Atlas Journal of Biology 2017, pp. 376–383
doi: 10.5147/ajb.2017.0153

Additional Polymorphisms Linked to Soybean Cyst Nematode Resistance At The *Rhg4* Locus

Oleksandra Zatserklyana[1,2], Khalid Meksem[3], and David A. Lightfoot[2,3,*]

[1] Department of Biochemistry and Molecular Biology, Southern Illinois University Carbondale, 1245 Lincoln Dr., Mail Code 4413, Carbondale, IL 62901; [2] Genomics Core Facility; Department of Plant Soil and Agricultural Systems, and Center for Excellence in Soybean Research, Teaching and Outreach, Southern Illinois University at Carbondale, Carbondale, IL 62901; [3] Department of Plant, Soil, and Agricultural Systems, Southern Illinois University Carbondale, 1205 Lincoln Dr., Mail Code 4415, Carbondale, IL 62901.

Received: March 29, 2017 / Accepted: April 20, 2017

Abstract

***Glycine max* L. Merr., (soybean) is one of the major crops in the United States, south America and Asia. Yet, the seed yield of soybean is significantly reduced due to *Heterodera glycines* (Ichinohe), the soybean cyst nematode (SCN). SCN is one of the most destructive pests and pathogens of soybean because the main methods for control have proven difficult. Lineages of soybean that are resistant to SCN have been developed but it is yet unknown what all the molecular causes of this resistance are. *Rhg1* and *Rhg4* loci are two of about 10 loci that underlie resistance. The *Rhg4* locus, is required for resistance to SCN race 3 in 'Peking'–type derived resistances. One gene in the locus has been isolated and a causative link shown, but the surrounding regions have not been fully analyzed. Focusing on the *Rhg4* locus, this study was aimed at uncovering other potential causes of resistance of soybean to SCN. In a bioinformatic analysis of the *Rhg4* locus, alleles of nine genes were analyzed. The set of large intergenic regions have key regulatory elements in them. Since partial resistances are often multigeneic, some of these nine genes could be candidates for causing and regulating resistance.**

Keywords: Soybean; nematode; resistance; *Rhg4*.

Introduction

SCN or soybean cyst nematode (*Heterodera glycines* Ichinohe,) is one of the most destructive pests of *Glycine max* L. Merr. (soybean) worldwide (Matthews, 2004; Kurle et al., 2015; Yu Q. 2015). Due to the importance of *G. max* as an agricultural crop in the United States, the need for better management of SCN-infested fields is pronounced, as crop yield can be reduced by up to 80% in the presence of SCN in a field (Smolik and Draper, 2007; Kazi et al., 2017).

Two loci involved in resistance to SCN, are *rhg1* and *Rhg4* (Liu et al., 2012; Liu et al., 2011). Both *rhg1* and *Rhg4* combine to encode resistance to the same type of SCN race 3, HgType0 (Afzal et al., 2012; 2013; Liu et al., 2012). *Rhg4* encodes resistance (Brucker et al., 2005; Liu et al., 2012) only in the 'Peking' – type of resistance that is the chief subject of this study.

Among the resistant lines that have been developed, none have been fully sequenced (Schmutz eet al., 2010). Among the partially sequenced resistant cultivars is 'Forrest' (unpublished) that was used extensively in this study. The cultivars used that were susceptible to *H. glycines* were 'Williams 82' [GenBank: EF623856] and 'Asgrow 3244' [GenBank: GP062386.1]. Deletions in these sequences can be seen (Zatserklyana, 2015; Tuteja and Vodkin 2008; Hauge et al., 2009a,b), introducing the problem of three sequences possibly not aligning correctly during genome mapping and recombination events. The origin of the Forrest cultivar was from the Peking lineage of soybeans (Hartwig and Epps 1973; Liu et al., 2009). Peking is a semi-wild, or semi-domesticated, black seeded *G. max*. The Forrest cultivar was derived from Peking lineage by introgression, of CNVs,

* Corresponding author: ga4082@siu.edu

SNPs, insertions and deletions at both the *rhg1* and *Rhg4* loci (Cook et al., 2012; Liu et al 2012; Srour et al., 2012).

Forrest cultivar sequence derived from two BAC libraries using BamHI and HindIII enzymes (Meksem et al., 2000). The construction of BAC libraries and the physical map helped put the fragmented genome sequence of Forrest together (Shultz et al., 2006). Although the Forrest cultivar library and the BAC libraries for the Asgrow 3244 and the Williams 82 susceptible cultivars were available, complete BAC libraries did not exist for either cultivar (Wu et al., 2004). Wu et al. (2004) reported a complete physical map from three BAC libraries derived from the Forrest cultivar. As the BAC libraries created complement each other, they expanded the number of clones used soybean genome analysis.

The Williams 82 susceptible cultivar, in contrast, arose largely through intra-cultivar genetic heterogeneity following or during introgression of phythophtora resistances (Haun et al., 2011). Not all the genes at or near or in the *Rhg4* locus have had their functions discovered (Yi et al., 2010; Center for Integrative Genomics 2017; Liu et al., 2017). Among those that have is CHS7, which is the seventh gene in the chalcone synthase family, a hydroxyl-proline serine transferase and the I gene for seed coat color.

The difficulty in fully mapping the *Rhg4* locus and the molecular markers associated with it arises out of HgType variability, genomic variability in the soybean and the very small distance between this locus and I locus of less than 0.01cM (Weisemann et al., 1992; Lewers et al., 2001; Wu et al., 2011). This relatively small genomic distance between the loci makes it difficult to map the *Rhg4* locus relative to the I locus at high resolution (Weisemann et al., 1992). For *rhg1* and *Rhg4* loci, both mapping and annotation are necessary to fully elucidate the full set of causes of resistance to all HGTypes of SCN. Genetic mapping alone may not be sufficient to show this (Lewers et al., 2002). The annotation to date has shown extensive duplications that have taken place in BAC 56G2, the BAC where the Williams 82 *Rhg4* susceptible sequence is located (Lewers et al., 2002). Here, nine genes were found in a BAC derived from Forrest that aligned poorly with Williams 82 sequences, some of which could be candidates for causing and regulating resistance. Further, large variations in intergenic regions were present that could have key regulatory elements within them.

Materials and Methods

BAC sequences from Forrest, Asgrow 3244, and Williams 82 soybean cultivars, and BLASTn, tBLASTx, tBLASTn, BLASTp, ORF Finder, and GeneMark.hmm were used. Sequence from Forrest cultivar was provided by Dr. Lightfoot from a BAC sent to JCVI for sequencing. The other cultivar sequences were obtained through BLAST searches from NCBI (used for Asgrow 3244) and Phytozome (for Williams 82). The nucleotide BLAST program at NCBI was run to differentiate between the two files containing the Forrest sequence. The alignment of each of the files containing the Forrest sequence separately gave the accession number of the BAC Williams 82 sequence. It was the accession number of the BAC Williams 82 sequence that was used in the alignments.

BLAST and ClustalW were used to do multiple alignments. With the BLASTn program, the alignment was being shown as broken up into sections. After the large gaps were found, Needleman-Wunsch was used to complete the alignment. Completing the alignment included trimming the Williams 82 56G2 BAC sequence to the length of the Forrest BAC (92 kbp) and doing alignments with the trimmed Williams 82 56G2 BAC sequence.

Pairwise alignments and ORF Finder were used to search for genes and for what could be genes in the *Rhg4* region. The pairwise alignments were done using BLASTn, tBLASTx, and tBLASTn (McGinnis and Madden, 2004; Ye et al., 2006; Johnson et al., 2008) using the Forrest B100B10 BAC DNA sequence and the nr database, limiting the search to G max. The reason for this was to see what known genes, if any, could align with the Forrest B100B10 BAC DNA sequence. The ORF Finder at NCBI (Wheeler, 2003) complemented these BLAST searches by showing what other genes may be encoded in the Forrest B100B10 BAC DNA sequence and where they may be located.

GeneMark.hmm and BLASTp were used in annotating the Forrest B100B10 BAC sequence. GeneMark.hmm was used to predict the genes that are located on the Forrest B100B10 sequence using the GFF mechanism of the program and having GeneMark.hmm give the protein sequences as part of the output. BLASTp was then used to look for the possible amino acid changes in the protein sequence.

Results

InDels and SNPs

All the insertions and deletions were looked at in Forrest in comparison to both susceptible cultivars and to either of the susceptible cultivars as was the case in the region of alignment where Williams 82 and Asgrow 3244 overlap (Supplementary Table 1). This was the region from 850 bp to 7,466 bp.

Longer Insertions and Deletions

The region of alignment between the Forrest and Williams 82 sequences showed a protein coding region (1,284 – 5,720 bp) in the deletions in Forrest (Supplementary Table 1). Out of the 1603 amino acids in this protein, nineteen amino acids were deleted from Forrest in the process of the Forrest cultivar becoming resistant to SCN. Sixteen of the nineteen deleted amino acids were nonpolar and hydrophobic, suggesting that this protein lies somewhere in one of the membranes in the cells. More than sixty percent of the deletions where codons have been deleted from the Forrest cultivar were more AT rich than GC rich (Table 1).

Outside of the protein-coding region in the stretch of DNA where Forrest and Williams 82 align, this becomes a pattern. 99.8% of the deletions in this region are AT rich. It is possible these deletions result from Forrest cultivar undergoing mutation while being strongly selected for resistance to SCN during backcrossing. Table 2 shows the deletions and how AT rich they are over the entire Forrest-Williams 82 region.

Table 1. Here are listed the deletions in the Forrest-Williams 82 protein-coding region and the percentage of AT or GC that is present in each deletion. The deletions, which are indicated in column 1 by number, all come from the Forrest sequence compared to the Williams 82. Columns 2 and 3 show the starting and ending base position of each deletions and while the numbers in columns 2 and 3 correspond to the base positions in Asgrow, they are of use as this sequence overlaps with Williams 82 in this region. Columns 4 and 5 show the percentage of AT and GC, respectively, and the last column shows codon.

Number	Start	End	Length	% AT	Codon
1	96623	96625	3	33	TGG
2	96634	96636	3	66.7	GAA
3	96749	96751	3	66.7	ATG start
4	96877	96879	3	33	GGT
5	96985	96987	3	67	TTG
6	97123	97125	3	67	CTA
7	97168	97170	3	67	ATT
8	97274	97276	3	67	GTA
9	97725	97727	3	66.7	not a codon
10	98135	98137	3	66.7	TAC
11	98409	98411	3	66.7	CAT
12	98554	98556	3	33.3	CAC
13	98680	98682	3	33.3	CAG
14	99114	99116	3	33.3	AGC
15	99127	99129	3	100	TTT
16	99379	99381	3	33.3	GTG
17	100089	100091	3	33.3	GGA
18	100655	1000657	3	33.3	GGA
19	101159	101161	3	66.7	TTC
20	101256	101258	3	33	GAG
21	101322	101324	3	100	ATT
22	101337	101339	3	66.7	GTA
23	101436	101438	3	66.7	CTT
24	101556	101558	3	100	TAA stop

SNPs

The distribution of SNPs forms a more or less normal distribution through the Forrest-Williams 82 and the Forrest-Asgrow 3244 aligned DNA. In the same region (through the 92.5% of the region of DNA where the Williams 82 sequence and the Asgrow 3244 sequence align), the number of the insertions and deletions in SNPs differ by whether a nucleotide is a purine or a pyrimidine. A few SNPs that are insertions are found in pyrimidine nucleotides, the opposite is seen with the SNPs that are deletions (Table 3). However, when the alignment of Forrest B100B10 BAC DNA sequence and Williams 82 56G2 BAC DNA sequence was analyzed, there were 508 SNPs that are result of a single nucleotide substitution.

There is great diversity between the Forrest B100B10 BAC DNA sequence and the Williams 82 56G2 BAC DNA sequence. Graphically, in the alignment between the Forrest B100B10 BAC DNA sequence and the Williams 82 56G2 BAC DNA sequence the two sequence appear to align well in small regions (Figure 1-2). Analysis of this alignment, by BLAST, showed the alignment as sections with gaps. Hence the need for Needleman-Wunsch alignments.

Table 2. Here is listed the deletions in the entire Forrest-Williams 82 region and the percentage of AT or GC that is present in each deletion. The deletions, which are indicated in column 1 by number, all come from the Forrest sequence in comparison to the Williams 82. Columns 2 and 3 show the starting and ending base position of each deletions and while the numbers in columns 2 and 3 correspond to the base positions in Asgrow, they are of use as this sequence overlaps with Williams 82 in this region. Column 4 shows the length of each deletion and was the basis of sorting the values of the table for the graph. Column 5 shows how AT rich each of the deletions is while the percentage of how GC richness each deletion is shown in column 6.

Number	Start	End	Length	% AT
1	96623	96625	3	33
2	96634	96636	3	66.7
3	96749	96751	3	66.7
4	96877	96879	3	33
5	96985	96987	3	67
6	97123	97125	3	67
7	97168	97170	3	67
8	97274	97276	3	67
9	97725	97727	3	66.7
10	98135	98137	3	66.7
11	98409	98411	3	66.7
12	98554	98556	3	33.3
13	98680	98682	3	33.3
14	99114	99116	3	33.3
15	99127	99129	3	100
16	99379	99381	3	33.3
17	100089	100091	3	33.3
18	100655	100657	3	33.3
19	101159	101161	3	66.7
20	101256	101258	3	33
21	101322	101324	3	100
22	101337	101339	3	66.7
23	101436	101438	3	66.7
24	101556	101558	3	100
25	101999	102001	3	100
26	102046	102048	3	66.7
27	102253	102255	3	100
28	102727	102729	3	67
29	103201	103203	3	66.7
30	101890	101893	4	50
31	96436	96439	4	25
32	96792	96795	4	50
33	97640	97644	4	60
34	97903	97906	4	75
35	98216	98219	4	25
36	98424	98427	4	25
37	98647	98650	4	50
38	98662	98665	4	75
39	98938	98941	4	50
40	99183	99186	4	60
41	99800	99804	4	40
42	100034	100037	4	75
43	101092	101095	4	50
44	101105	101108	4	50
45	101293	101296	4	100
46	101924	101927	4	50
47	101946	101949	4	50
48	102124	102127	4	50
49	102231	102234	4	50
50	103126	103129	4	50

Table 2. Continued.

Number	Start	End	Length	% AT
51	96314	96318	5	40
52	96447	96451	5	100
53	97284	97288	5	60
54	97315	97319	5	60
55	97411	97415	5	80
56	97864	97868	5	80
57	97965	97969	5	60
58	98330	98334	5	0
59	98704	98708	5	60
60	98752	98756	5	80
61	98881	98885	5	60
62	100834	100838	5	40
63	101391	101395	5	60
64	101527	101531	5	60
65	101690	101694	5	60
66	102389	102393	5	80
67	102423	102427	5	80
68	102990	102994	5	100
69	97196	97201	6	83.3
70	97970	97975	6	83.3
71	97985	97990	6	66.7
72	98715	98720	6	66.7
73	103174	103179	6	66.7
74	97253	97259	7	85.7
75	97336	97342	7	57.1
76	97813	97819	7	57.1
77	101303	101309	7	100
78	101580	101586	7	100
79	96378	96385	8	75
80	96465	96472	8	75
81	97385	97392	8	75
82	97710	97717	8	50
83	98381	98388	8	25
84	100609	100616	8	50
85	101673	101680	8	87.5
86	101982	101989	8	75
87	96402	96410	9	22.2
88	96981	96989	9	87.5
89	98121	98129	9	55.6
90	98617	98625	9	27.8
91	100780	100788	9	66.6
92	101265	101273	9	88.9
93	101281	101289	9	66.7
94	97268	97277	10	60
95	100710	100719	10	90
96	101444	101453	10	60
97	97678	97688	11	81.2
98	101535	101545	11	81.8
99	102951	102962	12	75
100	96330	96343	14	46.2
101	102885	102898	14	14.3
102	97474	97488	16	50
103	96850	96866	17	68.8
104	98584	98600	17	68.8
105	99412	99429	18	94.1

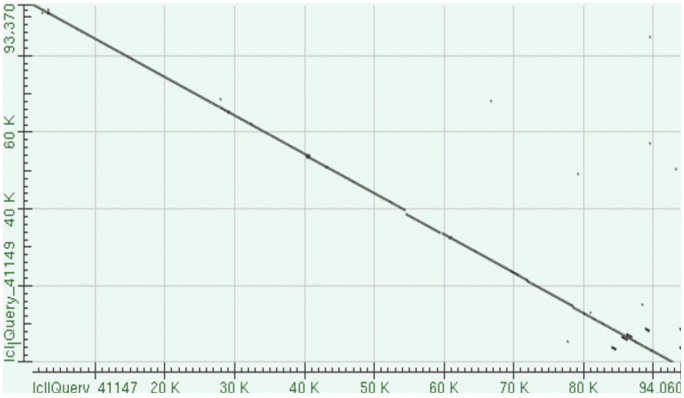

Figure 1. The alignment between the Forrest B100B10 BAC DNA sequence (y axis) and the Williams 82 56G2 BAC DNA sequence (x axis) that had been trimmed to the length of the Forrest B100B10 BAC DNA sequence of 93.7bp.

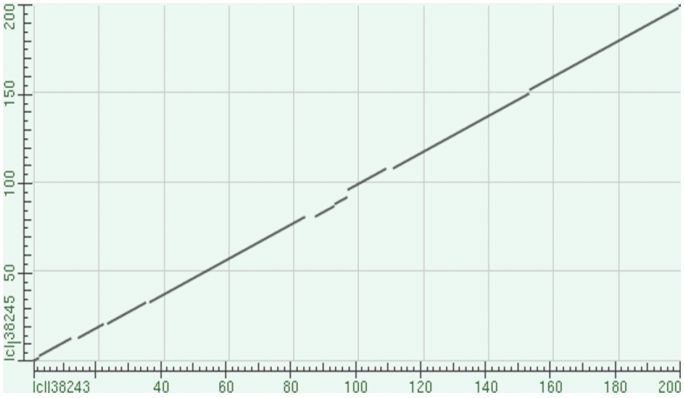

Figure 2. The alignment of the Rhg4 gene between the Forrest B100B10 BAC DNA sequence (x axis) and the Williams 82 56G2 BAC DNA sequence (y axis) from 1-200 bp.

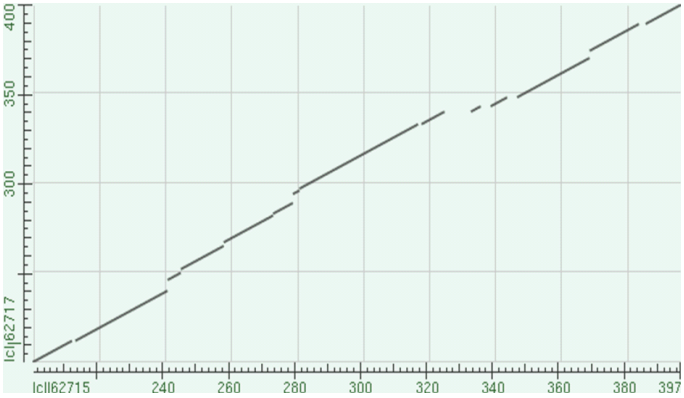

Figure 3. The alignment of the Rhg4 gene between the Forrest B100B10 BAC DNA sequence (x axis) and the Williams 82 56G2 BAC DNA sequence (y axis) from 201-400 bp.

Needleman-Wunsch

With the subrange for the Query and Subject sequences set to 10 kbp for the first nine sub-regions, and then to 1336 bp the tenth sub region, the Needleman-Wunsch algorithm was able to make the alignment. Between 1 and 200 bp, 201 and 400 bp, 1 and 1000 bp, 90,001 and 94,336bp there were many gaps in

Table 3. SNPs that result from single nucleotides substitutions and the changes in amino acids they cause, along with the genes these SNPs are expected to be located in. Red letters indicate insertion and blue letters indicate deletion. (The information in this table is up to date as of August 14, 2015).

Gene #	Gene Name	Strand	Start	End	SNPs	Amino Acid changes
1	Serine hydroxymethyltransferase	-	1441	2757	1593 A to T	GNIYRIGSLPSGFDLL QMSINLTCSLCDCFA
2	Adenosylhomocysteinase	+	10044	12138	no substitutions	V238F, M239D, S240N, F242Y, P243G, CRHSL, L24G6, V247L, SFLC
3	Predicted heat stress transcription factor A-5-like isoform X2	-	20983	21828	no substitutions	no amino acid changes
4	Unknown *Glycine max* gene	+	25127	25432	no substitutions	R10F, K11P, T12Q, Y13T, V14I, G15D, A19S, A22T, K24E, V25I, L28F, R32G, E33K, K34E, E35R, A36D, S37G, G40S, R41D, P43G, H44D, E46K, V49T, G50A, G51T, R54T, N55D, V57R, V58I, G59S, G60V, G61A, I62V, V63F, L66I, G68T, L69F, E70F, K71Q, R75H, T7R6, T77N, G79T, L80R, W81T, R82V, T83L, H85R
5	Predicted putative pentatricopeptide repeat-containing protein At1g12700 mitochondrial isoform X2	-	35243	36895	no substitutions	no amino acid changes
6	Predicted microfibrillar-associated protein 1	+	37980	39182	38375 G to A, 38436 G to C, 38459 G to A	no amino acid changes
7	Receptor-like kinase, Leucine-rich repeat receptor-like kinase	-	73853	74395	73866 G to A	no changes
8	Leucine-rich receptor-like kinase, Rhg4-like receptor kinase, receptor-like kinase	-	75072	77252	76596 A to T, 77124 C to A	G722A
9	Predicted anthocyanidin 3-O-glucoside 5-O-glucosyltransferase 1-like	-	83975	85411	84209 T to C, 84232 C to T, 84239 G to C, 84242 C to T, 83245 C to A, 84251 G to T, 84257 C to T, 84260 A to G, 84265 T to A, 84279 G to A, 84287 T to G, 84290 A to C, 84302 A to C, 84308 T to C, 84326 C to A, 84335 T to C, 84337 T to C, 84340 A to C, 84386 C to A, 84395 T to C, 84397 T to A, 84416 T to C, 84504 C to G, 84520 G to T, 84534 A to G, 84552 T to A, 84563 A to C, 84574 C to T, 84584 C to G, 84589 T to C, 84590 C to G, 84611 T to C, 84621 T to C	no amino acid changes

the alignment (Figures 3-5).

Alignments inferred the exact (already known) location of *Rhg4* gene (Liu et al., 2012). Using the trimmed and untrimmed version of the Williams 82 BAC sequence, the Query Coverage percentages were compared. For the alignment of Forrest as the Query Sequence and the trimmed Williams 82 BAC sequence as the Subject Sequence, the Query Coverage was 97% while for the alignment of Forrest as the Query Sequence and the untrimmed Williams 82 BAC sequence (starting from the point in the Williams 82 BAC sequence where Forrest sequence would end) as the Subject Sequence it was only 3%.

ORF Finder

An unknown was how many of the deleted or inserted codons (Table 1) were in the coding region of proteins. In total, in the Forrest B100B10 BAC sequence the ORF Finder found 327 ORFs (open reading frames), six of which are greater than 1,000 bp in length (Figure 6). One open reading frame was similar to serine hydroxyl-methyl serine transferase mRNA (1,317 bp for the open reading frame and 1,416 bp for the mRNA) but was incomplete on the BAC.

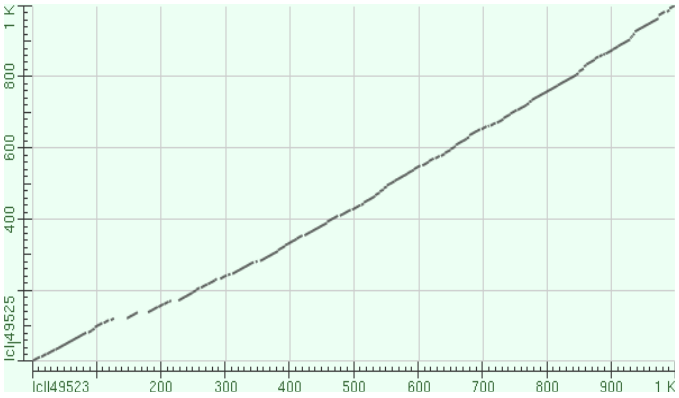

Figure 4. The alignment of the Rhg4 gene between the Forrest B100B10 BAC DNA sequence (x axis) and the Williams 82 56G2 BAC DNA sequence (y axis) from to 1-1000 bp.

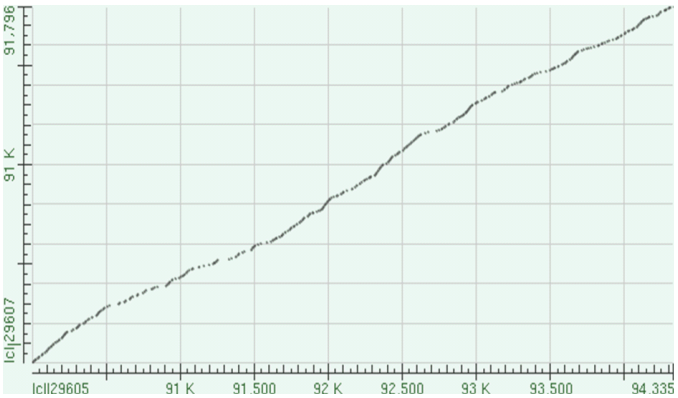

Figure 5. The alignment of the Rhg4 gene between the Forrest B100B10 BAC DNA sequence (x axis) and the Williams 82 56G2 BAC DNA sequence (y axis) from 89 kbp to 91 kbp.

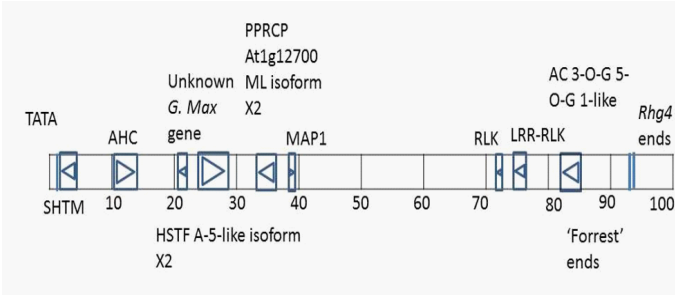

Figure 6. The overview of the Rhg4 region showing the seven genes, regulatory elements and extent of the Rhg4 region in relation to Forrest B100B10 BAC DNA sequence. If we look at these genes as though they would be located on one strand, if transcription would occur, we would see AHC, Unknown G. max gene, and MAP 1 genes being transcribed to the right and SHTM (Rhg4 gene) , HSTF A-5-like isoform X2, PPRCP At1g12700 ML isoform X2, RLK, LRR-RLK, and AC 3-O-G 5-O-G 1-like genes being transcribed to the left. (The information in this figure is up to date as of August 14, 2015).

Annotation

Another unknown was which SNPs, caused changes in the amino acid sequences. GeneMark.hmm predicted ten genes within the Forrest B100B10 BAC DNA sequence, but manual annotation inferred there were nine genes in the Forrest B100B10 BAC DNA sequence, as genes 2 and 3 may be one gene. In contrast to the large number of SNPs that were found in the Forrest B100B10 BAC DNA sequence that are the result of single nucleotide substitutions, the number of amino acid changes and the number of SNPs in the nine coding regions were 56 and 40, respectively. The results of the GeneMark.hmm prediction of genes in the Forrest B100B10 BAC DNA sequence and BLASTp, along with the SNPs and amino acid changes were shown in Table 3.

Discussion

Forrest cultivar sequence resistant to *H. glycines* and the susceptible Asgrow 3244 and Williams 82 cultivar sequences were used in this project. Deletions in these sequences can be seen (Tuteja and Vodkin 2008; Hauge et al., 2009a,b; Zatserklyana, 2015) which can caused the cultivar sequences to not align correctly, partly because the cultivars retain heterologous regions (Haun et al., 2011) and partly because BACs are single alleles cloned from within that variation. Multiple sequence alignments and Needleman-Wunsch Global Alignments were the key tools in the project.

A deleted peptide was noted and the size of insertions and/or deletions was notably larger compared to the alignment around *Rhg1* (Hemmati et al., 2017). Transposons often cause indels and may be responsible for altering the degree and type of SCN resistance. In the *Rhg4* region are predicted to be nine genes, four of which are genes with predicted functions (heat stress transcription factor A-5-like isoform X2, putative pentatricopeptide repeat-containing protein At1g12700 mitochondrial isoform X2, microfibrillar-associated protein 1, and anthocyanidin 3-O-glucoside 5-O-glucosyltransferase 1-like; Figure 6).

The major gene at the *Rhg4* region was at the end of the Forrest BAC sequence, a serine hydroxymethyltransferase. It is a gene proven to be involved in the resistance of soybean to SCN (Liu et al., 2012; Lakhssassi et al., 2012; 2017; Liu et al., 2017). This study confirmed those earlier findings and showed the complexity of the introgressed region. Strong selection pressures on wild and semi-domesticated soybeans to develop resistance to SCN were inferred (Han et al 2016).

Authors' Contributions

DAL and OZ did the analysis, with OZ actively carrying out the bioinformtics (DAL generated the GeneMark.hmm file) and DAL guided the work. KM isolated the BAC and wasCo-PI on the grant that sequenced it. The text of this article was based on the MS thesis written by OZ that DAL helped edit.

Acknowledgements

We thank Dr. Khalid Meksem for creating the clone. He persisted, away from home, at TAMU, to pick the 100th plate of the second library. We thank Dr. Hogbin Zhang for hosting Khalid and teaching the BAC library creation methods. We thank Dr. Naghmeh Hemmati for her help with downloading the 'Williams 82' sequence and submitting 'Forrest' sequence to GenBank as HQ008938.1. The Sanger sequencing was paid for by the United Soybean Board. Dr. Christopher Town of JCVI is thanked for supervising the sequencing and assembly. Dr. Matthew JB Geisler is thanked for providing lists of newly discovered CRE sequences prior to his own publications.

Conflicts of Interest

None exist for any co-author.

References

Afzal AJ, A Srour, A Goil, S Vasudaven, T Liu, R Samudrala, N Dogra, P Kohli, A Malakar, and DA Lightfoot (2013) Dimerization, ligand binding and structural prediction validation with a leucine rich repeat domain associated with soybean resistance to cyst nematodes and Fusarium sp. BMC Plant Biology 13:43-53

Afzal AJ, A Srour, N Saini, N Hemmati, HA El Shemy, and DA Lightfoot (2012) Recombination suppression at the dominant Rhg1/Rfs2 locus underlying soybean resistance to the cyst nematode. Theor Appl Genet 124:1027-1039 doi10.1007/s00122-011-1766-6.

Brucker E, S Carlson, E Wright, T Niblack, and B Diers (2005) Rhg1 alleles from soybean PI 437654 and PI88788 respond differentially to isolates of Heterodera glycines in the greenhouse. Theor Appl Genet 111:44-49.

Center for Integrative Genomics (2017) Genetic Genome Browser, version 2.39. In: Phytozome. http://www.phytozome.net/cgi-bin/gbrowse/soybean/. Accessed 5 May 2015.

Grant D, RT Nelson, MA Graham, and RC Shoemaker (2008) Bioinformatic Resources for Soybean Genetics and Genomic Research. In: Stacey G, editor. Genetics and Genomics of Soybean. New York: Springer-Verlag LLC, New York, pp144-63.

Han Y, X Zhao, D Liu, Y Li, DA Lightfoot, Z Yang, L Zhao, G Zhou, Z Wang, L Huang, and Z Zhang (2016) Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytologist, 209 (2): 871-884

Hartwig EE and JM Epps (1973) Registration of 'Forrest' Soybeans. Crop. Sci. 13:287-287doi 10.2135/cropsci1973.001183X00130 00200047x.

Hauge BM, ML Wang, JD Parson, and LD Parnell (2009a) Methods of introgressing nucleic acid molecules associated with soybean cyst nematode resistance into soybean. United States patent US7485770-A 4.

Hauge BM, ML Wang, JD Parson, and LD Parnell (2009b) Sequence 4 from patent US 7485770. http://www.ncbi.nlm.nih.gov/nucleotide/223623398?report=genbank&log$=nucltop&blast_rank=1&RID=JGTUG796016. In: NCBI GenBank, accessed 22 June 2015.

Haun WJ, DL Hyten, WW Xu, DJ Gerhadt, TJ Albert, T Richmond, JA Jeddeloh, G Jia, NM Springer, CP Vance, and RM Stupar (2011) The compositions and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol 155: 645-655.

Hemmati NH, MJB Geisler and DA Lightfoot (2017) Annotation of Cultivar Variations at the Multigeneic Rhg1/Rfs2 Locus: Polymorphisms Underling Alterations of Root Development and Pest Resistance. Teor Appl Genet (in revision).

Johnson M, I Zaretskaya, Y Raytselis, Y Merezhuk, S McGinnis, and TL Madden (2008) NCBI BLAST: A better web interface. Nucleic Acids Res 36:W5-W9 doi 10.1093/nar/gkn201.

Kazi S, J Shultz, J Afzal, Y-C Lee, and DA Lightfoot (2017) Identification of Loci Underlying Seed Yield in Recombinant Inbred and Near Isogeneic Soybean Lines Derived from Flyer by Hartwig. Atlas Journal of Biology 2017, pp. 355-363. doi: 10.5147/ajb.2017.0149.

Kurle J, D Malvick, B Potter, and J Orf (2015) Soybean Cyst Nematode Management Guide. http://www.extension.umn.edu/agriculture/soybean/soybean-cyst-nematode/EFANS-Soybean-Soybean-CystNematode-WebQuality.pdf Accessed 13 October 2015.

Lakhssassi N, S Liu, S Bekal, Z Zhou, V Colantonio, K Lambert, A Barakat, and K Meksem (2017) Characterization of the Soluble NSF Attachment Protein gene family identifies two members involved in additive resistance to a plant pathogen. Scientific Reports 7, 45226.

Lakhssassi N, VG Doblas, A Rosado, AE Del Valle, D Pose, AJ Jimenez, AG Castillo, V Valpuesta, O Borsani, and MA Botella (2012) The Arabidopsis tetratricopeptide thioredoxin-like gene family is required for osmotic stress tolerance and male sporogenesis. Plant Physiol 158: 1252-1266.

Lewers K, R Heintz, H Beard, L Marek, and B Matthews (2002) A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg4 loci. Theor Appl Genet 104: 253-260.

Lewers KS, SD Nelmalgoda, AL Warner, HT Knap, and BF Matthews (2001) Physical mapping of resistant and susceptible soybean genomes near the soybean cyst nematode resistance gene Rhg4. Genome 44: 1057-1064.

Liu S, PK Kandoth, N Lakhssassi, J Kang, V Colantonio, R Heinz, G Yeckel, Z Zhou, S Bekal, J Dapprich, B Rotter, S Cianzio, MG Mitchum, and K Meksem (2017) The soybean GmSNAP18 gene underlies two types of resistance to soybean cyst nematode. Nature Communications, 8, 14822.

Liu S, PK Kandoth, N Lakhssassi, J Kang, V Colantonio, R Heinz, G Yeckel, Z Zhou, S Bekal, J Dapprich, B Rotter, S Cianzio, MG Mitchum, and K Meksem (2017) The soybean GmSNAP18 gene underlies two types of resistance to soybean cyst nematode. Nature Communications 8, 14822.

Liu S, PK Kandoth, SD Warren, G Yecke, R Heinz, J Alden, C Yang, A Jamai, T El-Mellouki, PS Juvale, J Hill, TJ Baum, S Cianzio, SA Whitham, D Korkin, MG Mitchum, and K Meksem (2012) A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature 492: 256–260.

Liu X (2009) Molecular characterization of soybean resistance to soybean cyst nematode https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/9861/research.pdf?sequence=3. Accessed 2012.

Liu X, S Lui, A Jamai, A Bendahmane, DA Lightfoot, MG Mitchum, and K Meksem (2011) Soybean cyst nematode resistance in soybean is independent of the Rhg4 locus LRR-RLK gene. Funct Integr Genomics doi 10.007/s10142-011-0225-4.

Matthews B (2004) Genomic Approaches for Developing Soybeans with Resistance to Pests (Soybean Cyst Nematode) in Stalker TE, Brummer C, and Wilson RF, eds., Legume Crop Genomics, AOCS Press. p. 312-25.

McGinnis S and TL Madden (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32: W20-W25 doi 10.1093/nar/gkh435.

Meksem K, K Zobrist, E Ruben, D Hyten, T Quanzhou, HB Zhang, and DA Lightfoot (2000) Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosomes walking and genome wide physical mapping. Theor Appl Genet 101: 747-755.

Shultz JL, D Jayaraman, KL Shopinski, MJ Iqbal, S Kazi, K Zobrist, R Bashir, S Yaegashi, N Lavu, AJ Afzal, CR Yesudas, MA Kassem, C Wu, HB Zhang, CD Town, K Meksem, and DA Lightfoot (2006) The soybean genome database (SoyGD): A browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max. Nucleic Acid Research 34: D758-765.

Smolik JD and MA Draper (2015) Soybean Cyst Nematode. http://pubstorage.sdstate.edu/AgBio_Publications/articles/FS902A.pdf. Revised 2007. Accessed 8 July 2015.

Srour A, AJ Afzal, N Saini, L Blahut-Beatty, N Hemmati, DH Simmonds, H El Shemy, CD Town, H Sharma, X Liu, W Li, and DA Lightfoot (2012) The receptor like kinase transgene from the Rhg1/Rfs2 locus caused pleiotropic resistances to soybean cyst nematode and sudden death syndrome. BMC Genomics 13:368-377

Tuteja JH and LO Vodkin (2008) Structural features of the endogenous CHS silencing and target loci in the soybean genome. Crop Sci 48: S49-S68 doi:10.2135/cropsci2007.10.0542tpg.

Weisemann JT, BF Matthews, and TE Devine (1992) Molecular markers located proximal to the soybean cyst nematode resistance gene, Rhg4. Theor Appl Genet 85: 136-138.

Wheeler DL, DM Church, S Federhen, AE Lash, TL Madden, JU Pontius, GD Schuler, LM Schriml, E Sequeira, TA Tatsunova, and L Wagner (2003) Database resources of the National Center for Biotechnology. Nucleic Acids Res 3: 28-33.

Wu CC, P Nimmakayala, FA Santos, R Springman, C Scheuring, K Meksem, DA Lightfoot, and HB Zhang (2004) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet 109: 1041-1050. doi: 10.1007/s00122-004-1712-y.

Wu X, TD Vuong, JL Leroy, JG Shannon, DA Sleper, and HT Nguyen (2011) Selection of a core set of RILs from Forrest × Williams 82 to develop a framework map in soybean. Theor Appl Genet 122: 1179-1187. doi 10.1007/s00122-010-1522-3.

Ye J, S McGinnis, and TL Madden (2006) BLAST: improvements for better sequence analysis. Nucleic Acids Res. 34: W6-W9. doi: 10.1093/nar/gkl164.

Yi J, MR Derynck, L Chen, and S Dhaubhadel (2010) Differential expression of CHS7 and CHS8 genes in soybean. Planta 231: 741-753 doi 10.108/s00425-009-1079-z.

Yu Q (2015) Soybean Cyst Nematode (Heterodera glycines Ichinohe) (2015) http://cdn.intechopen.com/pdfs/22782/InTech-Soybean_cyst_nematode_heterodera_glycines_ichinohe_.pdf. Accessed 1 October 2015.

Zatserklyana O, K Meksem, and DA Lightfoot (2015) Glycine max cultivar Forrest clone BAC 73P06 genomic sequence. In: NCBI GenBank. http://www.ncbi.nlm.nih.gov/nuccore/HQ008938.1. Accessed 10 September 2015.

Atlas Journal of Biology - ISSN 2158-9151. Published By Atlas Publishing, LP (www.atlas-publishing.org)