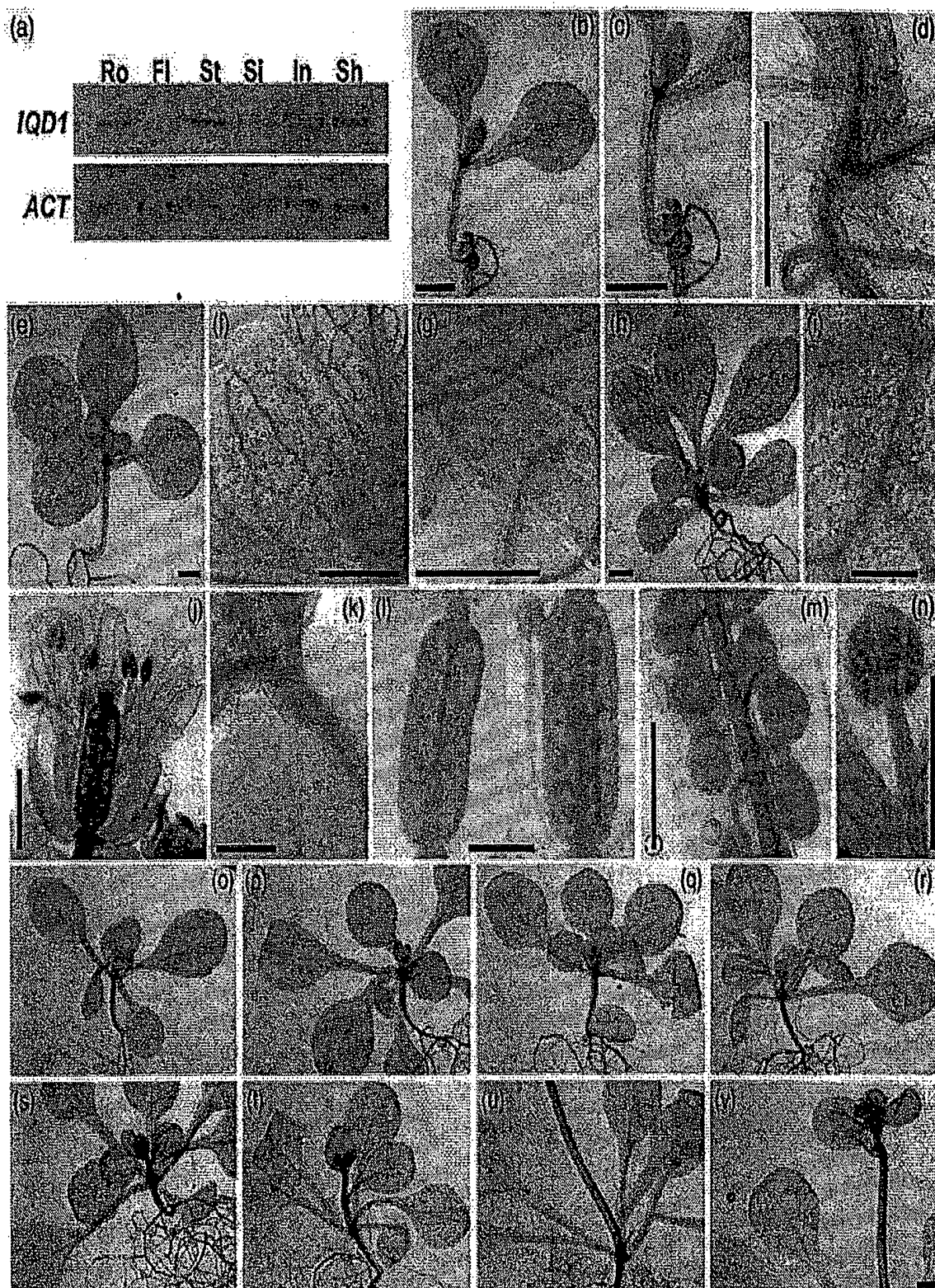


Figure 6

**Figure 7**

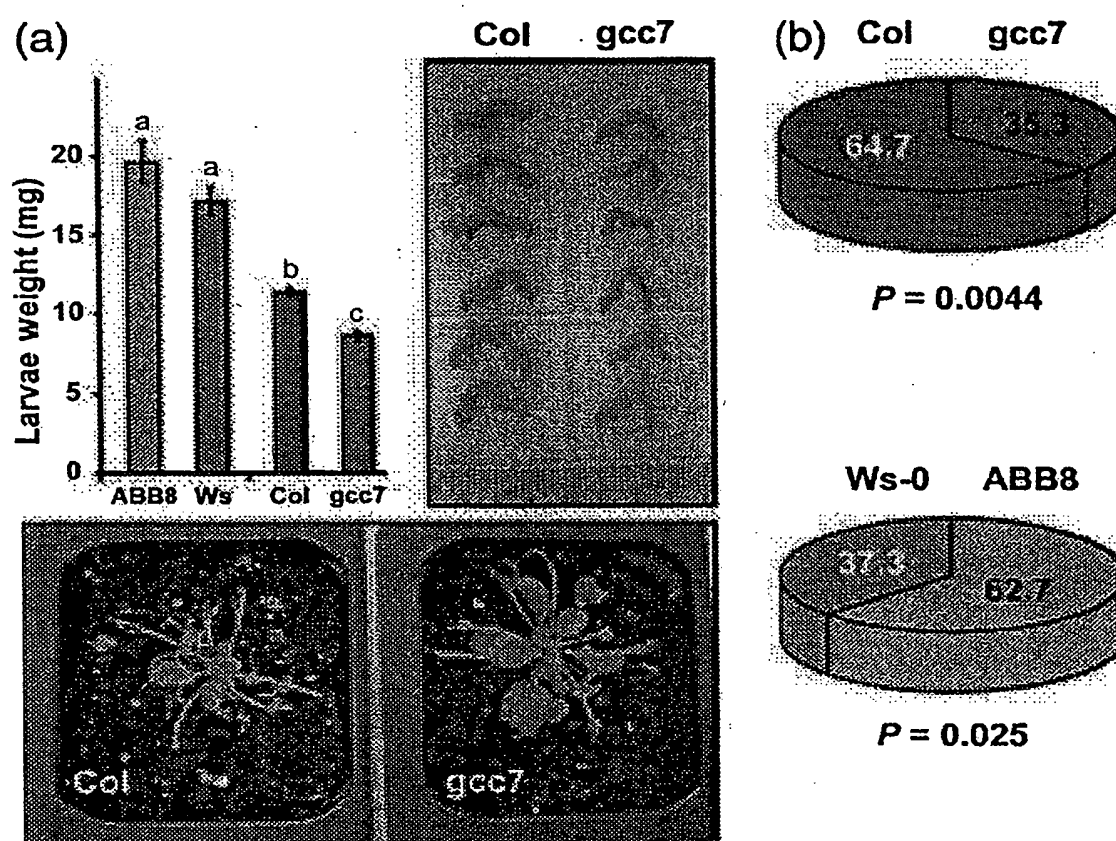


Figure 8

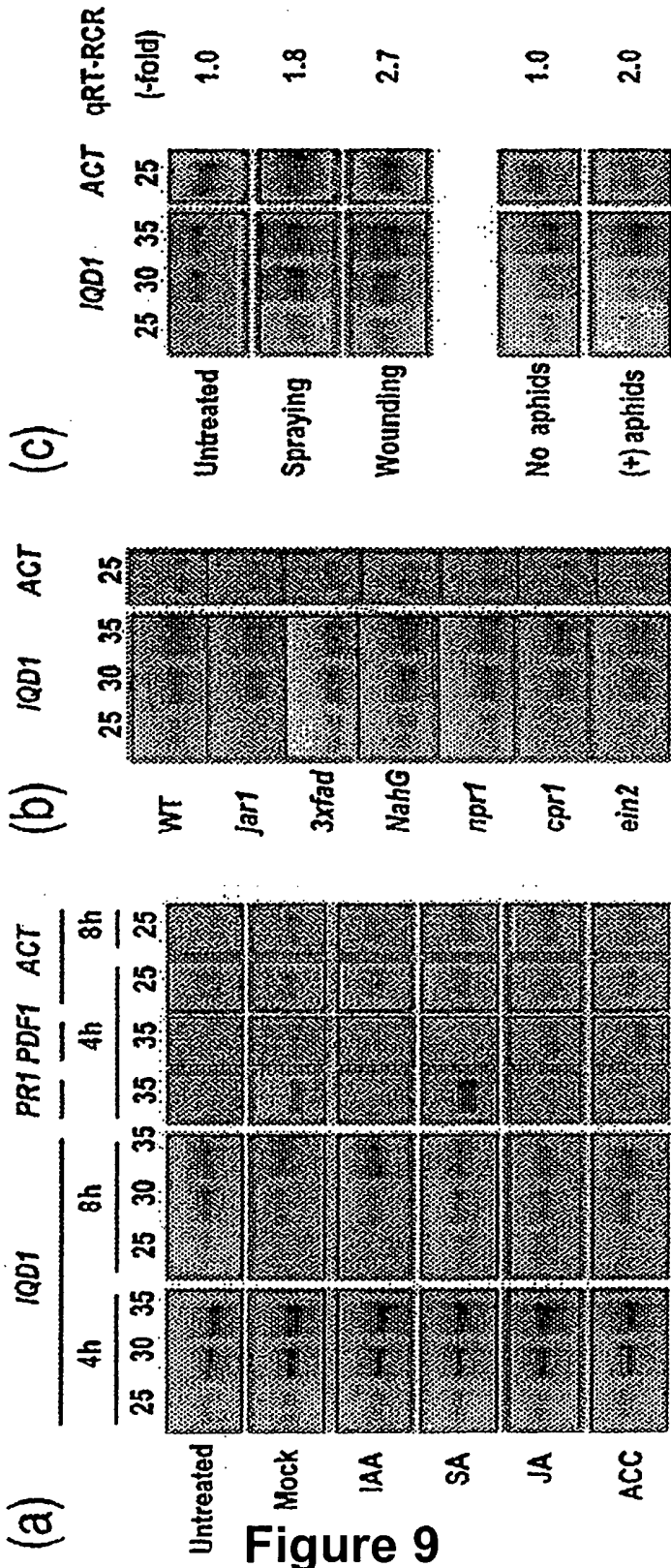


Figure 9