Figure S1

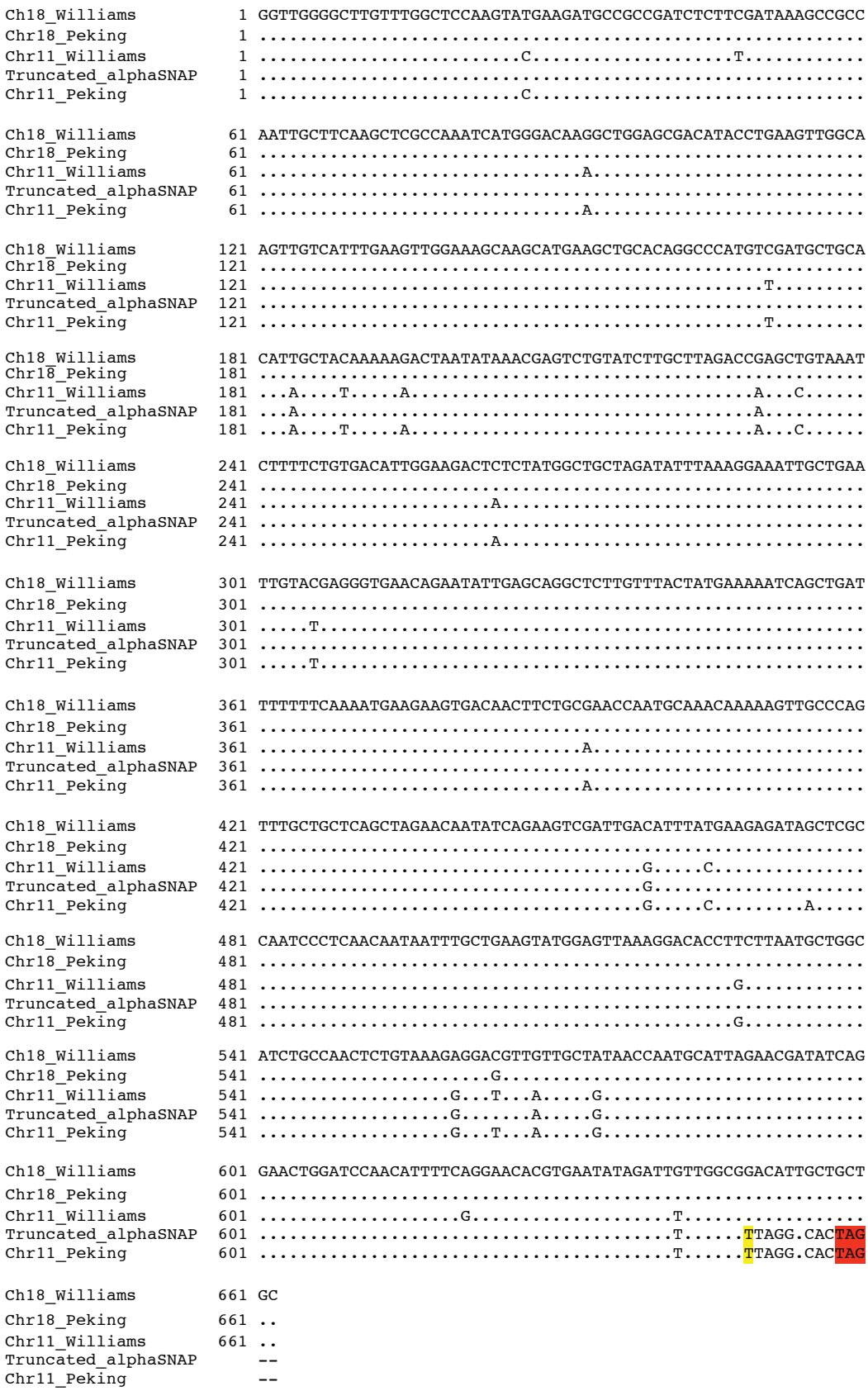

Figure S1. Previously reported truncated allele of -SNAP shares higher sequence similarity to the paralog encoded on chromosome 11 and is likely not encoded by *Glyma18g02590* at *Rhg1*.

Nucleic acid alignment for the first 661 bases of -SNAP encoded by chromosome 18 (*Rhg1*) and 11 (paralog) from Williams82 and Peking, and the previously reported truncated allele

Figure S1 cont'd
sequence in Matsye et al. (2012). Sequence from Williams82 is shown and positions with an identical sequence are listed as (.) The sequence reported in Matsye et al. (2012) for the truncated allele of *Glyma18g02590* is most similar to the Williams 82 and Peking paralogs encoded on chromosome 11. The polymorphism reported to change the exon-intron boundary and cause the splice variant is highlighted in yellow, and the resulting in frame stop codon is highlighted in red.

Figure S2

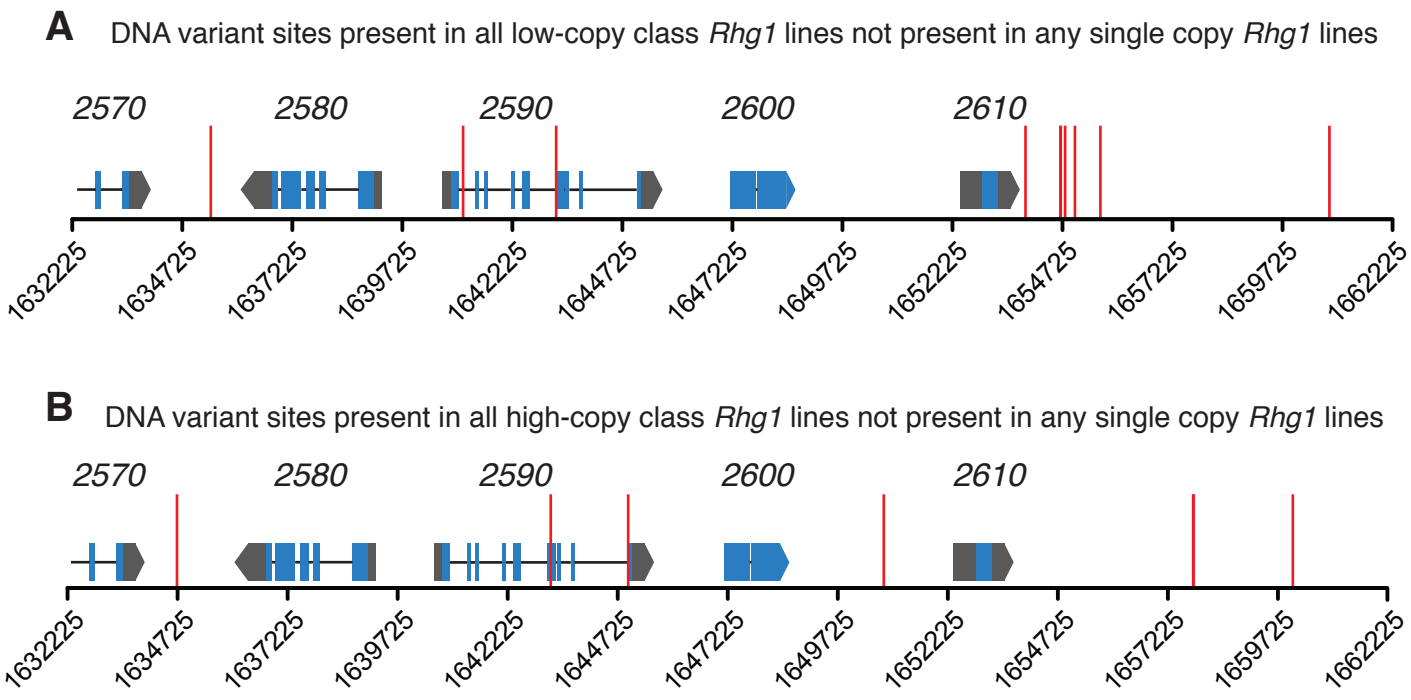

Figure S2. The DNA sequence of *Rhg1* repeats has continued to diverge between the low- and high-copy containing lines.
(A) DNA variant sites present in low-copy *Rhg1* Hg Type Test lines indicate that following divergence from the high-copy group, the locus is continuing to evolve. There are 10 DNA variants present in all low-copy *Rhg1* lines, not present in the high-copy or single-copy lines. (B) DNA variant sites are shown as in (A), but instead only DNA variant sites present in high-copy Hg Type Test lines, not present in low-copy or single-copy lines. Red vertical bars represent the location of the DNA variants across the Rhg1 locus. *Rhg1* locus gene models shown at correct x-axis position for reference (blue exons, black line introns, grey untranslated regions); gene name is above gene model (e.g., *2570* = *Glyma18g02570*).

Figure S3

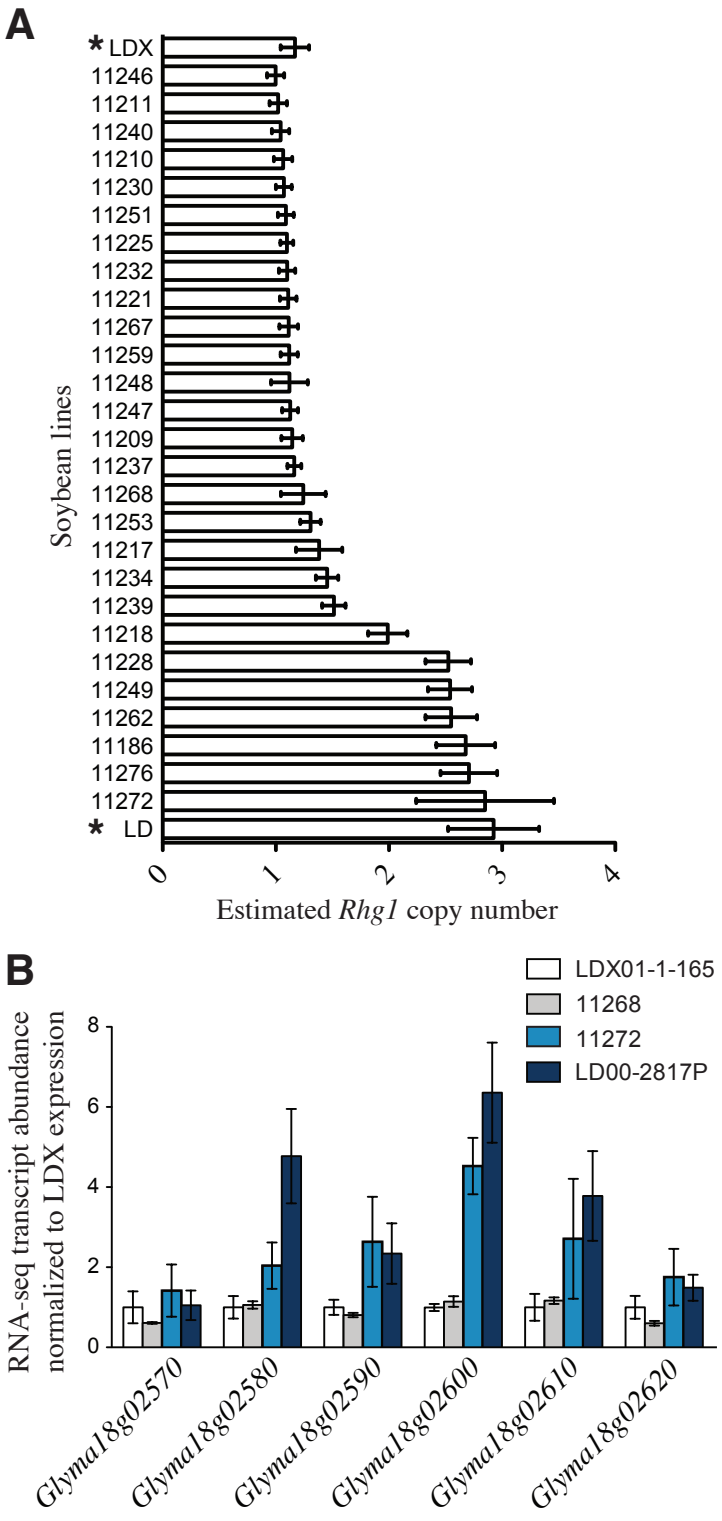

Figure S3. Copy number estimates and RNA-seq analysis indicate the presence of multiple copies of *Rhg1* in SCN resistant parent LD00-2871P and some progeny with a concomitant increase in transcription.

(A) Parental lines and 27 progeny were analyzed for *Rhg1* copy number estimates based on cytosine sequencing depth. The parental line LD00-2871P (LD) is estimated to contain 3 copies of the *Rhg1* repeat, consistent with its derivation of SCN resistance from PI 437654 described here. The two parental lines are denoted with *. (B) Relative transcript abundance based on RNA-seq reads indicates the 4 genes transcribed within the *Rhg1* repeat are expressed more highly in the parental line LD and progeny 11272 than in line LDX or 11268. There results are consistent with the *Rhg1* copy number estimates. RNA-sequencing is reported in reads per kilobase per million reads and normalized to expression from line LDX01-1-165.

Figure S4

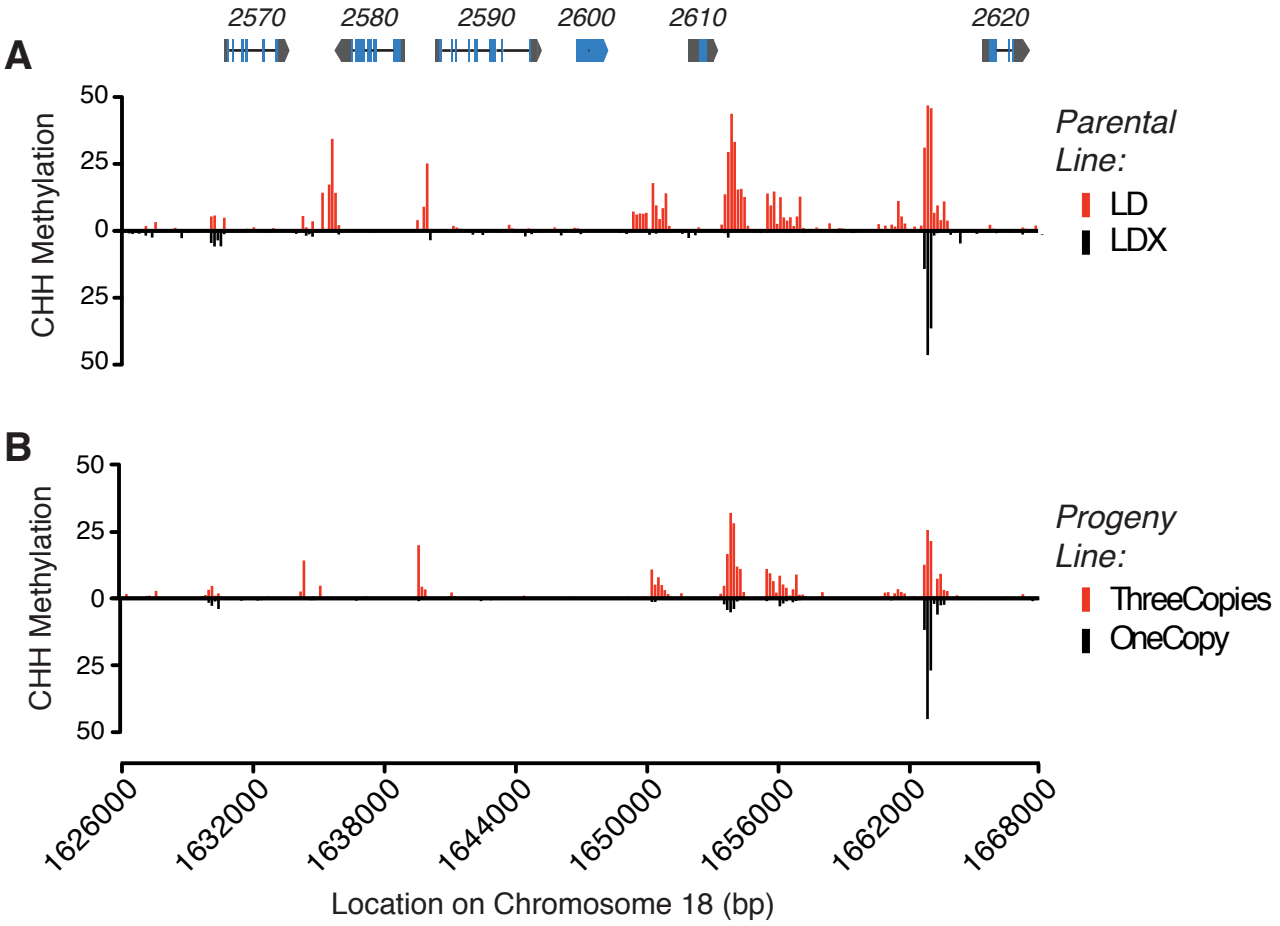

Figure S4. DNA methylome sequence in the CHH context from three-copy and single-copy *Rhg1* lines further support differential methylation at *Rhg1* SCN resistance genes.
Levels of DNA methylation reported as proportion of methylated cytosines detected from bisulfite sequencing. Data are for 150bp bins represented by a single vertical line. *Rhg1* locus gene models are shown at the top of panel (A) at correct x-axis position along chromosome 18 shown below panel (B) for reference (blue exons, black line introns, grey untranslated regions); gene name above gene model (e.g., *2570* = *Glyma18g02570*). (A) Levels of cytosine methylation for the sequence context CHH, showing differential methylation of parental line LD (three-copies of *Rhg1*, red vertical lines above x-axis) relative to parental line LDX (single copy of *Rhg1*, black vertical lines below x-axis). The greatest differential methylation is present up and downstream of the *Glyma18g02580* open read frame (ORF), in the common promoter for *Glyma18g02580* and *Glyma18g02590*, and both up and downstream of the *Glyma18g02610* ORF, with more methylation in the three-copy *Rhg1* SCN-resistant line. (B) Average CHH methylation in F3-derived progeny families of the cross between lines LD and LDX, either for all six progeny estimated to have an *Rhg1* copy number of 3 (red vertical lines above x-axis), or for all 16 progeny lines estimated to have an *Rhg1* copy number of 1 (black vertical lines below x-axis). Substantial similarities to the parental CHH methylation patterns are evident.

Figure S5

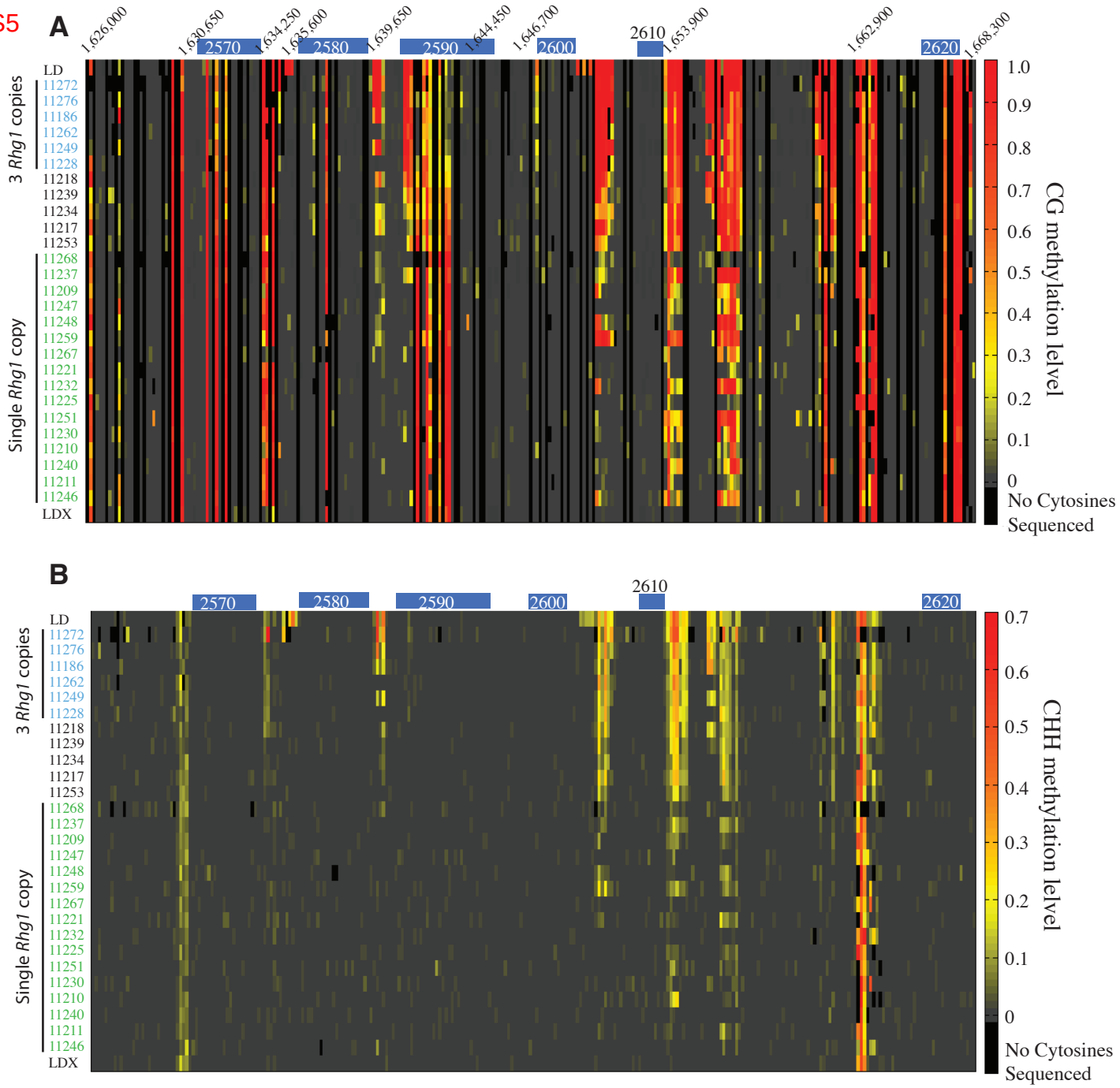

Figure S5. Soybean lines estimated to contain 3 copies of *Rhg1* display high levels of cytosine methylation in regulatory regions of genes shown to impact SCN resistance.

A heatmap depicting cytosine methlyation levels in 150bp bins at *Rhg1* shows high levels of (A) CG methlyation in line LD, estimated to contain 3 copies of *Rhg1*, relative to line LDX, estimated to contain a single copy of *Rhg1*. Their progeny were also assayed for cytosine methylation and progeny estimated to contain 3 copies of *Rhg1* (shown in blue) have a similarly high level of CG methylation compared to single copy progeny (shown in green). The progeny were selected in the F3 generation and likely contain lines with heterozygous *Rhg1* loci (shown in black). (B) Cytosine methylation was analyzed as in (A) but in the sequence context of CHH, where H is any nucleotide A, T, or C. Chromosome positions are shown for reference. Bins that did not have cytosine sequence data, either because of low coverage or because no cytosine exist in that seuence context within the bin, are shown as black. Gene models above each graph are shown in scale to chromosome 18 for reference.

Table S1. Summary statistics for SoyNAM whole genome sequencing

Genotype	Sequence (Mb)	Average Coverage	Reads	Quality Score	Read Length (bp)
LD00-3309	12,443	13.1	82,405,208	34.3	151
LG05-4292	11,725	12.3	77,646,454	31.05	151
4J105-3-4	12,112	12.7	80,213,570	32.9	151
CL0J095-4-6	9,641	10.1	63,847,612	32.58	151
LD02-4485	7,565	8.0	50,099,482	34.48	151
LD02-9050	6,191	6.5	40,999,452	34.19	151
Maverick	5,750	6.1	38,077,200	34.67	151
LD01-5907	5,915	6.2	39,174,146	31.51	151
LG05-4317	9,464	10.0	62,678,462	30.94	151
NE3001	8,535	9.0	56,524,252	34.51	151
PI518_751	14,566	15.3	96,464,146	31.88	151
LG92-1255	9,227	9.7	61,105,816	31.78	151
LG94-1128	6,025	6.3	39,903,618	31.93	151
LG94-1906	8,126	8.6	53,811,902	32.95	151
CL0J173-6-8	6,375	6.7	42,218,296	32.53	151
HS6-3976	9,026	9.5	59,774,700	32.4	151
LG03-3191	11,802	12.4	78,160,256	31.49	151
LG04-4717	9,997	10.5	66,208,394	31.23	151
PI398_881	10,717	11.3	70,971,440	29.14	151
PI427_136	13,702	14.4	90,738,654	28.98	151
PI507_681B	11,330	11.9	75,029,874	29.56	151
LG90-2550	5,441	5.7	36,035,936	31.5	151
Prohio	10,322	10.9	68,354,318	32.39	151
LG98-1605	5,344	5.6	35,391,702	32.82	151

Data are listed for each genotype and summarized as the total amount of sequence generated in megabases (Mb). Average Coverage is the genome wide average sequence coverage. Reads corresponds to the number of reads generated and the Quality Score is the average Phred based quality score for the total reads. All sequences were generated from a short insert library with paired-end sequencing of 151 bases.

Table S2. Summary statistics for Hg-Type Test whole genome sequencing

Genotype	Sequencing Library	Sequence (Mb)	Average Coverage	Reads	Quality Score	Read Length (bp)
Cloud	Paired-end	11,860	12.5	117,426,128	34.98	101
	Mate-pair	1,944	2.0	19,252,220	34.12	101
PI 209332	Paired-end	14,314	15.1	141,724,130	35.16	101
	Mate-pair	2,994	3.2	29,642,412	33.93	101
PI 437654	Paired-end	17,519	18.4	173,457,140	35.48	101
	Mate-pair	6,205	6.5	61,433,454	34.07	101
Peking	Paired-end	45,670	48.1	452,180,560	35.65	101
	Mate-pair	1,756	1.8	17,390,516	34.71	101
PI 89772	Paired-end	12,894	13.6	127,662,706	35.23	101
	Mate-pair	4,247	4.5	42,049,176	33.86	101
PI 90763	Paired-end	17,043	17.9	168,742,468	35.46	101
	Mate-pair	3,002	3.2	29,720,314	34.15	101

Data are listed for each genotype, separated into the sequencing for each library type. The total amount of sequence generated in megabases (Mb) and Average Coverage are presented genome wide. Reads corresponds to the number of reads generated and Quality Score is the average Phred based quality score for the total reads. All sequences had a read length of 101 bases.

Table S3. Estimated *Rhg1* copy number for SoyNAM lines using rapid mapping

	Copy Number Estimates	
Genotype	**Chromosome 18 (*Rhg1*)**	**Chromosome 11 (paralog)**
4J105-34	9.9 ± 1.9	1.0 ± 0.2
LD00-3309	9.9 ± 1.8	0.9 ± 0.2
LD02-4485	9.8 ± 2.2	1.0 ± 0.3
CL0J095-46	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
PI574486	1.3 ± 0.2	
LG05-4317	1.3 ± 0.2	
LG97-7012	1.2 ± 0.1	
LG04-4717	1.1 ± 0.6	
LG98-1605	1.1 ± 0.4	
PI427136	1.1 ± 0.3	
PI404188A	1.1 ± 0.3	
LG90-2550	1.1 ± 0.3	
U03-100612	1.1 ± 0.2	
PI398881	1.1 ± 0.2	
5M20-252	1.1 ± 0.2	
S06-13640	1.1 ± 0.2	
LG05-4832	1.1 ± 0.1	
LG94-1906	1.1 ± 0.1	
CL0J173-68	1.1 ± 0.1	
LG94-1128	1.1 ± 0.1	
PI518751	1.0 ± 0.3	
LG92-1255	1.0 ± 0.3	
HS6-3976	1.0 ± 0.3	
Prohio	1.0 ± 0.2	
PI561370	1.0 ± 0.2	
PI507681B	1.0 ± 0.2	
LG03-3191	1.0 ± 0.2	
LG03-2979	1.0 ± 0.2	
IA3023	1.0 ± 0.2	
NE3001	0.9 ± 0.3	
LG05-4464	0.9 ± 0.2	

Short reads from whole genome sequencing were aligned to a portion of the reference genome to rapidly estimate *Rhg1* copy number (chromosome 18), results are listed by

Table S3 cont'd
genotype. Copy number was estimated by summing the total number of reads in three equally sized DNA intervals spanning the *Rhg1* repeat, and the 5' and 3' adjacent intervals. The total number of reads from the *Rhg1* interval was independently divided by the total number of reads from the two adjacent intervals to estimate copy number. The reported copy number is the average of the two estimates along with the standard error of the mean. Similar analysis was performed for the paralogous sequence on chromosome 11, and did not significantly deviate from a single copy.

Table S4. Sequenced cDNA products confirms the expression of multiple alleles of *Glyma18g02590* in the different multi-copy *Rhg1* classes

Genotype	Identified Allele	Sequenced cDNA products
PI 88788	High copy allele	26
	Low copy allele	0
	Splice isoform	0
	Williams-type	2
PI 209332	High copy allele	8
	Low copy allele	0
	Splice isoform	0
	Williams-type	0
Cloud	High copy allele	7
	Low copy allele	0
	Splice isoform	0
	Williams-type	1
Peking	High copy allele	0
	Low copy allele	8
	Splice isoform	1
	Williams-type	0
PI 90763	High copy allele	0
	Low copy allele	6
	Splice isoform	3
	Williams-type	0
PI 89772	High copy allele	0
	Low copy allele	6
	Splice isoform	1
	Williams-type	0
PI 437654	High copy allele	0
	Low copy allele	8
	Splice isoform	3
	Williams-type	0
Williams 82	High copy allele	0
	Low copy allele	0
	Splice isoform	0
	Williams-type	6

Products from cloning cDNA products are listed by genotype. The identified allele corresponds to different alleles of *Glyma18g02590* from the Hg Type Test lines and Williams 82. A zero indicates the transcript was not observed. A splice isoform was detected in all low-copy genomes not detected in Williams 82 or the high-copy genomes.

Table S5. Amino acid polymorphisms for *Rhg1* paralogous genes encoded on chromosome 11.

Position (bp)	Peking	PI 90763	PI 89772	PI 437654	LD01-5907	Cloud	LG05-4292	Maverick
Glyma11g35840 (Gm18.2570.Paralog)				No Polymorphism				
Glyma11g35830 (Gm18.2580.Paralog)				No Polymorphism				
Glyma11g35820 (Gm18.2590.Paralog)								
37418427	Isoform	Isoform	Isoform	Isoform	Isoform	Isoform	Isoform	Isoform
37418685	A179T	A179T	A179T	A179T	A179T	A179T	A179T	
Glyma11g35810 (Gm18.2600.Paralog)								
37413493	R320Q	R320Q	R320Q	R320Q	R320Q	R320Q	R320Q	R320Q
37413566	A296T	A296T	A296T	A296T	A296T	A296T	A296T	A296T
Glyma11g35800 (Gm18.2610.Paralog)				No Polymorphsism				
Glyma11g35790 (Gm18.2620.Paralog)				No Polymorphsism				

Position : Chromosome 11 base-pair position relative to Williams 82 reference genome, with gene name (and chromosome18 paralog) above relevant bp positions. Of the six genes analyzed, four did not contain polymorphisms relative to Williams 82 as indicated. Isoform: An mRNA splice isoform predicted to be caused by a SNP as reported (Matsye et al., 2012). Amino acid polymorphisms are reported as the amino acid present in Williams 82, the amino acid position, and the resulting new amino acid discovered. Genotypes not listed did not show polymorphic amino acid sequence for the genes analyzed.

Table S6. Sequence frequencies at DNA variant positions across the *Rhg1* repeat indicates varying sequence content between copies.

	Genotypes			
Position	LD00-3309	PI 209332	Cloud	Average Frequency
1633532	1.00	1.00	1.00	1.00
1633629	1.00	1.00	1.00	1.00
1633700	1.00	1.00	1.00	1.00
1633840	1.00	0.99	1.00	1.00
1633930	1.00	1.00	1.00	1.00
1634533	1.00	0.90	1.00	0.97
1634534	1.00	0.92	1.00	0.97
1634535	1.00	0.92	1.00	0.97
1634536	1.00	0.92	1.00	0.97
1634610	1.00	1.00	1.00	1.00
1634620	1.00	1.00	1.00	1.00
1634626	1.00	1.00	1.00	1.00
1634635	1.00	1.00	1.00	1.00
1634643	1.00	1.00	1.00	1.00
1634714	1.00	1.00	1.00	1.00
1634856	1.00	1.00	1.00	1.00
1635001	1.00	1.00	1.00	1.00
1635014	1.00	1.00	1.00	1.00
1635093	1.00	1.00	1.00	1.00
1635120	1.00	1.00	1.00	1.00
1635364	1.00	1.00	1.00	1.00
1635912	1.00	1.00	1.00	1.00
1636766	1.00	1.00	1.00	1.00
1639354	0.96	0.99	1.00	0.98
1640056	1.00	1.00	1.00	1.00
1640137	1.00	1.00	1.00	1.00
1640151	1.00	1.00	1.00	1.00
1640292	1.00	1.00	1.00	1.00
1640480	0.92	0.97	1.00	0.96
1640581	0.88	0.90	0.81	0.86
1640675	0.86	0.90	0.75	0.84
1641208	0.88	0.87	0.88	0.88
1641800	0.89	0.87	0.88	0.88
1642266	0.79	0.92	0.81	0.84
1642672	0.91	0.92	0.84	0.89
1642762	0.94	0.95	0.91	0.93
1642848	0.87	0.90	0.81	0.86
1643208	0.91	0.88	0.87	0.89
1643849	0.89	0.89	0.87	0.89
1644089	0.91	0.87	0.83	0.87
1644385	0.90	0.96	0.90	0.92
1644493	0.88	0.93	0.88	0.90
1644525	0.89	0.89	0.84	0.87
1644577	0.85	0.87	0.84	0.85
1644965	0.85	0.87	0.81	0.84
1644968	0.84	0.88	0.81	0.84
1644972	0.87	0.92	0.87	0.89
1644974	0.84	0.88	0.83	0.85

1645218	0.84	0.89	0.77	0.83
1645437	0.86	0.93	0.92	0.90
1645745	0.90	0.91	0.86	0.89
1645759	0.88	0.88	0.85	0.87
1645811	0.92	0.87	0.84	0.88
1645908	0.88	0.89	0.87	0.88
1645914	0.93	0.92	0.93	0.93
1646130	0.81	0.70	0.54	0.68
1646138	0.77	0.62	0.51	0.63
1646145	0.76	0.61	0.50	0.62
1646211	0.88	0.85	0.82	0.85
1646226	0.91	0.86	0.83	0.87
1648850	0.93	0.94	0.85	0.91
1649069	0.96	0.90	0.88	0.91
1649212	0.91	0.89	0.86	0.88
1649293	0.96	0.91	0.87	0.92
1649328	0.94	0.92	0.87	0.91
1649335	1.00	0.93	0.81	0.91
1649371	0.92	0.89	0.85	0.88
1649385	0.90	0.90	0.83	0.88
1649553	0.97	0.97	0.92	0.95
1649630	0.84	0.93	0.85	0.87
1649892	0.92	0.91	0.85	0.89
1649934	0.91	0.87	0.88	0.88
1650106	0.93	0.94	0.93	0.93
1650140	0.90	0.91	0.90	0.90
1650201	0.89	0.92	0.89	0.90
1650310	0.98	0.95	0.92	0.95
1650501	0.89	0.86	0.88	0.88
1650590	0.87	0.93	0.98	0.93
1650772	0.90	0.88	0.86	0.88
1651003	0.95	1.00	1.00	0.98
1651056	0.93	0.93	0.92	0.92
1651602	0.95	0.91	0.90	0.92
1652723	0.92	0.93	1.00	0.95
1653661	0.82	0.90	0.87	0.86
1654230	0.87	0.96	0.82	0.88
1654282	0.94	0.92	0.80	0.89
1655099	0.88	0.87	0.72	0.82
1655195	0.86	0.92	0.86	0.88
1655348	0.94	0.99	0.89	0.94
1655353	0.89	0.92	0.85	0.89
1655408	0.95	0.90	0.87	0.91
1655500	0.94	0.95	0.90	0.93
1655564	0.96	0.94	0.82	0.91
1655585	0.93	0.87	0.82	0.87
1655836	0.88	0.89	0.84	0.87
1656044	0.90	0.88	0.80	0.86
1656178	0.88	0.94	0.85	0.89
1656263	0.88	0.89	0.82	0.86
1656394	0.88	0.87	0.83	0.86
1656417	0.82	0.82	0.71	0.79
1656462	0.81	0.84	0.85	0.83
1656633	0.88	0.94	0.94	0.92

1656719	0.91	0.97	0.79	0.89
1656769	0.88	0.91	0.83	0.87
1656898	0.89	0.91	0.81	0.87
1656979	0.92	0.95	0.84	0.90
1657025	0.42	0.54	0.51	0.49
1657162	0.85	0.91	0.91	0.89
1657183	0.86	0.93	0.90	0.90
1657307	0.91	0.86	0.87	0.88
1657506	0.91	0.88	0.88	0.89
1657803	0.87	0.88	0.74	0.83
1657807	0.83	0.76	0.68	0.76
1657815	0.84	0.76	0.73	0.78
1657816	0.84	0.74	0.70	0.76
1658170	0.90	0.93	0.87	0.90
1658284	0.83	0.91	0.81	0.85
1658617	0.86	0.88	0.91	0.88
1658735	0.90	0.91	0.88	0.90
1659502	0.91	0.95	0.93	0.93
1659777	0.88	0.92	0.86	0.89
1659829	0.83	0.89	0.83	0.85
1659914	0.91	0.88	0.87	0.89
1659945	0.91	0.83	0.88	0.87
1660067	0.91	0.88	0.82	0.87
1660183	0.89	0.91	0.89	0.90
1660790	1.00	1.00	1.00	1.00
1661155	1.00	1.00	1.00	1.00
1661264	0.87	0.72	0.62	0.74
1661293	0.87	0.78	0.74	0.80
1661406	0.89	0.92	0.86	0.89
1661428	0.91	0.90	0.86	0.89
1661460	0.95	0.93	0.84	0.91
1662031	0.96	0.91	0.86	0.91
1662115	1.00	1.00	1.00	1.00
1662177	0.88	0.87	0.86	0.87
1662656	0.84	0.94	0.89	0.89
1662666	0.80	0.91	0.89	0.87
1662682	0.81	0.93	0.91	0.88
1662714	0.86	0.95	0.91	0.91
1662734	1.00	1.00	1.00	1.00
1662810	0.82	0.85	0.85	0.84
1662851	1.00	1.00	1.00	1.00
1662946	0.83	0.83	0.80	0.82
1662953	0.71	0.83	0.79	0.78
1663007	0.63	0.90	0.86	0.80
1663014	0.63	0.90	0.82	0.78
1663032	0.60	0.90	0.85	0.79
1663064	0.57	0.90	0.86	0.78
1663114	0.69	0.94	0.86	0.83
1663133	0.65	0.92	0.78	0.78
1663148	0.66	0.92	0.76	0.78
1663225	1.00	1.00	1.00	1.00
1663250	0.70	0.94	0.78	0.81

Table S6 cont'd

Position: Base pair position on chromosome 18 of Williams 82 corresponding to a DNA variant position. The variant allele frequency is reported below each genotype as the sum of the total number of reads supporting an alternate sequence at the position, divided by the total number of sequenced reads (wild-type plus variant) at the position. Average Frequency: The average frequency at a given variant site computed from the three genomes.