Atlas Journal of Biology 2017, pp. 407–416

Annotation of Cultivar Variations at the Multigeneic *Rhg1/Rfs2* Locus: Polymorphisms Underling Alterations of Root Development and Pest Resistance

Melody N. Hemmati[1], Matthew J. B. Geisler[2], Naresh Pola[3], Khalid Meksem[1], and David A. Lightfoot[1,2*]

[1] Department of Plant, Soil Science and Agricultural Systems, Southern Illinois University, Carbondale, IL 62901-4415, USA. Present Address; Fermentation Unit, Monsanto, Des Moines, Iowa, USA; [2] Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901, USA; [3] Department of Electrical and Computer Engineering, Southern Illinois University, Carbondale, IL 62901, USA.

Received: July 22, 2017 / Accepted: October 28, 2017

Abstract

Soybean (*Glycine max* (L.) Merr.) suffers yield loss due to root infection from soil infestation by *Heterodera glycine* I. (soybean cyst nematode SCN) and *Fusarium virguliforme* (Aoki; sudden death syndrome (SDS)). The major locus for SCN and SDS resistance has previously been identified as *Rhg1/Rfs2* (chr18; LG G) (site reference). The objective of this experiment was to compare the Sanger DNA sequence of a resistant cultivar ('Forrest') and two susceptible cultivars ('Williams 82' and 'Asgrow A3244'). Sequences were downloaded from GenBank for Williams 82, Phytzome for A3244 and a newly sequenced BAC-B73P06 (82,157 bp) encompassing the *Rfs2/Rhg1* locus. Using the resistant cultivars, 800 single nucleotide polymorphisms (SNPs) and 57 indels were identified. In contrast, the susceptible cultivars had just 12 SNPs and no indels between them. Polymorphisms were clustered within 59 kbp, divided into three sections. There were 5 predicted recombination breakpoints. The third and fourth breakpoints were located before gene 3 and after gene 5 (Glyma18g02680; the RLK at *Rhg1/Rfs2*) which were therefore inferred to be derived from Peking, within the *Rhg1/Rfs2* region. Comparisons of SNPs identified in Illumina sequences from 31 semi-domesticated genomes showed 80% of the total SNPs in Forrest were found among the genomes. Annotation and gene prediction showed the BAC gene prediction encoded 9-10 genes. There were 31 SNPs within exons and 137 among introns. Just 11 SNPs caused amino acid changes. There were 5 SNPs in cis regulatory elements (CREs) and 14 in promoters. Polymorphisms indicated the regions that were introgressed from Peking had defined limits. Proteins across the region were highly conserved compared to non-coding regions, suggesting purifying selection occurred.

Keywords: SNP; indel; introgression; recombination; gene annotation; BACs.

Abbreviations: Receptor like kinase (RLK); soybean cyst nematode (SCN); sudden death syndrome (SDS); bacterial artificial chromosome(BAC); MAFF Genebank System, National Institute of Agrobiological Sciences, Tsukuba, Japan; NRRL, The Agriculture Research Service Culture Collection, National Center for Agricultural Utilization Research, USDA/ARS, Peoria, IL USA.

* Corresponding author: ga4082@siu.edu.

Introduction

Soybean (*Glycine max* (L.) Merr.) cyst nematode (SCN; *Heterodera glycine* I.) and sudden death syndrome (SDS; caused by *Fusarium virguliforme* Aoki) are two of the major pathogens causing yield loss in soybean. The genetic basis of resistance has been studied using genome sequences (Hague et al., 2000; Schmutz et al., 2010) and genetic transformations (Cook et al., 2012; Srour et al., 2013; Liu et al., 2017).

Root infections caused by soil infestations by soybean cyst nematode (SCN) have been severe since the crop was first domesticated, consequently SCN has become the world's most widespread and damaging soybean pathogen (Wrather, 2001). Losses to SCN have been reduced by almost half since 1990, but it still remains a major problem in soybean (Lightfoot, 2015). SCN causes plant and root stunting and leaf chlorosis. Sudden death syndrome (SDS) has been shown to be a facultative hemibiotrophic fungus (Li, 2009; Roy, 1997) that causes yield loss in soybean. The amount of loss has doubled every decade in the US, since 1990 (Wrather, 2001; Lightfoot, 2015). SDS, as a syndrome, is disease complex of both a root rot and a leaf scorch.

Studies have shown that the response of the root system to the pathogen has been associated with light, temperature, soil moisture and genetics of both the host and the pathogen, (Arelli, 1994; Lighfoot, 2005) suggesting a complex genetic control. Partial resistance to SCN and SDS was significantly associated with the *Rhg1/Rfs2* region at a sub-telomeric region of the soybean chromosome 18, molecular linkage group G (Ruben et al. 2006; Srour et al. 2013). The locus has been shown to be responsible for resistance to all Hg Types (previously races Niblack et al. 2003) of SCN and about half of the total variation in resistance to SDS in resistant by susceptible crosses. It has previously been reported the resistance forms of the *Rhg1/Rfs2* region were associated with delayed seedling development, lower root mass and reduced seed yield (Afzal et al. 2012) .

'Forrest' has one of the 3 types of resistance encoded by the *Rhg1/Rfs2* locus (Hague et al. 2000; Ruben et al. 2005). It was one of several high yielding cultivars developed during the second cycle of intercross breeding and selection from 'Peking' by Edgar E. Hartwig at Mississippi Agricultural and Forestry Experiment Station in 1972, a product of a USDA breeding program (Hartwig and Epps, 1973; Lightfoot, 2008). It originated as an F5 line selected from the cross 'Dyer' × 'Bragg' and was highly resistant to SCN HgType 2.5.7, HgType 0 and HgType 7 (Niblack et al. 2004; previously one type of race 1 and the two kinds of race 3). The Forrest cultivar alone prevented crop losses of about $450 million from 1975-1980 as it was one of the first cultivars released with resistance to SCN (Lightfoot, 2008). BAC B73P06 has been shown to encompass most of the *Rhg1/Rfs2* locus(Ruben et al. 2005). This locus works along with the Rhg4 locus (Liu et al., 2012; Lakhssassi et al., 2017) encompassed by BAC B100B10 (Zatskayera et al., 2017).

'Williams 82' has one of the 5 types of susceptibility encoded by the *Rhg1/Rfs2* locus (Hague et al. 2000; Ruben et al. 2005). It originated as a composite of four lines resistant to phytophthora root rot (Bernard and Cremeens, 1988). It was selected from a 'Williams' 3 × 'Kingwa' BC6F3 (six backcross generations). Kingwa was used as the donor parent to introgress Phytophthora root rot resistance into the recurrent parent Williams. Williams 82 experienced one generation of single-seed descent following the six back-cross generations. Haun et al. (2012) showed heterogeneity had persisted, but not on chromosome 18, so not within the *Rhg1/Rfs2* region.

Asgrow soybean variety 'A3244' has a different one of the 5 types of susceptibility encoded by the *Rhg1/Rfs2* locus (Hague et al., 2000; Ruben et al., 2005). It is a non-transgenic conventional variety. It is known for its superior agronomic characteristics and high-yield. The *Rhg1/Rfs2* region was sequenced from a BAC contig (Hague et al., 2000).

The objectives of this study were: to compare the Sanger sequences from the region encompassed by BAC B73P06 from a domesticated cultivar resistant to SCN and SDS with two susceptible cultivars; and to compare those sequences to the single nucleotide polymorphisms (SNPs) found by IlluminaTM sequencing among the 31 semi-domesticated and domesticated genomes reported by Lam et al. (2010); then to use those sequence comparisons to infer which regions were introgressed into resistant domesticated cultivars from semi-domesticated cultivars; where the recombination breakpoints were; and which region, genes and polymorphisms were likely to be part of a multigenic *Rhg1/Rfs2* locus.

Materials and Methods

Cultivars Compared

To perform a comparative genomic study, Sanger based DNA sequence of three cultivars at the *Rhg1/Rfs2* region were selected. Forrest was selected to represent the cultivar resistant to SCN and SDS. Twenty plants were used to make the BAC library (Meksem et al., 2000). One plant provided a molecule that was cloned in the BAC B73P6. BAC library development, physical mapping and screening candidate genes were described previously (Triwitayakorn et al., 2005; Ruben et al., 2006; Srour et al., 2012).

A3244 and Williams 82 were chosen to represent the susceptible cultivars based on sequences available at GenBank and Phytozome prior to 2009. The sequences were downloaded. The Williams 82 sequence was derived from DNA isolated from multiple plants in pUC18 (Schmutz et al., 2010). Predicted gene models had been made where there was transcript data to support the models. The A3244 sequence was derived from several overlapped BACs each derived from a single region but likely from separate plants (Hauge et al., 2006). Gene models had been predicted previously (Triwitayakorn et al., 2005).

BAC Sequencing Methods

The resistance alleles that might be derived from the *Rhg1/Rfs2* region from Forrest were analyzed by sequencing the entire BAC B73P06, which was derived from one region (parental alleles) of a single plant. The sequencing accuracy was high due to an 8 fold redundancy. The sequence of the insert (82,157 bp) was submitted to GenBank as HQ008938 (Srour et

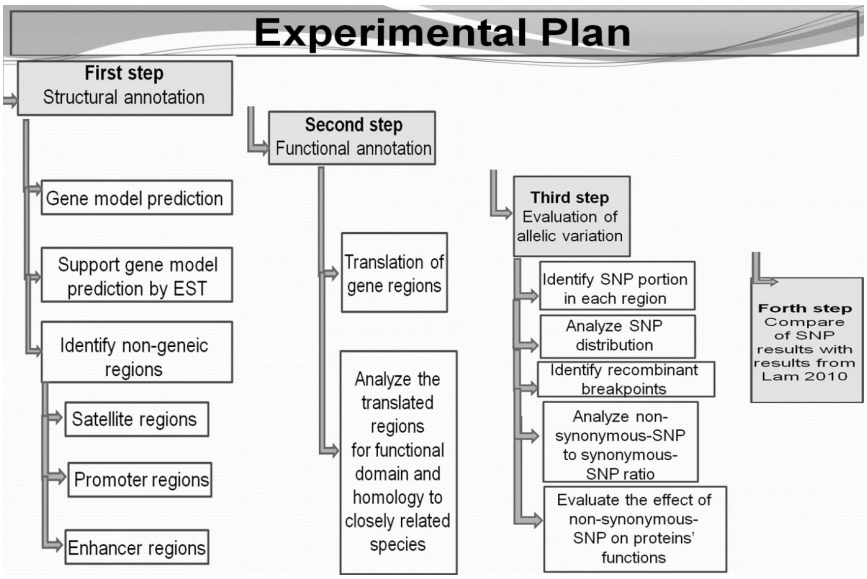

Figure 1. Flow chart of the experimental plan.

al. 2012). The sequencing method was the dideoxy chain-termination method using an ABI big dye cycle sequencing kitTM, on ABI3730 automated DNA sequencers at the J. Craig Venter Institute (JCVI; Ruben et al. 2006; Afzal et al. 2012; Srour et al. 2012).

Annotations

The annotation methods are summarized in Figure 1. There were four areas: structural annotation, functional annotation, searches for sequence variations and comparison of SNPs (Supplemental Table 1).

Structural Annotations

For structural annotation of genes, Eukaryotic GeneMark (Borodovsky et al. 2005) and a semi computational annotation, was applied to the Forrest sequence to predict gene models de novo. The accuracy of eukaryotic GeneMark was checked versus other gene predictor software: Augustus, Genscan, Geneid, Fgenesh-M, FGenesH and SNAP at DNA Subway (http://dnasubway.iplantcollaborative.org). Pairwise alignment based methods were applied to the Forrest BAC B73P06 sequence versus all available ESTs (Express Sequence Tags) to support the predicted genes. MASTER, an ExcelTM-based genomic SNP and indel comparative tool (Hemmati et al., 2014) was utilized for analysis of polymorphisms in and around the predicted genes.

Signals and Cis Regulatory Elements

The 112 most highly conserved cis-regulatory elements in plants (Dr. Matt Geisler, unpublished) were sought within the BAC sequence. These cis-regulatory elements were found in common between Arabidopsis and rice near (<500bp) promoter and enhancer regions. They were compared with the patterns of transcript abundance across many microarray experiments and were found in the 500 bp upstream region (not the 5'UTR) of Arabidopsis thaliana. Each of these 112 conserved cis-regulatory elements (enhancer like motifs) were 8 bp, they were all searched and mapped within the Forrest sequence by short sequence search software BLAT and PATMAT. The best result was prepared by using the alignment analysis tool "matcher" (Matcher, 2011).

Repetitive Elements and Genetic Markers

Microsatellites (simple sequence repeats; SSRs) were marked and mapped in the MASTER file for further evaluations. Dozens have previously been identified in earlier studies (Srour et al. 2012); new markers were also identified and listed.

Promoter Prediction

TSSP / Prediction of PLANT Promoters, RegSite Plant DB, Softberry Inc, (PLANT-Promoters, 2011) and putative eukaryotic Pol II promoter sequences (Pol-II-promoter, 2011) were used to predict the promoter regions which best matched with the predicted gene models. The best matches were mapped and recorded in MASTER and in the data table submitted to GenBank.

Functional Annotations

The putative genes were translated to amino acids. The translated regions were analyzed for functional domains and prediction of function by homology searches to closely related species.

Gene Ontology, Protein Searches, Domain Searches

Each predicted protein was analyzed by BLAST in NBCI-BLASTp and WU-BLAST. Pfam was also used for confirmation of possible domains and their functions. Information about

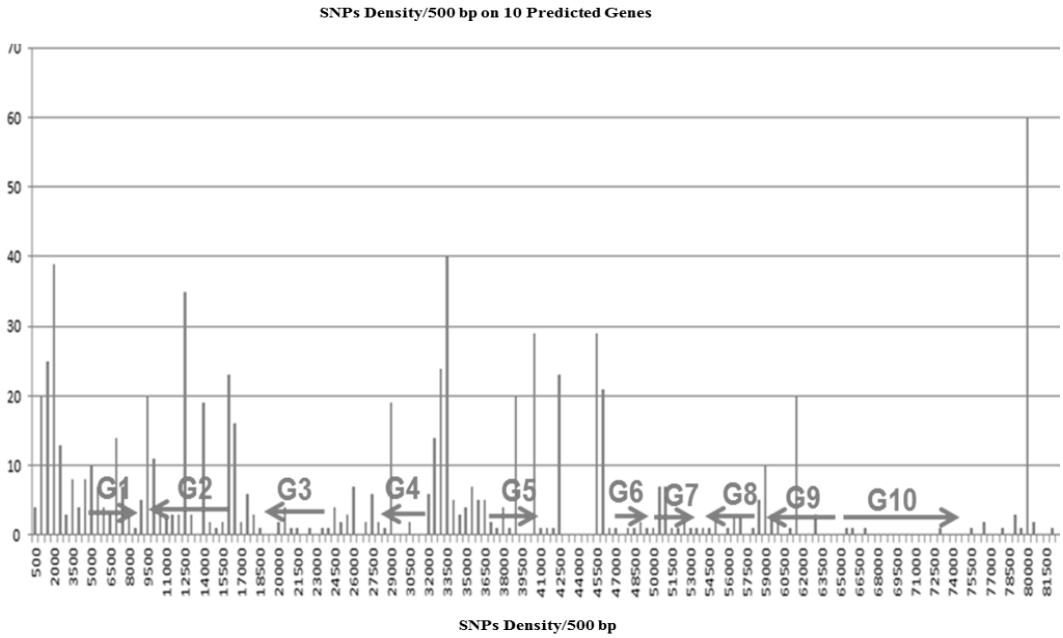

Figure 2. SNP density per 500 bp among the predicted genes and their intergenic regions. The X axis indicates B73P06 sequence (1-82,157 bp). SNP number was binned for every consecutive 500 bp intervals. The Y axis shows density of SNPs per 500 bp region. Black arrows are indications of the approximated positions of the predicted gene coding regions (total of 10 genes).

each possible protein, domain and its functions was recorded in the data table submitted to GenBank.

Homology Based Predictions of Function

NCBI-BLASTp search between Forrest versus the viridiplantae resulted in top hits with genes from *Ricinus communis* L., *Populus trichocarpa* L., *Medicago truncatula* L., and *Fragaria vesca* L. Sequence similarities were recorded in the data table submitted to GenBank.

Searches for Sequence Variations

All variations including single nucleotide polymorphisms (SNPs) or insertions and deletions, either geneic (within gene coding region) or non-geneic (within non-coding regions) were recorded in the data table submitted to GenBank for Forrest, A3244 and Williams 82 sequences. The SNP frequency was calculated per 500 bp region (Figure 2). SNP density was used to infer the distribution of recombination breakpoints within the Forrest sequence by using RDP4 (Recombination Detection Program version 4), a Windows 95/XP program (Martin et al. 2005). Also, Protein Variation Effect Analyzer (PROVEAN) software (Choi et al. 2012) was applied to predict if a protein sequence variation had an impact on protein function.

Comparison of SNP Results with 32 Genomes of Re-sequenced-SNPs

Using the SNPI-Tool MASTER the SNP motifs of 31 genomes (Lam et al., 2010) could be added by reference to the bp position of Williams 82. Where misalignments had occurred they were clear from the SNP motifs disagreements and could be logically nudged to the nearest, most likely position. Once aligned the SNPs were de-convoluted to single columns, each corresponding to a single genotype (Supplemental Table 2).

Data Management

Feature Table

The annotation was deposited in a feature-table-format according to the instructions provided by GenBank and published in GenBank database (Benson et al. 2013).

Visualization

The MASTER file contained all the predicted genomic features. It was a useful tool to actually observe and simultaneously perform data mining on any part of sequences of Forrest, A3244 and Williams 82.

Results

BAC B73P6 Composition

The BAC B73P6 insert was shown to encompass 82,157 bp and predicted to encode 9 genes. These genes comprised of 57 exons (2–8 per gene). Sequence composition (Figure 3) showed that 52%, was non-regulatory-non-coding regions. Intron regions comprised 24.5 %. Exon regions occupied 21.5% of the sequence. Enhancer, promoter and satellite regions together accounted for 2.3% of the sequence. The sequence composition

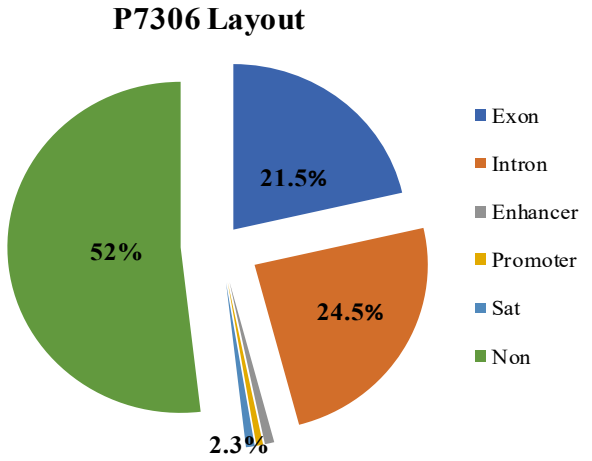

Figure 3. Compositions of the BAC 73P06 DNA encoded regions. Each portion of the pie chart indicates the percentages of each type of region encoded (exon, intron, enhancer, promoter, satellite DNA (Sat), non-regulatory-non-coding regions (Non)). All the genes are potentially involved in resistance.

was similar to earlier reports for A3244 (Hague et al. 2000) and Williams 82 (Schmutz et al. 2010).

Polymorphism Frequencies

However, sequence comparisons showed that across the entire BAC there were exactly 800 SNPs between the resistance region in Forrest and the sequences of both regions associated with SDS and SCN susceptibility. The number of SNPs found was as large as the number found among the 31 re-sequenced genomes in the same region (Lam et al. 2010). A few SNPs were found in the promoter and enhancer regions of all 9 genes (5 in potential cis-regulatory elements and 14 in core promoter regions). However there were only 31 SNPs within genes and only 11 of the 31 that caused amino acid changes. There were amino acid changes in just 6 of the 9 proteins caused by those 11 SNPs. Therefore, all of the protein coding sequences were inferred to be subject to purifying selection.

There was evidence for a large and highly polymorphic region within the BAC, 743 SNPs and 54 indels in 59 Kbp (from 1,500-60,500 bp). A highly polymorphic region was expected to be a characteristic of the region introgressed into Forrest from Peking. Equally, relatively monomorphic regions were common when comparing sequences of US cultivars. In fact, the same two regions of Williams 82 and A3244 were nearly identical with just 12 SNPs.

On the basis of high frequencies of polymorphism and transgenic plants, gene 5, Glyma18g02680, the receptor like kinase (RLK) (Srour et al. 2012; Afzal et al. 2013) was shown to be part of the Rfs2 allele and influence the Rhg1-a region derived from Peking. Here, on the basis of high rates of polymorphisms, 4 genes were inferred to be derived from Peking: gene 1 Glyma18g 2650, gene 2 Glyma18g 2660, gene 3 Glyma18g 2670, and gene 10 Glyma18g 2720. The coding regions were supported by 101 ESTs outside the 59 kbp central region. For example the region that encompassed gene 10 had 5 SNPs, across 18.25 kbp, among the 3 alleles. However, SNPs did not alter protein sequence, so there were no alloproteins.

Functional SNPs

Among the 800 SNPs between Forrest and the sequences of susceptible cultivars, there were only 31 SNPs within geneic regions. Exactly 20 of those 31 SNPs (64.5%) did not change amino acid sequence (synonymous SNPs). Only 11 SNPs (35.5%) caused amino acid changes (non-synonymous SNPs). Only one of these non-synonymous SNPs (C6615T / A152V), located at gene 1, arogenate dehydrogenase, was predicted to have a deleterious effect on protein function in susceptible cultivars. Interestingly, there were also 2 SNPs in the promoter region of gene 1 (Figure 4).

Recombination Break Point Prediction

There was a large and highly polymorphic region within the BAC, 743 SNPs in 59 Kbp, from 1.5–60.5 K bp. Figure 5, depicts all 9-10 genes that may be inferred from sequence and transcripts, including the genes located in *Rhg1/Rfs2* region. The 5 recombinant breakpoints predicted within *Rhg1/Rfs2* are shown. The only deleterious SNP located at gene 1 is marked as well. Remarkably, comparing the result of 1,068 SNPs found among 17 semi-domesticated and 14 domesticated genomes (Lam et. al. 2010), showed that 79% of those identified by the Forrest to Williams 82, comparison were found. Some SNPs were unique to Forrest and others to Williams 82, depending on the sequencing method used. Overall 63% of the 1,063 SNP were identified. The similarity between the SNPs in the current research and SNPs from Lam et al. (2010; Figure ; Supplemental Table 2) and polymorphism frequency in this region (Figure ; Supplemental Table 1) supports the breakpoints for the regions identified in Peking.

Discussion

This research study was able to shed more light on the complex *Rhg1/Rfs2* region by analysis of structural and functional annotation. The existence of predicted recombinant breakpoints in this region and the similarity between the SNPs in the current research and SNPs from 31 genomes indicated the regions are likely derived from Peking. It appears most of the SNPs in this region were introgressed into Forrest from the semi-domesticated Peking. In addition to the RLK proven involved by Srour et al. (2012) and the copy number variation in SNAP (Cook et al. 2013; 2014; Liu et al.2017) there are several other changes in the defense related gene alleles (Matsye et al. 2012; Matthews et al. 2013) and many may have been introgressed from Peking.

Possible functions and roles in defense mechanisms were found for all 9 predicted genes. Gene 1 (Glyma18g02650) encoded an arogenate dehydrogenase-like protein (AgDH; EC 1.3.1.43, NCBI Reference Sequence: XP_003552195.1) which is one of the enzymes in the shikimate pathway (Wink, 1999; Rippert and Matringe, 2002). The RNA Atlas at Soybase

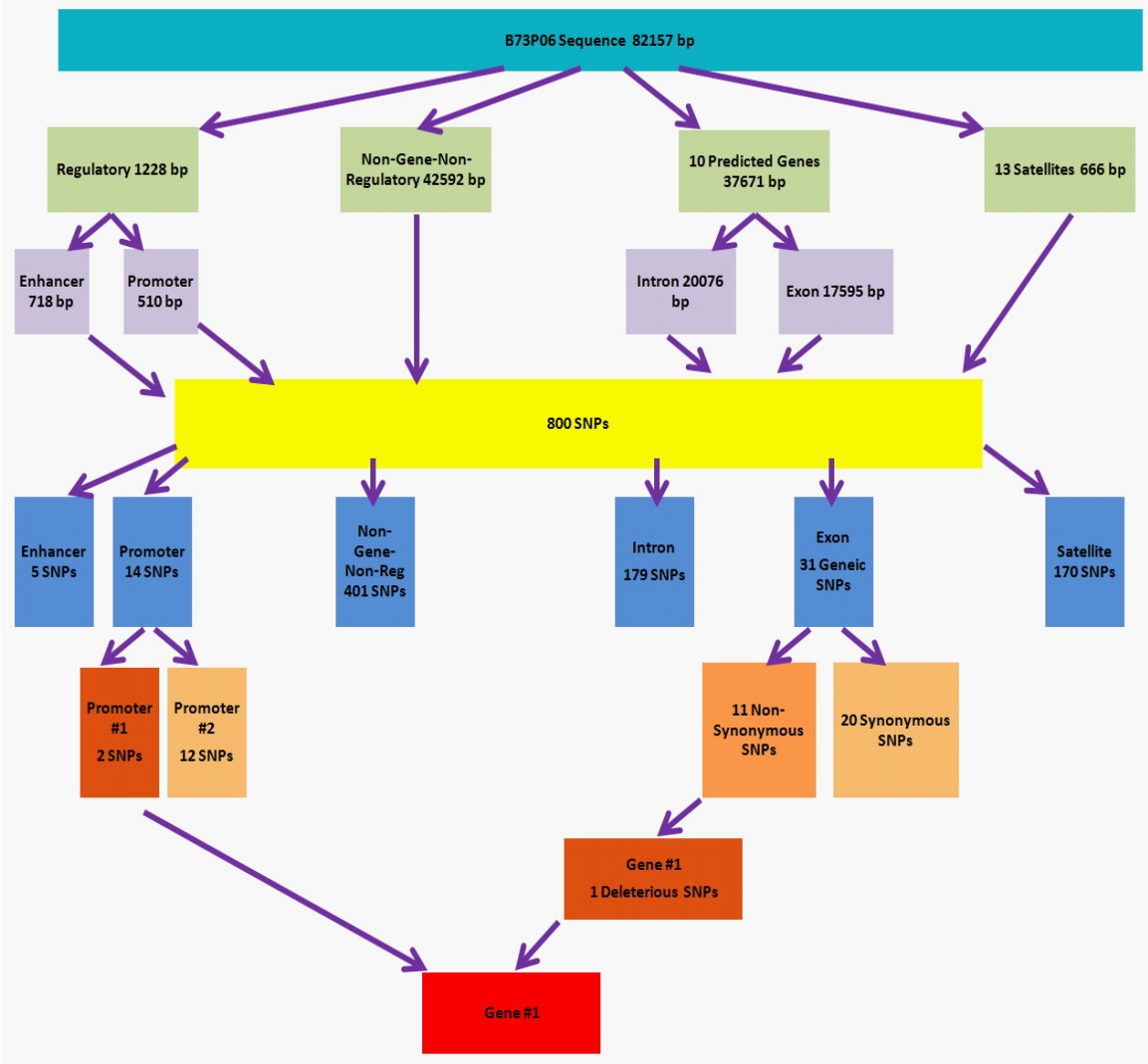

Figure 4. SNP detection flow chart.

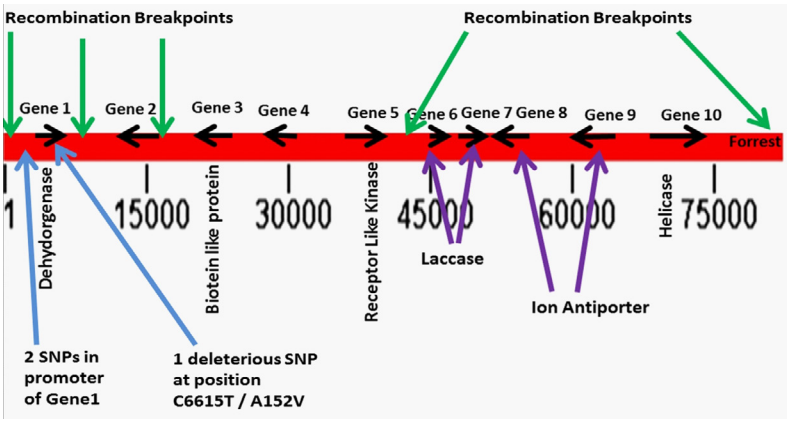

Figure 5. Ideogram of the structure of the regions predicted to be introgressed to Forrest from Peking. The black arrows indicate the positions of the 9-10 genes located in the BAC B73P6 *Rhg1/Rfs2* region. Putative gene functions are indicated. Recombinant breakpoints are shown as green arrows. A predicted deleterious SNP and 2 SNPs in the promoter of gene 1 are shown as blue arrows.

showed high transcript abundances in roots, flowers, pods and leaves (Supplemental Table 3). The products of shikimate pathway are precursors for many of aromatic and phenolic secondary metabolites. These metabolites are involved in numerous vital processes such as plant defense, formation of structural biopolymers and cell wall components including lignin, lignols, defense related phenolics and the synthesis of hormones and vitamins. Phenolic compounds are involved in PCD (Program Cell Death) in response to SCN and F. virguliforme infections (Mahalingam and Skorupska, 1996; Yuan et al. 2002; Iqbal et al. 2005; Iqbal et al. 2008; Kandoth et al. 2011).

Gene 2 (Glyma18g02660; NCBI CG_3016.1) encoded two putative conserved domains; the DUF 3411 superfamily (pfam 11891) and the DUF 399 (pfam 04187) superfamily both of which were domains of unknown functions. However, the transcript is abundant in flowers, nodules, roots, pods and leaves (Supplemental Table 3).

Gene 3 was a biotin-carboxyl-carrier-protein (Glyma18g02670; NCBI NP_567035.1). Biotin was particularly important in the SCN pathogenesis process (Liu et al. 2012). Biotin is a critical coenzyme which is needed for cell growth, the production of folate, the production of fatty acids, and the metabolism of fats and amino acids. Pathogens have a high demand for the macro- and micro-nutrients for their survival and proliferation during infection. Higher expression of proteins required for de novo biotin synthesis, uptake and metabolic adaption may be processes that allow pathogens to survive during infection. Biotin is required for both folate uptake and metabolism (Pendini et al. 2013) and folate starvation appears to be a key defense strategy of SCN resistant soybeans (Liu et al. 2012).

Gene 4 was unusual. It has no transcript support and did not present any similariies to known functional domains. However, a PSI-BLAST search through 8 iterations found 72 genes in the ortholog family, which inferred that the gene might be involved in cell and protein interactions. In addition, 20 of the 25 SNPs in that gene were in the C-terminal region, an ankyrin Pfam. Therefore, gene 4 might also be part of the *Rhg1/Rfs2* locus and the Rhg1-a allele derived from Peking.

Gene 5 (Glyma18g02680, NCBI: ACI05083.1, Gm-RLK18-1, Gene ID, symbol LOC547641) was a receptor-like-kinase shown to confer partial resistance to SCN and SDS in stable transgenic plants (Srour et al. 2012). The protein was shown to bind proteins and peptides in the nematode and plant secretomes (Afzal et al. 2013). Plants lack extra cellular antibodies for defense against pathogen attacks, but, they are able to establish other strategies for detection and adaptation to environmental changes. These strategies include a wide range of receptors both at cell surfaces and within the cells (Gomez-Gomez and Boller, 2000; Thordal-Christensen, 2003; Afzal, 2007; Haffani et al. 2004). The leucine-rich repeat extracellular domains of these receptors function as detectors and provide an early warning signal for the presence of potential pathogens. They commonly activate protective signaling cascades in plants and/or alter plant development (Matsushima and Miyashita, 2012). They often lead to activation of various host defense responses, including a specialized type of (PCD) programmed cell death known as the hypersensitive response (HR; Tao et al. 2000).

Predicted genes 6 and 7 (Glyma18g02690) were both parts of the diphenol oxidase laccase enzyme transcript (Iqbal et al. 2008). This enzyme (EC.1.10.3.2) is a blue copper-containing oxidase found in plants, fungi, bacteria and arthropods. Diphenol oxidase laccase enzymes catalyze the oxidation of wide variety of organic and inorganic substrates such as polyphenolic compounds. Phenolic compounds are synthesized from remote precursors as a response to pathogen attacks. Some antibiotic phenolics are stored in plant cells as inactive bound forms (preformed antibiotics) and are convertible into biologically active antibiotics by plant hydrolyzing enzymes in response to pathogen attacks. The normal anticipated amount of preformed antifungal phenolics in healthy plants may increase as a response to pathogens attack (Lattanzio et al. 2006). Moreover, diphenol oxidase laccase may participate in cell wall lignifications, by which lignifications of plant cell wall may contribute to reduction the frequency of SCN feeding site development (Lattanzio et al. 2006; Iqbal et al. 2008).

Genes 8 and 9 were ion antiporters (Glyma18g02700, Glyma18g02710). Na/H antiporters are key transporters in maintaining the homeostasis of actively metabolizing cells. The activity of Na+/H+ antiporters, Na+/H+ concentration and cell volume is critical for the viability of all cells. Typically, Na+/H+ antiporters are located at the plasma membrane (Uniport, 2012). Action of Na+/H+ antiporters is necessary for the uptake of most metabolites, plant growth and development and also for plant response to environmental stresses (Hunte et al. 2005).

Gene 10 was predicted to be a helicase (Glyma18g02620; DEXDc ,cd00046; HELICc, cd00079). Helicases are categorized as enzymes that use energy derived from the hydrolysis of a nucleotide triphosphate to unwind double-stranded structures. Increasing evidences suggest that the DEAD-box helicases play an important role in plant growth and development processes, possibly by regulation of RNA metabolism and gene expression (Wang et al. 2000; Li et al. 2008).

On the basis of copy number variation and transgenic hairy roots, 3 genes just distal to B73P06 were predicted to be part of the Rhg1-b region (Cook et al. 2012; Meksem et al. 2014; Liu et al. 2017). The genes Glyma18g 2570 and Glyma18g 2590 were shown to be present at 1 copy in susceptible cultivars but at 10 copies in cultivars with the rhg1-b allele. At the rhg1-a allele there appeared to be 3-5 copies of the 3 genes. There were again many SNPs, insertions and deletions that changed amino acid sequences of the alloproteins from Forrest compared to Williams 82 and Essex (Meksem et al. 2014). Using both virus induced gene silencing and hairy root assays these changes were shown to underlie part of the resistance to SCN. The alpha-SNAP gene had been identified to alter SCN resistance in hairy roots among dozens of other genes from whole genome screens (Matsye et al. 2012; Matthews et al. 2013; Meksem et al. 2014).

Finally, some studies found an *Rhg1/Rfs2*-like region at other locations in a few SCN resistant PIs (Concibido et al. 2004). Locations included linkage group (LG) B1, mid Lg G and Lg B2 (chromosomes 11, 18 and 14; Vierling et al. 1996; Wang et

al. 2001; Yue et al. 2001). The findings suggest the existence of functional paralogs of *Rhg1/Rfs2* among the duplicated regions of the soybean genome. Therefore, the *Rhg1/Rfs2* region is predicted to be multigeneic, each gene having paralogs, but to encompass several genes and polymorphisms in the region from Glyma18g 2570- Glyma18g 2720 in Forrest.

Data Depositions

TMD1 marker, FJ520231; Corrected RLK at *Rhg1/Rfs2* gene AF506516 and mRNA AF506517; SIUC-Satt122, bankit1155667; BAC pB73P06 complete sequence HQ008938 (corrected JN597009). Three supplementary tables will be published online and links to 5 sequences at Genbank are provided.

Conflicts of Interest

None

Statement of Contributions

Christopher D. Town of JCVI provided the BAC sequences from a BAC library made by Khalid Meksem. Naghmeh Hemmati, and David. A. Lightfoot analyzed BAC sequences and generated Figures and Supplemental Tables, wrote the manuscript and edited the final version. Naresh Pola extracted the SNPs for the 32 genomes. Matthew Geisler provided the list of CREs ahead of publication.

Acknowledgements

The physical map location of B73P06 was supported by the NSF under Grants No. 9872635. USB supported the sequencing of the BAC. Thanks are due to Christopher D. Town of JCVI for sequencing and Eliza Ruben for BAC isolation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

Afzal AJ and DA Lightfoot (2007) Inclusion bodies contain RHG1/RFS2 folding intermediates: A novel refolding protocol for protein purification. Protein Expr Purific 53: 346–355.

Afzal AJ, Srour A, Saini N, Hemmati N, El Shemy HA, Lightfoot DA (2011) Recombination suppression at the dominant Rhg1/Rfs2 locus underlying soybean resistance to the cyst nematode. Theor Appl Genet 124: 1027-1039.

Afzal AJ, A Srour, A Natarajan, N Saini, MJ Iqbal, M Geisler, HA El Shemy, and DA Lightfoot (2012) Resistance to soybean cyst nematode: Rhg1. J Plant Genom Sci 2012 39: 39-45.

Afzal AJ, AJ Wood, and DA Lightfoot (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Molec Plant Microb Int 21: 507-517.

Arabidopsis.org. (2013). http://www.arabidopsis.org/creon.

Arelli P (1994) Inheritance of resistance to Heterodera glycines race 3 in soybean accessions. Plant Dis 78: 898-900.

Benson DA, M Cavanaugh, K Clark, I Karsch-Mizrachi, DJ Lipman, J Ostell, and EW Sayers (2013) GenBank. Nucl Acids Res 41: 36-42.

Bernard RL and CR Cremeens (1988) Registration of Williams 82 soybean. Crop Sci 28: 1027–1028.

Borodovsky M, A Lukashin, A Lomsadze, V Ter Hovhannisyan, Y Chernoff, and M Borodovsky (2005). Gene identification in novel eukaryotic genomes by self-training algorithms. Nucl Acids Res 33: 6494-6506.

Choi Y, GE Sims, S Murphy, JR Miller, and AP Chan (2012) Predicting the functional effects of amino acid substitutions and indels. PLoS ONE 7(10): e46688. doi:10.1371/journal.pone.0046688.

Concibido VC , BW Diers, and PR Arelli (2004) A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci 44:1121-1131.

Cook DE, TG Lee, X Guo, A Melito, K Wang, AM Bayless, J Wang, TJ Hughes, DK Willis, TE Clemente, BW Diers, J Jiang, ME Hudson, and AF Bent (2012) Copy number variation of multiple genes at rhg1 mediates nematode resistance in soybean. Science 338: 1206-1209.

Cook DE, AM Bayless, K Wang, X Guo, Q Song, J Jiang, and AF Bent (2014) Distinct copy number, coding sequence, and locus methylation patterns underlie rhg1 mediated soybean resistance to soybean cyst nematode Plant Physiol. 165(2): 630–647.

Dixon R and LW Sumner (2003) Legume natural products: Understanding and manipulating complex pathways for human and animal health. Plant Physiol 91: 878–885.

DNA Subway http://dnasubway.iplantcollaborative.org/about/resources.html.

Flagel L, J Wendel, and J Udall (2012) Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC Genom 13: 302.

Haffani YZ, NF Silva, and DR Goring (2004) Receptor kinase signalling in plants. Can J Bot 82: 1–15.

Hartwig EE and JM Epps (1973) Registration of Forrest soybeans. Crop Sci. 13: 287.

Haun WJ, DL Hyten, WW Xu, DJ Gerhardt, TJ Albert, T Richmond, JA Jeddeloh, G Jia, NM Springer, CP Vance, and RM Stupar (2012) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol 5 (2): 645–55. doi: 10.1104/pp.110.166736. Epub 2010 Nov 29.

Hauge BM, ML Wang, JD Parsons, and LD Parnell (2006) Methods of introgressing nucleic acid molecules associated with soybean cyst nematode resistance into soybean. US Patent, 7154210.

Hemmati MN, LA Sholeh, and DA Lightfoot (2013) SNPI-Tool: A user friendly tool for sensitive and accurate SNP identification. ATLAS Biology 3: 206-211.

Hemmati N and DA Lightfoot (2011). Glycine max cultivar Forrest clone BAC 73P06 genomic sequence. GenBank: HQ008938.1. GenBank, HQ008938.1., http://www.ncbi.nlm.nih.gov/nuccore/JN597009.597001.

Huang X, Q Feng, Q Qian, Q Zhao, L Wang, A Wang, J Guan, D Fan, Q Weng, T Huang, G Dong, T Sang, and B Han (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19 (6): 1068–1076.

Hunte C, E Screpanti, M Venturi, A Rimon, E Padan, and H Michel (2005) Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. Nature 435: 1197-1202

Iqbal MJ and DA Lightfoot (2004) Application of DNA Markers: Soybean Improvement. Biotechnology in Agriculture and Forestry Edited by Lörz, Horst; Wenzel, Gerhard, Springer, New York, USA. pp. 475.

Iqbal MJ, S Yaegashi, R Ahsan, KL Shopinski, and DA Lightfoot (2005) Root response to Fusarium solani f. sp. glycines: Tempo-

ral accumulation of transcripts in partially resistant and susceptible soybean. Theor Appl Genet 110: 1429-1438.

Iqbal MJ, R Ahsan, AJ Afzal, A Jamai, K Meksem, HA El-Shemy, and DA Lightfoot (2008) Multigeneic QTL: the laccase encoded within the soybean Rfs2/rhg1 locus inferred to underlie part of the dual resistance to cyst nematode and sudden death syndrome. Curr Iss Molec Biol 1: 11-19.

Jinn TL, JM Stone, and JC Walker (2000) HAESA, an Arabidopsis leucine-rich repeat receptor kinase, controls floral organ abscission. Genes Dev 14: 108-117.

Kandoth PK, N Ithal, J Recknor, T Maier, D Nettleton, TJ Baum, and MG Mitchum (2011) The soybean Rhg1 locus for resistance to the soybean cyst nematode Heterodera glycines regulates the expression of a large number of stress- and defense-related genes in degenerating feeding cells1. Plant Physiol 155: 1960–1975.

Kawabe A, H Innan, R Terauchi, NT Miyashita (1997) Nucleotide polymorphism in the acid chitinase locus (chiA) region of the wild plant Arabiodopsis thaliana. Molec Biol Evol 14: 1303-1315.

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.

Lam HM, X Xu, X Liu, W Chen, G Yang, FL Wong, MW Li, W He, N Qin, B Wang, J Li, M Jian, J Wang, G Shao, J Wang, SSM Sun, and G Zhang (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nature Genet 42: 1053-1070.

Lattanzio V, VMT Lattanzio, and A Cardinali (2006) Role of phenolics in the resistance mechanisms of plants against fungal pathogens and insects: In Phytochemistry: advances in research, Ed F. Imperato, Research Signpost, Kerrala, India pp. 23-67.

Li S, GL Hartman, and Y Chen (2009) Evaluation of aggressiveness of Fusarium virguliforme isolates that cause soybean sudden death syndrome. J Plant Pathol 91: 77-86.

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 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 Meksem, K (2017) The soybean GmSNAP18 gene underlies two types of resistance to soybean cyst nematode. Nature Communications 8: 14822.

Lightfoot DA, VN Njiti, PT Gibson, MA Kassem, MJ Iqbal, and K Meksem (2005) Registration of the Essex by Forrest recombinant inbred line mapping population. Crop Sci 45: 1678-1681.

Lightfoot DA (2015) Two decades of molecular marker-assisted breeding for resistance to soybean sudden death syndrome. Crop Sci 55: 1460–1484 (2015). doi: 10.2135/cropsci2014.10.0721.

Mahalingam R and HT Skorupska (1996) Cytological expression of early response to infection by Heterodera glycines I. in resistant PI 437654 soybean. Genome 39: 986–998.

Martin DP, C Williamson, and D Posada (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21: 260-262.

Matcher (2011). alignment analysis tool: http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::matcher.

Matsye PD, GW Lawrence, RM Youssef, KH Kim, KS Lawrence, BF Matthews, VP Klink (2012) The expression of a naturally occurring, truncated allele of an α-SNAP gene suppresses plant parasitic nematode infection. Plant Mol Biol 80: 131–155.

Matthews BF, H Beard, MH MacDonald, S Kabir, RM Youssef, P Hosseini, and Brewer E (2013) Engineered resistance and hypersusceptibility through functional metabolic studies of 100 genes in soybean to its major pathogen, the soybean cyst nematode. Planta 14 (5): 1337–1357.

Meksem K, S Liu, and DA Lightfoot (2014) Soybean Resistant To Cyst Nematodes. Patent pending. U.S. Provisional Application No. 61/799, 912 filed 15 March 2013, Final Filing Date 02-24-2014.

Nagata T, H Lorz, JM Widholm, and G Wenzel (2008) Molecular Marker Systems in Plant Breeding and Crop Improvement: Springer.

NCBI-Genbank (2012) feature table. National Center for Biotechnology Information, http://www.ebi.ac.uk/embl/Documentation/FT_definitions/feature_table.html.

NCBI. (2013) BLAST. http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastnBLAST_PROGRAMS=megaBlastPAGE_TYPE=BlastSearch.

Niblack TL, GR Noel, KL Lambert (2003) The Illinois SCN type test: Practical application of the HG Type classification system. J Nematol 35: 355-345.

Pendini NR, MY Yap, SW Polyak, NP Cowieson, A Abell, GW Booker, JC Wallace, JA Wilce, and MC Wilce (2013) Structural characterization of Staphylococcus aureus biotin protein ligase and interaction partners: an antibiotic target. Protein Sci 22: 762-773.

Phytozome (2013) Soybean http://www.phytozome.net/soybean.php.

PLANT-Promoters (2011) TSSP / Prediction of PLANT Promoters, Using RegSite Plant DB, Softberry Inc. http://linux1.softberry.com/berry.phtml?topic=tsspgroup=programssubgroup=promoter

Pol-II-promoter (2011) Putative eukaryotic Pol II promoter sequences. http://www-bimas.cit.nih.gov/molbio/proscan/.

Purugganan MD and JI Suddith (1998) Molecular population genetics of the Arabidopsis CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral homeotic function. Proc Nat Acad Sci USA 95: 8130-8134.

Rippert P and M Matringe (2002) Molecular and biochemical characterization of an Arabidopsis thaliana arogenate dehydrogenase with two highly similar and active protein domains. Plant Molec Biol 48: 361-368.

Roy KW, JCD Rupe, DE Hershman, and TS Abney (1997) Sudden death syndrome of soybean. Plant Dis 81: 1100-1111.

Ruben E, A Jamai, J Afzal, VN Njiti, K Triwitayakorn, MJ Iqbal, S Yaegashi, R Bashir, S Kazi, PR Arelli, CD Town, H Ishihara, K Meksem, and DA Lightfoot (2006) Genomic analysis of the Rhg1/Rfs2 locus: candidate genes that underlie soybean resistance to the cyst nematode. Molec Genet Genom 276: 503-516.

Schmutz J, SB Cannon, J Schlueter, J Ma, T Mitros, W Nelson, DL Hyten, Q Song, J Thelen, J Cheng, D Xu, U Hellsten, May, G D, Yu, Y, Sakurai, T, Umezawa, T, Bhattacharyya, M K, Sandhu, D, Valliyodan, B, Lindquist, E, Peto, M, Grant, D, Shu, S, Goodstein, D, Barry, K, Futrell-Griggs, M, Abernathy, B, Du, J, Tian, Z, Zhu, L, Gill, N, Joshi, T, Libault, M, Sethuraman, A, Zhang, X, Shinozaki, K, Nguyen, H T, Wing, R A, Cregan, P, Specht, J, Grimwood, J, Rokhsar, D, Stacey, G, Shoemaker, R C, Jackson, S A (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178-183.

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 Genom 13: 368-377.

Triwitayakorn K, VN Njiti, MJ Iqbal, S Yaegashi, CD Town, DA

Lightfoot (2005) Genomic analysis of a region encompassing QRfs1 and QRfs2: genes that underlie soybean resistance to sudden death syndrome. Genome 48: 125-138.

Vierling RA, J Faghihi, VR Ferris, and JM Ferris (1996) Association of RFLP markers with loci conferring broad-based resistance to the soybean cyst nematode (Heterodera glycines). Theor Appl Genet 92: 83-86.

Walker JC and R Zhang (1990) Relationship of a putative receptor protein-kinase from maize to the s-locus glycoproteins of Brassica. Nature 345: 743-746.

Wang D, PR Arelli, RC Shoemaker, and BW Diers (2001) Loci underlying resistance to Race 3 of soybean cyst nematode in Glycine soja plant introduction 468916 Theor Appl Genet 103: 561-566.

Wink M (1999) Plant secondary metabolites from higher plants: biochemistry, function and biotechnology. Annual Plant Reviews 2 ed. Sheffield, UK: Sheffield Academics.

Wrather JA, TR Anderson, DM Arsyad, Y Tan, LD Ploper, A Porta-Puglia, HH Ram, and JT Yorinori (2001) Soybean disease loss estimates for the top ten soybean-producing countries in 1998. Can J Plant Path 23: 115-121.

Yuan J, VN Njiti, K Meksem, MJ Iqbal, K Triwitayakorn, MA Kassem, GT Davis, ME Schmidt, and DA Lightfoot (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Sci 42: 271-275.

Yue P, PR Arelli, and DA Sleper (2001) Molecular characterization of resistance to Heterodera glycines in soybean PI438489B. Theor Appl Genet 102: 921-928.

Zatserklyana O, K Meksem, and DA Lightfoot (2017) Glycine max cultivar Forrest clone BAC 73P06 genomic sequence. In: NCBI GenBank. HQ008938.1. Atlas Journal of Biology 2017: 376-383. doi:10.5147/ajb.v0i0.46.

Zhu YI, QJ Song, DL Hyten, CP van Tassell, LK Matukumalli, DR Grimm, SM Hyatt, EW Fickus, ND Young, and PB Cregan (2003) Single-nucleotide polymorphisms in soybean. Genetics 163: 1123-1134.