

(51) International Patent Classification:
C12N 15/82 (2006.01)(21) International Application Number:
PCT/US2013/040773(22) International Filing Date:
13 May 2013 (13.05.2013)

(25) Filing Language: English

(26) Publication Language: English

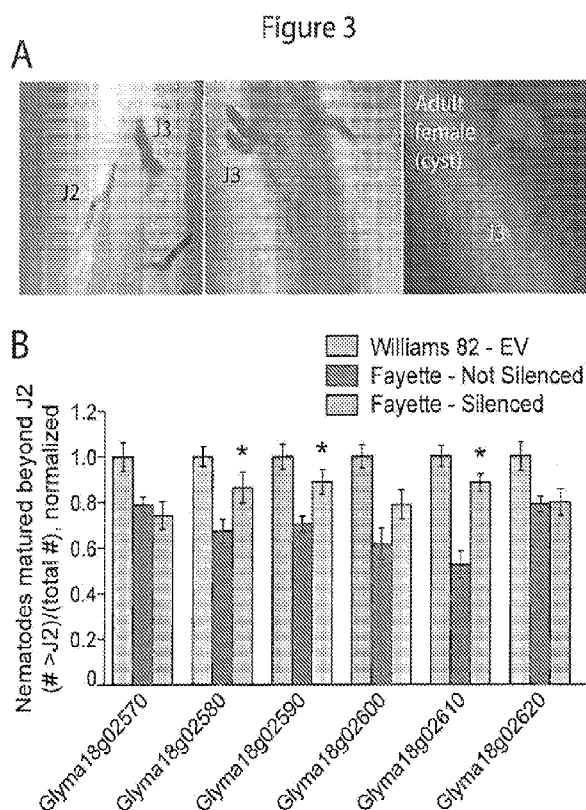
(30) Priority Data:

61/646,017	11 May 2012 (11.05.2012)	US
61/676,854	27 July 2012 (27.07.2012)	US
13/843,447	15 March 2013 (15.03.2013)	US

(71) Applicants: **WISCONSIN ALUMNI RESEARCH FOUNDATION** [US/US]; 614 Walnut Street, Madison, WI 53726 (US). **THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ILLINOIS** [US/US]; 352 Henry Administration Building, 506 South Wright Street, Urbana, IL 61801 (US).

(72) Inventors: **BENT, Andrew, Farmer**; 6214 North Highlands Avenue, Madison, WI 53705 (US). **HUDSON, Matthew**; Department of Crop Sciences, 334 NSRC, University of IL, Urbana, IL 61801 (US). **DIERS, Brian**; Department of Crop Sciences, 1101 West Peabody Drive, Urbana, IL 61801 (US). **MELITO, Sara**; Via dei Mille 61, I-07100 Sassari (IT). **COOK, David, Edward**; 5109 Flambeau Road, Madison, WI 53705 (US). **HUGHES, Teresa**; 423 N. 5th Street Apt. 1, Lafayette, IN 47901 (US). **BAYLESS, Adam**; 154 Proudfit Street, Unit 1, Madison, WI 53715 (US). **WANG, Jianping**; 1442 N.W. 104th Drive, Gainesville, FL 32606 (US). **LEE, Tong, Geon**; 310 Kenwood Road, Apt. 259, Champaign, IL 61821 (US). **GUO**,

[Continued on next page]

(54) Title: *RHG1* MEDIATED RESISTANCE TO SOYBEAN CYST NEMATODE

(57) Abstract: Methods of increasing the resistance of plants, in particular soybeans, to nematodes, in particular soybean cyst nematodes, are provided herein. The methods include increasing the expression of Glyma18g02580, Glyma18g02590 and/or Glyma18g2610 in cells of a plant and in particular in root cells of a plant to increase the resistance of the plant and plant cells to nematodes. The methods include increasing the expression using constitutive promoters or by increasing the copy number of the polynucleotides. Constructs for expressing these polypeptides, transgenic cells, transgenic plants and methods of generating the same are also provided. Methods of screening plant cells for resistance or susceptibility to nematodes are also provided.



Xiaoli; 308 Eagle Heights, Apt. J, Madison, WI 53705 (US).

(74) **Agents:** VANHEYNINGEN, Tambryn K. et al.; Andrus, Sceales, Starke & Sawall, LLP, 100 East Wisconsin Avenue, Suite 1100, Milwaukee, WI 53202 (US).

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

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

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

RHGI* MEDIATED RESISTANCE TO SOYBEAN CYST NEMATODE*CROSS-REFERENCE TO RELATED APPLICATIONS**

This patent application claims the benefit of priority of United States Provisional
5 Patent Application No. 61/646,017, filed May 11, 2012, and United States Provisional
Patent Application No. 61/676,854, filed July 27, 2012, and United States Utility
Application No. 13/843,447, filed March 15, 2013, all of which are incorporated herein
by reference in their entirety.

10 STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

This invention was made with government support under grant numbers 06-
CRHF-0-6055 and 10-CRHF-0-6055 awarded by USDA/NIFA. The government has
certain rights in the invention.

15 SEQUENCE LISTING

A Sequence Listing accompanies this application and is incorporated herein by
reference in its entirety. The Sequence Listing was filed with the application as a text file
on May 13, 2013.

20 BACKGROUND

Soybean cyst nematode (SCN) is currently the most economically damaging
disease for United States soybean production in most years. Estimates suggest that SCN
accounts for over \$700 million in reduced soybean production in the United States
annually. SCN also seriously impacts soybean production in other countries such as
25 Brazil, Argentina and China. Soybean varieties with increased resistance to SCN have
been identified, but resistance is quantitative and efficacy varies depending on nematode
genotypes, hence use of the more resistant varieties still can result in soybean yield loss
due to SCN.

The genetic basis for resistance to SCN has been partially defined, to the level of
30 genetic loci, and appropriate sources of the soybean locus *RhgI* make substantial
contributions to SCN resistance. Prior to the present work, the specific genes and gene

products controlling *Rhg1*-mediated SCN resistance have not been successfully documented.

SUMMARY

5 Methods of increasing resistance of a plant to nematodes, in particular increasing resistance of soybeans to SCN are provided herein. Several gene products from the *rhg1-b* locus are identified and the relationship of the gene products to resistance to SCN in soybeans is demonstrated. These genes and gene products may also increase resistance of other plants, including but not limited to, sugar beets, potatoes, corn, peas, or beans to
10 nematodes, in particular to cyst nematodes.

 In one aspect, methods of increasing resistance of a plant to nematodes, suitably cyst-forming nematodes, suitably SCN by increasing the expression of or altering the expression pattern or gene copy number of a polynucleotide encoding a Glyma18g02580 polypeptide, a Glyma18g02590 polypeptide, a Glyma18g02610 polypeptide, a
15 polypeptide having 90% or more identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO:6, or SEQ ID NO: 3, or a homolog or functional variant of any of the aforementioned polypeptides in cells of the plant are provided. Use of combinations of the polypeptides is envisioned. The polynucleotides encoding these polypeptide sequences may be derived from the Williams 82, PI88788 or Peking (PI 548402) soybean
20 varieties or other sources of the polynucleotides. The polypeptide sequences are provided and the polymorphisms between the sequences in different varieties are noted. Increased expression of the polynucleotides in cells of the plant increases the resistance of the plant to nematodes. Suitably expression is increased in cells of the root of the plant. Suitably expression of at least two of the polynucleotides is increased. Suitably,
25 expression of all three of the polynucleotides is increased.

 In another aspect, methods of increasing resistance of a plant to nematodes, suitably cyst-forming nematodes, suitably SCN by altering (increasing or decreasing) the expression in cells in the root of the plant of a polypeptide identical or similar to at least a portion of SEQ ID NO: 1 of Glyma18g02580, SEQ ID NO:2, 5 or 6 of Glyma18g02590
30 or SEQ ID NO: 3 of Glyma18g02610 relative to the expression in cells in the root of the plant of a polypeptide whose expression can be used as a control, such as

Glyma11g35820, are provided. Suitably expression of at least two of the polypeptides is increased. Suitably, expression of all three of the polypeptides is increased. Alternatively or in addition, expression of the polynucleotides encoding the polypeptides of Glyma18g02610, Glyma18g02590, and/or Glyma18g2580 may be increased as well.

5 In another aspect, methods of identifying plants that exhibit useful levels of resistance of a plant to nematodes suitably cyst-forming nematodes, suitably SCN by identifying plants that exhibit altered (increased or decreased) expression in cells in the root of the plant of a polypeptide identical or similar to at least a portion of SEQ ID NO: 1 of Glyma18g02580, SEQ ID NO:2, 5 or 6 of Glyma18g02590 or SEQ ID NO: 3 of
10 Glyma18g02610 relative to the expression in cells in the root of the plant of a polypeptide whose expression can be used as a control, such as Glyma11g35820, are provided. Suitably expression of at least two of the polypeptides is at a higher level than in plants that are more susceptible to SCN. Suitably, expression of all three of the polypeptides is at a higher level. Alternatively or in addition, expression of the
15 polynucleotides encoding the polypeptides of Glyma18g02610, Glyma18g02590, and/or Glyma18g2580 may be at a higher level as well.

In yet another aspect, a construct comprising a promoter operably linked to a polynucleotide encoding at least a portion of Glyma18g02580 polypeptide comprising SEQ ID NO: 1, a Glyma18g02590 polypeptide comprising SEQ ID NO: 2, 5 or 6, a
20 Glyma18g02610 polypeptide comprising SEQ ID NO: 3 or a polypeptide having at least 90% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6 or a homolog or functional portion of any of the aforementioned polypeptides or combinations thereof is provided. The construct may be used to generate transgenic plants or seeds.

25 In still another aspect, a transgenic plant comprising an exogenous or non-native polynucleotide encoding at least a portion of Glyma18g02580 polypeptide comprising SEQ ID NO: 1, Glyma18g02590 polypeptide comprising SEQ ID NO: 2, 5 or 6, Glyma18g02600 polypeptide comprising SEQ ID NO: 4, Glyma18g02610 polypeptide comprising SEQ ID NO: 3 or a polypeptide having at least 90% identity to SEQ ID NO:
30 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO:5, SEQ ID NO:6, a homolog or a functional portion of any of the aforementioned polypeptides or

combinations thereof or the polypeptides described herein from either the PI88788 or Peking-source is provided. The transgenic plant has increased resistance to nematodes, suitably cyst-forming nematodes, suitably SCN. Suitably, the transgenic plant comprises at least one polynucleotide encoding at least two or at least three of the polynucleotides encoding the Glyma18g02580, Glyma18g02590, and Glyma18g02610 polypeptides.

In a further aspect, a transgenic cell comprising a polynucleotide encoding a polypeptide capable of increasing resistance to nematodes, suitably cyst-forming nematodes, suitably SCN is provided. The polypeptide includes at least a portion of a polypeptide having at least 90% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3 or similar sequences derived from PI88788 (such as SEQ ID NO: 5) or Peking-source (such as SEQ ID NO: 6) or combinations thereof. Suitably, the polynucleotide includes at least two or three of the polypeptides having at least 90% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3.

In another aspect, methods of generating a transgenic plant by introducing an exogenous polynucleotide encoding at least a portion of a Glyma18g02580 polypeptide having at least 90% identity to SEQ ID NO: 1, Glyma18g02590 polypeptide having at least 90% identity to SEQ ID NO: 2, 5 or 6, or Glyma18g02610 polypeptide having at least 90% identity to SEQ ID NO: 3, or homologs or combinations thereof are provided. The transgenic plant has increased expression of Glyma18g02610, Glyma18g02590, and/or Glyma18g02580 in a cell in a root of the plant. The transgenic plant has increased resistance to nematodes, suitably cyst-forming nematodes, suitably SCN, as compared to a control plant. Suitably, the transgenic plant has increased expression of at least two of the polynucleotides or all three of the polynucleotides encoding the Glyma18g02610, Glyma18g02590, and/or Glyma18g02580 polypeptides.

In yet a further aspect, methods of identifying molecules that interact with the *Rhg1* locus, *Glyma18g02610*, *Glyma18g02590* and/or *Glyma18g02580* RNA transcripts, or the Glyma18g02610, Glyma18g02590 and/or Glyma18g02580 polypeptide are provided. The methods include detecting molecules capable of binding the *Rhg1* locus, *Glyma18g02610*, *Glyma18g02590* or *Glyma18g02580* RNA transcripts, or Glyma18g02610, Glyma18g02590 or Glyma18g02580 polypeptides.

In a still further aspect, methods of identifying the resistance or susceptibility phenotype of a plant to cyst nematodes are provided. The method includes detecting a genetic marker associated with cyst nematode resistance or susceptibility in a first plant cell and comparing the genetic marker in the first plant cell to the genetic marker in a second plant cell with a known resistance or susceptibility phenotype or a control plant cell. The genetic marker may be sequence variations, methylation differences, mRNA expression differences, small RNA production or other differences identified herein. Suitably, the genetic marker is associated with characteristics of the *Rhg-1* locus, such as those reported herein. Suitably, the genetic marker is the genomic copy number of at least one of *Glyma18g02600*, *Glyma18g02610*, *Glyma18g02590* or *Glyma18g02580*. Suitably the plant is a soybean and the nematodes are SCN.

In still a further aspect, methods of increasing resistance of a plant to nematodes comprising expressing a polynucleotide encoding a Glyma18g02610 polypeptide, a Glyma18g02590 polypeptide, or a Glyma18g02580 polypeptide, a polypeptide having 90% or more identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 3, or a homolog or functional variant or combinations of any of the aforementioned polypeptides in a cell. Suitably, the polynucleotide encodes at least two or all three of the Glyma18g02610, Glyma18g02590 or Glyma18g02580 polypeptides. The polypeptides or a cell encoding the polypeptide may then be applied to the plant, seeds of the plant or to soil in which the seeds may be planted. The application increases the resistance of the plant to nematodes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a pictorial depiction of the lifecycle of SCN.

Figure 2 is a pictorial depiction of one gene silencing strategy that uses artificial microRNA sequences to target a gene of interest.

Figure 3 shows a there are three genes at *rhg1-b* that contribute to SCN resistance. Figure 3A is a photograph showing representative SCN-infested roots; root vascular cylinder and nematodes stained with acid fuchsin. Fewer nematodes progress from J2 to J3, J4, adult male or egg-filled adult female (cyst) stages in SCN-resistant roots. Figure 3B is a graph showing that SCN development beyond J2 stage in transgenic roots of

soybean variety Fayette with the designated gene silenced, relative to Williams 82 (SCN-susceptible) and non-silenced Fayette (SCN-resistant) controls. Mean $\pm$ std. error of mean. *: Fayette (silenced) significantly different from Fayette (not silenced) based on ANOVA $p < 0.05$. EV: transformed with empty vector.

5 Figure 4 is a set of graphs showing that nematode development is impacted by level of silencing. Figure 4A and 4C show that nematode development on Williams 82 and Fayette roots transformed with empty vector (EV), or Fayette transformed with silencing constructs (2580RNAi or ami2590) was dependent on level of silencing. Transgenic roots with reduced target transcript abundance (+) displayed nematode
10 development similar to Williams 82 (SCN-susceptible), while transgenic roots with non-silenced transcript level (-) had nematode development similar to Fayette (SCN-resistant). Figure 4B and 4D show the transcript abundance of target genes in roots from (A) or (C) respectively, measured by qPCR. *SKP16* transcript used as reference and normalized to Fayette-EV. The results of Figure 4B and 4D were used to place roots in
15 the 'well-silenced' (+) or 'not well-silenced' (-) categories shown in Figure 4A and 4C. Figure 4A and 4B are *Glyma18g02580*. Figure 4C and 4D are *Glyma18g02610*. Bars represent mean $\pm$ std. error of mean.

Figure 5 shows a 31.2 kb repeat that elevates expression of the encoded genes is present in SCN-resistant haplotypes of the *Rhg1* locus. Figure 5A is a schematic of *Rhg1*
20 locus of Williams 82 (top), and five fosmid inserts from *rhg1-b* haplotype. DNA sequences of soybean reference genome shown for the two designated locations. Numbers and block icons refer to soybean genes (e.g., *Glyma18g02540*). Fosmids #3, 4 and 5 carry *rhg1-b* genome segments that span repeat junctions. Figure 5B shows the *Rhg1* repeat junction sequence from four different sources of SCN resistance (compare to
25 reference genome sequences in (Figure 5A)). Figure 5C is a graph showing the number of whole-genome shotgun sequencing reads corresponding to reference genome region shown in green in Figure 5A was ten-fold greater than for genome regions adjacent to *rhg1-b* on chromosome 18 or for *Rhg1*-homeologous loci on chromosomes 11 and 2. Figure 5D is a graph showing transcript abundance of genes encoded in the 31 kb repeat
30 region is much greater in roots from SCN-resistant soybean varieties relative to SCN-

susceptible varieties. Mean $\pm$ std. error of mean shown for qPCR; results for *Glyma18g02600* were at limit of detection

Figure 6 shows Fiber-FISH detection of *Rhg1* copy number variation in widely used soybean lines. Figure 6A is a schematic showing the two adjacent probes isolated from a single PI88788 (*rhg1-b*) genomic DNA fosmid clone whose insert spans a repeat junction, generating a 25.2 kb probe (green label) and an adjacent 9.7 kb probe (red label). DNA for green-labeled and red-labeled fiber-FISH probes are shown under the corresponding sequence regions of Williams 82. The 25.2 kb fragment from *rhg1-b* haplotype used for green probe was a single continuous DNA fragment that spans a repeat junction. Figure 6B shows a composite of four Fiber-FISH images (four DNA fibers) per genotype, and probe diagram. Alternating pattern of red and green hybridization on single genomic DNA fibers indicates ten and three direct repeat copies of the 31 kb block at *Rhg1* locus of SCN-resistant Fayette (*rhg1-b* derived from PI 88788) and Peking (PI 548402) respectively, and one copy per *Rhg1* haplotype in SCN-susceptible Williams 82. White bars = 10 μ m, which correspond to approximately 32 kb using a 3.21 kb/ μ m conversion rate.

Figure 7 shows that multiple SCN-resistant varieties contain the DNA junction indicative of a repeat within the *Rhg1* locus, and exhibit elevated expression of genes fully encoded within the repeat. Figure 7A is a schematic of PCR primers used in Figure 7B (see also Figure 5). Figure 7B is a photograph of a gel showing the results of PCR using outward-directed oligonucleotide primers shown in Figure 7A that match sequences at the outer edges of the 31 kb segment of *Rhg1* locus that is repeated in some soybean varieties. R indicates SCN-resistant and S indicates SCN-susceptible soybean variety. For primers 81 and 82 see Table 4. Figure 7C shows the DNA Sequence from 11 SCN-resistant varieties and reveals identical sequence for the repeat junction indicating a shared origin. Red bar indicates repeat junction (see also Figure 5). Figure 7D is a graph showing the transcript abundance for genes encoded at *Rhg1* (normalized to *SKP16*), revealing elevated expression of genes fully encoded within the repeats of *Rhg1* from PI 88788 or Peking sources, relative to expression of the same genes in SCN-susceptible varieties. Bars represent mean $\pm$ std. error of mean. *Glyma18g02600* is expressed below 0.01% of *SKP16* (CT > 35 cycles). Figure 7E is an RNA blot analysis for

Glyma18g02570 using RNA collected from roots of whole plants of Fayette and Forrest (SCN resistant) and Williams 82 (SCN susceptible). * denotes the band corresponding to the expected transcript size of *Glyma18g02570* (1.2 kb). The band at 1.8kb corresponds to non-specific ribosomal binding. Cultivars Fayette and Forrest (that contain repeats of the 31kb DNA segment) display the same banding pattern as Williams 82 (that contains a single copy of the 31kb DNA segment); no alternative transcripts for *Glyma18g02570* were detected as a result of the repeated DNA in Fayette and Forrest. RACE PCR from plants carrying *rhg1-b* confirmed full-length transcripts (with transcript ends as annotated in the reference genome) for *Glyma18g02580*, -2590 and -2610.

Figure 8 is a graph showing qPCR for genes in and outside of *Rhg1* repeat. RNA collected from roots of 3 individual plants grown in pots, 5 days post emergence. Dark gray bars are estimated to be high copy number lines based on gDNA qPCR and cDNA sequencing. Light grey bars are low copy number containing lines that also require *Rhg4* for full resistance.

Figure 9 contains example gel photographs and a table summarizing many experiments showing that resistant and susceptible cultivars have differential DNA methylation at or adjacent to the genes in the duplicated region, especially in the promoter regions. In *McrBC* experiments, methylated genomic DNA is cleaved by *McrBC*, which reduces the abundance of the PCR product, while in *HpaII* experiments, methylated genomic DNA is not cleaved by *HpaII* and it is the non-methylated DNA that is cleaved, leading to reduced abundance of the PCR product.

Figure 10 is a photograph of a Western blot showing that an epitope-tagged version of the *Glyma18g02610* protein, produced from an introduced polynucleotide in transgenic roots, is expressed in both Williams 82 and Fayette transgenic roots and the products are similar in size.

Figure 11A is a graph showing the quantitative PCR gene expression analysis for genes at the *Rhg1* locus in susceptible and resistant roots showing that some of these genes not only are more highly expressed in resistant cultivars (as is also shown in Figure 11), but also exhibit some upregulation after inoculation with SCN.

Figure 11B is a graph showing the quantitative PCR gene expression analysis following methyl jasmonate or water treatment, which reveals that *Glyma18g02610* is expressed more highly in response to elevated levels of methyl jasmonate.

Figure 12 is a set of photographs showing the histochemical staining of promoter-GUS expression in Fayette hairy root with (B, D, F and H) or without (A, C, E, and G) nematode inoculation. A and B show *Glyma18g02580*. C and D show *Glyma18g02590*. E and F show *Glyma18g02610*. G and H show *Glyma14g06080*.

Figure 13 provides the nucleotide and amino acid sequences for *Glyma18g2590* from the indicated varieties.

Figure 14 is a computer generated schematic of the three-dimensional structure of *Glyma18g2590* showing the polymorphisms among the varieties in the structure.

Figure 15 is a graph showing elevated SCN resistance conferred by simultaneous overexpression of multiple genes rather than overexpression of individual genes from the 31 kb *rhg1-b* repeat. SCN development beyond J2 stage is reported for transgenic soybean roots (variety Williams 82) overexpressing the designated single genes, or overexpressing all genes encoded within the 31 kb repeat (*Glyma18g02580*, -2590, -2600 and -2610), relative to Williams 82 (SCN-susceptible) and Fayette (SCN-resistant) controls. Mean $\pm$ std. error of mean for roots transformed with empty vector (EV) or gene overexpression constructs (OX). *: Williams 82 - OX significantly different from Williams 82 - EV based on ANOVA $p < 0.05$.

Figure 16 is a set of graphs showing that expressing the native Fayette *Glyma18g02590* allele in Williams 82 does not alter SCN development. Figure 16A is a graph showing similar nematode development on transgenic roots of Williams 82 expressing empty vector (EV) or Williams 82 expressing the Fayette (*rhg1-b*-type) allele of *Glyma18g02590* under control of Fayette *Glyma18g02590* promoter sequences (*2590_{FayP}*; *2590_{FayP}*). Williams 82 transformed with either construct allowed a greater proportion of nematodes to advance beyond the J2 stage compared to Fayette-EV. Figure 16B is a graph showing transcript abundance for *Glyma18g02590* in roots from Figure 16A, measured by qPCR. *SKP16* transcript used as reference; data normalized to Williams 82 - EV. Bars in Figure 16A and Figure 16B represent mean $\pm$ std. error of mean.

Figure 17 is a graph showing that qPCR reveals elevated transcript abundance of the intended genes in roots transformed with the multiple gene simultaneous overexpression construct of Figure 15, and no significant elevation of PR-1 expression. Transgenic roots carried either the multiple-gene construct (OX) or empty vector (EV). Similar results obtained in second independent experiment with different transgenic events, except PR-1 abundance was more similar (closer to 1.0) between Williams 82 - EV, Fayette-EV and Williams-OX roots in second experiment. Bars represent mean $\pm$ std. error of mean. Data for *Glyma18g02600* are less dependable for Williams-EV and Fayette-EV because their qPCR signal was at the limit of accurate qPCR detection (CT > 33).

Figure 18 is a set of graphs showing that overexpression of Glyma18g2580, Glyma18g2590 and Glyma18g2610 in combination can confer resistance on a susceptible Williams 82 variety.

DETAILED DESCRIPTION

Methods of identifying plants resistant or susceptible to cyst nematodes, such as the soybean cyst nematode (SCN), methods of assessing a plant's level of resistance or susceptibility to nematodes, such as SCN, methods of increasing resistance of a plant or plant cells to cyst nematodes and methods of generating transgenic plant materials, including transgenic cells and plants, are provided herein. In addition, constructs including polynucleotides encoding the *Rhg1* polypeptides described herein or homologs or variants thereof are provided herein as SEQ ID NO: 1-6. Transgenic plants or transgenic plant cells with increased resistance to cyst nematodes, particularly SCN, carrying a transgene encoding a non-native or exogenous *Rhg1* derived polynucleotide encoding the polypeptides of SEQ ID NOs: 1-6 are provided herein. Non-transgenic plants carrying the polypeptides or bred or otherwise engineered to express increased levels of the polypeptides or the polynucleotides encoding the polypeptides are also disclosed.

SCN is caused by the nematode *Heterodera glycines*. The life cycle of the nematode is shown in Figure 1. Once a field is infested with this nematode, no economically feasible means of eliminating SCN from that field presently exists. Current

management of SCN often focuses on crop rotation and planting of SCN-resistant varieties of soybeans to control *H. glycines* populations across multiple years, as well as use of SCN-resistant and/or SCN-tolerant soybeans to facilitate acceptable yield of the present year's crop. Practitioners have adopted "Race" and "Hg Type" terminologies to describe *H. glycines* populations according to their ability to overcome known sources of plant SCN-resistance. Several races and Hg Types exist and soybean resistance to one type may offer little to no protection against another type of the nematode. In addition, *H. glycines* are outcrossing organisms for which local populations are genetically heterogeneous (and new nematode genotypes can be introduced), hence local populations can undergo shifts in race or Hg Type such that previously effective plant SCN resistance can lose efficacy. Thus, the ability to identify which soybeans are resistant to which *H. glycines* nematode populations and the further ability to genetically engineer soybean plants with increased resistance to more than one type of nematode population is needed.

Soybeans with increased resistance to SCN are available and have been used in cross-breeding experiments to generate soybeans that are more resistant to SCN. The soybean *rhg1* locus of Peking was previously identified, mapped to a region of chromosome 18 (formerly known as linkage group G), and a gene at that locus encoding a product carrying leucine-rich repeats and a protein kinase domain (LRR-kinase) was hypothesized to account for the increased resistance to SCN. In plants carrying the *rhg1-b* locus derived from soybean PI88788, SCN still penetrate and initiate feeding, but a high percentage of the syncytia do not persist and undergo the full sequence of nematode development (molting through the J3 and J4 stages to adulthood, sexual fertilization, and female transition to an embryo-filled and environmentally persistent cyst) (Li, Chen et al. 2004), (Colgrove and Niblack 2008). The molecular basis of this partial SCN-resistance is not understood. Despite 50 years of research on SCN, a pathogen causing hundreds of millions of dollars of economic losses in the U.S. annually, there were no confirmed public reports of a cloned soybean SCN resistance gene prior to the priority application.

Further fine genetic mapping of the *rhg1* locus, which is also known as the *Rhg1* locus, or by other more restricted designations such as *rhg1-b*, was completed in plants carrying the PI88788 source of *Rhg1*, and new markers associated with the resistance genotype were identified. See Kim, M., D.L. Hyten, A.F. Bent and B.W. Diers, 2010.

Fine mapping of the SCN resistance locus *rhg1-b* from PI 88788. Plant Genome 3:81-89, which is incorporated herein by reference in its entirety. PI88788 was chosen because it is the source of resistance in many cross-bred lines currently marketed as resistant to SCN. These markers are tightly linked with resistance or susceptibility to SCN and may be useful to identify or predict whether soybean breeding lines are likely to display a SCN-resistant or SCN-sensitive phenotype. The refined map of the *rhg1-b* locus from PI88788 suggested that the LRR-kinase gene that is very close to the *rhg1-b* locus does not make significant contributions to the SCN resistance phenotype. The study of Melito et al. 2010, which used transgenic roots expressing full-length transcripts or constructs that partially silence the expression of transcripts, also found no evidence to support a role for the *rhg1-b*-proximal Glyma18g02680 LRR-kinase in SCN resistance. Kim et al. demonstrated that the *rhg1-b* genetic components associated with the SCN resistance/susceptibility phenotype of PI88788 and its derivatives are located within the chromosomal interval defined by the termini BARCSOYSSR_18_0090 and BARCSOYSSR_18_0094. The most recent fine-structure genetic mapping defined an interval for *rhg1-b* that corresponds to a 67 kb interval carrying 11 predicted genes in the sequenced genome of SCN-susceptible Williams 82 soybean (Kim, Hyten et al. 2010). See Figure 5.

Here we report the identification and functional testing of multiple genes in the *rhg1-b* genetic interval. Within the *Rhg1* locus, multiple copies (ten, seven or three copies in the varieties investigated to date) of a chromosome segment encoding four identified genes within the *Rhg1* locus are present in SCN-resistant soybean varieties, while only one copy of this segment is present in the tested SCN-susceptible varieties that lack *Rhg1* alleles derived from the resistant varieties such as PI88788, PI437654 or Peking. See Figures 5 and 6. Silencing of any one of three genes within the multi-copy gene block using miRNA leads to increased susceptibility to SCN in transgenic soybean roots. In transgenic roots from a previously SCN-susceptible soybean variety, simultaneous overexpression of three or four of the *rhg1-b* genes from the multi-copy gene block leads to increased SCN resistance. The genes within this block are expressed at significantly higher levels in the tested SCN-resistant soybean varieties. Trans-acting factors in Fayette also are not sufficient to drive the elevated expression of transgenic

DNA sequences carrying these ~2kb of Glyma18g02590 or Glyma18g02610 promoter DNA sequence, when those sequences are integrated at loci other than *rhg1-b* in Fayette. DNA methylation at multiple sites within the *Rhg1* locus is polymorphic between SCN-resistant and SCN-susceptible lines, and this may contribute to the gene expression differences that correlate with SCN resistance. The number of copies of this locus also correlates to the levels of expression of the Glyma18g2580, Glyma18g2590 and Glyma18g2610 polypeptides and mRNAs and to the level of resistance to SCN. Thus gene dosing based on increasing the number of copies of the repeated region of the DNA may be a key factor mediating increased expression of the polypeptides and increased resistance to SCN. Many portions of these findings were reported in Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A., Wang, J., Hughes, T.J., Willis, D.K., Clemente, T., Diers, B.W., Hudson, M.E. and Bent, A.F. 2012. Copy Number Variation of Multiple Genes at *Rhg1* Mediates Nematode Resistance in Soybean. Science 338:1206-1209 and the associated Supporting Online Material (Supplementary Materials) found at www.sciencemag.org/content/suppl/2012/10/10/science.1228746.DC1.html, which are incorporated herein by reference in their entirety.

The resistance or susceptibility phenotype of a plant can be predicted with valuable accuracy by comparing a genetic marker in the plant to the same genetic marker or selectable marker in a second plant with known resistance or susceptibility phenotype. Thus methods of screening a first plant or plant cell for resistance or susceptibility to cyst nematodes is provided herein. The methods include detecting a genetic marker or selectable marker associated with cyst nematode resistance or susceptibility to cyst nematodes in the first plant cell and using that marker to predict the resistance or susceptibility of the first plant or plant cell to nematodes. Prediction does not mean a 100% guarantee of the phenotype regarding resistance or susceptibility of the plant to cyst nematodes. The predicting step may include comparing the marker in the first plant or plant cell to the marker in a second plant or plant cell with a known resistance or susceptibility phenotype. The marker phenotype or genotype of the second cell is predictive of the cyst nematode resistance phenotype in the first cell. The prediction may be used to select resistant soybeans or resistant plant cells for use in generating resistant

plant lines. The plants include but are not limited to sugar beets, potatoes, corn, peas or beans.

A plant includes any portion of the plant including but not limited to a whole, plant, a portion of a plant such as a part of a root, leaf, stem, seed, pod, flower, cell, tissue or plant germplasm or any progeny thereof. Germplasm refers to genetic material from an individual or group of individuals or a clone derived from a line, cultivar, variety or culture. Plant refers to whole plants or portions thereof including, but not limited to, plant cells, plant protoplasts, plant tissue culture cells or calli. For example, soybean plant refers to whole soybean plant or portions thereof including, but not limited to, soybean plant cells, soybean plant protoplasts, soybean plant tissue culture cells or calli. A plant cell refers to cells harvested or derived from any portion of the plant or plant tissue culture cells or calli.

The *rhg1* locus is a chromosomal region identified as a region important for resistance to SCN. A locus is a chromosomal region where one or more trait determinants, genes, polymorphic nucleic acids, or markers are located. A quantitative trait locus (QTL) refers to a polymorphic genetic locus where the underlying gene controls a trait that is quantitatively measured and contains at least two alleles that differentially affect expression of a phenotype or genotype in at least one genetic background, with said locus accounting for part but not all of the observed variation in the overall phenotypic trait that is being assessed. A genetic marker is a nucleotide sequence or amino acid sequence that may be used to identify a genetically linked locus, such as a QTL. Examples of genetic markers include, but are not limited to, single nucleotide polymorphisms (SNP), simple sequence repeats (SSR; or microsatellite), a restriction enzyme recognition site change, genomic copy number of specific genes or target sequences or other sequence based differences between a susceptible and resistant plant.

Genetic or selectable markers can be detected using a variety of analytic methods, including RFLP, AFLP, sequence analysis, hybridization such as allele specific hybridization analysis, differential PCR or other methods such as those known to those of skill in the art. A list of single nucleotide polymorphisms between resistant and susceptible soybeans in the *Rhg1* multi-gene copy region is provided in Table 3 in the

Examples. In another embodiment, the marker is the genomic copy number, or an estimate of the genomic copy number, of at least one of the genes or DNA sequences found in the replicated region of the resistant lines. In yet another embodiment the marker is the genomic DNA segment carrying the border between the replicated region at
5 *Glyma18g02610* and *Glyma18g02570* as shown in Figures 5 and 6. Selection methods may also include analysis of traits, phenotype polymorphisms or selectable markers not defined by DNA or RNA sequence differences, such as differences in methylation of a DNA sequence, or polypeptide expression levels or in gene expression levels. As shown in the Examples the soybean SCN resistance *Rhg1* locus, in particular the promoter
10 regions of *Glyma18g02610*, *Glyma18g02590* and *Glyma18g02580*, was highly methylated in the resistant plants as compared to susceptible plants. Methylation distinctions in and adjacent to these genes, for example in the promoter and upstream regions of the genes, may be used to distinguish between resistant and susceptible lines. In addition, resistant plants had higher mRNA levels for *Glyma18g02610*,
15 *Glyma18g02590* and *Glyma18g02580* than susceptible plants. See Figure 5. Thus methods of detecting the gene expression levels of any of these genes, for example by monitoring mRNA abundance, may be used in the methods described herein. In another embodiment, the marker may be the protein expression level of at least one of *Glyma18g02610*, *Glyma18g02590* and *Glyma18g02580*. Any of these differences may
20 be used as a screen to test whether a plant or plant cell is likely to be resistant or susceptible to nematodes.

The markers described above are linked to the phenotype of increased resistance to cyst nematodes or alternatively to susceptibility to cyst nematodes. The methods of detecting may comprise amplifying the marker or a portion thereof to produce an
25 amplified product. The presence of the product may be indicative of the marker or the amplified product may be sequenced. The amplified product may also be assessed via differential sensitivity to a restriction endonuclease. The marker may be detected using allele specific hybridization analysis, quantitative PCR, Northern blot analysis, Western blot analysis or another methodology. Methods of detecting or evaluating genetic or
30 phenotypic markers of traits such as those described herein are available to those of skill in the art, many such methods are provided in the Examples, and it is anticipated that new

methods may be developed in the future to detect the *Rhg1* polymorphisms described herein. For example, the markers can be used to detect the presence or absence of the multi-copy *Rhg1* region during breeding selection processes.

A linked locus describes a situation in which a genetic marker and a trait are closely linked chromosomally such that the genetic marker and the trait do not independently segregate and recombination between the genetic marker and the trait does not occur during meiosis with a high frequency. The genetic marker and the trait may segregate independently, but generally do not. For example, a genetic marker for a trait may only segregate independently from the trait 5% of the time; suitably only 5%, 4%, 3%, 2%, 1%, 0.75%, 0.5%, 0.25%, or less of the time. Genetic markers with closer linkage to the trait-producing locus will serve as better markers because they segregate independently from the trait less often because the genetic marker is more closely linked to the trait. Genetic markers that directly detect polymorphic nucleotide sites that cause variation in the trait of interest are particularly useful for their accuracy in marker-assisted plant breeding. Thus, the methods of screening provided herein may be used in traditional breeding, recombinant biology or transgenic breeding programs or any hybrid thereof to select or screen for resistant varieties.

In the methods described herein the SCN resistance or susceptibility phenotype of a first soybean is identified by comparing the genetic marker in the first soybean to that in a second soybean with a known resistance phenotype. The second soybean may be known to be resistant to SCN. Thus a first soybean having the same genetic marker as the second soybean is likely to also be resistant to SCN. Resistant soybeans are known in the art and include but are not limited to PI88788, Peking, Hartwig, Fayette, Forrest, LD02-5320, LD02-5025, and LD01-7323 or lines carrying loci that contributed to or were derived from these cultivars such as those provided in Table 2. In particular, the methods allow identification of soybean plants having increased resistance to Race 3 SCN and other nematode populations, similar to PI88788. Alternatively, the second soybean may be known to be susceptible to SCN. Thus a first soybean having the same genetic marker as the second soybean is likely to be susceptible to SCN. Susceptible soybeans are known in the art and include, but are not limited to, 'Williams 82', Essex, Thorne, Sturdy, LG03-1672, and LG00-3372 or lines carrying loci that contributed to or

were derived from one of these cultivars such as those provided in Table 2. In particular, the methods allow identification of soybean plants having susceptibility to SCN similar to that of 'Williams 82.' Although resistance to SCN is widely observed to be a quantitative trait, the terms susceptibility and resistance as used in the preceding paragraphs refer to qualitative traits, such that identification as a resistant soybean indicates that the soybean is more resistant than the susceptible soybean line to which it is being compared. Likewise, identification of a soybean as a susceptible soybean indicates that the soybean is more sensitive than the resistant soybean line to which it is being compared.

Resistance (or susceptibility) to SCN can be measured in a variety of ways, several of which are known to those of skill in the art. In the examples, soybean roots were experimentally inoculated with SCN and the ability of the nematodes to mature (molt and proceed to developmental stages beyond the J2) on the roots was evaluated as compared to a susceptible and/or resistant control plant. A SCN greenhouse test is also described in the Examples and provides an indication of the number of cysts on a plant and is reported as the female index. Increased resistance to nematodes can also be manifested as a shift in the efficacy of resistance with respect to particular nematode populations or genotypes. Additionally but not exclusively, SCN-susceptible soybeans grown on SCN-infested fields will have significantly decreased crop yield as compared to a comparable SCN-resistant soybean. Improvement of any of these metrics has utility even if all of the above metrics are not altered.

As demonstrated in the Examples a set of three genes found on a tandemly repeated segment of chromosome 18 were identified whose silencing led to increased susceptibility to SCN in a resistant variety. The three genes are found along with a fourth gene, part of a fifth gene, and other DNA sequences in a chromosome segment approximately 31 kb in length that is present in 10 copies in the soybean varieties that carry the *rhg1-b* allele or haplotype of *Rhg1* that is in widespread commercial use for control of SCN disease of soybean. This *Rhg1* chromosome segment is found in at least three copies in all SCN resistant varieties tested to date. Various resistant varieties carry three, seven or ten copies and the higher copy number versions of *Rhg1* express higher levels of transcripts for the three genes. Higher copy number versions of *Rhg1* also confer more resistance to SCN on their own (exhibit less reliance on the simultaneous

presence of desirable alleles of other SCN resistance QTL such as *Rhg4* in order to effectively confer SCN resistance, relative to *Rhg1* haplotypes with lower *Rhg1* repeat copy numbers). In the Examples, over-expression of the three genes in a susceptible variety made roots more resistant to SCN. Methods of increasing resistance of a plant to cyst nematodes by selecting plants carrying genetic markers associated with *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* alleles that are present within the *Rhg1* locus are described. As shown in the Examples, genetic polymorphisms ranging from single nucleotide polymorphisms to gene rearrangements (i.e. gene duplications) and differences in methylation may occur in other *Glycine max* plant lines and other *Glycine* species, which may alter the expression or biological impact of one or more genes linked to the *Rhg1* locus, and careful selection of desirable alleles of particular genes at the *Rhg1* locus may be desirable to allow selection of plants with increased resistance to SCN.

Methods of increasing resistance of a plant to cyst nematodes, including but not limited to SCN, by increasing the expression of or altering the expression pattern of or increasing the copy number of a polynucleotide encoding the *Glyma18g02610* (SEQ ID NO:3), *Glyma18g02590* (SEQ ID NOs: 2, 5 and 6), and/or *Glyma18g02580* (SEQ ID NO: 1) polypeptides or functional fragments or variants thereof in cells of the plant are also provided. The polypeptide may be 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to the sequences provided. We have sequenced these genes from both resistant and susceptible varieties and found few polymorphisms within the coding regions and few changes that result in an amino acid change. The *Glyma18g02590* polypeptide does have some significant polymorphisms between the resistant and susceptible varieties that appear to be functionally related to SCN resistance as shown in the Examples.

Suitably the expression of the polypeptides encoded by *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* is increased in a root of the plant. Suitably, the expression of the polypeptides encoded by *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* is increased in root cells of the plant. The plant is suitably a soybean plant or portions thereof. The polynucleotides may also be transferred into other non-soybean plants, or homologs of these polypeptides or polynucleotides encoding the polypeptides from other plants, or synthetic genes encoding products similar to the

polypeptides encoded by *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be overexpressed in those plants. Other plants include but are not limited to sugar beets, potatoes, corn, peas, and beans. The overexpression of the genes may increase the resistance of plants from these other species to nematodes and in particular cyst

5 nematodes, such as the soybean cyst nematode *Heterodera glycines*, the sugar beet cyst nematode *Heterodera schachtii*, the potato cyst nematodes *Globodera pallida* and related nematodes that cause similar disease on potato such as *Globodera rostochiensis*, the corn cyst nematode *Heterodera zea*, and the pea cyst nematode *Heterodera goettingiana*.

The expression of the polynucleotides may be increased by increasing the copy
10 number of the polynucleotide in the plant, in cells of the plant, suitably root cells, or by identifying plants in which this has already occurred. These plants may then be used in traditional breeding. Suitably, the polynucleotide is present in three, seven or even ten copies. Suitably at least two or all three of the polynucleotides encoding the polypeptides or the polypeptides of *Glyma18g02610*, *Glyma18g02590*, and *Glyma18g02580* are
15 expressed. Alternatively the expression may be increased using recombinant DNA technology, e.g., by using a strong promoters to drive increased expression of one or more polynucleotides.

In addition, methods of increasing resistance of a plant to cyst nematodes may be achieved by cloning sequences upstream from *Glyma18g02610*, *Glyma18g02590*, and/or
20 *Glyma18g02580* from resistant lines into susceptible lines. For these methods, nucleotide sequences having at least 60%, 70% or 80% identity to nucleotide sequences that flank the protein-coding regions of *Glyma18g02610*, *Glyma18g02590* or *Glyma18g02580* (or sequences having at least 80%, 85%, or 90% identity to those protein-coding regions), said flanking regions including 5' and 3' untranslated regions of the mRNA for these
25 genes, and also including any other genomic DNA sequences that extend from the protein coding region of these genes to the protein coding regions of immediately adjacent genes may be used.

The increase in expression of *Glyma18g02610*, *Glyma18g02590*, and/or
Glyma18g02580 in the plant may be measured at the level of expression of the mRNA or
30 at the level of expression of the polypeptide encoded by *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580*. The level of expression may be increased

relative to the level of expression in a control plant as shown in the Examples. The control plant may be an SCN-susceptible plant or an SCN-resistant plant. For example, a susceptible plant such as 'Williams 82' may be transformed with an expression vector such that the roots of the transformed plants express increased levels of *Glyma18g02610*,
5 *Glyma18g02590*, and/or *Glyma18g02580* as compared to an untransformed plant or a plant transformed with a construct that does not change expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580*, resulting in increased resistance to nematodes. Alternatively, the control may be a plant partially resistant to nematodes and increased expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may result in
10 increased resistance to nematodes. Alternatively, the plant may be resistant to nematodes and increasing expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may result in further increased resistance to nematodes. Alternatively, the plant may be more resistant to certain nematode populations, races, Hg types or strains and less resistant to other nematode populations, races, Hg types or strains, and increasing
15 expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may result in increased resistance to certain of these nematode populations, races, Hg types or strains.

In the Examples, a decrease in expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* is shown to increase the susceptibility of a SCN-resistant soybean to SCN maturation. In addition, roots of the susceptible 'Williams 82' soybean
20 are shown to have lower levels of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* mRNA as compared to the resistant Fayette line. Because low levels of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* mRNA correlate with nematode susceptibility, and increased levels correlate with resistance, and direct lowering of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* mRNA is
25 causally associated with greater nematode susceptibility of previously resistant tissues, increasing the levels of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* in a soybean should in many instances increase the resistance of the soybean to nematodes, in particular SCN. In Figure 15, increased expression of a combination of *Glyma18g02600*, *Glyma18g02610*, *Glyma18g02590*, and *Glyma18g02580* was shown to
30 increase resistance to SCN of a susceptible line. Increased expression of three genes, *Glyma18g02610*, *Glyma18g02590*, and *Glyma18g02580* was also shown to increase

resistance of an SCN susceptible variety in Figure 18. Increased expression of fewer than these three polynucleotides or of the polypeptides encoded by the polynucleotides may be similarly effective to increase resistance.

Expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be increased in a variety of ways including several apparent to those of skill in the art and may include transgenic, non-transgenic and traditional breeding methodologies. For example, the expression of the polypeptide encoded by *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be increased by introducing a construct including a promoter operational in the plant operably linked to a polynucleotide encoding the polypeptide into cells of the plant. Suitably, the cells are root cells. Alternatively, the expression of the polypeptide encoded by *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be increased by introducing a transgene including a promoter operational in the plant operably linked to a polynucleotide encoding the polypeptide into cells of the plant. The promoter may be a constitutive or inducible promoter capable of inducing expression of a polynucleotide in all or part of the plant, plant roots or plant root cells. In another embodiment, the expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be increased by increasing expression of the native polypeptide in a plant or in cells of the plant, such as the plant root cells. In another embodiment, the expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be increased by increasing expression of the native polypeptide in a plant or in cells of the plant such as the nematode feeding site, the syncytium, or cells adjacent to the syncytium. In another embodiment, the expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be increased by increasing expression of the native polypeptide in a plant or in cells of the plant such as sites of nematode contact with plant cells. In another embodiment, expression may be increased by increasing the copy number of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580*. Other mechanisms for increasing the expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* include, but are not limited to, increasing expression of a transcriptional activator, reducing expression of a transcriptional repressor, addition of an enhancer region capable of increasing expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580*, increasing mRNA stability,

altering DNA methylation, histone acetylation or other epigenetic or chromatin modifications in the vicinity of the relevant genes, or increasing protein or polypeptide stability.

In addition to the traditional use of transgenic technology to introduce additional
5 copies or increase expression of the genes and mediate the increased expression of the polypeptides of Glyma18g02610, Glyma18g02590, and/or Glyma18g02580 in plants, transgenic or non-transgenic technology may be used in other ways to increase expression of the polypeptides. For example, plant tissue culture and regeneration, mutations or altered expression of plant genes other than Glyma18g02610,
10 Glyma18g02590, and/or Glyma18g02580, or transgenic technologies, can be used to create instability in the *Rhg1* locus or the plant genome more generally that create changes in *Rhg1* locus copy number or gene expression behavior. The new copy number or gene expression behavior can then be stabilized by removal of the variation-inducing mutations or treatments, for example by further plant propagation or a conventional cross.
15 In one of the examples, although a transgenic plant was used to create the change in copy number, the result would be a non-transgenic line (and conceivably regulated as such) with enhanced resistance due to increased copy number of the locus. Examples of transgenic technologies that might be used in this way include targeted zinc fingers, ribozymes or other sequence-targeted enzymes that create double stranded DNA breaks
20 at or close to the *Rhg1* locus, the Cre / loxP system from bacteriophage lambda or other similar systems like *frt/Flp*, Transcription Activator-Like Effector Nucleases (TALENs), artificial DNA or RNA sequences designed to recombine with *Rhg1* that can be introduced transiently, or enzymes that "shuffle" DNA such as the mammalian *Rag1* enzyme or DNA transposases. Mutations or altered expression of endogenous plant
25 genes involved in DNA recombination, DNA rearrangement and/or DNA repair pathways are additional examples.

The screening methods described above could also be used to screen soybean isolates (*Glycine max*) and closely related species (*Glycine soja*, *Glycine tomentella* or other *Glycine* species) for resistance markers and then resistant lines can be crossed
30 naturally or artificially with soybean to develop a soybean with a variant copy number or sequence at the *Rhg1* site. Any useful alleles identified in such screens could then be

introduced using traditional breeding or transgenic technology into soybeans. Similar systems are available for other plants

Non-transgenic means of generating plant varieties carrying traits of interest such as increased resistance to nematodes, such as SCN, are available to those of skill in the art and include traditional breeding, chemical or other means of generating chromosome abnormalities, such as chemically induced chromosome doubling and artificial rescue of polyploids followed by chromosome loss, knocking-out DNA repair mechanisms or increasing the likelihood of recombination or gene duplication by generation of chromosomal breaks. Other means of non-transgenetically increasing the expression or copy number of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* include the following: screening for mutations in plant DNA encoding miRNAs or other small RNAs, plant transcription factors, or other genetic elements that impact *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* expression; screening large field or breeding populations for spontaneous variation in copy number or sequence at *Rhg1* by screening of plants for nematode resistance, *Rhg1* copy number or other *Rhg1* gene or protein expression traits as described in preceding paragraphs; crossing of lines that contain different or the same copy number at *Rhg1* but have distinct polymorphisms on either side, followed by selection of recombinants at *Rhg1* using molecular markers from two distinct genotypes flanking the *Rhg1* locus; chemical or radiation mutagenesis or plant tissue culture/regeneration that creates chromosome instability or gene expression changes, followed by screening of plants for nematode resistance, *Rhg1* copy number or other *Rhg1* gene or protein expression traits as described in preceding paragraphs; or introduction by conventional genetic crossing of non-transgenic loci that create or increase genome instability into *Rhg1* - containing lines, followed by screening of plants for either nematode resistance or *Rhg1* copy number. Examples of loci that could be used to create genomic instability include active transposons (natural or artificially introduced from other species), loci that activate endogenous transposons (for example mutations affecting DNA methylation or small RNA processing such as equivalent mutations to *met1* in *Arabidopsis* or *mop1* in maize), mutation of plant genes that impact DNA repair or suppress illegitimate recombination such as those orthologous or similar in function to the *Sgs1* helicase of yeast or *RecQ* of *E. coli*, or overexpression of genes such as *RAD50*

or *RAD52* of yeast that mediate illegitimate recombination. Those of skill in the art may find other transgenic and non-transgenic methods of increasing expression of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580*.

The polynucleotides and/or polypeptides described and used herein may encode the full-length or a functional fragment of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* from the *rhg1-b* locus, or a naturally occurring or engineered variant of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580*, or a derived polynucleotide or polypeptide all or part of which is based upon nucleotide or amino acid combinations similar to all or portions of *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* or their encoded products. Additional polynucleotides encoding polypeptides may also be included in the construct such as *Glyma18g02600* (which encodes the polypeptide of SEQ ID NO:4). The polypeptide may be at least 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to the sequences provided herein. The polynucleotides encoding the polypeptides may be at least 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% identical to the sequences available in the public soybean genetic sequence database.

The expression of the polypeptide encoded by *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* may be increased, suitably the level of polypeptide is increased at least 1.2, 1.5, 1.7, 2, 3, 4, 5, 7, 10, 15, 20 or 25 fold in comparison to the untreated, susceptible or other control plants or plant cells. Control cells or control plants are comparable plants or cells in which *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* expression has not been increased, such as a plant of the same genotype transfected with empty vector or transgenic for a distinct polynucleotide.

Increased resistance to nematodes may be measured as described above. The increased resistance may be measured by the plant having a lower percentage of invading nematodes that develop past the J2 stage, a lower rate of cyst formation on the roots, reduced SCN egg production within cysts, reduced overall SCN egg production per plant, and/or greater yield of soybeans on a per-plant basis or a per-growing-area basis as compared to a control plant grown in a similar growth environment. Other methods of measuring SCN resistance also will be known to those with skill in the art. In the methods of increasing resistance to nematodes described herein, the resulting plant may have at least 10% increased resistance as compared to the untreated or control plant or

plant cells. Suitably the increase in resistance is at least 15%, 20%, 30%, 50%, 100%, 200%, 500% as compared to a control. Suitably, the female index of the plant with increased resistance to nematodes is about 80% or less of the female index of an untreated or control plant derived from the same or a similar plant genotype, infested with a similar nematode population within the same experiment. More suitably, the female index after experimental infection is no more than 60%, 40%, or 20% of that of the control plant derived from the same or a similar plant genotype, infested with a similar nematode population within the same experiment. Suitably, when grown in fields heavily infested with SCN (for example, more than 2500 SCN eggs per 100 cubic centimeters of soil), soybean grain yields of field-grown plants are 2% greater than isogenic control plants. More suitably, the grain yield increase is at least 3%, 4%, or 5% over that of isogenic control plants grown in similar environments.

Also provided herein are constructs including a promoter operably linked to a *Glyma18g02610*, *Glyma18g02590*, and/or *Glyma18g02580* polynucleotide encoding a polypeptide comprising SEQ ID NO: 1-3 or 5-6 or a fragment or functional variant thereof. Also included are homologs or variants of these sequences from other soybean varieties. The constructs may further include *Glyma18g02600* or other genes. The constructs may be introduced into plants to make transgenic plants or may be introduced into plants, or portions of plants, such as plant tissue, plant calli, plant roots or plant cells. Suitably the promoter is a plant promoter, suitably the promoter is operational in root cells of the plant. The promoter may be tissue specific, inducible, constitutive, or developmentally regulated. The constructs may be an expression vector. Constructs may be used to generate transgenic plants or transgenic cells. The polypeptide may be at least 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to the sequences of SEQ ID NO: 1-3 or 5-6. The constructs may comprise all three polynucleotides and may mediate expression of all three polypeptides.

Transgenic plants including a non-native or exogenous polynucleotide encoding the *rhg1-b* polypeptides identified and described herein are also provided. Suitably the transgenic plants are soybeans. The soybean polynucleotides and polypeptides identified herein as associated with resistance to nematodes may also be used to generate transgenic sugar beets, potatoes, corn, peas, or beans capable of expressing the soybean genes

described herein. Alternatively, homologous genes or polypeptides from these plants may be identified by comparison to the soybean genes and polypeptides identified herein and these genes may be used to generate transgenic plants. The transgenic plants express increased levels of Glyma18g02610, Glyma18g02590, and/or Glyma18g02580

5 polypeptide as compared to a control non-transgenic plant from the same line, variety or cultivar or a transgenic control expressing a polypeptide other than Glyma18g02610, Glyma18g02590, and/or Glyma18g02580. The transgenic plants also have increased resistance to nematodes, in particular SCN, as compared to a control plant. Portions or parts of these transgenic plants are also useful. Portions and parts of plants includes, but
10 is not limited to, plant cells, plant tissue, plant progeny, plant asexual propagates, plant seeds.

Transgenic plant cells comprising a polynucleotide encoding a polypeptide capable of increasing resistance to nematodes such as SCN are also provided. Suitably the plant cells are soybean plant cells. Suitably the cells are capable of regenerating a
15 plant. The polypeptide comprises the sequences of SEQ ID NOs:1-3 or 5-6 or fragments, variants or combinations thereof. The polypeptide may be 85%, 90%, 95%, 97%, 98%, 99% or 100% identical to the sequences provided. The transgenic cells may be found in a seed. A plant, such as a soybean plant, may include the transgenic cells. The plant may be grown from a seed comprising transgenic cells or may be grown by any other means
20 available to those of skill in the art. Chimeric plants comprising transgenic cells are also provided.

The expression of the polypeptide and the polynucleotides encoding the polypeptides in the transgenic plant is altered relative to the level of expression of the native polypeptides in a control soybean plant. In particular the expression of the
25 polypeptides in the root of the plant is increased. The transgenic plant has increased resistance to nematodes as compared to the control plant. The transgenic plant may be generated from a transgenic cell or callus using methods available to those skilled in the art.

The Examples provided below are meant to be illustrative and not to limit the
30 scope of the invention or the claims. All references and appendices cited herein are hereby incorporated by reference in their entireties.

EXAMPLES

To identify the gene conferring resistance to SCN in the PI88788 soybean within the locus identified by Kim et al., 2010 (within the chromosomal interval defined by the termini BARCSOYSSR_18_0090 and BARCSOYSSR_18_0094), a candidate gene testing approach was used. This approach is described in Melito et al. (BMC Plant Biology 2010, 10:104), which is incorporated herein by reference in its entirety. Briefly, this candidate gene approach was completed with various genes at the *Rhg1* locus defined above using a resistant soybean variety Fayette, which carries the PI88788-derived *rhg1-b* allele of the *Rhg1* locus, to make transgenic soybean roots that carry gene-silencing constructs and then testing these transgenic roots for loss of SCN resistance. The silencing strategy used is depicted in Figure 2. The artificial microRNA used in the Melito et al. reference was replaced with artificial microRNA sequences directed against various candidate or putative genes within the *Rhg1* locus. The expression of the artificial microRNAs was driven by the soybean Ubi3 promoter. The construct also contained a GFP reporter such that transformed roots could readily be identified by GFP expression. Transgenic soybean roots expressing artificial micro-RNA (amiRNA) or hairpin (RNAi) constructs were produced using *Agrobacterium rhizogenes*. Roots expressing GFP were selected for further analysis. Transgenic roots were inoculated with SCN to test for decreased or increased resistance to SCN caused by candidate gene silencing conditioned by artificial microRNA expression.

Soybean resistance to SCN was measured two weeks after root inoculation by determining the proportion of the total nematode population that had advanced past the J2 stage in each root (Fig. 3A), relative to known resistant and susceptible controls. Silencing any of three closely linked genes, namely *Glyma18g2580*, *Glyma18g2590*, and *Glyma18g2610*, at the *rhg1-b* locus of the SCN-resistant soybean variety Fayette significantly reduced SCN resistance (Fig. 3B). Depletion of resistance was dependent on target transcript reduction (Fig. 4). Silencing of other genes in and around the locus did not impact SCN resistance (e.g., Fig 3B, genes *Glyma18g02570* and *2620*).

The predicted *Glyma18g02610* protein product contains a Wound-Induced protein domain (Pfam domain PF07107; *M. Punta, et al., (2012) The Pfam protein families database. Nucleic Acids Research Database Issue 40:D290-D301 and Logemann et al.,*

(1988) *Differential expression of genes in potato tubers after wounding*. *Proc Natl Acad Sci USA* 85: 1136–1140) and a homologous (55% identical) protein in ice plant (*Mesembryanthemum crystallinum*) was previously shown to be responsive to both biotic and abiotic stimuli (Yen *et al.*, *Environmental and developmental regulation of the wound-induced cell wall protein W112 in the halophyte ice plant*. *Plant Physiol.* 127:517–528). The annotated protein product of *Glyma18g02610* does not have other widely known protein domains or inferred biochemical functions that, at the present time, are obvious to those with normal skill in the art. However, the above results indicate that *Glyma18g02610* is necessary for full *Rhg1*-mediated SCN resistance.

10 A genomic duplication of four genes at *Rhg1* in *Glycine max* is present in the tested SCN-resistant lines

Concurrent study of the physical structure of the *rhg1-b* locus revealed an unusual genomic configuration. A 31.2 kb genome segment, encoding the above three genes that contribute to SCN resistance, is present in multiple copies in SCN resistant lines (Figs. 5, 6). The DNA sequence of fosmid clone inserts carrying genomic DNA from the *rhg1-b* genetic interval identified a unique DNA junction, not present in the published Williams 82 soybean genome, in which a 3' fragment of *Glyma18g02570* is immediately adjacent to the intergenic sequence downstream of (centromeric to) *Glyma18g02610* (Fig. 5A). The genomic repeat contains full copies of *Glyma18g02580*, -2590, -2600 and -2610 as well as the final two exons of *Glyma18g02570*. Whole-genome shotgun sequencing of a line containing *rhg1-b* revealed ten-fold greater depth of coverage of this interval relative to surrounding or homologous regions (Fig. 5B), suggesting the presence of multiple repeats.

Sequencing and PCR amplification confirmed the presence of the *Glyma18g02610*-2570 junction in DNA from multiple SCN-resistant soybean accessions, including accessions that carry the commercially important PI 88788, Peking and PI 437654 haplotypes of the *Rhg1* locus (Fig. 5C and Fig. 7). The junction was not detected in four tested SCN-susceptible varieties including Williams 82 (Fig. 7B). This constitutes a direct test for economically desirable alleles of the *Rhg1* locus. The shared identity of the junction sites from disparate sources of SCN resistance suggests a shared origin of the initial resistance-conferring event at *Rhg1*.

Fiber-FISH (fluorescence in situ hybridization) was utilized to directly determine the number of copies and arrangement of the 31 kb repeat segment in different haplotypes of the *Rhg1* locus. The hybridization pattern and DNA fiber length estimates generated using these probes (Fig. 6 and Table 1) are consistent with the presence of a single copy of the repeat in Williams 82, as in the reference soybean genome. In Fayette, fiber-FISH revealed ten copies per DNA fiber of the predicted 31 kb repeat segment, in the same configuration throughout the multiple nuclei sampled, in a pattern indicating ten direct repeats abutting in a head-to-tail arrangement (Fig. 6 and Table 1). No additional copies (e.g., at other loci) were evident. In samples from soybean line Peking, three copies per DNA fiber were present in apparent direct repeat orientation (Fig. 6). Although fiber-FISH cannot resolve small sequence differences, the single size of all junction-amplification PCR products and the consistency of all junction sequences assembled from fosmid or genomic DNA sequencing (Fig. 7) further suggest the presence of adjacent direct repeat copies.

15

Table 1. Length estimates for Fiber-FISH hybridization signals. N: number of DNA fibers analyzed.

	Length of Fiber-FISH signals (μm)	Estimated Length of Fiber-FISH signals (kb)	N
Williams 82	8.60 ± 0.49	27.60 ± 1.58	20
Fayette	96.56 ± 7.54	309.94 ± 24.2	20
Peking	34.42 ± 2.91	110.48 ± 9.34	20

5

To discover additional copy number and DNA polymorphisms we analyzed whole genome re-sequencing (WGS) data for 41 soybean lines, the “NAM parents” from a current nested association mapping study, using sequence data provided by Dr. Perry Cregan. The data set consists of whole genome shotgun sequencing reads produced using Illumina Hiseq equipment and protocols, with average depth of sequencing coverage ranging from 5 to 60 fold. Reads were mapped to the Williams 82 reference genome (Phytozome, assembly 189) and analyzed for read depth (RD), INDELs and SNPs relative to the SCN susceptible line Williams 82. When this Illumina read depth was used to estimate copy number at the *Rhg1* locus, using methods analogous to those used to generate Figure 5B, 8 of the 41 soybean lines analyzed had a normalized read depth for the approximately 31 kb *Rhg1* region (corresponding to the *rhg1-b* repeat segment described above) that differed by greater than 5 standard deviations of the mean from the read depth of the two 30 kb regions immediately flanking the region corresponding to the *rhg1-b* repeat segment. See Table 2. Seven of those lines had an estimated copy number ranging from 9.2 to 9.9 copies. These lines have PI 88788 in their pedigree, where pedigree are available, and have been classified as SCN-resistant in laboratory and/or field tests. The other genotype predicted to carry *Rhg1* repeats had an estimated copy number of 2.9. Its pedigree contains both Peking and PI 437654. The *Rhg1* loci derived from Peking and PI 437654 are widely recognized to be much less effective at conferring SCN resistance if they are not coupled with preferred alleles of an unlinked locus, *Rhg4*. All other genotypes (33) were estimated to contain one copy of the approximately 31 kb of *Rhg1* DNA described in this document. All twelve of the 33 lines that had an available, previously determined SCN resistance phenotype were listed as SCN

25

susceptible, while information on the SCN resistance phenotype was not readily available for the other 21 lines. As a control, read depth was used to estimate copy number at the homologous region on Chromosome 11. The estimated copy number was approximately 1 in all tested genotypes. In addition to those lines shown in Table 2, we recently

5 determined that a variety known as Cloud (PI 548316), that displays intermediate levels of SCN resistance, carries seven copies of the *Rhg1* locus repeat segment.

Table 2: Analysis of copy number of the *Rhg1* locus in various soybean varieties.

Genotype	Copy Number Estimate	
	Chromosome 18 (<i>Rhg1</i>)	Chromosome 11 (homolog)
LD00-3309	9.9 ± 1.8	0.9 ± 0.2
4J105-3-4	9.9 ± 1.9	1.0 ± 0.2
LD02-4485	9.8 ± 2.2	1.0 ± 0.3
CLOJ095-4-6	9.6 ± 1.5	0.9 ± 0.2
LD02-9050	9.4 ± 3.4	1.0 ± 0.4
LG05-4292	9.4 ± 1.7	1.0 ± 0.2
Maverick	9.2 ± 3.3	0.9 ± 0.3
LD01-5907	2.9 ± 0.9	1.1 ± 0.3
PI427_136	1.1 ± 0.3	
PI404_188A	1.1 ± 0.3	
LG98-1605	1.1 ± 0.4	
U03-100612	1.1 ± 0.2	
LG90-2550	1.1 ± 0.3	
PI398_881	1.1 ± 0.2	
PI518_751	1.0 ± 0.3	
IA3023	1.0 ± 0.2	
PI561_370	1.0 ± 0.2	
HS6-3976	1.0 ± 0.3	
LG03-2979	1.0 ± 0.2	
LG92-1255	1.0 ± 0.3	
Prohio	1.0 ± 0.2	
LG03-3191	1.0 ± 0.2	
PI507_681B	1.0 ± 0.2	
LG05-4464	0.9 ± 0.2	
NE3001	0.9 ± 0.3	

The source of the first duplication event to arise at *Rhg1* is not known, but was possibly the result of nearby Ty1/copia-like retrotransposon RTvr1 or RTvr2 activity. Later copy number expansion may have occurred by rare unequal exchange events between homologous repeats during meiotic recombination.

5 Genes within the duplicated gene block at *rhg1-b* are expressed at higher levels than their homologs from SCN-susceptible *Rhg1* haplotypes

Gene expression analysis using quantitative PCR (qPCR) determined that the three genes found to impact SCN resistance exhibit significantly more transcript abundance in roots of SCN-resistant varieties relative to susceptible lines (Fig. 5D and
 10 Fig. 7D). In contrast, the transcript abundance for genes immediately flanking the SCN-impacting genes did not differ significantly between SCN-resistant and susceptible varieties (Fig. 5; *Glyma18g02600* expression in roots is at or below the limits of detection of qPCR, cDNA cloning and RNAseq methods; See Cook et al. 2012 methods and Severin et al. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome.
 15 Bmc Plant Biology, 2010). Full-length transcripts were confirmed for *Glyma18g02580*, -2590 and -2610, and no hybrid repeat-junction transcripts were detected for *Glyma18g02570* (Fig. 7E). The above suggested that elevated expression of one or more of the SCN-impacting genes could be a primary cause of elevated SCN resistance.

Quantitative real-time PCR (qPCR) was also used to examine and compare the
 20 mRNA transcript abundance of five genes at the *Rhg1* locus in non-inoculated roots of the Hg-typing soybean lines. These lines have been established and accepted by researchers as representing a useful and diverse set of SCN resistant soybean lines (Niblack et al., 2002, J. of Nemat. 34(4): 279-288; T. L. Niblack, K. N. Lambert, G. L. Tylka, 2002, Annu. Rev. Phytopathol. 44:283-303). Transcript abundance for three of
 25 the genes, *Glyma18g02580*, *Glyma18g02590*, and *Glyma18g02610* are all expressed more highly in each of the 7 tested SCN differentials relative to the SCN susceptible line Williams 82 as shown in Figure 8. Another gene at the locus, *Glyma18g02600*, was also more highly expressed in the SCN resistant lines, but the data for *Glyma18g02600* may be less accurate and the absolute measured transcript level of *Glyma18g02600* was near
 30 the limit of detection (consistent with published RNA-seq data from soybean roots). As a control, a neighboring, but not duplicated gene, *Glyma18g02570* shows similar

expression pattern for all tested genotypes. In a separate experiment, two additional genes, *Glyma18g02620* and *Glyma18g02630*, flanking the repeat to the centromeric side, also show similar transcript abundance across SCN resistant and susceptible lines.

Four of the SCN resistant genotypes (Peking, PI 90763, PI 89772, and PI 437654) are similar to each other in their level of mRNA abundance for the four genes *Glyma18g02580*, *Glyma18g02590*, *Glyma18g02610*, and *Glyma18g02600*. In these genotypes, transcript abundance is 1.5 to 5 fold higher for the four repeated genes relative to Williams 82. Separately, the SCN-resistant genotypes Cloud, PI 88788, and 209332 are similar to each other in their levels of elevated mRNA abundance for *Rhg1* genes compared to Williams 82 and the previous four genotypes. Transcript abundance ranged from 4 to 20 fold higher for the repeated genes in the Cloud, PI 88788, and PI 209332 genotypes. These data show that the increased DNA copy number encoding these four genes increases the transcript abundance. There is also a strong grouping for DNA copy number and transcript abundance, making at least two classes with genotypes Peking, 90763, 89772, and 437654 together, and genotypes Cloud, PI 88788, and 209332 together. We note the correlation of these *Rhg1* genotype (copy number) and *Glyma18g02580*/*Glyma18g02590*/*Glyma18g02610* expression level groupings with the SCN resistance phenotype groupings reported by Colgrove et al. 2008 (Colgrove, A.L., and T.L. Niblack. 2008. Correlation of female indices from virulence assays on inbred lines and field populations of *Heterodera glycines*. Journal of Nematology 40:39-45). While *Glyma18g02600* is more highly expressed in SCN resistant lines, there is not a clear relationship between copy number and transcript abundance as found for the other three genes in the repeat. This suggests that increased DNA content does increase transcript abundance, but not in a dosage dependent fashion for *Glyma18g02600*.

***Rhg1* DNA methylation state is cultivar-dependent for genes within the duplicated gene block.**

To address the mechanism leading to the observed higher gene expression in SCN-resistant cultivars, we assessed DNA methylation of the *Rhg1* locus using methylation sensitive restriction enzyme digestion and PCR. McrBC is an endonuclease that specifically cleaves DNA containing 5-methylcytosine (5-mC) while leaving unmethylated DNA intact. DNA incubated with McrBC and then subjected to PCR fails to

produce a product if the product spans methylated cytosines. We identified significant and reproducible differences between SCN-resistant and SCN-susceptible cultivars when soybean genomic DNA was tested for methylation at the *Rhg1* locus. For example, three different primer pairs for the *Glyma18g02610* promoter or coding regions either failed to
 5 amplify a product, or the product was greatly reduced, between McrBC-digested and undigested genomic DNA from resistant cultivars (Figure 9). The same primer pairs used for PCR with DNA collected from susceptible cultivars produced similar products whether the genomic DNA template had been digested or not, indicating little or no DNA methylation.

10 Interestingly, a consistent correlation between hypermethylated DNA and elevated gene expression was discovered in the SCN-resistant cultivars tested. The promoter region for *Glyma18g02580*, *Glyma18g02590* and *Glyma18g02610* were all methylated and showed higher transcription. A neighboring gene, *Glyma18g02620*, did not display polymorphic methylation or altered gene expression between resistant and
 15 susceptible cultivars (Figure 9).

Further characterization of 2580, 2590 and 2610 genes

RACE PCR for *Glyma18g02590* and *Glyma18g02610* from Fayette (not inoculated with SCN) revealed that the transcripts derived from the SCN-resistant PI88788 haplotype have identical start and stop sites to the annotated transcripts
 20 associated with the Williams 82 (SCN-susceptible) genome sequence that is available at Phytozome (<http://www.phytozome.net/cgi-bin/gbrowse/soybean/>). As an initial test for readily detectable protein degradation or post-translational modification differences between SCN-resistant as opposed to SCN-susceptible soybean lines (not inoculated with SCN), protein immunodetection experiments by western blot using 1.5 kb of Fayette
 25 native promoter driving *Glyma18g02590*-HA (*2590_{Fayette}::2590*-HA) or using 3.2kb of Fayette native promoter driving *Glyma18g02610*-HA (*2610_{Fayette}::2610*-HA) revealed a detectable protein product that migrated at approximately the predicted size for the respective proteins, and did not reveal protein size differences in Williams 82 as opposed to Fayette (Figure 10).

30 To explore the possibility that impacts of *Glyma18g02580*, *Glyma18g02590* and *Glyma18g02610* on SCN development in soybean correlate with SCN-inducible gene

expression, we analyzed gene expression in excised root tissue from heavily SCN-infested root segments of resistant and susceptible lines. Modest increases in the expression of *Glyma18g02580* and *Glyma18g02610* were observed in SCN-resistant Fayette, above the high levels of expression already present in non-inoculated Fayette
 5 compared to SCN-susceptible Williams 82 (Figure 11).

To further explore the possibility that impacts of *Glyma18g02580*, *Glyma18g02590* and *Glyma18g02610* on SCN development in soybean roots correlate with SCN-inducible expression, promoter-GUS fusion constructs (Prom₂₅₈₀::GUS, Prom₂₅₉₀::GUS, Prom₂₆₁₀::GUS) were made and expressed in transgenic roots.

10 Transgenic roots were stained for GUS activity 5 days after inoculation with SCN (Figure 12). Prom₂₅₈₀::GUS roots had a moderate level of background staining and appeared to have brighter staining of swollen plant vascular tissue at the head of infecting J2 SCN (apparent developing syncytia). Prom₂₅₉₀::GUS roots showed very low levels of background staining and consistently greater GUS staining at apparent syncytia, along
 15 with strong staining of the tips of emerging lateral roots. Prom₂₆₁₀::GUS roots had very high levels of background staining but also appeared to be more highly stained at areas likely to form syncytia. The levels of elevated GUS staining in apparent syncytia were similar to those we observed in positive control promoter-GUS experiments using the previously characterized syncytium-inducible promoter for *Glyma14g06080*. As noted
 20 above, the predicted *Glyma18g02610* encoded protein contains a Wound-Induced protein domain (Pfam07107) and a homologous (55% identical) protein in ice plant (*Mesembryanthemum crystallinum*) was shown to be responsive to both biotic and abiotic stimuli. Non-transgenic Fayette soybean plants exposed to MeJA had significantly elevated *Glyma18g02610* transcript abundance compared to neighboring genes, further
 25 documenting stress-associated expression of this gene (Figure 11).

Amino acid polymorphism or overexpression of any one of the three identified *rhg1-b* genes did not account for SCN resistance on its own. From all available *rhg1-b* sequence reads (across multiple repeat copies), no predicted amino acid polymorphisms relative to Williams 82 were identified for *Glyma18g02580*, *Glyma18g02600* or
 30 *Glyma18g02610*. Some copies of *Glyma18g02590* from *rhg1-b* resemble the Williams 82 sequence, while others contain a set of polymorphisms, notably at the predicted C-

terminal six amino acids of the predicted α -SNAP protein (Table 3, confirmed by cDNA sequencing).

Table 3. Amino acid polymorphisms within the 31 kb repeat at *rhg1-b*. The position of the protein-coding genes was predicted by comparing the fosmid clone sequences to the Glyma1 version of the soybean genome assembly. (L) or (R) indicate either the left or right side of the junction. * indicates an insertion of 1 amino acid (3 bp of DNA sequence) between amino acid position 287 and 288 based on the Williams 82 genome assembly (Glyma1; www.phytozome.net).

Gene ID	Amino Acid Position	W82	#1	#3	#4(L)	#5(L)	#5(R)
2590	203	Q	K	Q	K	-	K
2590	285	E	Q	E	Q	Q	Q
2590	286	D	H	D	H	H	H
2590	287	D	E	D	E	E	E
2590	287-288*	-	A	-	A	A	A
2590	288	L	I	L	I	I	I

The whole-genome sequencing (WGS) data were also analyzed for DNA polymorphisms such as insertions or deletions (INDELs) and single nucleotide polymorphisms (SNPs). In the seven genotypes with an estimated copy number ranging from 9 to 10, a number of SNPs were identified relative to the reference Williams 82 sequence. One gene contained in the repeated *rhg1-b* segment of DNA, *Glyma18g02590*, contains DNA polymorphisms relative to Williams 82 as defined above. There is a single SNP, C to A (Williams 82 to PI 88788 derived), at position 1,643,192 that results in a Q to K amino acid substitution. There are 3 SNPs present at the C-terminus, occurring at positions 1,644,965 (G to C), 1,644,968 (G to C), and 1,644,974 (C to A), and a 3 bp insertion after base 1,644,972 (GGC) that collectively change the final 5 amino acids of the Williams 82 protein from EEDLT to QHEAIT. The nucleotide and amino acid sequences are shown in Figure 13. The 3 bp insertion causes an extra amino acid in the PI 88788 derived lines. All base pair positions correspond to the Williams 82 genome version 1.1, assembly 189. Numerous SNP and INDEL polymorphisms were observed within the approximate 31 kb *Rhg1* repeat DNA region, between Williams 82 and PI 88788-source *Rhg1*, in the nucleotide regions outside of those that directly comprise the final open reading frame of *Glyma18g02580*, *Glyma18g02590*, *Glyma18g2600*, and

Glyma18g02610. Analyses of Illumina sequencing read depth, in the seven soybean lines from the NAM sequencing project with an estimated copy number ranging from 9 to 10, indicated that there were 9 very similar copies of the PI 88788-type repeat at *rhg1-b*, and one partial copy of a Williams 82-like repeat at *rhg1-b*.

5 The soybean line LD01-5907 from the soybean NAM parent sequencing project, which carries an estimated copy number of 3, also contains DNA polymorphisms affecting the amino acid sequence for *Glyma18g02590*. The DNA polymorphisms are different than those found in PI 88788 derived lines, but occur at similar positions. There is a SNP at position 1,643,225 that results in a D to E amino acid substitution. There are
10 2 SNPs present at the C-terminus, occurring at positions 1,644,968, (G to T) and 1,644,974 (C to A) and a 3 bp insertion after base 1,644,972 (GGT) that collectively change the final 5 amino acids of the Williams 82 protein from EEDLT to YEVIT. The 3 bp insertion causes an extra amino acid in the *Glyma18g02590* protein product in lines with an *Rhg1* locus derived from Peking or PI 437654 sources.

15 The DNA polymorphisms for *Glyma18g02590* identified through WGS analysis were confirmed to be expressed using 3' RACE and cDNA sequencing. In SCN resistant genotypes Cloud, PI 88788, and PI 209332, two different *Glyma18g02590* transcripts were identified. One of the sequences corresponded to the Williams 82 reference type sequence, and the other corresponded to the sequence from PI 88788-derived resistant
20 sources (from NAM parents). The proportion of PI 88788-derived versus Williams 82-type cDNA sequence follows that observed for DNA sequence. That is, the cDNA of PI 88788 derived *Glyma18g02590* is roughly 90% of the total transcripts sequenced. This is consistent with the data that these genotypes contain 8 or 9 copies of the 31kb DNA segment derived from PI 88788. A SNP present in the 5' UTR of *Glyma18g02610* was
25 also analyzed in PI 88788. The proportion of the sequence types fits the other observations.

In SCN-resistant genotypes Peking, PI 90763, PI 89772, and PI 437654 two different transcripts were identified for *Glyma18g02590*. One of the sequences corresponds to the sequence from the Peking/PI 437654-derived resistant source LD01-
30 5907 from the NAM sequencing project. See Figure 13. An alternative form of cDNA was also detected from each of the four SCN-resistant genotypes Peking, PI 90763, PI

89772, and PI 437654, with the same type of polymorphism across all four sources. This apparent mRNA splicing isoform had 36 nucleotides deleted resulting in a *Glyma18g02590* isoform with 12 fewer amino acids as shown in Figure 13. The deletion occurs at the end of exon 6 and splices back into frame in exon 7. None of the sequenced products from Peking, PI 90763, PI 89772, and PI 437654 contained the Williams 82-type *Glyma18g02590* sequence, consistent with the WGS analysis. Based on the proportion of cDNAs sequenced, very approximately 70% to 90% of the *Glyma18g02590* transcript is the full-length version in these lines.

The various polymorphisms may result in functional differences in the *Glyma18g2590* polypeptide and are modeled three-dimensionally in Figure 14 which relies on the solved crystal structure of the yeast Sec17 protein. The deleted alpha-helix is shown in light gray. It is noted that these polymorphisms are clustered in one general area near the C-terminus of the predicted folded *Glyma18g02590*, which is an alpha-SNAP protein homolog, and that substantial functional data are available for eukaryote alpha-SNAP proteins that suggest particular functions for this region of the protein (e.g., Barnard RJ, Morgan A, & Burgoyne RD (1996) Domains of alpha-SNAP required for the stimulation of exocytosis and for N-ethylmaleimide-sensitive fusion protein (NSF) binding and activation. *Molecular biology of the cell* 7(5):693-701; Barnard RJ, Morgan A, & Burgoyne RD (1997) Stimulation of NSF ATPase activity by alpha-SNAP is required for SNARE complex disassembly and exocytosis. *J Cell Biol* 139(4):875-883; Jahn R & Scheller RH (2006) SNAREs - engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* 7(9):631-643.)

However, expressing only the Fayette polymorphic *rhg1-b*-type *Glyma18g02590* downstream of a strong constitutive promoter or native promoter sequence did not increase the SCN resistance reaction of Williams 82 transgenic roots (Fig. 15 and Fig. 16), suggesting that *rhg1-b* SCN resistance requires more than this 2590 amino acid polymorphism. But such polymorphisms may play a contributing role in SCN resistance that was not detected in these experiments. Overexpression of *Glyma18g02580* or *Glyma18g02610* also failed to increase SCN resistance (Fig. 13) when expressed alone using a strong constitutive promoter. These data are preliminary and may indicate that

the resistance phenotype requires more than a single gene or that some other factor is necessary to mediate resistance.

Given the above, simultaneous overexpression of the set of genes within the 31 kb repeat segment was tested as a possible source of SCN resistance. A single recombinant DNA construct was made in which each of the genes *Glyma18g02580*, *Glyma18g2590*, *Glyma18g2600* and *Glyma18g2610* was fused to a strong promoter. In two separate experiments that together tested >25 independent transgenic events for each DNA construct, resistance to SCN was significantly increased in SCN-susceptible Williams 82 by simultaneous overexpression of this set of genes (Fig. 15). Increased SCN resistance was conferred despite the fact that three of the genes being overexpressed encode predicted amino acid products identical to those of SCN-susceptible Williams 82, and the polymorphic Fayette *rhg1-b Glyma18g02590* gene that was used was not sufficient to cause a detectable change in SCN resistance when overexpressed on its own (Fig. 15). Of note, there was no significant elevation of *PR-1* in these transgenic roots, which could have indicated non-specific elevation of defenses (Fig. 16). We also tested the impact of simultaneously over-expressing *Glyma18g02580*, *Glyma18g02590_{Fayette}*, and *Glyma18g02610* on SCN resistance in Williams 82. We observed increased resistance to SCN in transgenic roots of Williams 82 over-expressing *Glyma18g02580*, *Glyma18g02590_{Fayette}*, and *Glyma18g02610* relative to Williams 82 empty vector roots as shown in Figure 17. These data indicate that over-expressing this combination of genes results in enhanced SCN resistance.

These results reveal a novel mechanism for disease resistance: an expression polymorphism for multiple disparate but tightly linked genes, derived through copy number variation at the *Rhg1* locus. This knowledge suggests future approaches to enhance the efficacy of *Rhg1*-mediated quantitative resistance to the highly important SCN disease of soybean, for example through isolation of soybean lines that carry more copies of the 31 kb *Rhg1* repeat, or through transgenic overexpression of the relevant genes. These approaches may be applicable in other species as well, for resistance to other endoparasitic nematodes.

The biochemical mechanisms of *Rhg1*-mediated resistance remain unknown. Other sequenced plant genomes do not carry close homologs of the predicted

Glyma18g02610 protein, although a wound-inducible protein in ice plant with 55% identity has been studied. Modeling of the Glyma18g02610 predicted tertiary structure using Phyre2 indicated, with 98% confidence, similarity of 48% of Glyma18g02610 to the PhzA/B subfamily of Delta(5)-3-ketosteroid isomerase/nuclear transport factor 2 family proteins. Hence Glyma18g02610 may participate in the production of phenazine-like compounds that are toxic to nematodes. Thus application of Glyma18g02610 to plants, soil or seeds may inhibit nematodes in susceptible plants. Secretion of the Glyma18g02610 protein or other plant products that contribute to disease resistance may be impacted by the Glyma18g02590 α -SNAP protein. Because it is one of at least five α -SNAP homologs encoded in the soybean reference genome, Glyma18g02590 may have undergone subfunctionalization or neofunctionalization. Fully sequenced plant genomes carry from two dozen to over five dozen annotated amino acid transporters of many subtypes (www.phytozome.net), which can be involved in amino acid import and/or export between cells or between subcellular organelles. The Glyma18g02580 protein and its most closely related transporters of soybean and other species are not functionally well-characterized, so the concept that Glyma18g02580 alters nematode success by altering the levels of specific amino acids or amino acid derivatives at the feeding site is only one of many viable hypotheses for future study regarding the SCN-detering function of Glyma18g02580.

Copy number variation (CNV) of a block of dissimilar genes, rather than CNV for a single gene family, confers *Rhgl*-mediated SCN resistance. Recent analyses of genome-architecture in sorghum, rice, and soybean have reported high levels of CNV, and a tendency for overlap of regions of CNV with postulated biotic and abiotic stress-related genes. The present work provides a concrete example of CNV conferring a valuable disease resistance trait. In humans and insects, adaptive traits have been associated with CNV for specific single genes. Single-copy clusters of functionally related but non-homologous genes are highly unusual in multicellular eukaryotes, but these have been reported in association with plant secondary metabolism. We provide a unique example of CNV involving more than two repeats, with the repeat encoding multiple gene products that are necessary for adaptation to the same important environmental constraint. Given the highly repetitive nature and plasticity of plant

genomes and the relatively underexplored association between CNV and phenotypes, it seems likely that a number of other complex traits are controlled by the general type of CNV we report for soybean *Rhgl*.

Materials and Methods:

5 *Agrobacterium rhizogenes* Soybean Root Transformation

A. rhizogenes strain Arqual was transformed by freeze-thaw as previously reported by Wise, A.A., Z. Liu, and A.N. Binns, Three methods for the introduction of foreign DNA into *Agrobacterium*. *Methods Mol Biol*, 2006. **343**: p. 43-53 and Hofgen, R. and L. Willmitzer, Storage of competent cells for *Agrobacterium* transformation.

- 10 Nucleic Acids Research, 1988. **16**(20): p. 9877-9877. The cells were plated on selective media with the appropriate antibiotic and incubated at 28°C for two days. *A. rhizogenes* strain Arqual was received from Dr. Jean-Michel Ane, University of Wisconsin Madison. Soybean seeds lacking macroscopic signs of fungal or viral contamination were surface-sterilized for 16-20 h in a desiccator jar with chlorine gas generated by
- 15 adding 3.5 ml 12N HCl into 100 ml household bleach (6% sodium hypochlorite). At least 20 seeds per experiment were plated onto germination media (Gamborg's B5 salts (3.1g/L), 2% sucrose, 1X Gamborg's B5 vitamins, 7% Noble agar, pH 5.8) in 100 x 25 mm Petri plates. Plates were wrapped with Micropore tape (3M, St. Paul, MN) and incubated at 26°C in a growth chamber (18/6 light/dark hours) for approximately one
- 20 week. Soybean cotyledons were harvested 5-7 days after germination by gently removing them from the hypocotyls with sterile forceps. With a sterile forceps and Falcon #15 scalpel, several shallow slices were made across the abaxial surface of the cotyledons after dipping the scalpel in *A. rhizogenes* suspension (OD₆₀₀ 0.6 - 0.7 in sterile ddH₂O). The cotyledons were then placed abaxial-side down on a co-culture medium (CCM)
- 25 (0.31g/L Gamborg's B5 salts, 3% sucrose, 1X Gamborg's B5 vitamins (BioWorld, Dublin OH), 0.4g/L L-cysteine, 0.154g/L dithiothreitol, 0.245g/L sodium thiosulfate, 40mg/L acetosyringone, 5% Noble agar, pH 5.4) in 100 x 15 mm Petri plates with a piece of 70 mm filter paper (Whatman, Piscataway, NJ) on the surface of the agar to prevent *A. rhizogenes* from overgrowing. Plates were wrapped with parafilm and incubated in the
- 30 dark at room temperature for three days. The explants were then transferred to a hairy root medium (HRM) of 4.3g/L MS salts (Sigma Co., St. Louis, MO), 2% sucrose, 1X

Gamborg's B5 vitamins (BioWorld, Dublin, OH), 7% Noble agar, 0.15g/L cefotaxime, 0.15g/L carbenicillin, pH 5.6 in 100 x 15 mm Petri plates, wounded side up. Plates were wrapped with Micropore tape and incubated in the dark at room temperature until roots emerged, usually in around 2 weeks. Transgenic soybean roots were detected based on plasmid vector-encoded GFP expression, using a fluorescence stereomicroscope (LEICA MZ FL III with GFP2 filter). Transgenic soybean root tip segments (2-3 cm) were transferred to HRM. Roots that were expressing incomplete strips of fluorescence (chimeras) or exhibiting overall low levels of GFP fluorescence were avoided. Independent transgenic events, generated from different inoculation sites or different cotyledons, were maintained separately for RNA extraction and nematode demographic assays.

Nematode maintenance

An SCN population from Racine, Wisconsin (Hg type 7), collected by Ann MacGuidwin (University of Wisconsin-Madison), was maintained on the susceptible soybean cultivar Williams 82. Seeds were germinated between two damp pieces of paper towel that were rolled-up and placed vertically in a glass beaker with a small amount of water at the bottom for 2-4 days. Germinated seeds were then planted in autoclaved 4:1 sand:soil mixture and inoculated with 2000 eggs of *H. glycines* per plant, and grown in a 28°C growth chamber. Cysts were collected ~50 days after infection when soybeans were at R2 (full flowering) and extracted from soil and roots using sieves and centrifugation. Briefly, soil and roots from infected pots was placed in a pitcher of water and agitated. The soil-cyst-water slurry was passed over a 710µm - 250 µm sieve tower, and the mixture from the 250 µm sieve was backwashed into a 50mL plastic conical tube. The tubes were centrifuged at 2000 rpm for 4 minutes then the supernatant was poured off. A 60% sucrose solution was added to the tubes, stirred, and centrifuged at 2000 rpm for 2 min. Cysts in the supernatant were then collected over a 250 µm sieve. Collected cysts were stored at 4°C in sealable plastic bags containing twice-sterilized flint sand.

Nematode demographics assay

Nematode demographics assays were performed as in Melito et al, *infra*. Vigorous new root segments (2-3 cm including root tip) were utilized. All roots (all genotypes within an experiment) were coded with a random number prior to inoculation,

to mask root genotype information from the investigators who stained roots two weeks later and determined the number of nematodes in each nematode development category. For inoculum, *H. glycines* eggs were collected by breaking open cysts with a large rubber stopper and collecting the eggs on a sieve stack consisting of 250 μm - 75 μm - 25 μm 5 sieves (USA Standard Testing Sieve). Eggs were collected from the 25 μm sieve and rinsed. Eggs were placed in a hatch chamber with 3 mM ZnCl_2 for hatching at room temperature in the dark for 5-6 days. See Wong, A.T.S., G.L. Tylka, and R.G. Hartzler, Effects of 8 herbicides on in-vitro hatching of *Heterodera glycines*. Journal of Nematology, 1993. **25**(4): p. 578-584. Hatched J2 nematodes were surface-sterilized for 10 3 min in 0.001% mercuric chloride and washed three times with sterile distilled water, then suspended in room temperature 0.05% low-melting point agarose to facilitate even distribution. Baum, T.J., M.J.E. Wubben, K.A. Hardy, H. Su, and S.R. Rodermel, A screen for *Arabidopsis thaliana* mutants with altered susceptibility to *Heterodera schachtii*. Journal of Nematology, 2000. **32**(2): p. 166-173. The number of active 15 nematodes was determined by viewing an aliquot under a stereomicroscope at least one-half hour after surface-sterilization and washing, and 200-250 active J2s were inoculated onto each fresh root segment. Inoculated roots with nematodes were maintained on HRM media at 28°C; substantial root growth typically occurred during the subsequent two weeks. Nematode infection and development within these root systems was monitored 20 by clearing and staining with acid fuchsin, typically 15 days post inoculation (dpi). Bybd, D.W., T. Kirkpatrick, and K.R. Barker, An improved technique for clearing and staining plant-tissue for detection of nematodes. Journal of Nematology, 1983. **15**(1): p. 142-143. The nematode demographic assay was then completed by recording the number of nematodes in each root system that exhibited a morphology resembling either J2 (thin), 25 J3 (sausage-shaped), elongated male, or J4/adult female nematodes, as noted in text and figures. Typically, 20-80 nematodes were present in each root; roots containing fewer than ten nematodes were excluded from further analysis. Results were expressed as % of nematodes that had developed beyond J2 stage ($[\text{J3} + \text{adult males} + \text{adult females}] / [\text{J2} + \text{J3} + \text{adult males} + \text{adult females}]$). Each data point was normalized to the mean for 30 Williams 82 roots transformed with empty vector, from the same experiment. All

reported data are based on at least two independent biological replicate experiments (n > 12 independently transformed roots for each bar on a bar graph).

Primer Table

Primer sequences used to perform this research are listed in Table 4 and referred to by number in this document.

Table 4. DNA sequences of oligonucleotide primers used for PCR.

ID	Primer	Sequence	SEQ ID
Silencing constructs			
1	2570 hpRNAi _F	AGGATCCATTTAAATCAAGTACTCTTCCCCACAAAAGCT	19
2	2570 hpRNAi _R	ACCTAGGAGGCGCGCCTGGGGCCATTTAGTAATTAGGTC	20
3	2580 hpRNAi _F	acctaggaggcgcgccTCATGAAGGTTCTCGGCGTAG	21
4	2580 hpRNAi _R	aggatccatttaaataCCACCAGTGAATTCCAAACCA	22
5	2590 hpRNAi _F	GAcctaggcgcgccGGACTTGGTCGTCAACACAGTC	23
6	2590 hpRNAi _R	GCggatccatttaaataGAGCAGCAAACCTGGGCAACT	24
7	2600 hpRNAi _F	acctaggaggcgcgccGCCAAATTCAAAAGGCTTGCT	25
8	2600 hpRNAi _R	aggatccatttaaataCACCATTCAACATGCCTGTCA	26
9	2610 hpRNAi _F	taacctaggaggcgcgccACAACCTCCTTCGATTTCGTTCCG	27
10	2610 hpRNAi _R	caggatccatttaaataAGATACAACCACCTGAATACGCCC	28
11	2620 hpRNAi _F	AGGATCCATTTAAATCTCGCAACACCATATCCAGAGTA	29
12	2620 hpRNAi _R	ACCTAGGAGGCGCGCCGGTGTTAAGGTGCAACCTGCGAA	30
13	2590-1 I miR-s	gaTATTGGTTATAGCAACACCGTtctctcttttgattcc	31
14	2590-1 II miR-a	gaACTTTGCTATAACCAATAtcaaagagaatcaatga	32
15	2590-1 III miR*s	gaACAGTGTTGCTATTACCAATTtcacaggtcgtgatatg	33
16	2590-1 IV miR*a	gaAATTGGTAATAGCAACACTGTtctacatatatattcct	34
17	2610-1 I miR-s	gaTATTTCCCGACCCGACGGGACtctctcttttgattcc	35
18	2610-1 II miR-a	gaGTCCCGTCGGGTTCGGGAAATAtcaaagagaatcaatga	36
19	2610-1 III miR*s	gaGTACCGTCGGGTTCGGGAAATTtcacaggtcgtgatatg	37
20	2610-1 IV miR*a	gaAATTTCCCGACCCGACGGTACtctacatatatattcct	38
21	2610-2 I miR-s	gaTATCCAGTCACCGCGACGTGGtctctcttttgattcc	39
22	2610-2 II miR-a	gaCCACGTCGCGGTGACTGGATAtcaaagagaatcaatga	40
23	2610-2 III miR*s	gaCCCCGTGCGGGTGTCTGGATTtcacaggtcgtgatatg	41
24	2610-2 IV miR*a	gaAATCCAGACACCGCGACGGGGTctacatatatattcct	42
Transcript Abundance using qPCR			
25	gm18: 2570 F_qPCR	TGAGATGGGTGGAGCTCAAGAAC	43
26	gm18: 2570 R_qPCR	AGCTTCATCTGATTGTGACAGTGC	44
27	gm18: 2580 F_qPCR	CGTGTAGAGTCCTTGAAGTACAGC	45
28	gm18: 2580 R_qPCR	ACCAGAGCTGTGATAGCCAACC	46
29	gm18: 2590 F_qPCR	TCGCCAAATCATGGGACAAGGC	47
30	gm18: 2590	CAATGTGCAGCATCGACATGGG	48

	R_qPCR		
	gm18: 2600		49
31	F_qPCR	GCTTCAGTCAAGAAAATGTGCATG	
	gm18: 2600		50
32	R_qPCR	CACCCGAAACCGCGACACAAATG	
	gm18: 2610		51
33	F_qPCR	AGGTCACGTGTTGCCGTTG	
	gm18: 2610		52
34	R_qPCR	AAACCACACCAATAACAACAAAGCTCT	
	gm18: 2620		53
35	F_qPCR	AAGCCCAACAGGCCAAAGAGAG	
	gm18: 2620		54
36	R_qPCR	ACACCAAATGGGTTTCGCACTTC	
	gm18: 2630		55
37	F_qPCR	TTGTGGAAGTGAAAGTCGGTTTGC	
	gm18: 2630		56
38	R_qPCR	GTTGTCACGTTTCCCGTAACAATG	
39	EF1b_FoR.qRT	CCACTGCTGAAGAAGATGATGATG	57
40	EF1b_Rev.qRT	AAGGACAGAAGACTTGCCACTC	58
41	SKIP16_FoR.qRT	GAGCCCAAGACATTGCGAGAG	59
42	SKIP16_Rev.qRT	CGGAAGCGGAAGAACTGAACC	60
43	UKN2_FoR.qRT	GCCTCTGGATACCTGCTCAAG	61
44	UKN2_Rev.qRT	ACCTCCTCCTCAAACCTCCTCTG	62
45	ACT11_FoR.qPCR	ATCTTGACTGAGCGTGGTTATTCC	63
46	ACT11_Rev.qPCR	GCTGGTCCTGGCTGTCTCC	64
47	UNK1_FoR.qPCR	TGGTGCTGCCGCTATTTACTG	65
48	UNK1_Rev.qPCR	GGTGGAAGGAAGTCTAACAATC	66
49	TIP41_FoR.qPCR	AGGATGAAGTCTGCTGATAATGG	67
50	TIP41_Rev.qPCR	CAGAAACGCAACAGAAGAAACC	68
51	PR-1 (6790) F	TGCTTGGTCACCTGGAAGTTGG	69
52	PR-1 (6790) R	AACTTCCTGCGAGCTGCGATAC	70
53	PR-1 (6800) F	AGTCATTGTGGGTGATCATGCTG	71
54	PR-1 (6800) R	GCAGCGTTGTGTGCATTAACAAAG	72

Expression Vectors

55	Ox2610-SalF2	gtcgacATGCGCATGCTCACCGG	73
56	P2610fused-R	TATTGCGAGAACCAAACCGG	74
57	Ox2590Sal-F	GGgtcgacATGGCCGATCAGTTATCGAAGG	75
58	Ox2590fused-R	AGTAATAGCCTCATGCTGCTCAAGTT	76
59	TerXba-R	ACtctagaGCGCATGTCTTGCGTTGATG	77
60	GmubiXba-F	GCtctagaGGGCCCAATATAACAACGACG	78
61	TerKpn-R	TCggtaccGCGCATGTCTTGCGTTGATG	79
62	PPA Linker_Top	GatgtcTTAATTAAtatctgtGGGCCcactatGGCGCGCCaatgtaaA	80
	PPA	AGCTTTtacattGGCGCGCCatagtGGGCCCacagata	81
63	Linker_Bottom	TTAATTAAGacatCTGCA	
64	Ox2600-F	ATGGTTTCGGTTGATGATGGG	82
65	Ox2600-R	TTTTTGTCATATAAGGGGTTTCAT	83
66	NosHind-F	GCaagcttGATCATGAGCGGAGAATTAAGGG	84
67	Nos2600-R	CCCATCATCAACCGAAACCATAGATCCGGTGCAGATTATTTGG	85
68	Nos2600-F	CCAAATAATCTGCACCGGATCTATGGTTTCGGTTGATGATGGG	86
69	NosAsc-R	TCggcgcgccGCGCATGTCTTGCGTTGATG	87

70	Ox2580-F	ATGTCTCCGGCCGCGCCG	88
71	Ox2580-R	TGACTTGCTACTAAAAGCATTATATATGTTG	89
72	NosAsc-F	CAGgcgcgccGATCATGAGCGGAGAATTAAGGG	90
73	Nos2580-R	CGGCGGCCGGAGACATAGATCCGGTGACAGATTATTTGG	91
74	Nos2580-F	CCAAATAATCTGCACCGGATCTATGTCTCCGGCCGCGC	92
75	NosSbf-R	TGcctgcaggGCGCATGTCTTGCGTTGATG	93
76	M13 F	GTAAAACGACGGCCAG	94
77	M13 R	CAGGAAACAGCTATGAC	95
78	pSM101 seq	GTCTTGATGAGACCTGCTGCG	96
79	g2590pHind-F	CTTaagcttGAATGGTTTTTGTGTTGTCTCTCAC	97
80	g2590pSal-R	TTGGTCGACCGTATCATCCAATG	98
Bridge PCR			
	SCN_Res Bridge		99
81	F	TTTAGCCTGCTCCTCACAAATTC	
	SCN_Res Bridge		100
82	R	TTGGAGAATATGCTCTCGGTTGT	
Probes for Northern Analysis and Alt Transcript			
83	2570F_qPCR	TGAGATGGGTGGAGCTCAAGAAC	101
84	2570 UTR Rev	CAAGTACTCTTCCCCACAAAAGC	102
85	2570 put exon F	TGCAGTTTTAGTGGAAAGGCC	103
86	2570 exon 6 R	TCATCAAGCTCAACTTGAATCCC	104
RACE PCR			
87	2590-5GSP	GATCGGCCATTTTCCTCCGATCGAAACA	105
88	2590-5NGSP	GACGACCAAGTCCAAATCCAAAACCCGC	106
89	2590-3GSP	AAGCCAAAGAACTTGAGCAGCATGAGGC	107
90	2590-3NGSP	CTGTCCAGTTGTTCTGTTACACATCCA	108
91	2610-5GSP	GGCGACGATCTTGACGACGGCGTT	109
92	2610-5NGSP	TCATACAGTGCAACCACCAGCCGCG	110
93	2610-3GSP	GGACGAGGTCACGTGTTGCCGTTGCT	111
94	2610-3NGSP	TTCACCACTATGGGCGTATTCAGGTGGT	112
95	2580-3GSP	CCTGGGGGATTCCAAAGGAACGC	113

Vector Construction for Soybean transformation

Binary vectors pSM101 and pSM103 for soybean transformation were constructed as previously described in Melito et al. To generate and clone soybean amiRNAs, the Web microRNA Designer (<http://wmd3.weigelworld.org>) and protocols were used. The concept is more thoroughly documented in other references. See Schwab, R., S. Ossowski, M. Riester, N. Warthmann, and D. Weigel, Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell*, 2006. **18**(5): p. 1121-1133. Soybean DNA was extracted from either expanding soybean trifoliates or soybean roots using a previously reported CTAB method. Doyle, J.J. and E.E. Dickson,

Preservation of plant-samples for DNA restriction endonuclease analysis. *Taxon*, 1987. 36(4): p. 715-722. PCR fragments for amiRNA construction were TA cloned using pCR8/GW/TOPO TA cloning kit (Life Technologies Corp., Carlsbad CA) (Table 4 13-24). Binary vectors pGRNAi1 and pGRNAi2 for soybean transformation were a gift
5 from Wayne Parrot, University of Georgia (unpublished). For each hairpin, a 300-600bp DNA fragment was PCR amplified (Table 4 1-12) using Phusion HF polymerase (New England Biolabs, Ipswich, MA) and iScript cDNA synthesis kit (Biorad, Hercules, CA) as a template, as per manufacturer's instructions. PCR products were TA cloned as previously described. Primers used to generate the DNA fragments were designed to
10 contain restriction sites AvrII/AscI (forward primer) and BamHI/SwaI (reverse primer) to allow cloning into pGRNAi1 and pGRNAi2. To generate the first arm of the hairpin, the insert and vector were sequentially digested with restriction endonucleases SwaI and AscI using manufacturer's recommended protocol (New England Biolabs, Ipswich, MA). DNA was separated on a 1.0% agarose gel stained with ethidium bromide, and respective
15 DNA fragments were gel purified using Qiaquick gel extraction kit (Qiagen, Valencia, CA) and ligated together overnight at 4°C using T4 DNA ligase (Promega, Madison, WI). The same procedure was used to insert the second arm of the hairpin construct using the restriction endonucleases BamHI and AvrII. To construct single gene overexpression vectors for *Glyma18g02580* (Table 4 70, 71), *Glyma18g02590* (Table 4
20 57, 58) and *Glyma18g02610* (Table 4 55, 56), full-length ORFs were PCR amplified from cDNA of Fayette using Phusion HF polymerase and TA cloned in pCR8/GW/TOPO as previously described. *Glyma18g02600* (Table 4 67, 68) was cloned from genomic DNA by similar methods, as no *Glyma18g02600* cDNA could be detected in root cDNA libraries. The *Glyma18g02610* and *Glyma18g02590* ORFs were recombined with
25 pGWB14 (CaMV 35S promoter, 6X HA-NOS terminator) using LR clonase reaction (Life Technologies Corp., Carlsbad, CA) per manufactures instructions. See Nakagawa, T., T. Kurose, T. Hino, K. Tanaka, M. Kawamukai, Y. Niwa, K. Toyooka, K. Matsuoka, T. Jinbo, and T. Kimura, Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng*,
30 2007. **104**(1): p. 34-41. *Glyma18g02610* (Table 4 55, 59) was PCR amplified from pGWB14 and TA cloned into pCR8. This vector and pSM103 were digested with

XbaI/KpnI and ligated to yield *GmUbi_{prom}:2610-HA:NOS_{term}* (OE:2610-HA). The same procedure was used for *Glyma18g02590* (Table 4 57, 59), except the amplicon contained XbaI/Sall sites and was TA cloned into pCR8. *2590-HA:NOS_{term}* and pSM103 were digested with XbaI/Sall and ligated to yield *GmUbi_{prom}:2590-HA:NOS_{term}* (OE:2590-HA). The full OE:2590-HA was also digested (XbaI/Sall) and ligated into pSM103 containing OE:2610-HA to yield OE:2610-OE:2590. To generate the four gene overexpression construct, the restriction sites PacI, PspOMI, and AscI were added to pSM101 between sites PstI/HindII by annealing oligos (Table 4 62, 63) to generate pSM101+. The two gene overexpression cassette (OE:2610- OE:2590) was moved to the new pSM101+ using the restriction enzymes PstI/KpnI and ligation. A Nos promoter was added to *Glyma18g02600* in pCR8 using overlap PCR (Table 4 65-68) and TA cloned into pCR8. This vector was recombined with pGWB16 (no promoter, 4xMyc-NOS terminator) in an LR clonase reaction to yield *Nos_{prom}:2600-myc:Nos_{term}* (OE:2600-myc). OE:2600-myc was PCR amplified (Table 4 66, 72) and TA cloned into pCR8, and subcloned into pSM101+ (OE:2610-OE:2590) using restriction enzymes HinIII/AscI to yield the three gene overexpression vector (OE:2610-OE:2590-OE:2600). A Nos promoter was added to *Glyma18g02580* in pCR8 using overlap PCR with primers 71-74 and TA cloned into pCR8. This vector was used with pGWB16 in an LR clonase reaction to yield *Nos_{prom}:2580-myc:Nos_{term}* (OE:2580-myc). OE:2580-myc was amplified (Table 4 72, 75) and TA cloned, then subcloned into the three gene overexpression vector resulting in the four gene overexpression vector pSM101+ OE:2610-OE:2590-OE:2600-OE:2580. The native Fayette *Glyma18g02590* (*2590_{FayP}:2590_{Fay}*) construct for Williams 82 complementation was subcloned from a fosmid containing the desired allele. A 6.5 kb DNA fragment containing the PI 88788 *Glyma18g02590* was isolated from a fosmid following Sall digestion and cloned into pSM101 using the Sall restriction site. This sequence contained approximately 1kb of 5' regulatory DNA sequence. An additional 600bp of 5' regulatory sequence directly upstream of the subcloned region was added to the construct by amplifying a PCR product (Table 4 79, 80) from the fosmid and inserted using the restriction enzymes HindIII/Sall. The resulting construct contained approximately 1.6 kb of naturally occurring 5' regulatory sequence of the Fayette *Glyma18g02590* allele. Vector sequences were confirmed at various steps using Sanger

sequencing with ABI Big Dye cycle sequencing kit (dideoxy chain-termination) and ABI 3730xl DNA Analyzers (Life Technologies Corp., Carlsbad, CA), using the DNA sequencing service at the University of Wisconsin-Madison Biotechnology Center.

Quantitative Real Time PCR

5 Quantitative PCR (qPCR) was performed using either the MyIQ or CFX96 real-time PCR detection system (BioRad, Hercules, CA). cDNA was synthesized from RNA using iScript cDNA synthesis kit (Biorad, Hercules, CA) per manufactures protocol by adding 0.825 ug to 1.0 ug of RNA depending on the experiment. Total RNA was extracted from root tissue of conventional and transgenic soybeans. RNA was extracted
10 from conventional soybean plants grown in Metro mix for two weeks at 26°C and 16 hours light prior to tissue collection. Roughly 200 mg of tissue was collected from each plant, immediately flash-frozen in liquid nitrogen and stored at -80C. Transgenic root material was collected from roots actively growing on HRM as previously described. Roughly 50-100 mg of tissue was collected from each root, flash frozen in liquid nitrogen
15 and stored at -80C. RNA was extracted using either the RNeasy Mini Kit (Qiagen, Valencia, CA) or TRIzol reagent (Life Technologies Corp., Carlsbad, CA) following manufactures protocols. RNA concentrations were determined using the NanoDrop-1000 spectrophotometer (Thermo Scientific, Waltham, MA). DNA was removed from RNA samples using either RNase-free DNase I (Qiagen, Valencia, CA) or DNA-free (Life
20 Technologies Corp., Carlsbad, CA) following manufacture protocols. RNA integrity was determined using the 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA) or 500 ng of total RNA was run on a 1.2% agarose gel stained with ethidium bromide and visualized under UV-light to ensure RNA quality following extraction. qPCR reactions were carried out using either IQ SYBR Green Supermix or SsoFast EvaGreen Supermix
25 (Biorad, Hercules, CA). Primer concentrations for all reactions were between 0.2 μ M and 0.3 μ M. Two technical replicates were run per RNA. Efficiency curves were generated for qPCR primer pairs using cDNA from the cultivar Fayette or Williams 82 following a 3-4 step, 3-5 fold dilution. Following amplification, a melt curve program was performed. To ensure qPCR fluorescent signal was not the results of DNA, 100 ng
30 of RNA extraction was added directly to IQ SYBR Green Supermix or SsoFast EvaGreen Supermix with primers. DNA contamination was considered negligible if CT values

were not detected until after 32-35 cycles. A control reaction was run in parallel using a known cDNA sample. Transcript abundance for genes at *Rhg1* was measured using primers X-X. A total of six primer pairs were tested as reference genes (*EF1B*, *SKIP16*, *UNK2*, *ACT11*, *UNK1*, *TIP41*) (Table 4 39-50). Hu, R.B., C.M. Fan, H.Y. Li, Q.Z.

5 Zhang, and Y.F. Fu, Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *Bmc Molecular Biology*, 2009, **10**: p. 12. Reference genes were validated using Bestkeeper analysis. Pfaffl, M.W., A. Tichopad, C. Prgomet, and T.P. Neuvians, Determination of stable housekeeping genes, differentially regulated target genes and sample integrity:

10 BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters*, 2004, **26**(6): p. 509-515. Primer pairs *SKP16* and *TIP41* were selected and used in subsequent experiments. Transgenic roots expressing empty-vector constructs analogous to the vectors carrying gene silencing or gene expression constructs were included in the experiments as controls and used to standardize gene expression. Results were
15 considered to be at the limits of detection if CT values were >35 (i.e., for *Glyma18g02600* transcripts).

DNA repeat junction analysis

The presence of a repeat junction was confirmed using PCR (Table 4 81, 82) and soybean genomic DNA from SCN resistant cultivars Fayette, Hartwig, Newton and SCN
20 susceptible cultivars Williams 82, Essex, Thorne and Sturdy. DNA extraction and PCR were performed as previously described. Possible impacts of retrotransposons on *Rhg1* locus evolution were investigated by searching for sequences with similarity to known plant retrotransposons. A 185 bp sequence with 75% identity to the 5' and 3' long terminal repeat (LTR) regions of Ty1/copia-like retrotransposons RTvr1 and RTvr2 is
25 present within 400 bp of the *rhg1-b* duplication junction.

Statistical analysis

Data were analyzed by ANOVA using Minitab (v.14) with the General Linear Model and Tukey Simultaneous Test.

Fosmid library construction

30 Seed of soybean Plant Introduction (PI) 88788 was obtained from the USDA soybean germplasm collection. Plants were grown in a growth chamber set at a photocycle

of 18/6 hr (day/night), 23/20°C (day/night), and 50% relative humidity for 1-2 weeks. Young leaf tissue was collected from six to 15 individuals for each line. Genomic DNA was extracted using cetrimonium bromide (CTAB). Plant samples were ground to fine powder in liquid nitrogen, transferred to 20 ml of CTAB extraction buffer (2% CTAB, 100 mM Tris pH 9.5, 1.4 M NaCl, 1% PEG 6000, 20 mM EDTA, 2% polyvinylpyrrolidone, 2.5% β -mercaptoethanol), and placed at 65 °C for 1 hr. After incubation, an equal volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1, pH 6.7) was added to the tube, then centrifuged at 8,000 g at 10°C for 10 min. The aqueous (top) phase was transferred to a new tube and an equal volume of chloroform:isoamyl alcohol (24:1) was added to the aqueous phase and centrifuged. The aqueous (top) phase was then transferred to a new tube and 0.7 volumes of isopropyl alcohol was added to the aqueous phase. After mixing well, the aqueous phase was centrifuged and the pellet resuspended in 70% EtOH, centrifuged at 7,500 g for 10 min. After centrifugation, the pellet was resuspended in 100 μ l of TE (10 mM Tris pH 7.5, 1mM EDTA). The DNA was treated with RNase A by incubating in 20 μ g/ml RNase A at 37°C for 1 hr. The PI 88788 fosmid library was constructed using the CopyControl™ Fosmid Library Production Kit (Epicentre, Madison, WI) following the manufacturer's protocol. Briefly, 20 μ g of the size-fractionated DNA was used for end-repair. 35-45 kb fragment pools of DNA were cloned in the pCC1FOS™ Vector. Ligated DNA was packaged using the MaxPlax™ Lambda Packaging Extracts and transformed into the Phage T1-Resistant EPI 300™-T1^R *E. coli* strain.

Fosmid clone sequencing and assembly

Five candidate fosmid clones were identified by PCR-based pool screening using primers based on the *rhg1-b* interval of the Williams 82 reference sequence. Once it was confirmed that end sequences matched the anticipated region of the reference soybean genome sequence, they were sequenced using both the Roche 454/GS FLX+ system (Roche) and Illumina MiSeq (Illumina). 1-3 μ g of fosmid clone DNA was used for making paired-end sequencing libraries for 454/GS FLX+. After library construction, pooled barcoded libraries were loaded onto one lane of the sequencing flow cell and sequenced. The average read length was 463 bp. The number of reads generated from 454/GS FLX+ is as follows: fosmid clone #1 in Fig. 2A: 10,865, #2: 6,271, #3: 6,648, #4:

6,520, and #5: 9,390. The reads were assembled using Phrap/Cross_match (www.phrap.org) and CAP3. Huang, G.Z., R. Allen, E.L. Davis, T.J. Baum, and R.S. Hussey, Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(39): p. 14302-14306. For the MiSeq, 0.3-2 ug of DNA was used for making the sequencing library. Average DNA fragment size was 550 bp (range from 430 to 720 bp). 154 cycles from each end of the fragments were performed using a TruSeq SBS sequencing kit version 1 and analyzed with Casava1.8 (pipeline 1.8). Throughout the reads, the average quality scores for each base were over 30. The number of reads generated from MiSeq is as follows: fosmid clone #1 in Fig. 2A: 1,067,403, #2: 814,728, #3: 1,156,784, #4: 1,091,852, and #5: 946,028. ABySS was used to assemble the reads from MiSeq. Simpson, J.T., K. Wong, S.D. Jackman, J.E. Schein, S.J.M. Jones, and I. Birol, ABySS: A parallel assembler for short read sequence data. Genome Research, 2009. **19**(6): p. 1117-1123. The result was visualized using Geneious. Homopolymeric sequences and other problematic regions were manually sequenced using Sanger primer walking.

Whole-genome shotgun sequencing and read depth in duplicated region

Whole-genome shotgun sequencing of a soybean breeding line LD09-15087a, a near-isogenic line (NIL) that harbors *rhg1-b* from PI 88788, was conducted using Illumina technology. 1.5 ug of genomic DNA was sequenced using the Illumina HiSeq 2000 instrument with 100 bp paired-end sequencing at the University of Illinois Biotechnology Center. The DNA fragment size for the soybean whole-genome shotgun sequencing library was 600 bp; the library was loaded onto one lane of a flow cell and sequenced using version 3 of sequencing kits and Casava 1.8 (pipeline 1.9). 312,909,668 reads (about 28 × coverage of the 1.1 gb soybean genome) were generated with all positions having average quality scores 30 or higher. To examine the depth of the coverage within the duplicated region, reads from the sequencing were aligned to the Glyma1 version of the soybean genome assembly. Novoalign (v 2.08.01) (<http://www.novocraft.com>) with paired end options (PE 600,120) was used to align the reads to the reference genome. Approximately 95.1% of reads were aligned to the reference sequence. The number of reads aligned to the target interval was counted from

a BAM file using SAMtools (v 0.1.18). Target interval is as follows: “Block” in Fig. 5B: a 31.2 kb region (1,632,225-1,663,455 on chromosome 18), “Block-1”: the same size region as region of interest upstream, and “Block+1”: the same size region as region of interest downstream. Homeologous regions on chromosome 11 (“Block” in Fig. 5B: 37,392,345-37,434,356 bp) and 2 (“Block”: 47,772,323-47,791,521 bp) were identified using BLASTN. Analogous approaches were used with the soybean NAM parent sequence data. Table 5 shows the amino acid sequences for the four genes.

Table 5: Amino acid sequences for Glyma18g2580, Glyma18g2590, Glyma18g2600 and Glyma18g2610

>Glyma18g02580.1|PACId:16307711 (SEQ ID NO: 1)
 MSPAAGVSVPLLGDSKGTTPPASVPGAVFNVATSIVGAGIMSIPAIMKVLGVVPA
 FAMILVAVLAELSVDFLMRFTHSGETTTYAGVMREAFSGGALAAQVCVIITN
 VGGLILYLIHGDVLSGKQNGGEVHLGILQQWFGIHWNSREFALLFTLVFVMLP
 15 LVLYKRVESLKYSSAVSTLLAVAFVGICCGLAITLVQGKTQTPRFLPRLDYQTS
 FFDLFTAVPVVVTAFTHFENVHPIGFELAKASQMTTAVRLALLCAVIYLAIGLFG
 YMLFGDSTQSDILINFQDQNAAGSAVGSLNSLVRVSYALHIMLVFPLNFSRLRTNID
 EVLFPKKPMLATDNKRFMILTLLVLLVFSYLAALAIIPDIWYFFQFLGSSSAVCLAFIF
 PGSIVLRDVKGISTRDKIHALIMILAVVTSVLAISTNIYNAFSSKS*

>Glyma18g02590.1|PACId:16307712 (SEQ ID NO: 2)
 MADQLSKGEEFEKKAEEKLSGWGLFGSKYEDAADLFDKAANCFKLAKSWDKA
 GATYLKLASCHLKLESKHEAAQAHVDAAHCYKKTNINESVSCLDRAVNLFCDIG
 RLSMAARYLKEIAELYEGEQNIQALVYYEKSADFFQNEEVTTSANQCKQKVAQ
 25 FAAQLEQYQKSIDIYEEIARQSLNNLLKYGVKGHLLNAGICQLCKEDVVAITNA
 LERYQELDPTFSGTREYRLLADIAAAIDEEDVAKFTDVVKEFDSTPLDSWKTTL
 LLRVKEKLKAKELEEDDLT*

>Glyma18g02610.1|PACId:16307714 (SEQ ID NO: 3)
 MRMLTGDSAADNSFRFVPQSIAAFGSTVIVEGCDSARNIAWVHAWTVTDGMITQ
 30 IREYFNTALTVTRIHDSGEIVPARSG

>Glyma18g02600.1|PACId:16307713 (SEQ ID NO: 4)
 MVSVDGIVNPNDIEKSNKSVNEFASMDISATQKSYLNSEDQRRRLQGTLISSS
 35 VTNRINFLKFGSASAKFKRLATERDQVSISVSPRSKSLRSRFSGMFAQKLDWAS
 VKKMCMEWIRNPVNMAFVWIIICVAVSGAILFLVMTGMLNGVLPRKSKRNAW
 FEVNNQILNAVFTLIPNDISSLRKVYCKNVTYKPHWTHMMVVVILLHVNCFAQ
 YALCGLNLGYKRSEPAIGVGICISFAIAGLYTILSPLGKDYDCEMDEEAQVQITA
 SQGKEQLREKPTTEKKYSFASKDQQRVVENRPKWGGILDINDISLAYLSLFCFTF
 40 CVLGWNMKRLGFGNMYVHIAIFMLFCMAPFWIFLLASVNIDDDNVRQALAAVG
 IILCFLGLLYGGFWRIQMRKRFLPAYDFCFGKPSASDCTLWLPCWCWSLAQEAR
 TRNNYDLVEDKFSRKETDTSQPSISPLAREDDVVSTRSGTSSPMGSTSNSSPYMM
 KTSSSPNSSNVLKGYYSPDKMLSTLNEDNCERGQDGTMNPLYAQK*

Fiber-FISH

Soybean nuclei were lysed to release large chromosomal segments and, in contrast to more standard FISH methods, the chromosome segments were decondensed to generate extended DNA fibers before fixing to microscope slides and hybridizing to fluorescently labeled DNA probes. Young leaf tissues were collected from fast growing plants of Williams 82, Peking, and Fayette. Nuclei isolation, DNA fiber preparation, and fiber-FISH were performed following published protocols. Jackson, S.A., M.L. Wang, H.M. Goodman, and J. Jiang, Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. Genome, 1998. 41(4): p. 566-72. A fosmid clone spanning an *rhg1-b* repeat from PI 88788 was digested using the exonuclease SmaI (New England Biolabs, Ipswich, MA). The products of the restriction digestion were separated in a 0.7% gel and isolated using the Qiaex II gel extraction kit (Qiagen, Valencia, CA). DNA probes were labeled with either biotin-16-UTP or digoxigenin-11-dUTP (Roche Diagnostics, Indianapolis, IN) using a standard nick translation reaction. The fiber-FISH images were processed with Meta Imaging Series 7.5 software. The final contrast of the images was processed using Adobe Photoshop CS3 software. The cytological measurements of the fiber-FISH signals were converted into kilobases using a 3.21 kb/ μ m conversion rate.

Transcript analysis

To confirm the annotation of transcripts at *Rhg1*, rapid amplification of cDNA ends (RACE) PCR was performed for *Glyma18g02580* (Table 4 95), *Glyma18g02590* (Table 4 87- 90) and *Glyma18g02610* (Table 4 91- 94) using the SMARTer RACE cDNA kit per manufacturer protocols (ClonTech, Mountain View, CA). Following RACE, PCR products were TA cloned into pCR8/GW/TOPO as previously mentioned. Randomly chosen colonies were sequenced (Table 4 76, 77) as described to confirm the 5' and 3' ends of individual transcripts. To detect potential transcript isoforms, northern analysis was conducted using standard methods. Probes were generated for *Glyma18g02570* (Table 4 83, 84). Absence of truncated *Glyma18g02570* transcripts (Table 4 85, 86) derived from 31.2 kb repeat junctions was also confirmed by PCR from cDNA, using a 2570 reverse primer and a forward primer in the most strongly predicted exon upstream of the repeat junction. Hebsgaard, S.M., P.G. Korning, N. Tolstrup, J.

Engelbrecht, P. Rouze, and S. Brunak, Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Research*, 1996. **24**(17): p. 3439-3452. Transcript abundance studies using qPCR also indicated that there is not a *Glyma18g02570*-like transcript produced by the repeated DNA

5 insertion. *Glyma18g02570* transcript abundance was measured using primers (Table 4 25, 26) that amplify the final two exons and hence should amplify both the reference genome (full-length; Williams 82-like) *Glyma18g02570* transcript and possible hybrid *Glyma18g02570* transcripts that are transcribed from DNA that spans the repeat junction. If the repeated DNA produced an alternative transcript, these primers would amplify
10 additional product from genotypes with the repeat. However, no differences in transcript abundance were detected between SCN-resistant vs SCN-susceptible varieties using *Glyma18g02570* primers 25 and 26.

Protein structure prediction and comparison

The protein structure for the predicted *Glyma18g02610* gene product was
15 modeled and proteins with the most homologous structures were identified using Phyre2, with default settings. Kelley, L.A. and M.J.E. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*, 2009. **4**(3): p. 363-371.

Methylation Analysis

Locus specific DNA methylation was analyzed using the methylation specific
20 endonuclease McrBC, or the methylation sensitive endonuclease HpaII followed by PCR. McrBC (New England Biolabs, Ipswich, MA) digests DNA with methylated cytosines in a sequence-independent manner while unmethylated DNA is not digested. HpaII (New England Biolabs, Ipswich, MA) digests DNA at the recognition sequence CCGG, but HpaII endonuclease activity is blocked by cytosine methylation. Restriction digestions
25 were performed using 600-700ng of DNA and manufacturer's protocols. Control reactions were set up by adding the same amount of DNA to the reaction buffer with no restriction enzyme. Samples, with and without the restriction enzyme, were incubated at 37°C for 90 minutes, and heat inactivated at 65°C for 20 minutes. DNA was visualized in a 0.8% ethidium bromide stained gel to ensure DNA digestion. Both digested and
30 control DNA samples were used for subsequent PCR using GoTaq Flexi DNA polymerase (Promega, Madison WI). For DNA treated with McrBC, PCR primers that

spanned methylated DNA would not produce the intended product following PCR because the template DNA would be digested by McrBC. DNA that was not methylated or not treated with the enzyme yielded a product of the expected size. For DNA treated with HpaII, PCR primers that spanned the DNA sequence CCGG in which either cytosine was methylated yielded a PCR product of the expected size. DNA sequence CCGG that was not methylated was cleaved by HpaII and failed to yield a PCR product. DNA incubated in buffer without HpaII yielded expected PCR products. See table in Appendix F and figure for primer details and results.

Western Blot Analysis

Protein size and abundance were measured using Western blot and immunodetection procedures (Ausubel et al. 1997). Briefly, protein was extracted from roots of transgenic soybeans by homogenizing frozen root tissue and re-suspending the material in 2x Tricine sample buffer (0.1M TrisCl/0.3%SDS pH6.8, 24% glycerol, 8% SDS, 0.2M DTT) at 1:1 w/v ratio. An equal volume of each protein sample was separated in a Tris-Tricine polyacrylamide gel (9.8% separation gel, 3.9% stacking gel) using electrophoresis in the Biorad Mini Protean3 cassette (Biorad, Hercules CA). The samples are separated at 35 volts for roughly one hour, followed by another hour at 160 volts. The gel was moved to a transfer cassette and aqueous transferred to a Protran nitrocellulose membrane (Whatman, Piscataway, NJ). The transfer was run for an hour at 80 volts at room temperature. Following transfer, membranes were stained for total protein using 0.1% Ponceau S (Sigma-Aldrich, St. Louis, MO) in 5% acetic acid and imaged. Ponceau S was destained in ddH₂O, and the membrane was blocked over night at 4°C in TBST (20mM Tris pH7.5, 8g/L NaCl, 0.1% Tween) carrying 5% milk. New 5% milk in TBST was added to the membrane and placed on shaker at room temperature for 30 minutes. The membrane was incubated with HA primary antibody directly conjugated to horse radish peroxidase (HRP) at a 1:1000 concentration in 5% milk TBST for 90 minutes. The membrane was washed 3x in TBST at room temperature on a shaker for 20 minutes each. Supersignal West Dura Extended Duration Substrate (Thermo Scientific, Waltham, MA) ECL kit was used following manufacture protocols to detect the HA-HRP antibody on the membrane. The membrane was exposed to Cl-Xposure film (Thermo Scientific, Waltham, MA) and developed.

CLAIMS

We claim:

1. A method of increasing resistance of a plant to nematodes comprising increasing expression of, altering the expression pattern of or increasing the copy number of a polynucleotide encoding a Glyma18g02580 polypeptide, a Glyma18g02590 polypeptide, or a Glyma18g02610 polypeptide, a polypeptide having 90% or more identity to SEQ ID NO: 1 (2580), SEQ ID NO: 2 (2590), SEQ ID NO: 3 (2610), SEQ ID NO: 5 (2590-88788), SEQ ID NO: 6 (2590-Peking) or homologs, functional variants or combinations of any of the aforementioned polypeptides in cells of the plant, wherein increased expression of the polynucleotide in cells of the plant increases the resistance of the plant to nematodes.
2. The method of claim 1, wherein the expression is increased via genetic transformation of the plant.
3. The method of claim 1 or 2, wherein the plant is selected from soybean, sugar beets, potatoes, corn, peas and beans.
4. The method of any one of claims 1-3, wherein expression is increased in cells in the root of the plant.
5. The method of any one of claims 1-4, wherein expression of the polypeptide is increased by increasing expression of the native polynucleotide.
6. The method of any one of claims 1-4, wherein expression of the polypeptide is increased by introducing a construct comprising a promoter operably linked to a polynucleotide encoding the polypeptide into cells of the plant.
7. The method of any one of claims 1-4, wherein expression of the polypeptide is increased by incorporation of a transgene comprising a promoter operably linked to a polynucleotide encoding the polypeptide in the plant.
8. The method of any one of claims 1-4, wherein expression of the polypeptide is increased by introducing one or more copies of a polynucleotide encoding at least two of a Glyma18g02580 polypeptide, a Glyma18g02590 polypeptide and a Glyma18g02610 polypeptide.

9. The method of claim 8, wherein the copies are introduced in the *Rhgl* locus of the plant.
10. The method of any one of the preceding claims, wherein at least three copies of the polynucleotide are introduced in the plant.
- 5 11. The method of any one of the preceding claims, wherein at least ten copies of the polynucleotide are introduced in the plant.
12. The method of any one of the preceding claims, wherein the expression of a Glyma18g02580 polypeptide, a Glyma18g02590 polypeptide and a Glyma18g02610 polypeptide are increased.
- 10 13. The method of any one of the preceding claims, further comprising increasing the expression of or altering the expression pattern of a polynucleotide encoding a Glyma18g02600 polypeptide or a polypeptide having 90% or more identity to SEQ ID NO: 4 (02600).
14. The method of any one of the preceding claims, wherein the plant has
15 increased resistance to cyst nematodes, or soybean cyst nematodes (SCN) as compared to a control plant.
15. The method of claim 14, wherein increased resistance is measured by the plant having a lower percentage of invading nematodes that develop past the J2 stage, a lower rate of cyst formation on the roots, reduced nematode female index of a plant exposed to nematodes, reduced SCN egg production within
20 cysts, reduced overall SCN egg production per plant, or greater yield of soybean seeds on a per-plant basis or a per-growing-area basis as compared to a control plant grown in a similar growth environment.
16. The method of any one of claims 1-15, wherein the nematode female index of the plant after exposure to nematodes is lower than that of a control plant.
25
17. The method of any one of claims 1-16, wherein sequences derived from PI88788-, PI437654-, or Peking-derived sources are used in place of the Williams 82 sequences provided in Table 5.
18. A construct comprising a promoter operably linked to a polynucleotide
30 encoding a Glyma18g02580 polypeptide comprising SEQ ID NO: 1, a Glyma18g02590 polypeptide comprising SEQ ID NO: 2, 5 or 6, a

Glyma18g02610 polypeptide comprising SEQ ID NO: 3 or a polypeptide having at least 90% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6, or a homolog or a functional portion of any of the aforementioned polypeptides or combinations thereof.

- 5 19. The construct of claim 18, wherein the promoter is a plant promoter.
20. The construct of claim 18 or 19, wherein the polynucleotide encodes all three of the polypeptides.
21. The construct of claim 18 or 19, wherein the polynucleotide encodes at least two of the polypeptides.
- 10 22. The construct of any one of claims 18-21, wherein polynucleotide further encodes a Glyma18g02600 polypeptide.
23. A transgenic cell comprising a polynucleotide encoding a polypeptide capable of increasing resistance to nematodes, the polypeptide having at least 90% identity to a Glyma18g02580 polypeptide comprising SEQ ID NO: 1, a
- 15 Glyma18g02590 polypeptide comprising SEQ ID NO: 2, 5 or 6, a Glyma18g02610 polypeptide comprising SEQ ID NO: 3 or a polypeptide having at least 90% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 6, or a homolog or functional portion thereof or combinations thereof.
- 20 24. The transgenic cell of claim 23, wherein the polynucleotide encodes at least two of the polypeptides selected from Glyma18g02610, Glyma18g02590 and Glyma18g02580.
25. The transgenic cell of claim 23, wherein the polynucleotide encodes at least a Glyma18g02610 polypeptide, a Glyma18g02590 polypeptide and a
- 25 Glyma18g02580 polypeptide.
26. The transgenic cell of any one of claims 23-25, wherein the polynucleotide is present in at least three copies in the plant.
27. The transgenic cell of any one of claim 23-25, wherein the polynucleotide is present in at least ten copies in the plant.
- 30 28. The transgenic cell of any one of claims 23-27, wherein polynucleotide further encodes a Glyma18g02600 polypeptide.

29. The transgenic cell of any one of claims 23-28, wherein the plant is selected from soybean, sugar beets, potatoes, corn, peas and beans.
30. A seed comprising the transgenic cell of any one of claims 23-29.
31. A plant grown from the seed of claim 30.
- 5 32. A transgenic plant comprising the cell of any one of claims 23-29.
33. A part, progeny or asexual propagate of the transgenic plant of claim 32.
34. A method of generating a transgenic plant with increased resistance to nematodes comprising introducing an exogenous polynucleotide encoding a Glyma18g02580 polypeptide having at least 90% identity to SEQ ID NO: 1, 10 Glyma18g02590 polypeptide having at least 90% identity to SEQ ID NO: 2, 5 or 6, or Glyma18g02610 polypeptide having at least 90% identity to SEQ ID NO: 3, or a homolog, functional variant or combinations thereof into a soybean plant cell or progeny thereof, whereby expression of the polypeptide is increased in a root cell of the soybean plant and whereby the plant has 15 increased resistance to nematodes as compared to a control plant.
35. The method of claim 34, wherein the polynucleotide encodes at least two of the polypeptides selected from Glyma18g02610, Glyma18g02590 and Glyma18g02580
36. The method of claim 34, wherein the polynucleotide encodes at least a 20 Glyma18g02610 polypeptide, a Glyma18g02590 polypeptide and a Glyma18g02580 polypeptide.
37. The method of any one of claims 34-36, wherein polynucleotide further encodes a Glyma18g02600 polypeptide.
38. The method of any one of claims 34-37, wherein the polynucleotide is present 25 in at least three copies in the plant.
39. The method of any one of claims 34-37, wherein the polynucleotide is present in at least ten copies in the plant.
40. The method of any one of claims 34-38, wherein the plant is selected from soybean, sugar beets, potatoes, corn, peas and beans.
- 30 41. A method of screening a first plant cell for resistance or susceptibility to nematodes comprising:

detecting a genetic marker or selectable marker in the first plant cell, the genetic marker or the selectable marker associated with cyst nematode resistance or susceptibility to cyst nematodes;
and predicting the resistance or susceptibility of the first plant cell to nematodes.

42. The method of claim 41, wherein the genetic marker is the methylation status of the promoter, upstream region or gene body of at least one of *Glyma18g02610*, *Glyma18g02590*, or *Glyma18g02580*.
43. The method of claim 41, wherein the genetic marker is the level of RNA transcript, transcription or protein expression of at least one of *Glyma18g02610*, *Glyma18g02590*, *Glyma18g02580* or *Glyma18g02600*.
44. The method of claim 43, wherein the level is measured in root cells.
45. The method of claim 41, wherein the genetic marker is the genomic copy number of at least one of *Glyma18g02610*, *Glyma18g02590*, *Glyma18g02580*, *Glyma18g02600* or any portion of the 31.2 kb *Rhg1* region identified as being present in multiple copies in resistant varieties.
46. The method of claim 41, wherein the genetic marker is the genomic DNA segment carrying a repeat junction between *Glyma18g02610* and *Glyma18g02570*.
47. The method of claim 41, wherein the genetic marker is the transcription or expression level of at least one of *Glyma18g02610*, *Glyma18g02590*, *Glyma18g02600*, or *Glyma18g02580* after or upon contact with a nematode.
48. The method of claim 41, wherein the genetic marker is a single nucleotide polymorphism in at least one of *Glyma18g02610*, *Glyma18g02590*, *Glyma18g02580*, *Glyma18g02600* or within the 31.2 kb *Rhg1* repeat region.
49. The method of claim 41, wherein the genetic marker is the presence or absence of more than one copy of a genomic region comprising at least one of *Glyma18g02600*, *Glyma18g02610*, *Glyma18g02590*, or *Glyma18g02580*.
50. The method of any one of 41-49, wherein the predicting step comprises comparing the marker in the first plant cell to the marker in a second plant cell

with a known resistance or susceptibility phenotype, wherein the phenotype of the second cell is predictive of the phenotype in the first cell.

51. The method of any one of claims 41-50, wherein the second plant cell is SCN-resistant, is PI88788, Peking or carries an *Rhg1* locus derived from PI88788 or Peking or other sources of *Rhg1*-mediated SCN resistance.

52. The method of any one of claims 41-51, wherein the second plant cell is SCN-susceptible is 'Williams 82', 'Lee 84', Essex or carries an *Rhg1* locus derived from 'Williams 82', 'Lee 84', Essex or other varieties in which the *Rhg1* locus is less effective at mediating SCN resistance than the *Rhg1* locus of PI88788 or Peking.

53. The method of any of claims 41-52, wherein the SCN is controlled by *rhg1-b*, is a Race 3 SCN, an SCN Hg Type lacking the "2" designation in its Hg Type designation or an SCN Hg Type lacking the "1" designation in its Hg Type designation.

54. The method of any one of claims 41-53, wherein detecting comprises amplifying the marker or a portion of the marker to produce an amplified product and determining all or part of the DNA sequence of the amplified product or assessing the marker using differential sensitivity to a restriction endonuclease, allele specific hybridization analysis, allele specific PCR, high density nucleotide array analysis, quantitative PCR, Northern blot analysis, microsatellite analysis, ELISA or Western blot analysis.

55. The method of any one of claims 41-54, wherein the prediction is used to select resistant plant cells for use in developing resistant soybean lines.

56. The method of any one of claims 41-54, wherein the prediction is used to select resistant plant cells for use in developing resistant sugar beets, potatoes, corn, peas or beans.

57. A method of increasing resistance of a plant to nematodes comprising expressing a polynucleotide encoding a Glyma18g02580 polypeptide of SEQ ID NO:1, a Glyma18g02590 polypeptide of SEQ ID NO: 2, 5 or 6, or a Glyma18g02610 polypeptide of SEQ ID NO: 6, a polypeptide having 90% or more identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO:

5 or SEQ ID NO: 6, or a homolog, a functional variant or combination of any of the aforementioned polypeptides in a cell, and applying the polypeptide or a cell comprising the polypeptide to the plant, seeds of the plant or soil in which the seeds may be planted, wherein the application increases the resistance of the plant to nematodes.

58. The method of claim 57, wherein the polynucleotide encodes at least two of a Glyma18g02610 polypeptide, a Glyma18g02590 polypeptide, or a Glyma18g02580 polypeptide

59. The method of claim 57, wherein the polynucleotide encodes a Glyma18g02610 polypeptide, a Glyma18g02590 polypeptide, and a Glyma18g02580 polypeptide.

60. The method of any one of claims 57-59, wherein the polynucleotide further encodes a Glyma18g02600 polypeptide

Soybean Cyst Nematode (SCN) Life Cycle

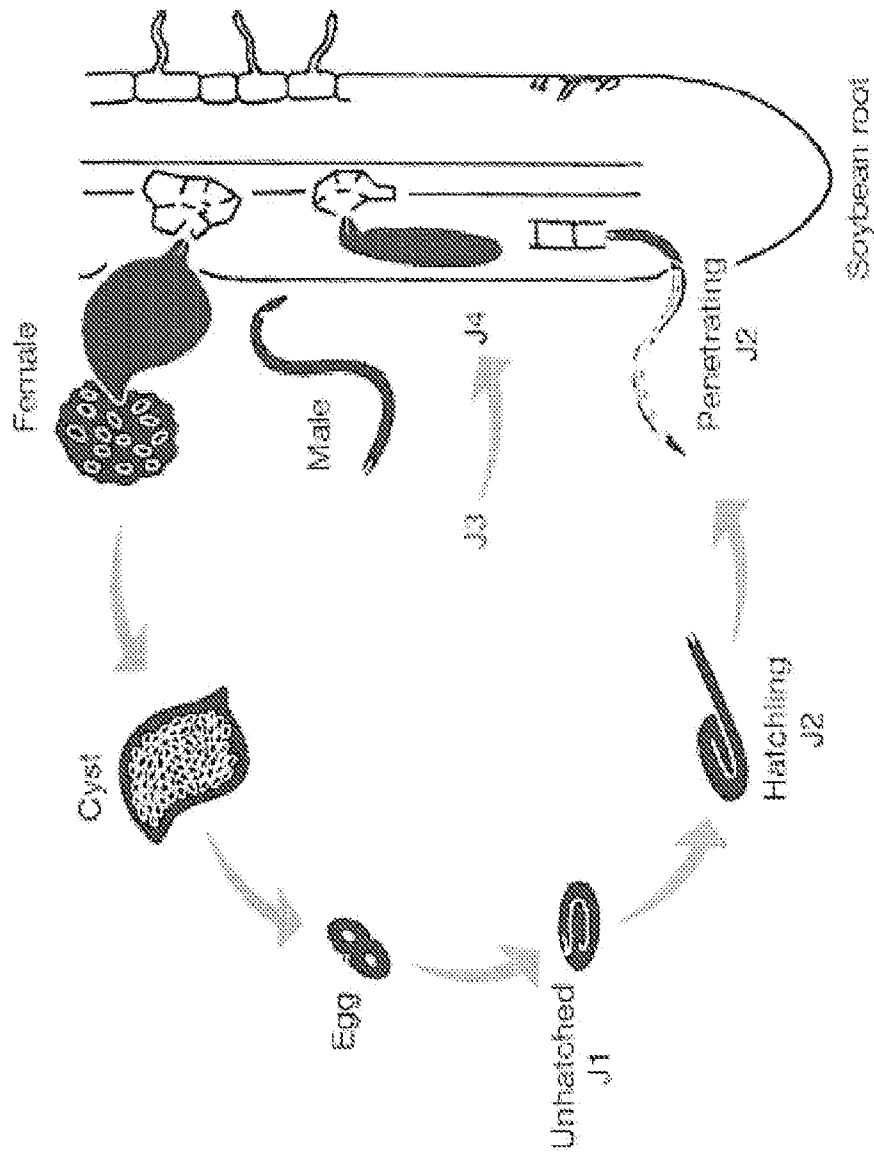
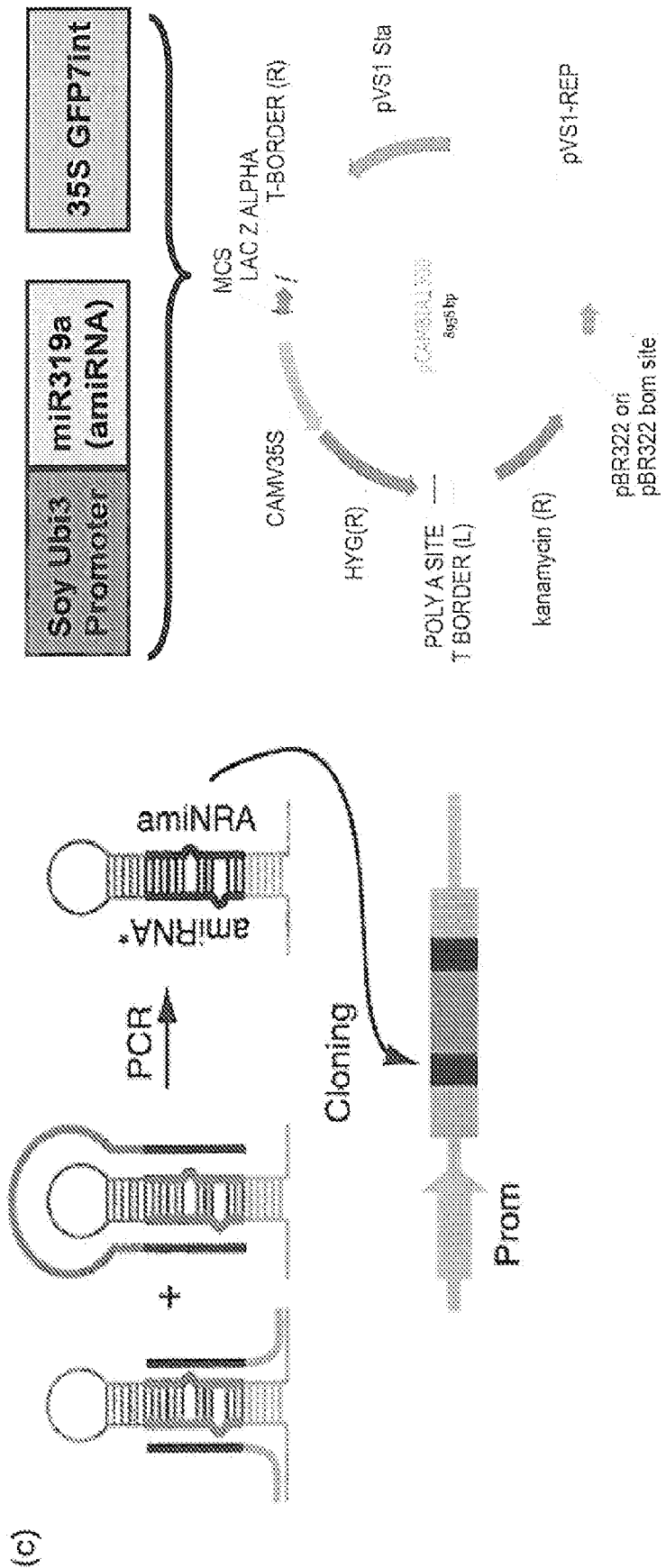


Figure 1

Figure 2: Silencing strategy: artificial miRNAs

Replace the miR319a sequence with amiRNA sequence
targeting gene of interest, under control of a strong soybean
promoter



REPLACEMENT SHEET
3 / 21

Figure 3

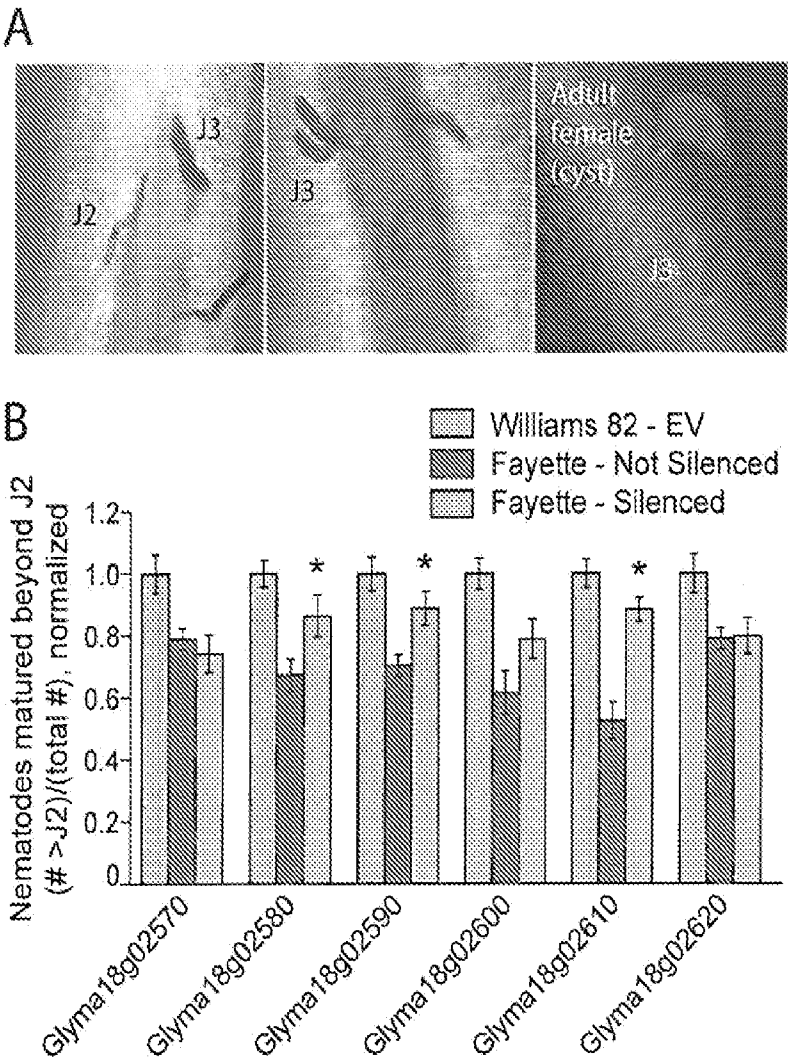


Figure 4

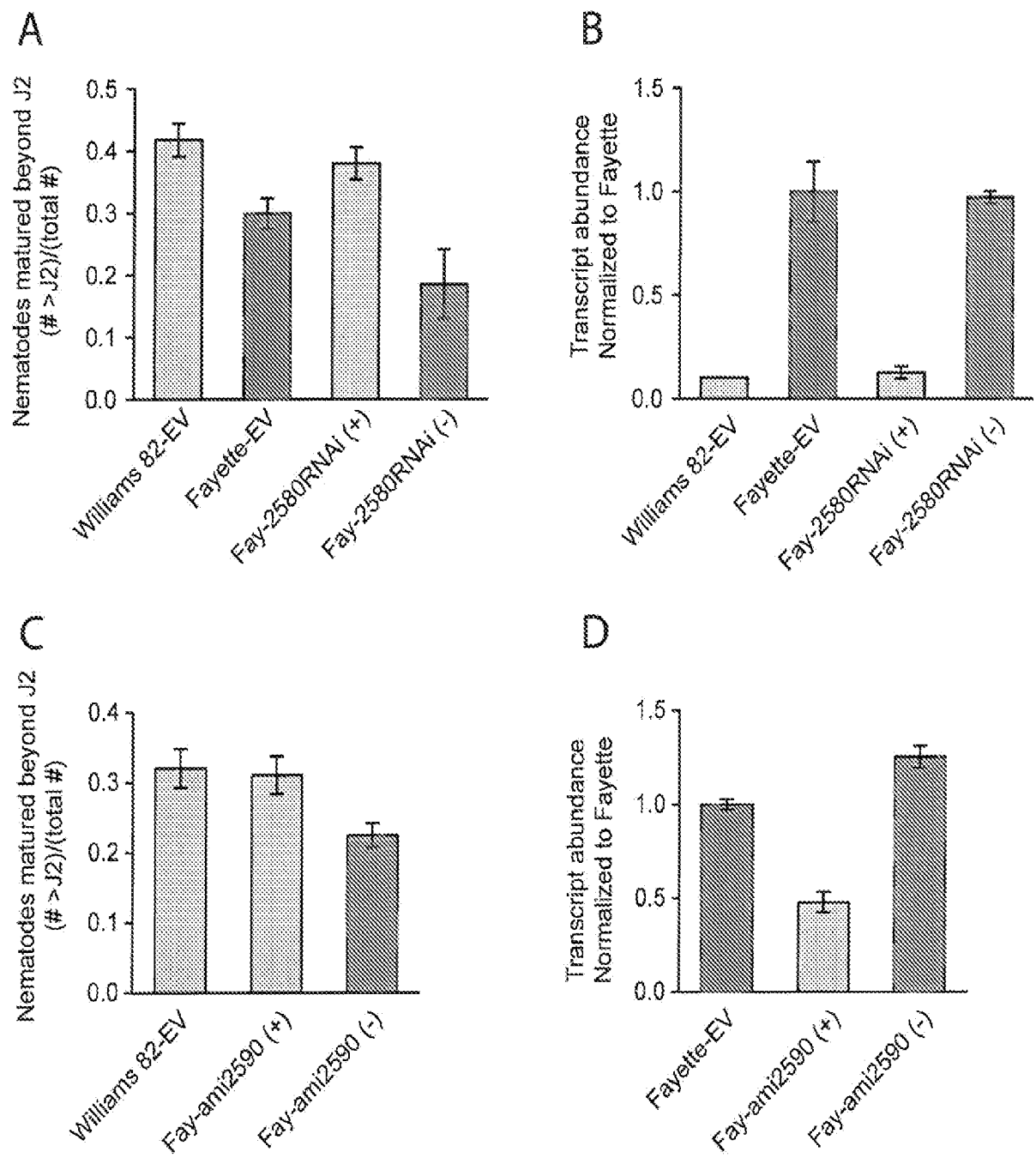


Figure 5

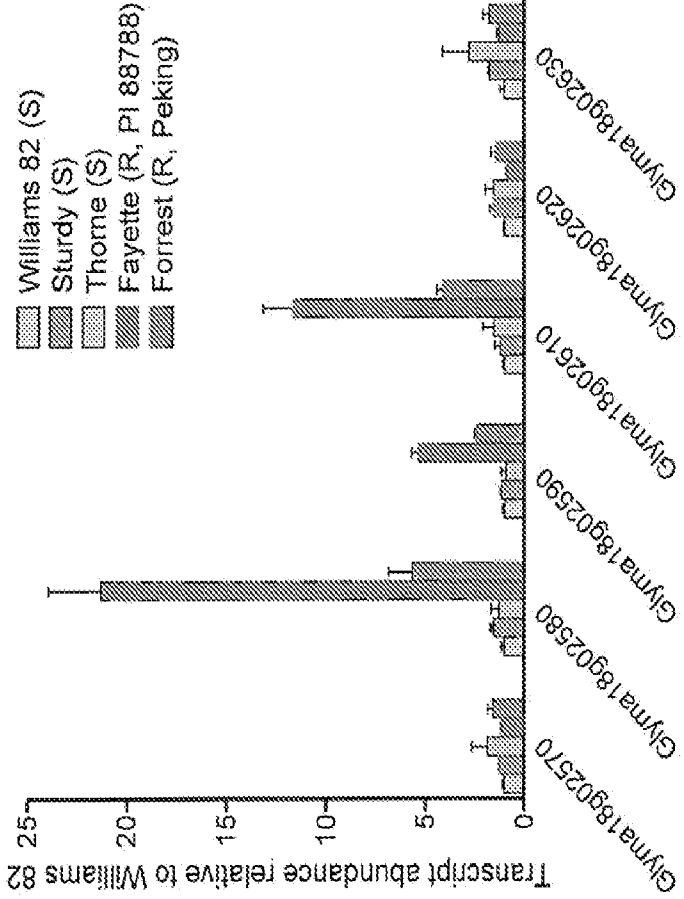
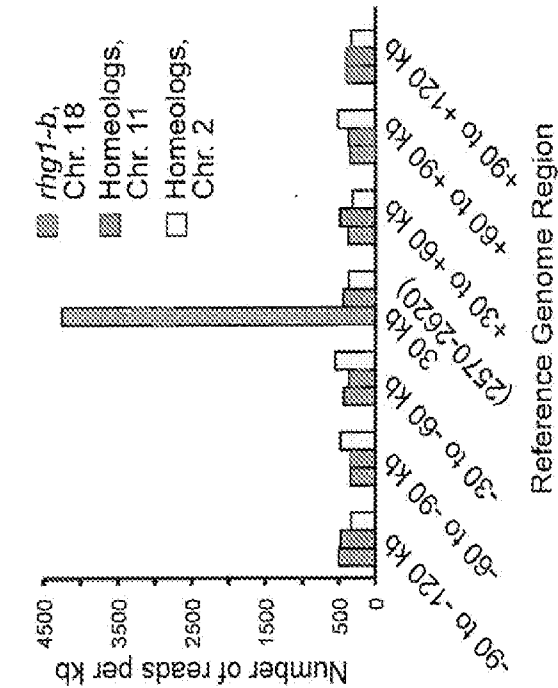
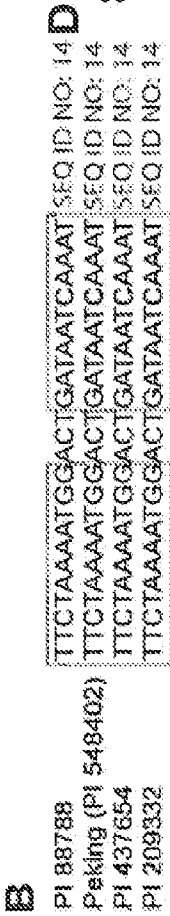
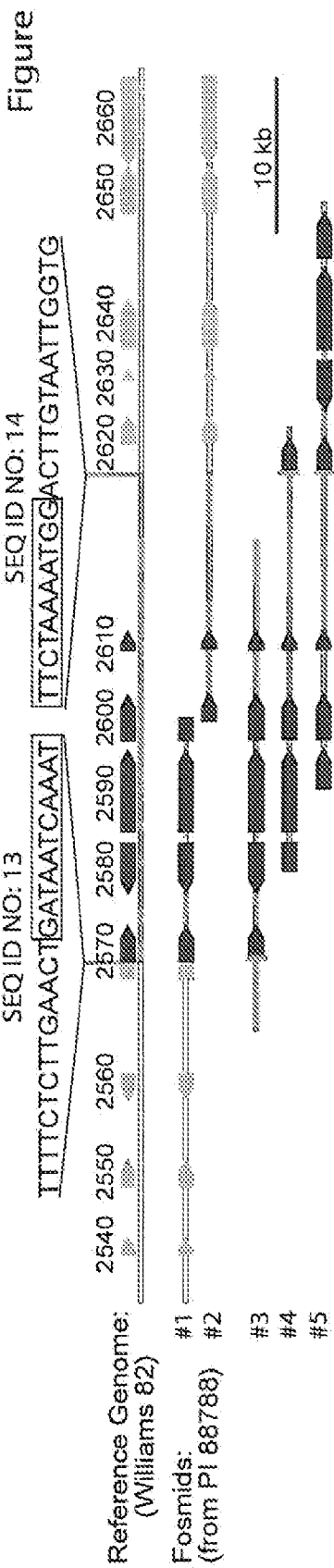
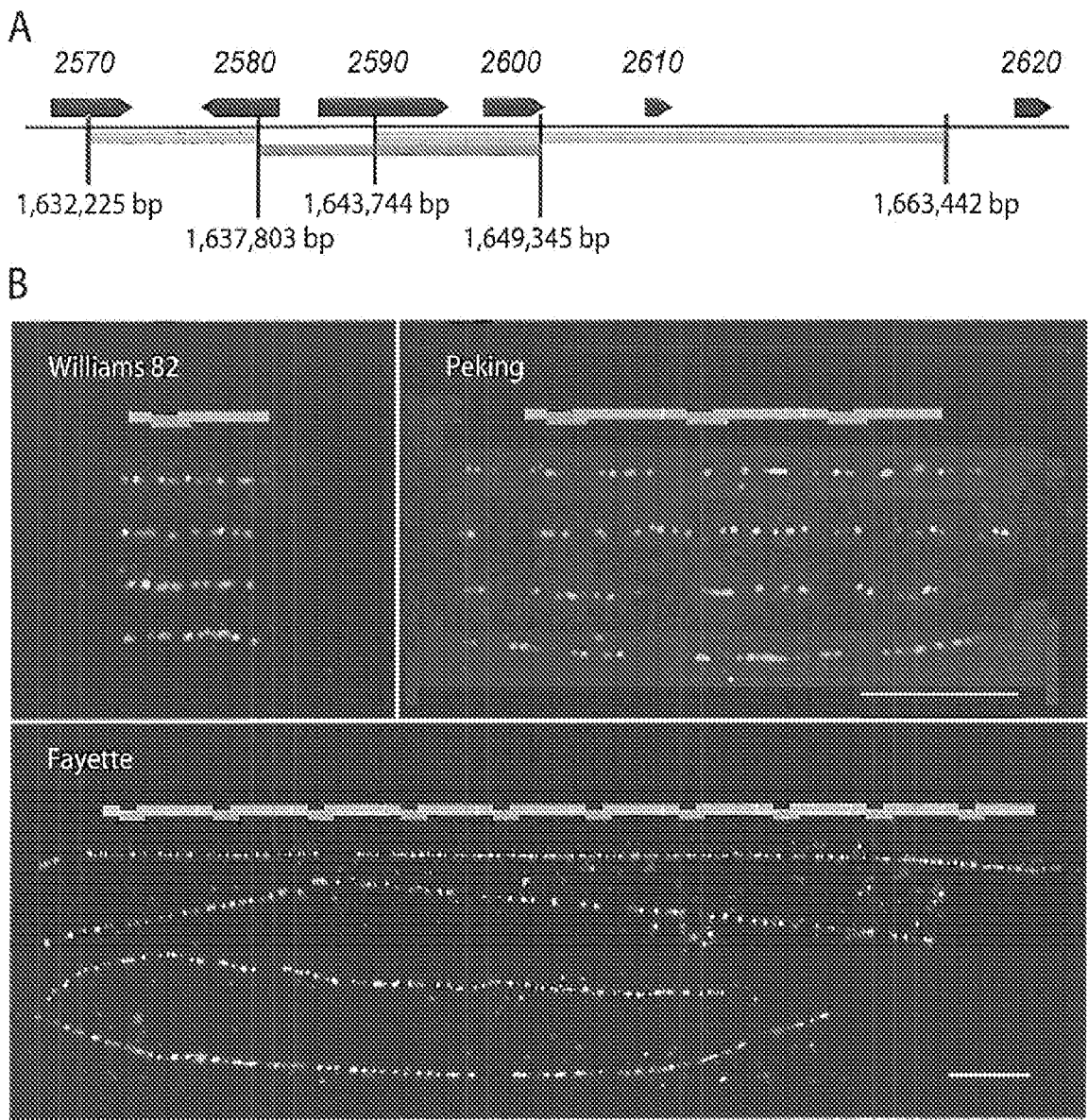


Figure 6



7 / 21

Figure 7

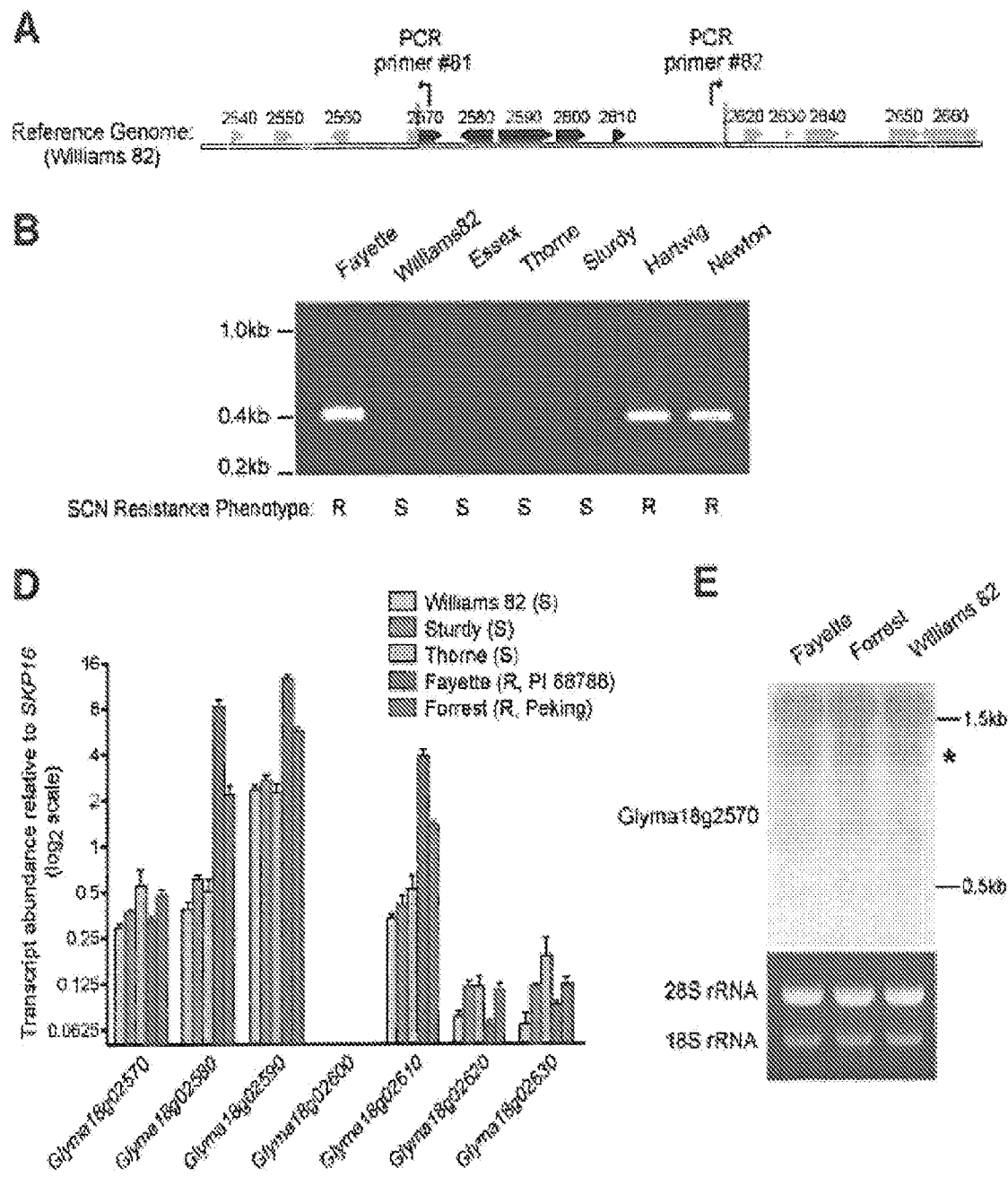
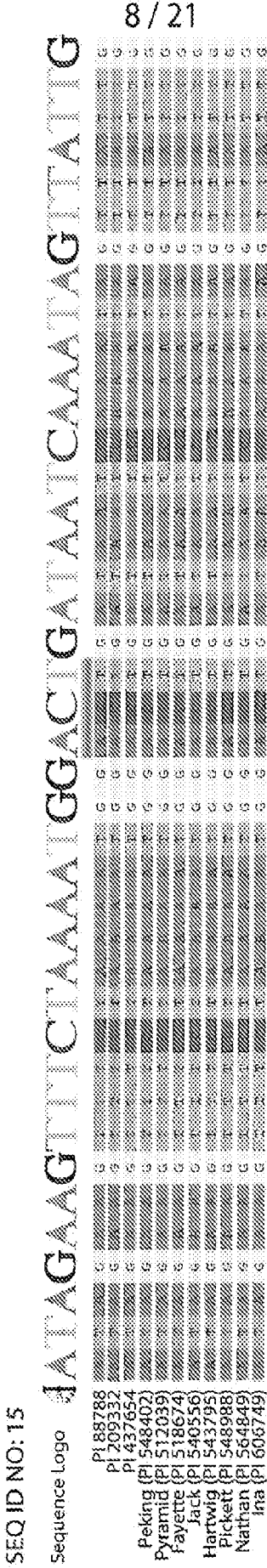
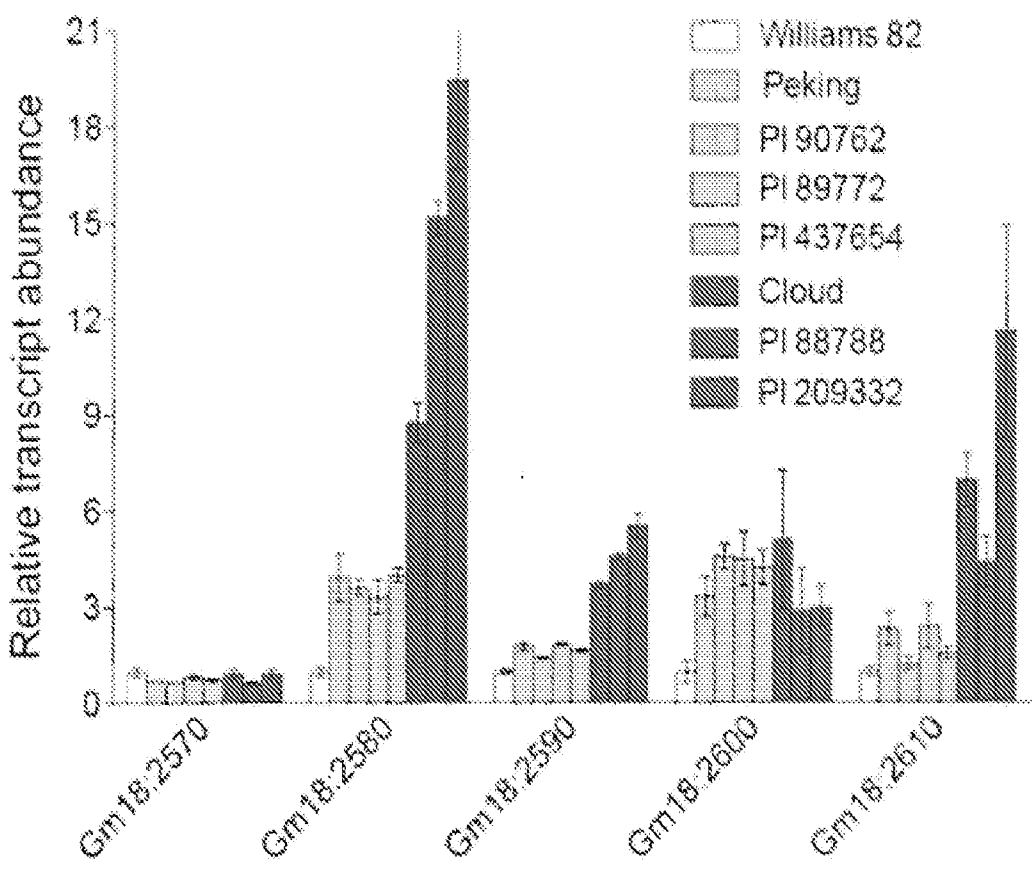


Figure 7C



9 / 21

Figure 8



10 / 21

Figure 9

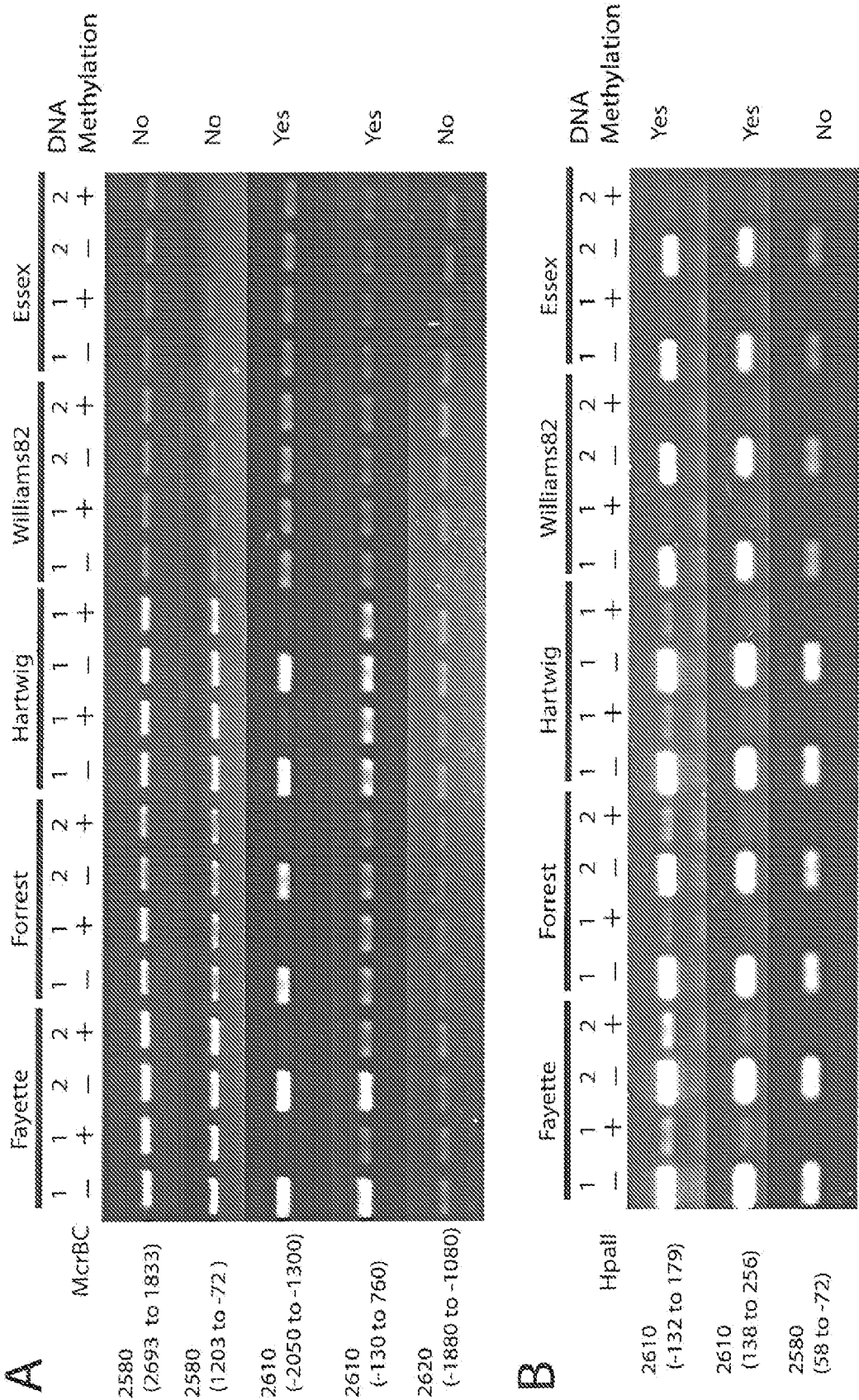


Figure 9

C

Region Analyzed (nt position relative to ATG)	DNA treatment	Presence of Methylation					Polymorphic Methylation
		Fayette	Forrest	Hartwig	Williams82	Essex	
Gm18:2570							
-770 to -1415	McrBC	Yes	Yes		Yes	Yes	No
-115 to -850	McrBC	Yes	Yes		Yes	Yes	No
Gm18:2580							
58 to -72	McrBC & HpaII	No	No	No	No	No	No
1203 to -72	McrBC	No	No	No	No	No	No
2639 to 1833	McrBC	No	No	No	No	No	No
Gm18:2580/2590							
(relative to 2590 ATG)							
-1692 to -1042	McrBC	Yes	No	No	No	No	Yes
-1255 to -709	McrBC	Yes	No		No	No	Yes
-798 to -154	McrBC	Yes			No		Yes
Gm18:2590							
3669 to 4230	McrBC	No	No	No	No	No	No
Gm18:2600							
-230 to 455	McrBC	No	No		No	No	No
Gm18:2610							
-2052 to -1301	McrBC	Yes	Yes	Yes	No	No	Yes
-1324 to -184	McrBC	Yes	Yes		No	No	Yes
-132 to 757	McrBC	Yes	No	No	No	No	Yes
-132 to 179	HpaII	Yes	Yes	Yes	No	No	Yes
-216 to 5	HpaII	Yes	Yes	Yes	Yes	Yes	No
138 to 256	HpaII	Yes	Yes	Yes	No	No	Yes
Gm18:2620							
-1880 to -1080	McrBC	No	No	No	No	No	No
-890 to -80	McrBC	No	No	No	No	No	No

11/21

12 / 21

Figure 10

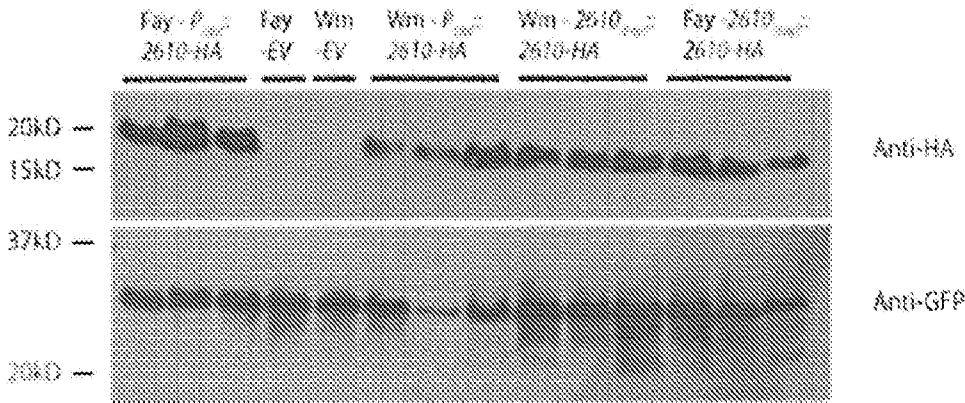


Figure 11

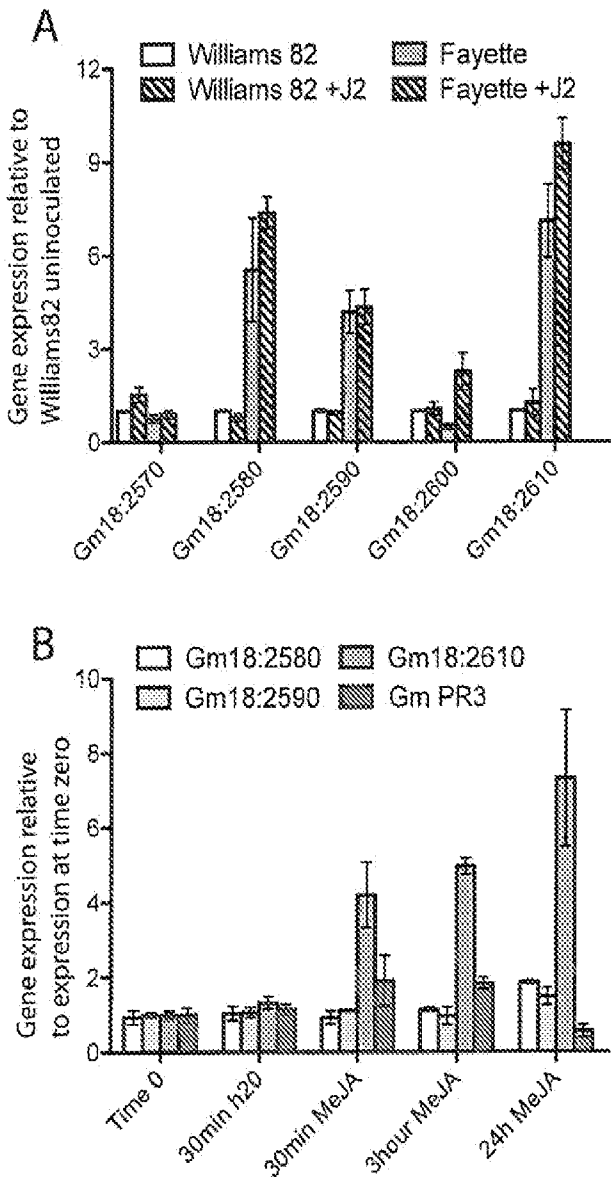
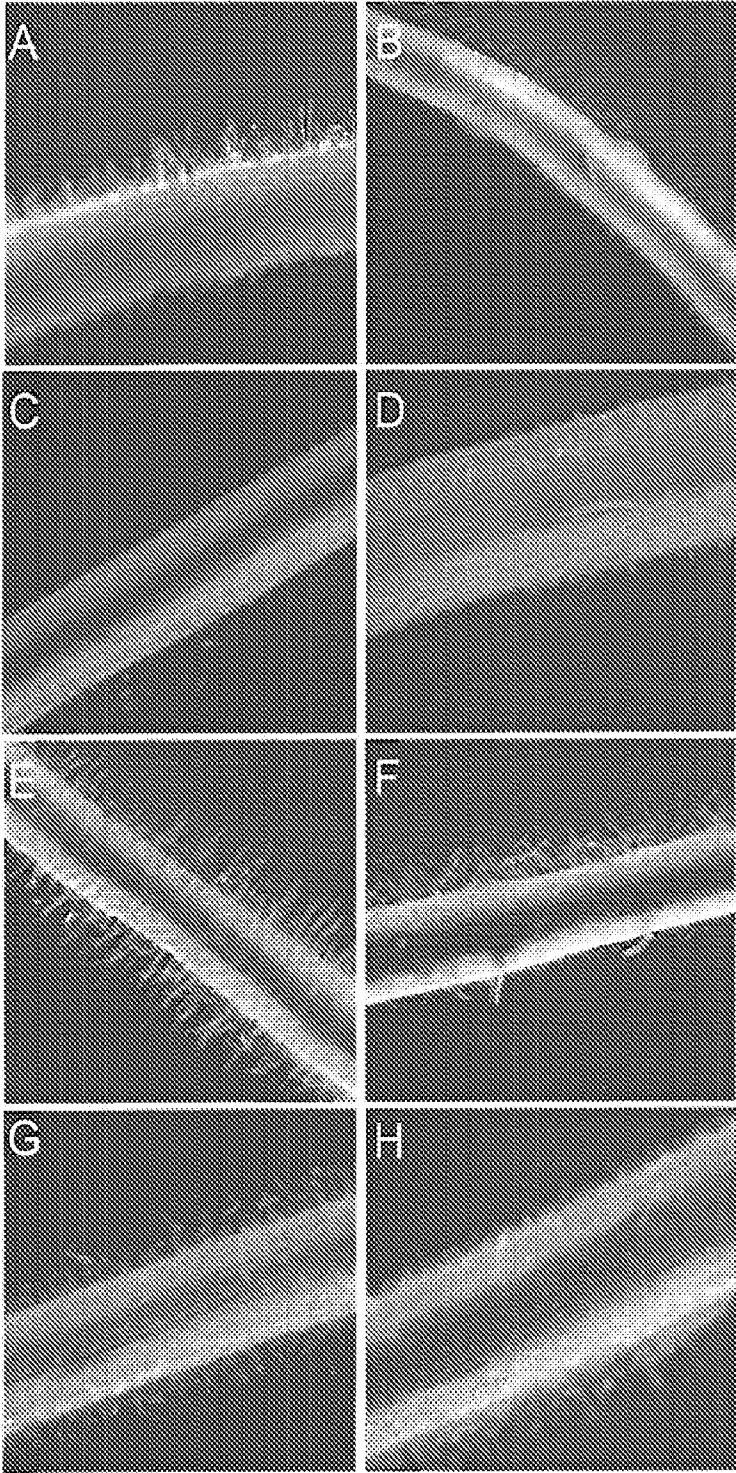


Figure 12



REPLACEMENT SHEET

14/21

Figure 13 (Continued on next page)

Williams	1	GGTTGGGGCTTGTTTGGCTCCAAGTATGAAGATGCCGCCGATCTCTTCGATAAAAGCCGCC
Pek.isoform	1	
Fayette	1	
Peking	1	
Williams	61	AATTGCTTCAAGCTCGCCAAATCATGGGACAAGGCTGGAGCGACATACCTGAAGTTGGCA
Pek.isoform	61	
Fayette	61	
Peking	61	
Williams	121	AGTTGTCATTTGAAGTTGGAAAGCAAGCATGAAGCTGCACAGGCCCATGTCGATGCTGCA
Pek.isoform	121	
Fayette	121	
Peking	121	
Williams	181	CATTGCTACAAAAAGACTAATATAAACGAGTCTGTATCTTGCTTAGACCGAGCTGTAAAT
Pek.isoform	181	
Fayette	181	
Peking	181	
Williams	241	CTTTTCTGTGACATTGGAAGACTCTCTATGGCTGCTAGATATTTAAAGGAAATTGCTGAA
Pek.isoform	241	
Fayette	241	
Peking	241	
Williams	301	TTGTACGAGGGTGAACAGAATATTGAGCAGGCTCTTGTTTACTATGAAAAATCAGCTGAT
Pek.isoform	301	
Fayette	301	
Peking	301	
Williams	361	TTTTTTCAAAATGAAGAAGTGACAACTTCTGCGAACCAATGCAAACAAAAGTTGCCAG
Pek.isoform	361	
Fayette	361	
Peking	361	
Williams	421	TTTGCTGCTCAGCTAGAACAATATCAGAAGTCGATTGACATTTATGAAGAGATAGCTCGC
Pek.isoform	421	
Fayette	421	
Peking	421	
Williams	481	CAATCCCTCAACAATAATTTGCTGAAGTATGGAGTTAAAGGACACCTTCTTAATGCTGGC
Pek.isoform	481	
Fayette	481	
Peking	481	
Williams	541	ATCTGCCAACTCTGTAAAGAGGACGTTGTTGCTATAACCAATGCATTAGAACGATATCAG
Pek.isoform	541	
Fayette	541	A.....
Peking	541	G.....
Williams	601	GAACTGGATCCAACATTTTCAGGAACACGTGAATATAGATTGTTGGCGGACATTGCTGCT
Pek.isoform	565	
Fayette	601	
Peking	601	

Figure 13 (Continued on next page)

Williams	661	GCAATTGATGAAGAAGATGTTGCAAAGTTTACTGATGTTGTCAAGGAATTTGATAGTATG
Pek.isoform	625	
Fayette	661	
Peking	661	
Williams	721	ACCCCTCTGGATTCTTGGAAGACCACACTTCTCTTAAGGGTGAAGGAAAAGCTGAAAGCC
Pek.isoform	685	
Fayette	721	
Peking	721	
Williams	781	AAAGAACTTGAGGAGGATGATCTTACTTGA--- (SEQ ID NO: 9)
Pek.isoform	745	T....GG...T.ACTTGA (SEQ ID NO: 10)
Fayette	781	C..C....GGC..T.ACTTGA (SEQ ID NO: 8)
Peking	781	T....GG...T.ACTTGA (SEQ ID NO: 11)

Figure 13
(Continued from previous pages)

Peking	MADQLSKGEEFEKKA	60
Pekingiso	MADQLSKGEEFEKKA	60
Fayette	MADQLSKGEEFEKKA	60
Williams	MADQLSKGEEFEKKA	60

Peking	SCHLKLESKHEAAQAHVDA	120
Pekingiso	SCHLKLESKHEAAQAHVDA	120
Fayette	SCHLKLESKHEAAQAHVDA	120
Williams	SCHLKLESKHEAAQAHVDA	120

Peking	LYEGEQNLIEQALVYVEKSADFFQNEE	180
Pekingiso	LYEGEQNLIEQALVYVEKSADFFQNEE	180
Fayette	LYEGEQNLIEQALVYVEKSADFFQNEE	180
Williams	LYEGEQNLIEQALVYVEKSADFFQNEE	180

Peking	QSENNNLLKYGVKGHLLNAGICQLCKEE	240
Pekingiso	QSENNNLLKYGVKGHLLNAGICQLCKEE	240
Fayette	QSENNNLLKYGVKGHLLNAGICQLCKEE	240
Williams	QSENNNLLKYGVKGHLLNAGICQLCKEE	240

Peking	AIDEEDVAKFTDVVK	290 (SEQ ID NO: 6)
Pekingiso	AIDEEDVAKFTDVVK	290 (SEQ ID NO: 6)
Fayette	AIDEEDVAKFTDVVK	290 (SEQ ID NO: 6)
Williams	AIDEEDVAKFTDVVK	290 (SEQ ID NO: 6)

17 / 21

Figure 14

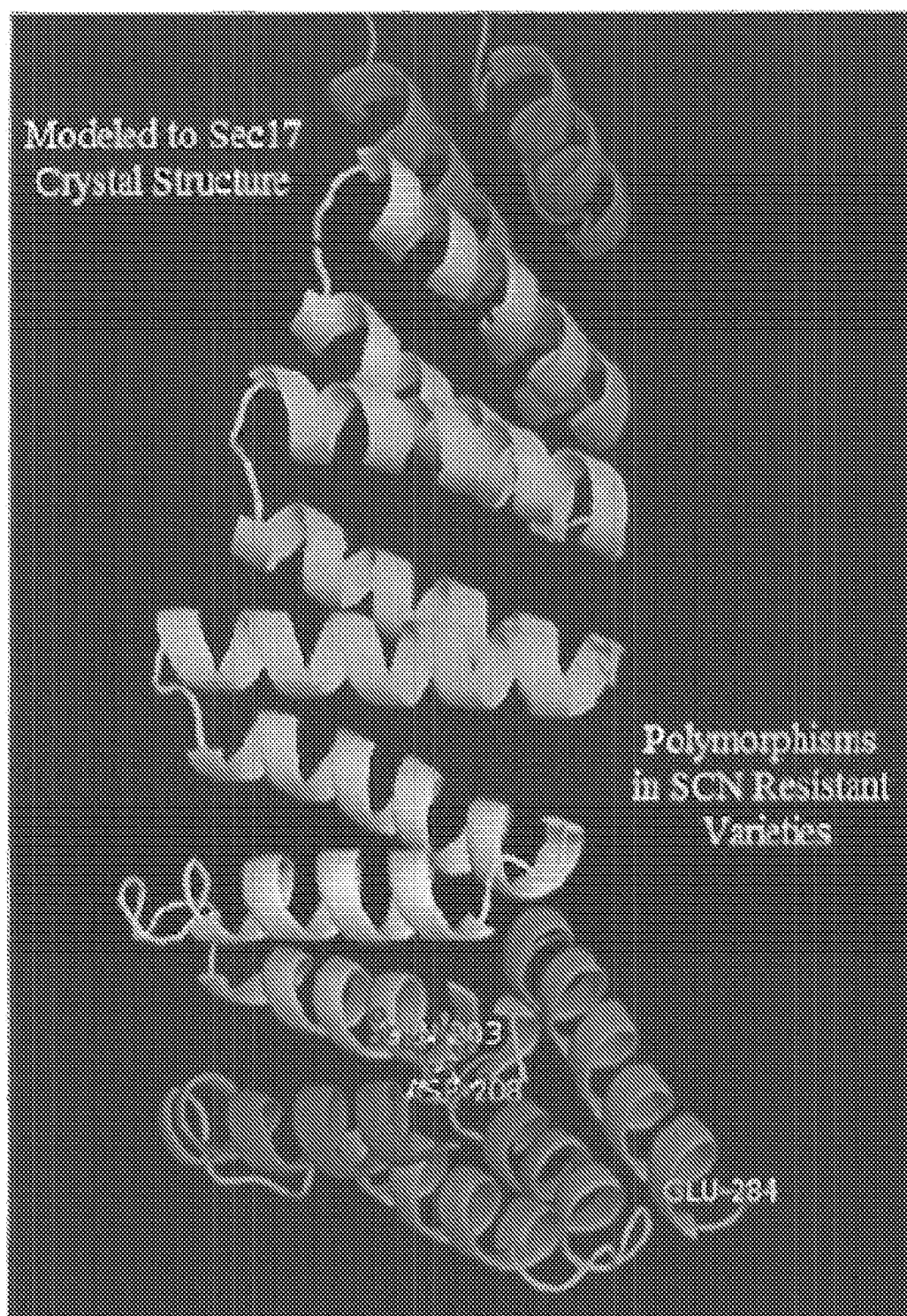


Figure 15

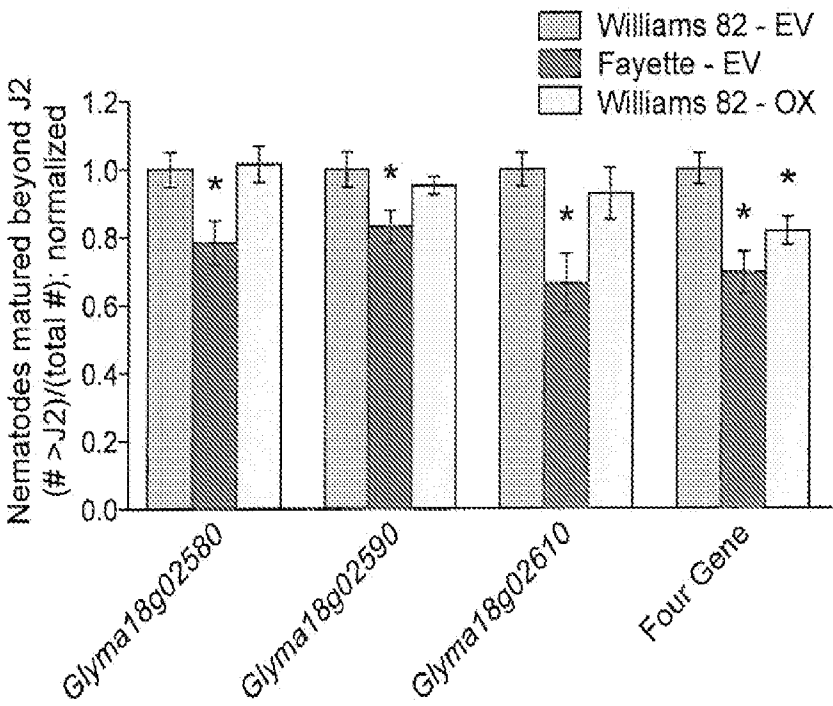


Figure 16

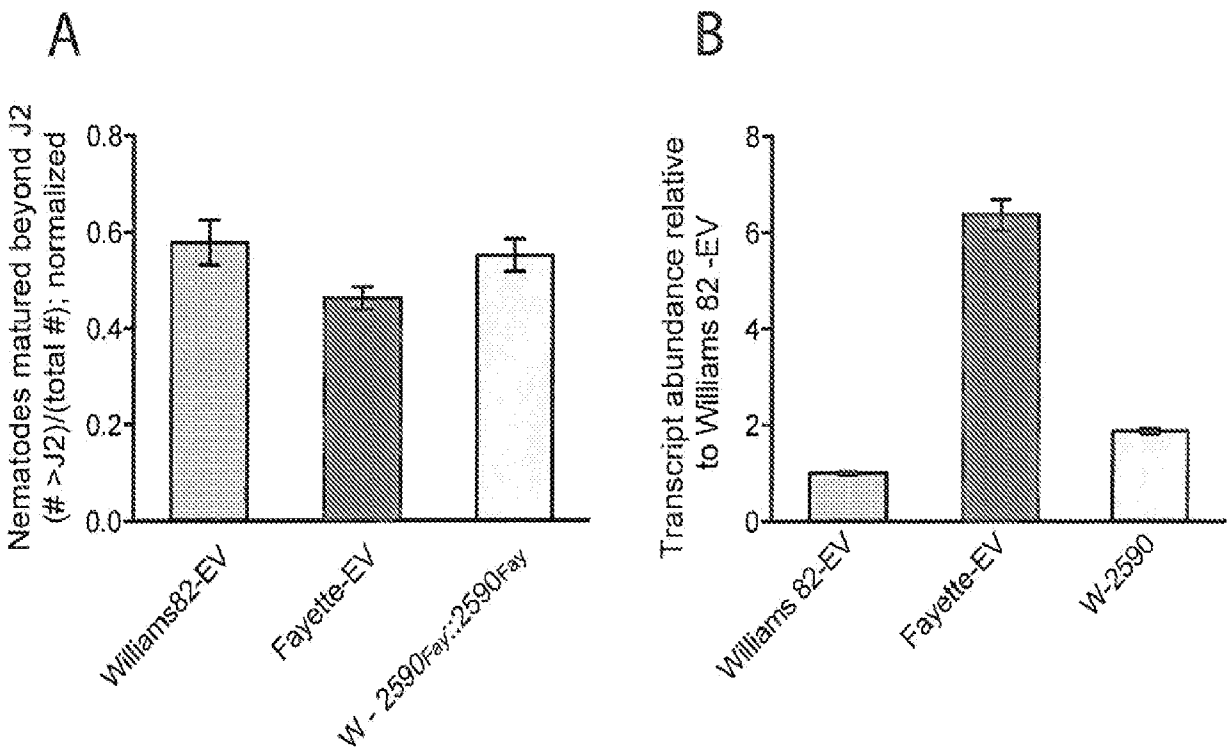


Figure 17

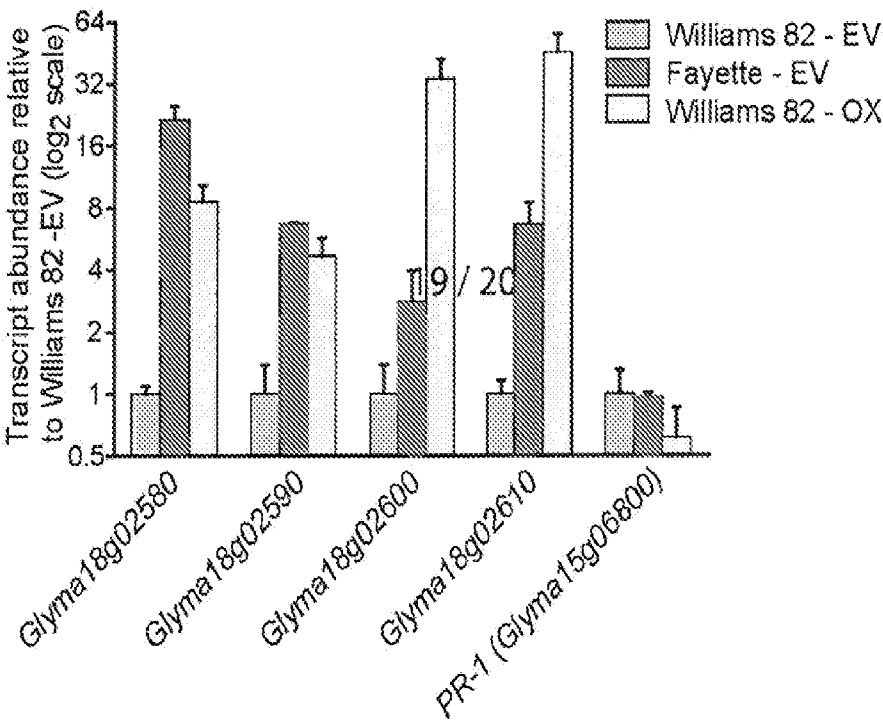


Figure 18

