MAP AND FINE MAP APHID RESISTANCE GENES IN SOYBEAN PLANT INTRODUCTION (PI) 567597C, 567585A AND 567537

By

Wenyan Du

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics and Biotechnology – Crop and Soil Sciences – Doctor of Philosophy

2016

ABSTRACT

MAP AND FINE MAP APHID RESISTANCE GENES IN SOYBEAN PLANT INTRODUCTION (PI) 567597C, 567585A AND 567537

By

Wenyan Du

The soybean aphid has become one of the most devastating pests to soybean in North America since its invasion in 2000. Soybean breeders have been seeking resistance sources and incorporating the resistance into elite cultivars since then. My research focuses on mapping and fine mapping soybean aphid resistance genes in three aphid resistance sources PI 567597C, PI 567537 and PI 567585A.

The first objective of this study was to characterize the inheritance pattern of soybean aphid resistance gene/genes in PI 567597C and pinpoint the resistance gene/genes with genetic markers. Four populations segregating for aphid resistance from PI 567597C were studied. Phenotypic analysis by Chi-square test showed that aphid resistance in PI 567597C was controlled by a single partially dominant gene. QTL analysis using population 050107 located the aphid resistance gene on chromosome 16 which explained 90% of the variation, named *Rag*3e. Data from population 050018 excluded possible QTLs on chromosomes with known aphid resistance QTLs. Data from two populations confirmed the aphid resistance gene in different genetic backgrounds.

The second objective of this study was to narrow the *Rag*3e QTL region in PI 567597C to facilitate marker assisted selection. residual heterozygous lines (RHL) were selected from two segregating populations. SNP markers discovered from the next generation sequencing data enabled the fine mapping process. *Rag*3e was pinpointed to a 60kb interval on chromosome 16 with seven candidate genes.

The third objective of this study was to use RHLs to fine map aphid resistance QTL, *Rag*3d, from PI 567585A, to develop single nucleotide polymorphism (SNP) markers for marker assisted selection and to find candidate genes for functional study. RHLs were selected from two segregating populations. Important RHLs were genotyped with the SoySNP50K chip to help identifying recombination breaking points. *Rag*3d was fine mapped to a 46kb interval on chromosome 16 with five candidate genes.

The fourth objective of this study was to use SNP markers to fine map aphid resistance QTL *Rag*3b from PI 567537, to identify candidate genes for *Rag*3b and to develop genetic markers for marker assisted selection. Three F2 populations derived from the cross between aphid resistant and the susceptible parents were used to validate *Rag*3b and screen for recombination breaking points. The *Rag*3b QTL was narrowed to a 199kb region on chromosome 16 with twelve candidate genes.

This dissertation is dedicated to my dear father (Yong Du), mother (Zeqin Luo), Ding Wang and my beloved grandparents.

ACKNOWLEDGEMENTS

My sincere thanks to my Ph. D. advisor, Dr. Dechun Wang, who treated me like his daughter and was the best advisor ever. To my committee members: Dr. Chris DiFonzo, who taught me entomology, teaching and communication skills; Dr. Amy Iezzoni, who showed me plant breeding as a career and how to deal with difficulties; Dr. Jin Chen, who not only enlighten me in bioinformatics but also taught me to think critically.

Appreciation to Wang lab members: Kate Zhang, Desmi Chandrasena, John Yuan, Umesh Rosyara, Chunyan Yang, Zhiming Dong, Yingdong Bi, Shichen Zhang, Zixiang Wen, Rujuan Tan, Paul Collins, Feng Lin, John Boyse, Randy Laurenz and Cherry Gu for their help and encouragement.

Gratitude goes to my other mentors, Drs Russell Freed, Cholani Weebadde, Guo-qing Song, Eric Olson, Walter Pett, Dave Douches, Tylor Johnston, Robin Buell, Shin-Han Shiu, Jianping Hu, Alan Prather, Federica Brandizzi, Melinda Frame and Katherine Osteryoung for their training and support. I also would like to extend my gratitude to the PSM, Plant Biology, CIPS, Horticulture, ASHS and ASA-CSSA-SSSA staff, PSM, PBGB, Plant Biology and Horticulture faculty and graduate students.

My sincere appreciation to PBGB, MSU graduate school, PSM, CANR, COGS and ASHS

for enrichment funding; and to Michigan Soybean Promotion Committee, the North Central Soybean Research Program, and the United Soybean Board for my Ph. D. research and assistantship fund support.

To all my dearest friends who are still in Michigan, not Michigan any more, in China, in Europe, in Australia, etc. for their emotional support and advise for life and work, especially when I was hurt.

To my family Yong Du, Zeqin Luo and Ding Wang for their unconditional love and support!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: LITERATURE REVIEW

Soybeans

Glycine max (L.) Merr., the cultivated soybean, and its wild relative, *Glycine soja*, are dicots and both belong to the subgenus *Soja* of the family *Leguminosae.* The *Glycine max* genome went through two rounds of ancient duplications and diploidized eventually with 2n=40 (Roulin et al. 2013). Soybean plants are bush-like and their flowers are normally selfing. Because soybean plants are photoperiod and temperature sensitive, latitude and day length determine if soybean can be cultivated in a particular region (from the 45° latitude in the north to near the equator) (Government of Canada 2012).

Soybean as A Major Crop

Soybean was one of the most important crops in East Asia long before written history. In 1765, soybean was introduced to the United States from China by a sailor. Nowadays, soybeans rank number two in acreage in the U. S. It is used as a rotation crop due to the ability to fix nitrogen, as well as used to feed livestock, consumed as meal, flour, infant formula, even meat and dairy substitutes. Most significantly, soybean is a source of oil production (Lusas and Riaz 1995; Miniello et al. 2003; Giampietro et al. 2004; Hoogenkamp 2005; Lim 2012). In fact, soybean oil dominates ninety percent of U. S.

vegetable oil production (USDA).

Soybean Aphids

Soybean aphid (SBA), *Aphis glycines* Matsumura, is in order Hemiptera, family Aphididae. SBA is native to eastern Asia. This small (<1/16" long when mature) and pale yellow crop pest is recognized by its black cornicles and pale cauda. The life cycle of SBA starts with male and female mating in the fall and producing eggs to overwinter on the primary host, buckthorn (*Rhamnus* spp.). These eggs hatch and develop into wingless fundatrices, which are the first generation in the spring. The second generation consists of mostly wingless females that produce the third and further generations of winged morphs on buckthorn preparing to colonize the secondary host, soybeans. In the summer, winged migrants start the colony on soybean plants by feeding for a short time and depositing a few nymphs. Then, several parthenogenetic generations are produced with both the winged and wingless morphs. SBA can be highly populated on soybeans with a reproduction time as short as 1.5 days at 25^0C (McCornack et al. 2004). During later summer and fall, sexual reproduction starts. The cycle continues as the eggs are laid on buckthorn to overwinter (Ragsdale et al. 2004).

SBA was discovered in the U. S. in 2000 (Hartman et al. 2001). By 2009, SBA has invaded thirty states in the United States and three provinces in Canada (Ragsdale et al. 2011). SBA damages soybean by sucking sap and transmitting viruses. It also secretes honeydew which results in the inhibition of photosynthesis and induces fungal infection.

They not only cause damage in soybeans but also other major crops (Mian et al. 2008), for example, potato, dry bean and vine crops, due to its ‘detecting by tasting’ nature. Damages of crops by the soybean aphid include yield reduction and severe quality decrease (Beckendorf et al. 2008). In 2003, it was estimated that 42 million acres of soybean in the north-central United States was greatly affected by an outbreak of the soybean aphid (Song et al. 2006). The resulting economic loss could be over 2.4 billion dollars (Song et al. 2006). Instead of chemical spray, an economical and environmentally favorable way to deal with soybean aphids was needed.

Host Plant Resistance to Insects

Over millions of years, insects coevolved with their host plants (Bruce 2014). Insects can feed on plants for food, shelter and egg laying. To fight against invasion, plants developed mechanisms to avoid, deter or kill insects. These mechanisms are called host plant resistance (HPR) to insects (Bosque-Pérez and Buddenhagen 1992). To survive, insects in turn evolved biotypes adapted to the HPR mechanisms. This is the ‘an arms race’ between insects and the host plants. HPR can be incorporated with other practices for Integrated Pest Management (IPM). Examples of HPR includes brown planthopper resistance in rice (Du et al. 2009), Hessian fly resistance in wheat (Dweikat et al. 1997) and potato leaf hopper resistance in alfalfa (Shade et al. 1979).

There are three types of host plant resistance, tolerance, antixenosis and antibiosis. Tolerance is the ability of a host plant to withstand injury due to an insect, without a reduction in growth and yield. An example for tolerance is corn tolerant to European corn

borers has a stronger stalk (Myers 1932). Insect tolerant varieties need no or fewer insecticide sprays. Because tolerance only involves traits of the plant itself and does not reduce insect feeding or reproduction, there is no selection pressure on the insect pest population. Thus, there is no chance an insect will be resistant to the plant. However, tolerance is generally not a target for breeding for insect host plant resistance. This is because tolerant crops can support a large number of insects that may migrate to nearby susceptible fields or spread plant viruses. Some entomologists and breeders do not consider tolerance as a form of resistance.

Antixenosis or non-preference resistance refers to plant traits which drive insects away or simply are not attractive to insects. Antixenosis literally means anti-guest. Plants with antixenosis are less attractive to insects for colonization, oviposition and feeding. Allelochemical nonpreference and morphological nonpreference are distinct mechanisms of antixenosis. Plants with allelochemical nonpreference use plant volatiles, attractants or host cues to fight against insects. A higher level of cucurbitacins, a very common chemical in cucurbit plants, causes bitterness that leads to resistance to two-spotted spider mites in cucumber (Ponti and Garretsen 1980). Interestingly, cucumber beetles are attracted to cucurbitacins (Pessarakli 2016). This reminds us that allelochemical non-preference is relative in terms of different insects. Allelochemical non-preference is also relative in the sense that even if a plant is not preferred, it often can be infested if nothing else is available. Hairs, trichomes, waxes or other physical structures contribute to morphological nonpreference. A frego-bract leaf in cotton, wrapped around the boll has a narrower leaf shape, holding fewer cotton bollworms (Sharma 2008). Another example of morphological

nonpreference is a wax bloom on cabbage, compared to cabbage with a glossy surface (Eigenbrode et al. 1996). Morphological non-preference is harder to overcome because it involves physical traits like hairs or wax. But these traits may be unacceptable on crops which are eaten fresh. And again, in the absence of a better host, an insect can still feed on plants with antixenosis resistance. Therefore, non-preference is a limited form of host plant resistance to insects. It is useful if there is no other alternative, or couple with antibiosis (below).

Antibiosis is described as plant traits that negatively impact insect biology. These traits reduce development, survival or reproduction of insects. Insects affected by antibiosis can have reduced growth rate or increased mortality for immatures, smaller or malformed adults, abnormal behaviors (like restlessness), reduced fecundity and shorter life span. Structural factors, allelochemicals or nutritional factors can result in antibiosis. For example, glandular trichomes on wild potato trap or even kill potato leafhoppers by secreting viscous exudate that rapidly darkens and hardens (Tingey and Gibson 1978). High benzyl alcohol (as a growth inhibitor) content in wheat and barley lower the reproduction rate of greenbug (Juneja et al. 1975). Rice with a high level of DIMBOA (as a toxin) is resistant to European corn borer (Abel 1998). Maize with high aspartic acid, low nitrogen and sugar (nutritional factors) is resistant to maize stem borer (Books 2014).

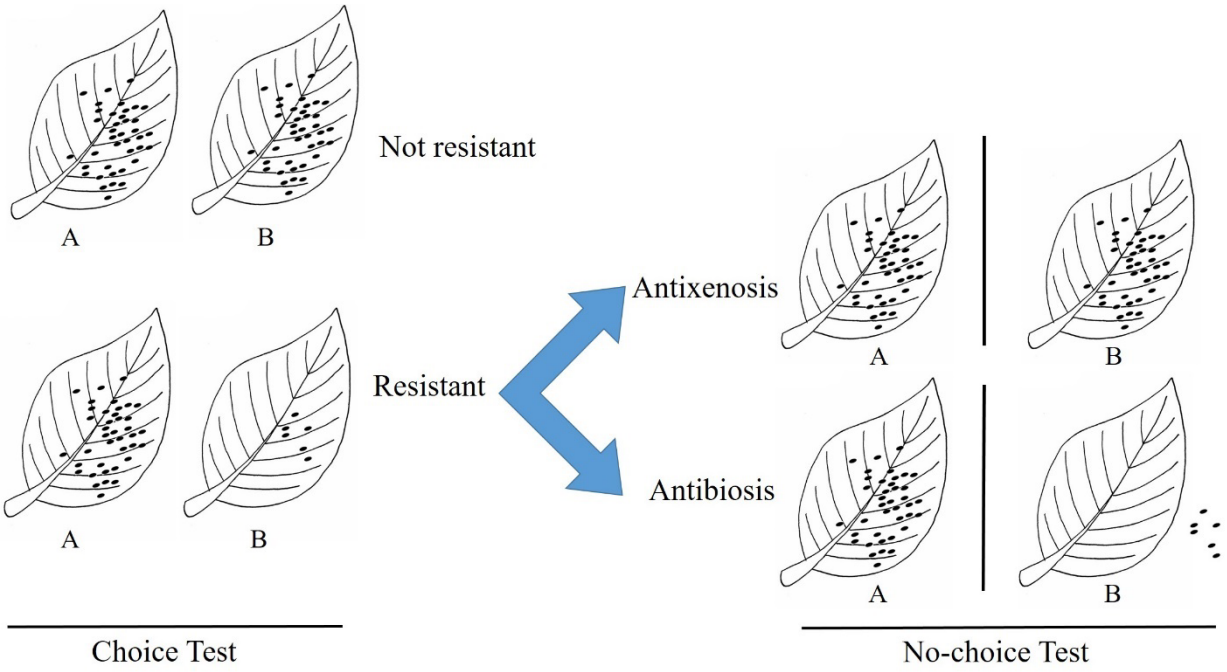
Antibiosis is true resistance. It can be almost a hundred percent effective. Therefore, antibiosis is the most important type of host plant resistance to insects. However, antibiosis puts strong selection pressure on an insect population, so, biotypes may evolve

quickly. Plant breeders need to be one step ahead of the insects. From the agricultural perspective, the preferred resistance is antibiosis combined with antixenosis (Emden and Harrington 2007).

Choice and No-Choice Tests

Choice and no-choice tests are experimental tools to determine the type of host plant resistance. For instance, to distinguish between antixenosis and antibiosis. Different parts of the plant or the whole plant can be used in choice and no-choice tests. For example, in figure 1.1 leaf samples are used. Two samples are put together for a choice test. The choice test indicates if the testing leaf sample (sample B) is resistant to the insect or not. In a no-choice test, two samples would be placed separately. The results demonstrate whether the testing sample (sample B) confers antixenosis or antibiosis resistance to insects (Figure 1.1). Choice test and no-choice test have been applied in multiple situations. Mensah et al. (2005) conducted choice and no-choice when screening aphid resistance source from two thousand one hundred and forty-six germplasm. This helped them to characterize and distinguish different germplasm. Chandrasena et al. (2012) used choice and no-choice test to discover the interesting fact that soybean with aphid resistance gene *rag*1b and *rag*3 were susceptible to Japanese beetle.

Figure 1.1: Choice and no-choice tests. Sample A is the control, which is susceptible to the insect. Sample B is the testing sample. Black dots on or around the leaf samples represent the insect being tested.



Breeding for Soybean Aphid Resistance

Breeding for host plant resistance to insects is challenging compared to breeding for agronomic traits, because the breeder needs:

1) manipulate both the plant and the insect,

2) understand the biology of the plant and the insect,

3) have a healthy, continuous and large supply of insects,

4) develop an efficient inoculation system,

5) develop an accurate and reproducible rating system,

6) monitor development of new biotypes, and

6) be aware that HPR could negatively impact other practices, for example, biocontrol.

Despite the challenges, breeding for host plant resistance to insects is worthwhile when the insect is a continuous and common problem across a wide region, especially if it is costly to control, resistant to insecticides, or controlled by no or few other methods. The soybean aphid is such an important pest. This is why soybean breeding program across the U. S. began to breed for SBA after its discovery in the U. S.

Since 2002, seven, three and two resistance sources were identified at Michigan State University, University of Illinois and Ohio State University, respectively (Mensah et al. 2005; Li et al. 2006; Chen et al. 2007; Mian et al. 2008; Liu 2010; Jun et al. 2012; Zhang

et al. 2013). These sources were all tested by choice and no-choice tests, to determine their resistance type (Table 1.1). Du et al. (2015) described three major steps to incorporate soybean aphid resistance, pinpoint, transfer and stack. First, pinpoint the resistance gene from the resistance source. Next, the resistance gene will be transferred into elite cultivars. Last and the most creative step, is to stack different traits based on demands. Traits like high yield, disease resistance, insect resistance, drought resistance, high protein content, etc. can all be stack into one cultivar. This ‘super’ cultivar would be ready for licensing and commercialization.

Pinpoint, in another word, map and fine map, is the most important and time consuming step. Through map and fine map procedures, markers tightly linked with the trait are identified. They can be used to enable marker assisted selection. This 1) enables precision breeding; 2) saves space, labor, time and financial cost, and 3) for some traits, there is no need to conduct phenotyping (Du and Wang 2014).

Table 1.1: A summary of genes for soybean aphid resistance in soybean.

Resistant source	Resistant gene/genes	Mapped region	Number & inheritance of the gene	Resistance to aphid*	Reference
Dowling	*Rag*1 [a]	Chr 7 [b]	Single, dominant	Antibiosis, biotype 1 [c]	Li et al. 2006. Mol Breeding 19: 25-34
Jackson	*Rag*	Chr 7	Single, dominant	Antibiosis, biotype 1, (no data for type4)	Li et al. 2006. Mol Breeding 19: 25-34
PI 567541B	*rag*1c, *rag*4	Chr 7, Chr 13	Two, recessive	Antibiosis, Biotype 1&2	Zhang et al. 2009. Theor Appl Genet 118: 473–482
PI 567301B	*Rag*5 *(Rag2), QTL on chr8*	Chr13, Chr8	NA	Antixenosis, Biotype 1&2(no data for type4)	Jun et al. 2012. Theor Appl Genet 124: 13-22
P203	*[Rag6]_P203*	Chr8	Single, dominant	Antixenosis	Xiao et al. 2013. Theor Appl Genet 126: 2279–2287
PI 243540	*Rag*2	Chr 13	Single, dominant	Antibiosis, Biotype 1&2	Mian et al. 2008. Theor Appl Genet 117: 955–962
PI 200538	*Rag*2	Chr 13	Single, dominant	Antibiosis, Biotype 1&2	Hill et al. 2009. Crop Sci 49:1193
PI 567543C	*Rag*3	Chr 16	Single, dominant	Antixenosis, Biotype 1, 2, 3,4	Zhang et al. 2010. Theor Appl Genet 120: 1183-1191
PI 567598B	*rag*1b, *rag*3	Chr 7, Chr 16	Two, recessive, partially dominant	Antibiosis, Biotype 1, 2, 3	Bales et al. 2013. Theor Appl Genet 126(8): 2081-91
PI 567585A	*Rag*3d	Chr 16	Single, partially dominant	Antixenosis & Antibiosis, Biotype 1, 2, 3, (no data for type4)	Liu 2010. MSU Ph. D. dissertation Chapter 2 & 3
PI 567597C	*Rag*3e	Chr 16	Single, partially dominant	Antixenosis, Biotype 1, 2, 3,4	Du et al. 2016 (in preparation)
PI 567537	*Rag*3b	Chr 16	Single, dominant	Antibiosis, Biotype 1, 2, 3, (no data for type4)	Zhang et al. 2013. Mol Breeding Volume 32, Issue 1, pp 131-138

Table 1.1 (cont'd)

[a] *Rag* stands for resistance to *aphis Glycine*. Numbers following *Rag* were used to distinguish aphid resistance gene discovered at different location of soybean genome.
[b] Chromosome with number in soybean genome
[c] Biotypes are aphid colonies distinguished by resistance to specific resistance sources.
*Information of resistance to 4 aphid biotypes was summarized from literatures (Kim et al. 2008; Hill et al. 2010; Alt and Ryan-Mahmutagic 2013).

Map and Fine Map Aphid Resistance Gene/Genes in Soybeans

Researchers have been studying the genetics of aphid resistance for more than ten years (Table 1.1). Three major quantitative trait loci (QTL) have been identified in different plant introductions (PI). *Rag*, resistance to *aphis Glycine*, was used to name the aphid resistance genes. *Rag*1 from Dowling, *Rag* from Jackson, *rag*1b from PI 567598B and *rag*1c from PI 567541B were mapped to the same genetic region on chromosome 7 (Hill et al. 2006a; Hill et al. 2006b; Zhang et al. 2009; Yan Li 2012); *Rag*2 from PI 243540 and PI 200538, and *Rag*5 from PI 567301B (later was claimed to be the same as *Rag*2) were mapped to chromosome 13 (Kang et al. 2008; Hill et al. 2009; Jun et al. 2012); *rag*4 from PI 567541B was also mapped to Chromosome 13 but a different region from *Rag*2 (Zhang et al. 2009); *Rag*3*/rag*3 from PI 567543C, PI 567598B, PI 567585A and PI 567537 were mapped to chromosome 16 (Zhang et al. 2010; Liu 2010; Zhang et al. 2013; Bales et al. 2013). There are also claims of an aphid resistance QTL on Chromosome 8 (Jun et al. 2012; Xiao et al. 2013). It is still unknown if the resistance QTL from different PIs that mapped to the same region are different genes closely linked or if they are different alleles of the same gene. In fact, the significant differences shown by the effect of the genes (whether it is dominant, recessive or co-dominant) as well as the resistance type to the soybean aphid (whether it is antibiosis, antixenosis or tolerance) reflects the complexness of soybean aphid resistance as an important trait for soybean breeding (Bales et al. 2013)

Among these soybean aphid resistance QTLs, fine mapping studies of *Rag*1, *Rag*2, *rag*1c and *[Rag6]_P203* have been published. Using single nucleotide polymorphism (SNP) markers to genotype eight hundred and twenty-four BC_4F_2 and a thousand BC_4F_3 plants, *Rag*1 from Dowling was fine mapped to a 115 kb region on chromosome 7. There were thirteen predicted genes within the 115 kb interval. Two of these were potential candidate genes for *Rag*1. *Rag*2 from PI 200538 was fine mapped into a 54 kb region on chromosome 13 using five thousand seven hundred and eight-three F_2 plants. Seven predicted genes were in the fine mapped interval including one nucleotide-binding site leucine-rich repeat (NBS-LRR) gene (Kim et al. 2010a; Kim et al. 2010b). Next, *rag*1c gene from PI 567541B was fine mapped into a 96kb region on chromosome 7, which was different from the interval of *Rag*1(Yuan 2014). *[Rag6]_P203* from line P203 has been fine mapped to a 192-kb interval with five candidate genes (Xiao et al. 2013). Compared with the initial QTL mapping results, fine mapping intervals are much narrower and have fewer candidate genes. These intervals are narrow enough so that there is almost no recombination happening within the interval. Therefore, the flanking markers of the fine mapped interval could be confidently used for marker assisted selection.

However, it must be pointed out that the resistance conferred by *Rag*1 (Kim et al. 2008), *Rag*1b (Bales et al. 2013) and *Rag*2 (Chandrasena et al. 2015) genes is overcome by certain biotypes of soybean aphids or have limited resistance to soybean aphids. Besides conducting genetic studies for the existing resistance sources, pyramiding different

resistance genes and discovering new sources of resistance would be an efficient and long lasting strategy to control soybean aphids (Bales et al. 2013; Chandrasena et al. 2015).

Utilization of high through-put genotyping platform, the SoySNP50K iSelect SNP beadchip (SoySNP50K chip) and the Next Generation Sequencing Technology in Soybean Studies

High through-put genotyping platform has been developed in Wang lab at Michigan State University since 2012 (Du and Wang 2014). The steps are: 1) tissue collection in a 96-well plate, 2) freeze dry, 3) sample grinding and adding glass beads, 4) DNA extraction using CTAB method with multiple channel pipette, 5) DNA dilution and distribution for PCR using a Biomek 2000 robot, 6) PCR with KASP enzyme mix using Bio-Rad PCR machine, 7) plate reading by Roche 480 plate reader, and 8) data analysis with Roche 480 software and output. Compared with traditional genotyping platform, it handles more samples at the same time. It is more accurate with robot distribution and plate reading. It also saves the time to run the electrophoresis gel (for SSR markers).

The SoySNP50K iSelect SNP beadchip (Song et al. 2013) is another way for massive genotyping. It focusses on a few samples with numerous markers whereas high through-put genotyping platform is targeting a large number of individual DNAs, with a few markers. Markers on the SoySNP50K chip have been carefully selected and tested (Song et al. 2013). Therefore, it has a great representation to cover the genetic pool of soybeans and

has a stable performance. It is applied widely in characterizing the genetic pool and important parents, genome-wide association mapping in disease resistance, etc. (Wen et al. 2014).

Next generation sequencing (NGS) mostly refers to Illumina based technology which does sequencing by synthesis. As sequencing technologies prospered and became affordable, next generation sequencing became commonly used in soybean breeding and genetic studies. Kim et al. (2012) sequenced *G. max* and *G. soja* genome and described many structural genomic differences between these two. Among these differences, genes that are important for domestication processes would be interesting to study in detail. SNP discovery is another application of the NGS data. Hyten et al. (2010) sequenced a reduced representation library and discovered 7,108 to 25,047 predicted SNPs. With this dataset, they were also able to anchor and orient scaffolds in the soybean whole genome sequence. More importantly, the NGS data helped greatly in QTL mapping and locating the candidate gene/genes. Xu et al. (2013) constructed a population of recombinant inbred lines (RIL) and they sequenced lines in the population (low coverage) as well as the parents. They were able to develop and validate SNPs from the dataset and used these SNPs to construct the linkage map. A major QTL was located in a 29.7 kb region with two candidate genes. Apart from sequencing the genomic DNA, researchers also sequenced cDNA which was reverse transcribed from RNA. Molina et al. (2012) carried out metatranscriptomic analysis of small RNAs in soybean deep sequencing libraries in order to understand how the soybean plant reacts to the environment. Severin et al. (2010) did RNA-seq for different development stages of different soybean tissues. This dataset

would help greatly in soybean genome annotation and functional studies.

For this study, the majority of the individuals were genotyped via high through-put genotyping platform. The important recombinants and parents were genotyped with the SoySNP50K chip. All the parents were also sequenced by next generation sequencing with 5x coverage. The data from the SoySNP50K chip were used to identify recombination breaking points for fine mapping. SNP markers in this study were either developed from the SoySNP50K chip or the NGS data.

Application of Residual Heterozygous Lines to Fine Map Aphid Resistance Gene

Residual heterozygous lines (RHL) are lines that have been advanced for several generations. Most of the loci in the genome are not segregating any more but very few loci are still heterozygous. These lines could be derived from a bi-parental cross or any of the selfing processes happened in nature. For example: a plant introduction. Because of this, RHL is useful in fine mapping. One way to use RHL is to focus on one of the candidate regions. In this case, lines that are genotyped to be heterozygous in the target region will be kept for every generation. After several generations, only the target region is segregating while others are already homozygous. In this process, the trait has to be confirmed segregating phenotypically as well. Another scenario to use RHL is when there is no particular region/regions to focus on. Under this circumstance, lines that are segregating phenotypically will be proceeded to the next generation. After generations, only the selected trait is still segregating while other traits are already fixed. At this point, when genotyping, only a few heterozygous regions will stand out for further confirmation

and selection.

In this study, one reason to use RHL is because there were no SNP markers developed to cover the whole initial mapping region before F5:6. Another reason is that more phenotypically segregating lines (RHLs) were needed. This is to make sure that if the QTL was located to a wrong place, phenotypically segregating RHLs could be retrieved and tested for other possible regions. Within the phenotypically segregating lines genotypically segregating lines would be selected to continue as the RHL for next generation and the genotypically fixed ones for progeny test in the next generation. In both cases, markers flanking and within the candidate regions were used.

Objectives of the present study

This study focused on PI 567585A, PI 567597C and PI 567537. The first objective was to map and fine map the soybean aphid resistance QTL from PI 567597C. A single partially dominant gene, *Rag*3e was found explaining 90% of the phenotypic (soybean aphid resistance) variation. Through fine mapping, *Rag*3e was narrowed to a 60kb intervals on chromosome 16. The possible candidate genes were identified.

The second objective was to fine map aphid resistance gene in PI 567585A and PI 567537. They all have a single gene on chromosome 16 controlling aphid resistance. Aphid resistance gene from PI 567585A and PI 567537 were narrowed to a 46kb and 199kb intervals on chromosome 16, respectively. The possible candidate genes were identified.

QTLs from PI 567585A, PI 567597C and PI 567537 differed by inheritance pattern of the resistance genes (dominant, partially dominant or recessive) and the distinct resistance types to soybean aphids (antixenosis or antibiosis) (Table 1.1). They would contribute diverse resistance sources to soybean breeding. Markers tightly linked to soybean aphid resistance in these PIs is essential for integrating the resistance genes into elite lines. With the candidate gene information from this study, further analysis of the *Rag*3/*rag*3 region on chromosome 16 would help unravel the mechanisms of host plant resistance to the soybean aphid. Table 1.2 is summary of previous studies on these three PIs.

Table 1.2: A summary of previous research on PI 567585A, PI 567597C and PI 567537.

PI	Susceptible Parents in Mapping Population	Susceptible Parents in Validation Population
PI 567585A	IA2070（86 F2:3）	Skylla （222 F2:3）
	Initially mapped to Chr16, *Rag3/rag3* region by Menghan Liu (Liu 2010) Antixenosis and antibosis	
PI 567597C	Skylla (population 050107) E00003 (population 050018)	Skylla (population 100049) Titan (population 100130)
	Phenotypically identified by Clarice (Mensah et al., 2005) Genetically characterized in this study Antixenosis	
PI 567537	E0003 (86 F4)	Skylla(233 F2)
	Initially mapped to Chr16, *Rag3/rag3* region by Guorong Zhang (Zhang et al. 2013) Antibiosis	

REFERENCES

REFERENCES

Abel C (1998) Introgressing a new source of host-plant resistance to European corn borer into two elite maize inbred lines. Ph.D. dissertation, Iowa State University

Alt J, Ryan-Mahmutagic M (2013) Soybean Aphid Biotype 4 Identified. Crop Sci 53:1491

Bales C, Zhang G, Liu M, et al (2013) Mapping soybean aphid resistance genes in PI 567598B. Theor Appl Genet 126:2081–2091

Beckendorf EA, Catangui MA, Riedell WE (2008) Soybean Aphid Feeding Injury and Soybean Yield, Yield Components, and Seed Composition. Agron J 100:237

Books V (2014) Intermediate Second Year Botany (English Medium) test papers: Model papers, Practice paper, Important Questions. Vikram Publishers Pvt Ltd

Bosque-Pérez NA, Buddenhagen IW (1992) The development of host-plant resistance to insect pests: outlook for the tropics. In: Proceedings of the 8th International Symposium on Insect-Plant Relationships. Springer Netherlands, pp 235–249

Bruce TJA (2014) Interplay between insects and plants – dynamic and complex interactions that have coevolved over millions of years but act in milliseconds. J Exp Bot 66(2):455-65

Chandrasena D, DiFonzo C, Wang D (2012) An Assessment of Japanese Beetle Defoliation on Aphid-Resistant and Aphid-Susceptible Soybean Lines. Crop Sci 52:2351

Chandrasena D, Wang Y, Bales C, et al (2015) Pyramiding 3, 1b, 4, and 1c Aphid-Resistant Genes in Soybean Germplasm. Crop Sci 55:2108

Chen CY, Gu C, Mensah C, et al (2007) SSR marker diversity of soybean aphid resistance sources in North America. Genome Natl Res Counc Can Génome Cons Natl Rech Can 50:1104–1111

Du B, Zhang W, Liu B, et al (2009) Identification and characterization of Bph14, a gene conferring resistance to brown planthopper in rice. Proc Natl Acad Sci 106:22163–22168

Du W, Wang D (2014) Marker Assisted Selection as a Powerful Strategy for Gene

Pyramiding in Soybean Breeding. WENYANs Teach. WORLD. http://wdmsu.weebly.com/marker-assisted-selection.html. Accessed 14 Mar 2016

Du W, Wang D (2015) Perspectives from Soybean Breeding: A Combination of Classic and New Technologies to Map and Fine Map Aphid Resistance in Soybean and Its Application in Horticulture. ASHS Search and Access Archived Conference Presentations. https://ashs.confex.com/ashs/2015/videogateway.cgi/id/1983?recordingid=1983. Accessed 5 Sep 2015

Dweikat I, Ohm H, Patterson F, Cambron S (1997) Identification of RAPD markers for 11 Hessian fly resistance genes in wheat. Theor Appl Genet 94:419–423

Eigenbrode S d., Castagnola T, Roux M-B, Steljes L (1996) Mobility of three generalist predators is greater on cabbage with glossy leaf wax than on cabbage with a wax bloom. Entomol Exp Appl 81:335–343

Emden HFV, Harrington R (2007) Aphids as Crop Pests. CABI, UK

Giampietro PG, Bruno G, Furcolo G, et al (2004) Soy protein formulas in children: no hormonal effects in long-term feeding. J Pediatr Endocrinol Metab JPEM 17:191–196

Government of Canada CFIA (2012) The Biology of Glycine max (L.) Merr. (Soybean). Canadian Food Inspection Agency. http://www.inspection.gc.ca/plants/plants-with-novel-traits/applicants/directive-94-08/biology-documents/glycine-max-l-merr-/eng/1330975306785/1330975382668. Accessed 18 Nov 2013

Hartman GL, Domier LL, Wax LM, et al (2001) Occurrence and Distribution of Aphis glycines on Soybeans in Illinois in 2000 and Its Potential Control. Plant Health Prog. http://naldc.nal.usda.gov/download/11854/PDF. Accessed 11 Apr 2016

Hill CB, Crull L, Herman TK, et al (2010) A new soybean aphid (Hemiptera: Aphididae) biotype identified. J Econ Entomol 103:509–515

Hill CB, Kim K-S, Crull L, et al (2009) Inheritance of Resistance to the Soybean Aphid in Soybean PI 200538. Crop Sci 49:1193

Hill CB, Li Y, Hartman GL (2006a) A Single Dominant Gene for Resistance to the Soybean Aphid in the Soybean Cultivar Dowling. Crop Sci 46:1601

Hill CB, Li Y, Hartman GL (2006b) Soybean Aphid Resistance in Soybean Jackson Is

Controlled by a Single Dominant Gene. Crop Sci 46:1606

Hoogenkamp HW (2005) Soy protein and formulated meat products. CABI Pub., Cambridge, MA

Hyten DL, Cannon SB, Song Q, et al (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38

Jun T-H, Mian MAR, Michel AP (2012) Genetic mapping revealed two loci for soybean aphid resistance in PI 567301B. Theor Appl Genet 124:13–22

Juneja PS, Pearcy SC, Gholson RK (1975) Chemical Basis for Greenbug Resistance in Small Grains: II. Identification of the Major Neutral Metabolite of Benzyl Alcohol in Barley. Plant Physiol 56:385–389

Kang S-T, Rouf Mian MA, Hammond RB (2008) Soybean Aphid Resistance in PI 243540 Is Controlled by a Single Dominant Gene. Crop Sci 48:1744

Kim K-S, Bellendir S, Hudson KA, et al (2010a) Fine mapping the soybean aphid resistance gene Rag1 in soybean. Theor Appl Genet 120:1063–1071

Kim K-S, Hill CB, Hartman GL, et al (2010b) Fine mapping of the soybean aphid-resistance gene Rag2 in soybean PI 200538. Theor Appl Genet 121:599–610

Kim K-S, Hill CB, Hartman GL, et al (2008) Discovery of Soybean Aphid Biotypes. Crop Sci 48:923

Kim MY, Van K, Kang YJ, et al (2012) Tracing soybean domestication history: From nucleotide to genome. Breed Sci 61:445–452

Li Y, Hill CB, Carlson SR, et al (2006) Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. Mol Breed 19:25–34

Lim DTK (2012) Glycine max. In: Edible Medicinal and Non-Medicinal Plants. Springer Netherlands, pp 634–714

Liu M (2010) Inheritance of resistance and molecular mapping of soybean aphid resistance genes in soybean PI 567585A; identification of aphid resistance genes in soybean using modified nested association mapping (MNAM). Ph.D. dissertation, Michigan State University

Lusas EW, Riaz MN (1995) Soy protein products: processing and use. J Nutr 125:573S–

580S

McCornack BP, Ragsdale DW, Venette RC (2004) Demography of soybean aphid (Homoptera: Aphididae) at summer temperatures. J Econ Entomol 97:854–861

Mensah C, DiFonzo C, Nelson RL, Wang D (2005) Resistance to Soybean Aphid in Early Maturing Soybean Germplasm. Crop Sci 45:2228

Mian MAR, Hammond RB, St. Martin SK (2008) New Plant Introductions with Resistance to the Soybean Aphid. Crop Sci 48:1055

Miniello VL, Moro GE, Tarantino M, et al (2003) Soy-based formulas and phyto-oestrogens: a safety profile. Acta Paediatr Oslo Nor 1992 Suppl 91:93–100

Molina LG, da Fonseca GC, de Morais GL, et al (2012) Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries. Genet Mol Biol 35:292–303

Myers KH (1932) Adjusting corn belt farming to meet corn-borer conditions. U.S. Dept. of Agriculture

Pessarakli M (2016) Handbook of Cucurbits: Growth, Cultural Practices, and Physiology. CRC Press, Boca Raton, FL

Ponti OMBD, Garretsen F (1980) Resistance in Cucumis sativus L. to Tetranychus urticae Koch. 7. The inheritance of resistance and bitterness and the relation between these characters. Euphytica 29:513–523

Ragsdale DW, Landis DA, Brodeur J, et al (2011) Ecology and Management of the Soybean Aphid in North America. Annu Rev Entomol 56:375–399

Ragsdale DW, VOEGTLIN DJ, O'NEIL RJ (2004) Soybean Aphid Biology in North America. Ann Entomol Soc Am 97:204–208

Roulin A, Auer PL, Libault M, et al (2013) The fate of duplicated genes in a polyploid plant genome. Plant J 73:143–153

Severin AJ, Woody JL, Bolon Y-T, et al (2010) RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biol 10:160

Shade RE, Doskocil MJ, Maxon NP (1979) Potato Leafhopper Resistance in Glandular-Haired Alfalfa Species1. Crop Sci 19:287

Sharma HC (2008) Biotechnological Approaches for Pest Management and Ecological Sustainability. CRC Press, Boca Raton, FL

Song F, Swinton S, DiFonzo C, et al (2006) Profitability Analysis of Soybean Aphid Control Treatments in Three North-central States. Staff paper at Michigan State University. http://msuent.com/assets/Song06SBAProfitablity.pdf. Accessed 11 Apr 2016

Song Q, Hyten DL, Jia G, et al (2013) Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. PLoS ONE 8(1):e54985

Tingey WM, Gibson RW (1978) Feeding and Mobility of the Potato Leafhopper Impaired by Glandular Trichomes of Solanum berthaultii and S. polyadenium. J Econ Entomol 71:856–858

USDA (2012) Soybeans & Oils Crops - Background. USDA Economic Research Service. http://www.ers.usda.gov/topics/crops/soybeans-oil-crops/background.aspx#.UoE_KRBFaLB. Accessed 11 Nov 2013

Wen Z, Tan R, Yuan J, et al (2014) Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics 15:809

Xiao L, Hu Y, Wang B, Wu T (2013) Genetic mapping of a novel gene for soybean aphid resistance in soybean (Glycine max [L.] Merr.) line P203 from China. Theor Appl Genet 126:2279–2287

Xu X, Zeng L, Tao Y, et al (2013) Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. Proc Natl Acad Sci 110:13469–13474

Yan Li CBH (2012) Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. Mol Breed 19:25–34

Yuan J (2014) Fine Mapping Soybean Aphid Resistance Gene *rag1c* in Soybean (*Glycine max* L.). Plant and Animal Genome XXII. https://pag.confex.com/pag/xxii/webprogram/Paper13372.html. Accessed 9 Dec 2015

Zhang G, Gu C, Wang D (2013) Mapping and validation of a gene for soybean aphid resistance in PI 567537. Mol Breed 32:131–138

Zhang G, Gu C, Wang D (2009) Molecular mapping of soybean aphid resistance genes

in PI 567541B. Theor Appl Genet 118:473–482

Zhang G, Gu C, Wang D (2010) A novel locus for soybean aphid resistance. Theor Appl Genet 120:1183–1191

CHAPTER II: MAP AND FINE MAP SOYBEAN APHID RESISTANCE GENE IN PI 567597C

Abstract

The soybean aphid (*Aphis glycines* Matsumura) has been an important pest on soybean [*Glycine max* (L.) Merr] in North America since 2000. PI 567597C was characterized as one of the soybean aphid resistance resources from our previous studies. However, little is known about the genetics of the soybean aphid resistance from PI 567597C. The first objective of this study was to characterize the inheritance pattern of soybean aphid resistance gene/genes from PI 567597C and pinpoint the resistance gene/genes with genetic markers. Four populations, 050107, 050018, 100049 and 100130 segregating for aphid resistance from PI 567597C were studied. Phenotypic analysis by Chi-square test showed that aphid resistance in PI 567597C was controlled by a single partially dominant gene. QTL analysis using population 050107 located the aphid resistance gene to an interval spanned a 29cM region on chromosome 16 which explained 90% of the variation, named *Rag*3e. Data from population 050018 excluded possible QTLs on chromosomes with known aphid resistance QTLs. Data from population 100049 and 100130 confirmed the aphid resistance gene in different genetic backgrounds. The second objective of this study was to narrow the QTL region. Residual heterozygous lines (RHL) were selected from population 100049 and 100130. SNP markers discovered from the next generation

sequencing data played an important role in fine mapping process. *Rag*3e was delimited to a 60kb interval on chromosome 16 with seven candidate genes. Genetic markers associated with *Rag*3e from this study would facilitate marker assisted selection for aphid resistance breeding in soybean.

Key words: Soybean, Aphid Resistance, Host Plant Resistance, QTL Mapping, Partial dominance, *Rag*3e, Fine Mapping, Residual Heterozygous Lines, Next Generation Sequencing

Introduction

Native to Asia, soybean aphid invaded the U. S. and quickly spread to more than ten states in the year 2000. By 2009, Soybean aphid was found in thirty states in the United States and three provinces in Canada (R. L. Blackman, V. F. Eastop 2000; Ragsdale et al. 2011). The soybean aphid became a major soybean pest in North America, affecting million acres of soybeans and costing billions to control (Song et al. 2006).

Soybean aphids damage crop by sucking plant sap and transmitting viruses (Hill et al. 2001). Honeydew secreted from the soybean aphid induce fungal infection, blocking photosynthesis. Soybean plants infested with soybean aphids have limited growth and

development in both vegetative and reproduction stages (Ragsdale et al. 2011). Despite up to 75% of yield loss (Catangui et al. 2009), soybean seed quality was also affected by soybean aphid (Beckendorf et al. 2008).

Economic threshold guided chemical spray has been a major practice for soybean aphid control (Ragsdale et al. 2007). The recommended economic threshold of 250 aphids per plant was proved effectively protecting yield even if the impact of natural enemies of soybean aphids was reduced (McCarville et al. 2011). However, once the economic threshold has been reached, it took much effort and billions of U. S. dollars to spray pesticides, which may also cause pollution to the environment (Song et al. 2006). A more effective, economical and environmentally favorable way, such as utilizing host plant resistance, to deal with soybean aphids was needed.

There are two types of host plant resistance to insects: antibiosis and antixenosis (Painter 1951). Plants with antixenosis are less attractive to insects for colonization, oviposition and feeding. Plants with antibiosis can negatively impact insect biology. More than forty soybean aphid resistant germplasm have been identified with either antixenosis or antibiosis or both resistances to the soybean aphid (Li et al. 2006; Hesler et al. 2007; Mian et al. 2008; Hill et al. 2009; Zhang et al. 2009; Zhang et al. 2010; Liu 2010; Jun et al. 2012; Bhusal et al. 2013; Bansal et al. 2013; Zhang et al. 2013; Bales et al. 2013; Xiao

et al. 2013; Hesler 2013; Bhusal et al. 2014; Liu et al. 2014; Hanson et al. 2016)

Three major quantitative trait loci (QTL) have been identified in different plant introductions (PI). *Rag*, resistance to *Aphis glycines*, was used to name the aphid resistance genes. *Rag*1 from Dowling, *Rag* from Jackson, *rag*1b from PI 567598B and *rag*1c from PI 567541B were mapped to the same genetic region on chromosome 7 (Hill et al. 2006a; Hill et al. 2006b; Zhang et al. 2009; Yan Li 2012); *Rag*2 from PI 243540 and PI 200538, and *Rag*5 from PI 567301B (later was claimed to be the same as *Rag*2) were mapped to chromosome 13 (Kang et al. 2008; Hill et al. 2009; Jun et al. 2012); *rag*4 from PI 567541B was also mapped to Chromosome 13 but a different region from *Rag*2 (Zhang et al. 2009); *Rag*3*/rag*3 from PI 567543C, PI 567598B, PI 567585A and PI 567537 were mapped to chromosome 16 (Zhang et al. 2010; Liu 2010; Zhang et al. 2013; Bales et al. 2013). There are also claims of an aphid resistance QTL on Chromosome 8 (Jun et al. 2012; Xiao et al. 2013)

Among these soybean aphid resistance QTLs, *Rag*1 from Dowling was fine mapped to a 115 kb region on chromosome 7 with two potential candidate genes; *Rag*2 from PI 200538 was fine mapped into a 54 kb region on chromosome 13 with one nucleotide-binding site leucine-rich repeat (NBS-LRR) gene (Kim et al. 2010a; Kim et al. 2010b); *rag*1c gene from PI 567541B was fine mapped into a 96kb region on chromosome 7, which was

different from the interval of *Rag*1 (Yuan 2014); and *[Rag6]_P203* from line P203 has been fine mapped to a 192-kb interval with five candidate genes (Xiao et al. 2013). Compared with the initial QTL mapping results, fine mapping intervals are much narrower and have fewer candidate genes. These intervals are narrow enough so that there is almost no recombination happening within the interval. Therefore, the flanking markers of the fine mapped interval could be confidently used for marker assisted selection.

During the co-evolution of the insect and the host plant, host plant resistance could be overcome by new insect biotypes (Diehl and Bush 1984). PI 567597C and its derived lines have strong antixenosis resistance to the soybean aphid and are resistant to all four biotypes of the soybean aphid (Mensah et al. 2005; Alt and Ryan-Mahmutagic 2013). It would be an appealing addition to the resistance source pool. This study characterized the soybean aphid resistance gene *Rag3e* from PI 567597C, mapped and fine mapped *Rag*3e on chromosome 16 and enabled marker assisted selection for soybean aphid resistance breeding in soybeans.

Materials and Methods

Mapping population

Two mapping populations, 050107 and 050018, segregating for aphid resistance from PI

567597C, were used in this study. Population 050107 consisted of 250 $F_{4:9}$ recombinant inbred lines (RILs) derived from the cross of Skylla x PI 567597C. Population 050018 contained 94 $F_{3:6}$ RILs from the cross of E00003 x PI 567597C. E00003 and Skylla are elite cultivars that are susceptible to soybean aphids.

Alternative population

To confirm the aphid resistance QTL in a different genetic background, two alternative populations, 100049 and 100130 with 229 and 109 F3:4 lines, respectively, were used in the study. Populations 100049 and 100130 were derived from the crosses of E09933 x Skylla and E09933 x Titan, respectively. E09933 is a breeding line selected from the progenies of the cross of Skylla x PI 567597C and carries the aphid resistance from PI 567597C. Titan is an elite cultivar susceptible to soybean aphids.

Fine mapping population

Population 100049 and 100130 were validation populations for QTL study of aphid resistance gene from PI 567597C. Recombinants were selected form two hundred and twenty-nine and one hundred and nine F3:4 lines respectively.

F3:4 and F4:5 were selected phenotypically based on the aphid rating. The seeds of segregating lines, F4:5 and F5:6, were kept for planting in the greenhouse in fall 2012 and spring 2013, respectively. F6:7, F7:8, F8:9, F9:10, F10:11 and F11:12 were planted in the field summer 2013, in greenhouse fall 2013, in greenhouse spring 2014, in field summer 2014, in greenhouse fall 2014 and in spring 2015, respectively. Since F5:6, the selected plants were both phenotyped (aphid rating) and genotyped individually with SNP markers. Table 2.1 is a summary of the fine mapping populations.

Table 2.1: Fine mapping population information for PI 567597C. R stands for resistant parent; S stands for susceptible parent.

Population	Female Parent	Male Parent	Generation	# of Lines
100049	E09933^{R} (SkyllaxPI 567597C)	Skylla^{S}	F3:4	229
100130	E09933^{R} (SkyllaxPI 567597C)	Titan^{S}	F3:4	112

Aphid resistance evaluation

The planting conditions followed these described by Bales et al. (2013). In greenhouse trials, eight seeds per line were planted in a plastic pot with 105mm in diameter and 125 mm deep. The greenhouse temperature was maintained at 26/15^{0}C day/night with sodium vapor lights supplementing light intensity during the day (14h). In summer field trials, with 10 to 15 seeds, each line was planted in a single-row plot, 60 cm long with a row spacing of 60 cm. Aphid resistance was evaluated either in an aphid cage [a 12.2 x

18.3 m aphid and predator-proof polypropylene cage with 0.49-mm size mesh (Redwood Empire Awning Co., Santa Rosa, CA, USA)] in the summer or in the greenhouse without the cage in the fall and spring. Summer aphid colonies were collected by Dr. Christina DiFonzo from state-wide scouting every year. The colonies were kept and propagated in a small field cage. The greenhouse aphid colonies were kept in the greenhouse aphid room all year around. By the end of every summer season, aphid colonies in the field will also be moved to the greenhouse aphid room. For inoculation, small paint brushes were used to transfer the aphid from the original colonies to the testing soybean plants. Each plant was infested at the V2 stage with two wingless soybean aphids. Aphid resistance evaluation took place three or four weeks after the infestation. Each plant was rated with a zero to four scale (zero is resistant, four is susceptible) developed by Mensah et al.(2005, 2008). The phenotypes of plants within each line were converted into damage index (DI): the sum of the scale value times number of plants in each scale category, divided by 4 times the total number of plants and then times 100 [DI=Σ (Scale value × No. of plants in the category)/ (4× Total No. of plants) ×100] (Mensah et al. 2005). Rating value and damage index corresponding to number of aphid per plants is summarized in Table 2.2.

Table 2.2: Rating value and damage index corresponding to number of aphid per plants.

No. of Aphid/plant	Rating Value	Damage Index
0	0	0
<10	0.5	12.5
11-100	1.0	25.0
101-150	1.5	37.5
151-300	2.0	50.0
301-500	2.5	62.5
501-800	3.0	75.0
>800 [a]	3.5	87.5
>800 [b]	4.0	100.0

[a] More than 800 aphids per plant, plants stunted, leaves curled and slightly yellow, no sooty mold and few cast skins more than 800 aphids

[b] More than 800 aphids per plant, plant stunted, leaves severely curled and yellow, covered with sooty mold and cast skins.

Population 050107 was phenotyped in an aphid cage in the summer of 2008 and in the greenhouse in the spring of 2009. In the summer of 2008, population 050107 was planted in 2 replications and rated at both week 3 and week 4. The final rating scores for both week 3 and week 4 were the average of 2 replications. Population 050018 was planted in the pots in the greenhouse and rated in 2008. The F3:4 populations (population 100049 and population 100130) were planted in the field cage in 2012. Phenotyping work for population 050107 and population 050018 were conducted by Dr. Zhenyu Yang.

DNA extraction and marker analysis

For each line in QTL mapping populations, a 1cm^2 leaf tissue from all the fifteen plants were collected in a 15 ml tube (centrifuge tubes, Corning Inc.). In each tube, all the leaf tissues together was called a bulk collection. Leaf tissue of individual plants from the fine mapping populations was collected in 96 well-plate. After one day in -80 C° freezer, the samples were freeze-dried, then ground. The DNA extraction process was conducted following the CTAB method (Bales et al. 2013). The original DNA were diluted fifty times for PCR reactions. Simple sequence repeat (SSR) markers, the PCR reactions and gel running and viewing systems were used for population 050107 and 050018 (Bales et al. 2013). Kompetitive Allele Specific PCR (KASP) assays were developed (Semagn et al. 2013) to genotype population 100049 and 100130. The complete list of SNP markers

used in this study were listed in table 2.9 and table 2.10. KASP assays were run with 3 μL reaction system including 1.5 μL KASP master mix (KBiosciences, Herts, England), 0.045 μL of primer mix and 1.5 μL of 10-25 ng/μL genomic DNA. The PCR conditions for KASP marker assay were 95°C for 15 min, followed by 10 cycles of 95°C for 20 seconds and 65°C for 1 minute, then followed by 32 cycles of 95°C for 20 seconds and 58°C for 1 minute. For the KASP markers, the PCR reactions are either running in the Bio-Rad PCR machine (model C1000 touch, Bio-Rad Laboratories, Inc., USA) or Roche 480 light cycler (Roche Diagnostics, Germany). The fluorescent level of the final PCR products was measured and analyzed with the Roche 480 light cycler. KASP assays were developed based on the SNP information from the SoySNP50K iSelect SNP beadchip (Song et al. 2013) and the whole genome re-sequencing data (Bales 2013).

Parent DNAs were sent for whole genome SNP genotyping analysis with the SoySNP50K iSelect SNP beadchip (Song et al., 2013). DNA concentration was determined by the Quant-iT™ Picogreen® dsDNA Assay Kit (Invitrogen, USA) and quantified using BioTek Multi-Detection Microplate Reader (Biotek, USA). After normalized to 50 ng/ul. Each DNA sample was prepared for Infinium assay following manufacture’s protocol. GenomeStudio Genotyping module was used for data analysis.

KASP assay developed from the whole genome re-sequencing data

SNP markers MSUSNP16-44, MSUSNP16-132 and MSUSNP16-136 (Table 2.7, Table 2.9 and Table 2.10) were developed from the whole genome re-sequencing data.

Genomic DNA of PI 567585A and Skylla were prepared and sequenced by Carmille Bales following protocol of Carmille Bales (2013)

The SNP calling process was following SNP discovery pipeline from Bales (2013). The steps are:

1) Get rid of Illumina adapter sequences (fastx_clipper program) by using FastQC for quality control (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) then use FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/)
2) Map the cleaned short reads into the reference genome Williams 82 (Gmax_109 assembly data obtained from www.phytozome.com/soybean) with Bowtie version 0.12.7. The parameters used include using the paired-end mode for paired end reads and –v mode for single end reads. Only two mismatches were allowed for a read to map to the reference sequence. Reads that map only once (unique alignments) were processed for SNP calling.
3) Alignments for reads that mapped uniquely to the chromosomes were processed

using the sort, index, and pileup programs within SAMtools version 0.1.12a to generate unfiltered pileup files that are then filtered for quality using the varFilter option. The SAMTools varFilter parameters considered for high quality SNPs are: a) should at least have three read depth coverage (minimum), b) should at most have twenty read depth coverage (maximum), c) per base SNP quality should be more than twenty phred score (at least 1/100 error rate).

100-bp flanking sequences of the SNPs were obtained from the reference genome by sfetch program from hmmer-2.3.2 (http://hmmer.wustl.edu/, Eddy 1998).

Statistical and QTL analysis

The phenotypic distributions of F3:4 populations and recombinants in the fine mapping populations were compared with the ideal distributions by Chi-square test using Excel 2013 (Microsoft Office Professional Plus 2013).

Genotyping data of population 050107, population 100049 and population 100130 were used to construct the linkage map. All the linkage maps were constructed with a LOD score of 3.0 and Kosambi function using the software JoinMap 4.0 (Van Ooijen 2006). Single marker analysis (SMA) and composite interval mapping (CIM) were conducted

using QTL cartographer V2.5 (Wang et al. 2012). For the CIM analysis, a 1000-permutation test was used to determine the LOD threshold at the 5% probability level. The forward and backward regression method with a walking speed of 1cM was used when running the method. All the QTL maps were drawn by Mapchart (Voorrips 2002).

Candidate gene search and annotation

A candidate gene search was conducted using both soybean genome assemblies Glyma.Wm82.a1 and Glyma.Wm82.a2 (Schmutz et al. 2010; Grant 2015). Because the SNP position for Glyma.Wm82.a1 was kept throughout this study, the corresponding position of the two flanking markers can be used directly to search in Glyma.Wm82.a1 from soybase (http://www.soybase.org/gb2/gbrowse/gmax1.01/). To match the fine mapping interval into Glyma.Wm82.a2, two methods were used. For SNP developed from the SoySNP50K chip, the corresponding position in Glyma.Wm82.a2 is already labelled by Song et al. (2013). After the corresponding new name of the SNP was found, the new position in Glyma.Wm82.a2 is searched in soybase (http://www.soybase.org/gb2/gbrowse/gmax2.0/). For the SNP marker developed from SNP discovery, the primer sequence of the SNP was used to BLAST against the Glyma.Wm82.a2 version genome to locate the marker position (http://www.soybase.org/ NCBI Blast report, blastn). Candidate genes and their annotations from Glyma.Wm82.a1

and Glyma.Wm82.a2 were compared.

Results

Phenotypic analysis

The phenotypic values of population 050107, 100049 and 100130 and its parents are summarized in Table 2.3. In both field and greenhouse trials, the susceptible parents had severe damage (>501-800 SBA per plant). In contrast, the resistant parents always had fewer aphids (<11-100 SBA per plant). All the populations segregated for aphid ratings, which were suitable populations to conduct genetic studies such as QTL mapping. For the mapping population 050107, aphid evaluation has been repeated in three trials, 2008 week three, 2008 week four and 2009 week three. As shown in figure 2.1, the phenotypic correlations between these three trials were strong, indicating a stringent standard across all the ratings and a good performance of the aphid infestation trials. Using damage index, the phenotypic distribution of lines from mapping populations 050107 and 050018 was drawn in figure 2.2. The distributions were continuous and could be potentially divided into three categories as shown by the dotted lines in figure 2.2. Very likely, aphid resistance from PI 567597C was controlled by a single partially dominant gene.

Table 2.3: Aphid damage index for population 050107, 100049, 100130, the resistant parent and the susceptible parent.

Trials	Parents [a]		F4-derived lines			
	Resistant parent	Susceptible parent	Mean [b]	Range [c]	SD	SE
Population 050107	PI 567597C	Skylla				
Field 2008						
3-week rating	26.7a*[d]	78.1b	61.2	10.2-92.2	18.8	1.2
4-week rating	27.5a*	85.0b	58.7	25.0-98.0	23.3	1.5
Greenhouse 2009						
3-week rating	23.0a*	87.9b	57.4	12.5-96.9	27.0	1.7
Population 100049	E09933	Skylla				
Field 2012						
3-week rating	12.5a*	86.4b	51.2	12.5-100.0	22.8	1.7
Population 100130	E09933	Titan				
Field 2012						
3-week rating	25.0a*	85.9b	59.6	23.9-100.0	16.0	1.5

[a] parents' average mean

[b] mean of the population

[c] range of the phenotypic values

[d] The original rating values were converted into a damage index. Multiple comparison with letter 'a' and 'b' indicating significant differences ($P<0.05$)

* Significant different in row

Figure 2.1: Correlation between 2009 week three rating, 2008 week three rating and 2008 week four rating. X and Y axles are rating value for corresponding trials. Blue dots demonstrate rating value for each individual line in different trials. Red line labels the expected trend: y = x. A. Correlation between 2008 week three rating and 2008 week four rating. B. Correlation between 2008 week three rating and 2009 week three rating. C. Correlation between 2008 week four rating and 2009 week three rating.

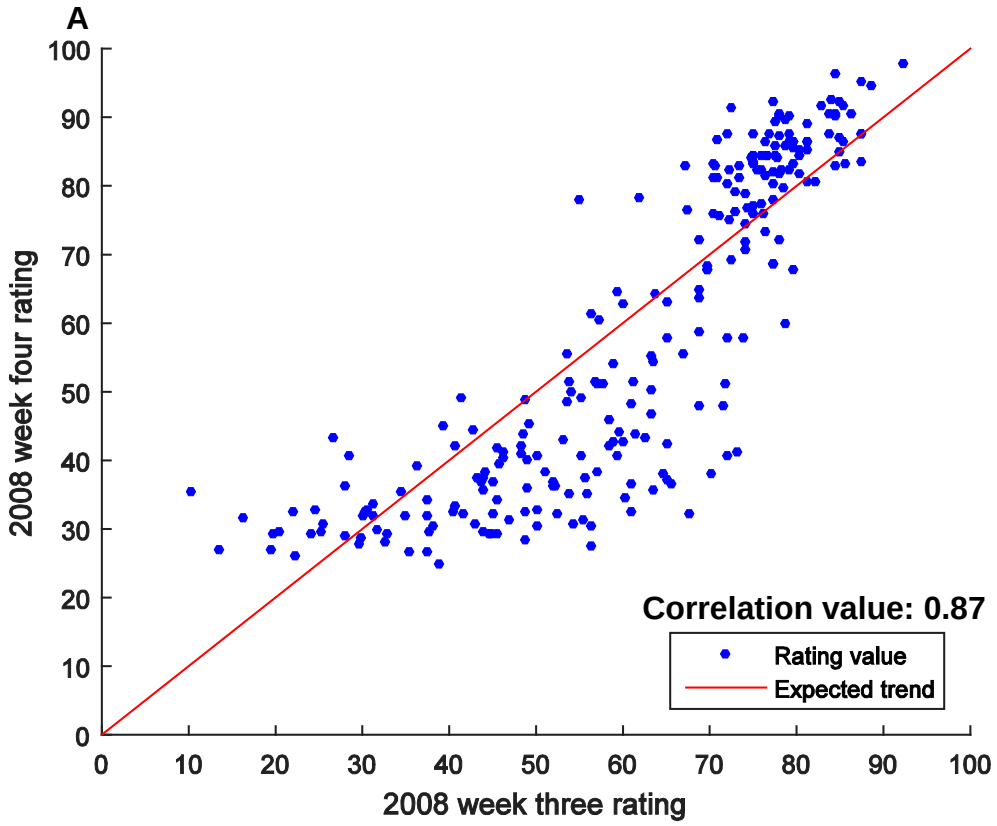
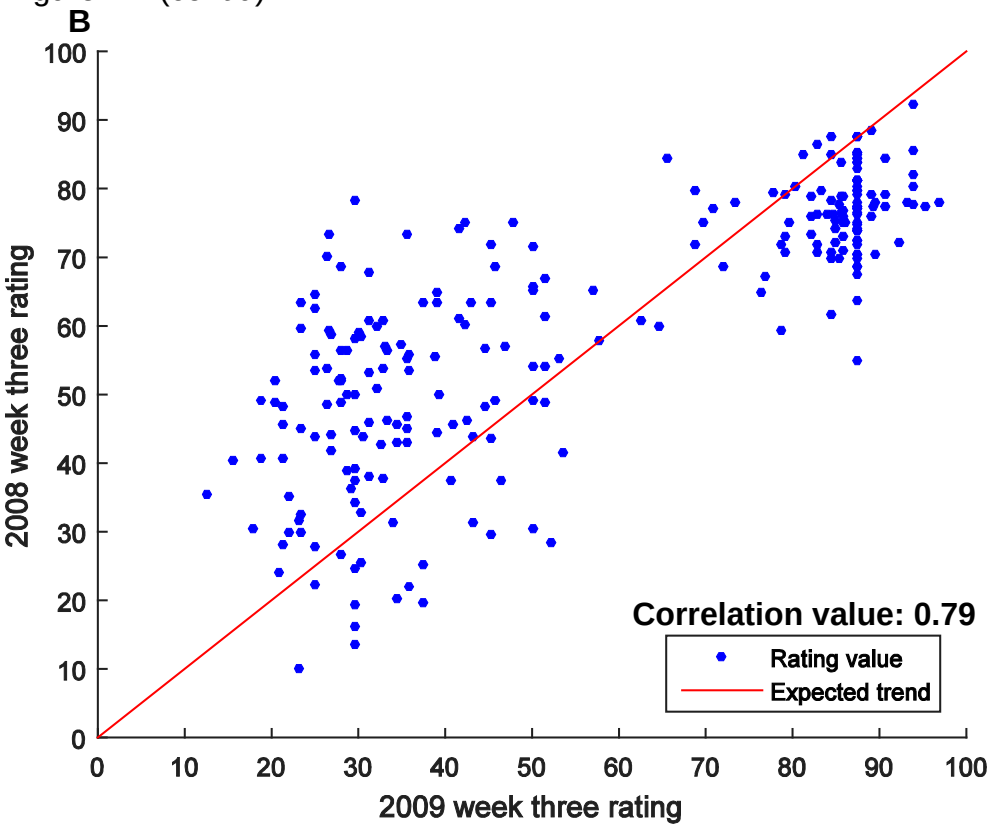


Figure 2.1 (cont'd)



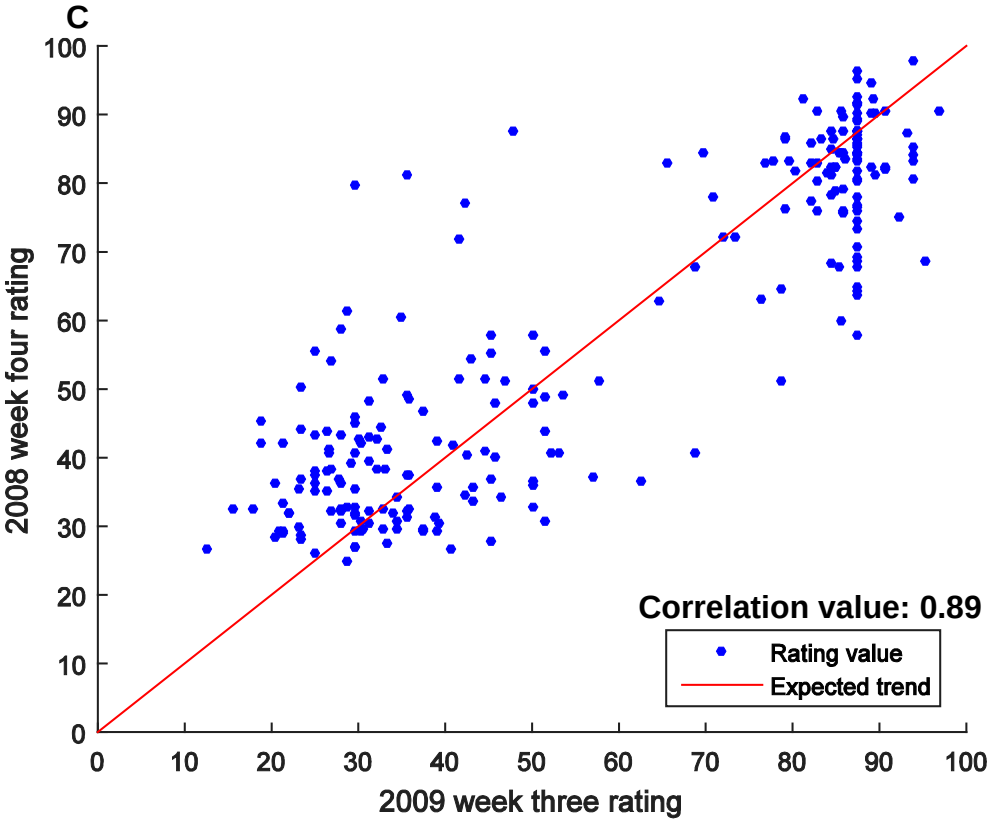
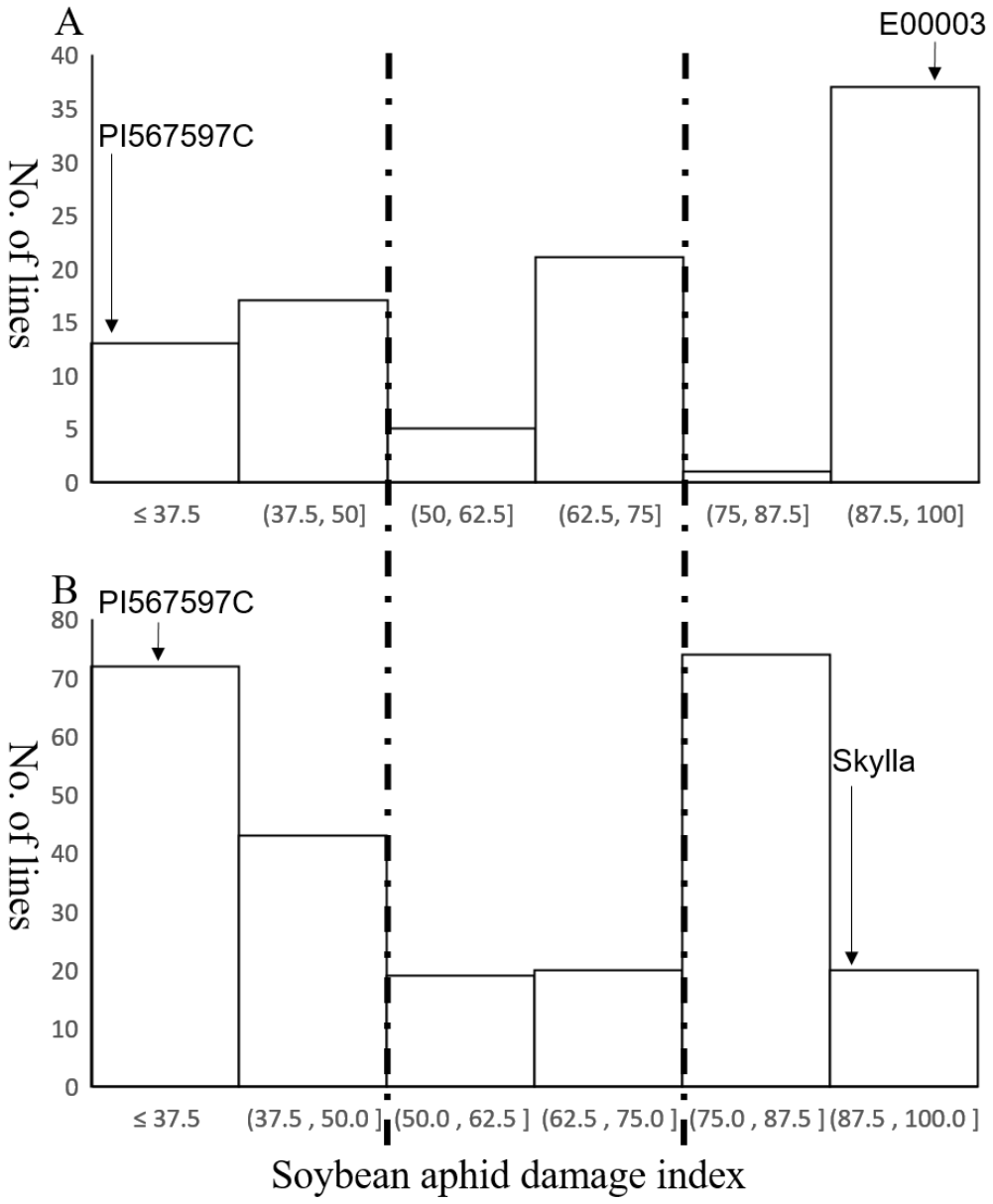


Figure 2.2**:** Frequency distribution of soybean aphid damage index. Parents are shown by arrows. Dotted lines indicated categorical boarders of damage index for resistant (0 - 50), segregating (50 - 75) and susceptible (75 - 100) lines. A. Mapping population 050018 (E00003 × PI 567597C, 94 lines in total) rated in the greenhouse, 2008; B: Mapping population 050107 (Skylla × PI 567597C, 249 liens in total) rated in the field, 2008.



To further investigate the inheritance pattern of the aphid resistance gene in PI 567597C, the damage index for each line from the mapping populations 050107 and 050018 were categorized into resistant (R, 0-50), heterozygous (H, 50-75) and susceptible (S, 75 -100), then, the actual phenotypic distributions were compared with the ideal distributions by Chi-square test. The expected ratio for a partially dominant gene to a certain generation was displayed in Table 2.4. All the *p* values were greater than 0.1, indicating the actual ratios did not significantly deviate from the expected ratio and the aphid resistance gene in PI 567597C followed the inheritance pattern of a partially dominant gene.

Table 2.4: Chi-square tests between the actual and ideal phenotypic distribution of populations 050018 and 050107.

Population ID	Resistant parent	Susceptible parent	Generation [a]	Expected ratio [b]	Total no. of plants [c]	Expected [d]			Observed [e]			*P* value of Chi-square test [i]
						R [f]	H [g]	S [h]	R	H	S	
050018	PI 567597C	E00003	F3 derived F6	3:2:3	94	35.3	23.5	35.3	30	26	38	0.532
050107	PI 567597C	Skylla	F4 derived F9	7:2:7	249	108.9	31.1	108.9	115	39	95	0.128

[a] The generation of each population when phenotyping and genotyping was conducted

[b] The expected segregation ratio of a certain generation

[c] Total number of plants within each population

[d] and [e] The expected and observed number of plants in each category (R, H, S)

[f], [g] and [h] Number of plants that been categorized into resistant, segregating, susceptible phenotypes respectively

[i] *P* value of the chi-square test results of comparing the observed and expected individual numbers, *p* value greater than 0.05 indicating insignificant difference

QTL analysis

From data of population 050107 (Figure 2.3, Table 2.5), a major QTL was identified on Chr16. Ten simple sequence repeat markers (SSR) spanning 100 cM were used to cover the whole chromosome. The marker order and marker positions were comparable to the Soybean Consensus Map 4.0 and marker order and positions from BARCSOYSSR_1.0 soybean SSR database (Song et al. 2010; Hyten et al. 2010). Using CIM, this QTL explained 70-90% of the phenotypic variation among three different trials (2008 week three, 2008 week four, 2009 week three). The peak QTL positions for the three trials were 27.01 cM, 27.01 cM and 26.01 cM, respectively. The flanking markers for the QTL were Satt249 and Satt596, which were constant among three trials. The consistency of phenotypic variation, QTL peak position and flanking markers among these trials indicated a reliable result (Table 2.5). Even though the peak position and the confidence interval for the SMA shifted from the CIM results, it was still in the same region. The additive effect of the QTL was negative, which meant that the allele from the resistant parent PI 567597C decreased the phenotypic value, the damage index. Thus, PI 567597C contributed the resistant allele. Because this QTL explained the majority of the phenotypic variation, aphid resistance in PI 567597C was controlled by a single gene. This aphid resistance gene in PI 567597C was designated as *Rag*3e.

To exclude other possible QTLs, 5 SSR markers on chromosome 7 for *Rag*1, and 2 SSR markers on chromosome 13 for *Rag*2, were used to genotype population 050018. As shown in Figure 2.4, none of them had significant peaks.

Table 2.5: Summary for aphid resistance locus, *Rag*3e, detected in the mapping and validation populations using the composite interval mapping method.

Trials	Chr/LG [a]	Peak pos. [b] (cM)	Flanking markers [c]	Genetic effect			
				LOD [d]	R^2 [e]	a [f]	d [g]
Population 050107							
Field 2008							
3-week rating	16/J	27.01	Satt249-Satt596	20.26	0.70	-16.28	-
4-week rating	16/J	27.01	Satt249-Satt596	44.29	0.86	-21.53	-
Greenhouse 2009							
3-week rating	16/J	26.01	Satt249-Satt596	54.60	0.90	-25.60	-
Population 100049							
Field 2012							
3-week rating	16/J	31.62	16-97 – 16-28	13.73	0.33	-15.83	-0.96
Population 100130							
Field 2012							
3-week rating	16/J	35.7?*+12	16-10 – 16-44	9.43	0.40	-12.05	2.48

[a] Chromosome and linkage group
[b] Peak position of the QTL, unit centiMorgan
[c] The right and left side markers next to the QTL peak
[d] The LOD score at the peak position of the QTL
[e] The percentage of phenotypic variation explained by the QTL
[f] Additive effect
[g] Dominant effect
*The question mark indicates an estimation of the genetic distance between the first marker and the beginning of the chromosome.

Figure 2.3: Map of chromosome 16 in mapping population 050107 with aphid resistance locus determined by both single marker analysis method and composite interval mapping method. 1-LOD and 2-LOD support intervals of each locus are marked by *thick* and *thin* bars, respectively. *Unfilled bars* represent loci for the week-three rating in the field cage trial. *Bars filled with parallel hatch lines* represent loci for week-four rating in the field cage trial. *Bars filled with diagonal hatch lines* represent loci for week-three rating in the greenhouse trial. Graph on the *right* shows the corresponding LOD scores. The solid arrow on the graph points out the LOD score threshold 3.0. The graph legend is presented in the box at the *rightmost*. *SMA* single marker analysis method, *CIM* composite interval mapping method, *08wk3* rating at 2008 three weeks after infestation, *08wk4* rating at 2008 four weeks after infestation, *09wk3* rating at 2009 three weeks after infestation

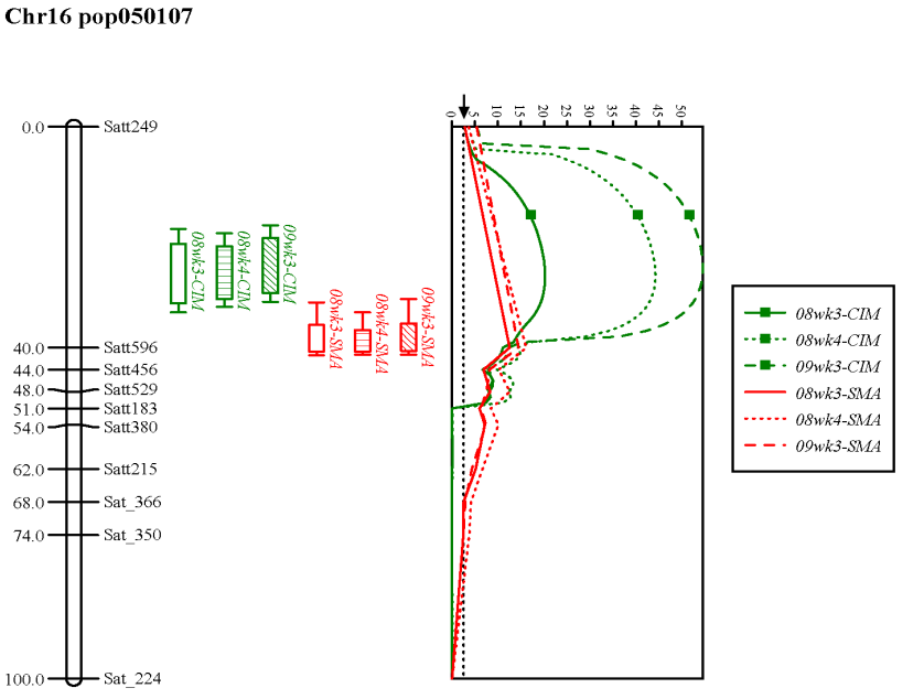
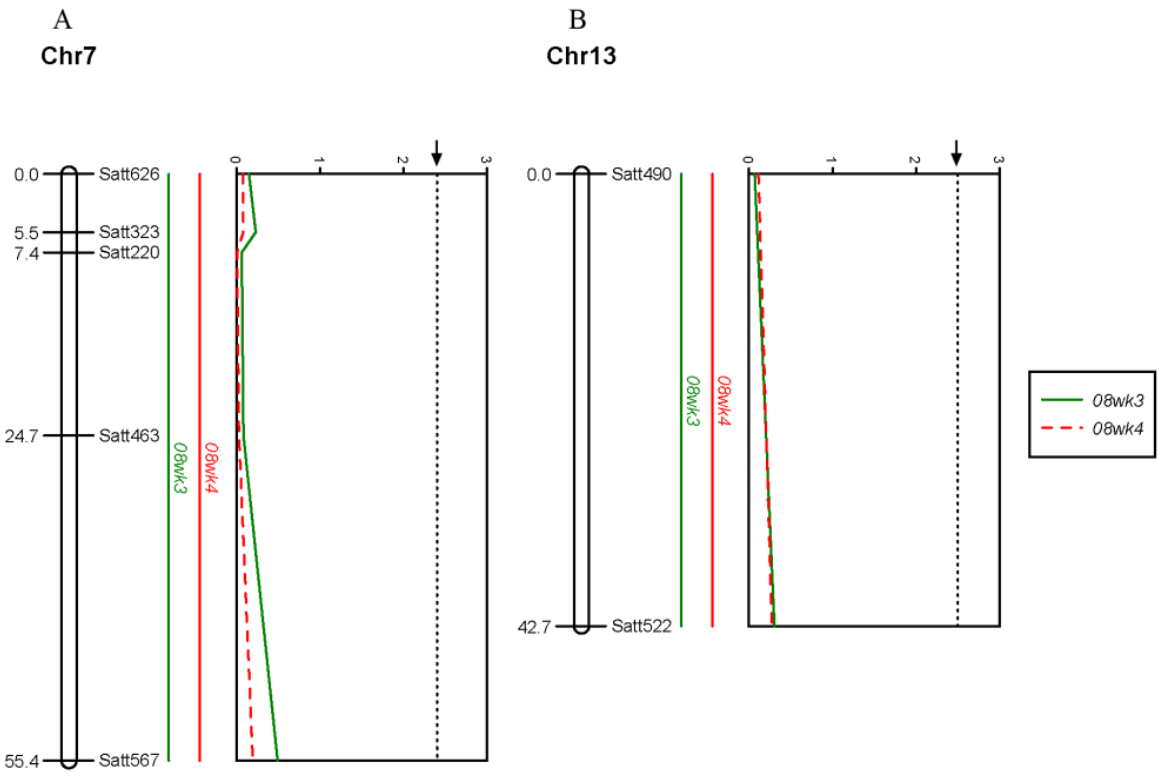


Figure 2.4: Map of chromosome 7 and chromosome 13 in mapping population 050018. Graph on the *right side* of each panel shows the corresponding LOD scores. The solid arrow on the graph points out the LOD score threshold 3.0. The graph legend for both panel A and B is presented in the box at the *rightmost*. The dotted line and solid line are overlapping in the graph of panel B for chromosome 13. Single marker analysis method was used to determine the association between aphid resistance value of the population and genotypic value on chromosome 7 and 13. As shown, there was no association between the marker value and the phenotypic value. *08wk3* rating at 2008 three weeks after infestation, *08wk4* rating at 2008 four weeks after infestation



QTL validation

Population 100049 and 100130 were validation populations with different genetic backgrounds than the QTL discovery population. Eleven single nucleotide polymorphism (SNP) markers spanning a 57.4 cM region and five SNP markers spanning a 32.8 cM region on chromosome 16 were used to genotype these two populations, respectively. Although both linkage maps were inflated, markers on both maps were in the same order as the consensus map. Markers used to genotype population 100049 covered the QTL interval of mapping results from population 050107. Markers used to genotype population 100130 overlapped, and were positioned nearby, the QTL interval of mapping results from population 100049. Figure 2.5 shows that for both populations, the confidence interval and the peak position of the QTL were overlapping using SMA and CIM. Also, the QTL was located within the QTL region detected by the mapping population 050107. This confirmed the major QTL *Rag*3e from PI 567597C on Chr16 is associated with aphid resistance.

All the QTLs conducted by CIM from different populations were drawn on one map in Figure 2.6. Clearly, this major QTL, *Rag*3e, on Chr16 in PI 567597C is a true QTL and it confers resistance in different genetic backgrounds (PI 567597C, Skylla and Titan).

Table 2.5 provides detailed information of the mapping results: 1) the peak position of the QTL was between 26 and 32 cM on Chromosome 16. 2) The QTL interval spanned a 29 cM region. 3) Negative additive effect indicated that the aphid resistance allele was contributed by the resistant parent, PI 567597C. 4) The peak position, the QTL interval and the additive effect in all the populations were close, illustrating a reliable result. 5) The F3:4 population had a lower R^2 and LOD value than the RIL population. 6) There was no dominant effect for the RIL populations.

Figure 2.5: Map of chromosome 16 in validation population 100049 and population 100130 with aphid resistance locus determined by both single marker analysis method and composite interval mapping method. 1-LOD and 2-LOD support intervals of each locus are marked by *thick* and *thin* bars, respectively. *Unfilled bars* represent loci resulted from composite interval mapping method. *Bars filled with parallel hatch lines* represent loci resulted from single marker analysis method. Graph on the *right side* of each panel shows the corresponding LOD scores. The solid arrow on the graph points out the LOD score threshold 3.0. The graph legend for both panel A and B is presented in the box at the *rightmost*. The dotted line and solid line are overlapping in the graph of panel B for population 100130. *SMA* single marker analysis method, *CIM* composite interval mapping method, *2012wk3* rating at 2012 three weeks after infestation

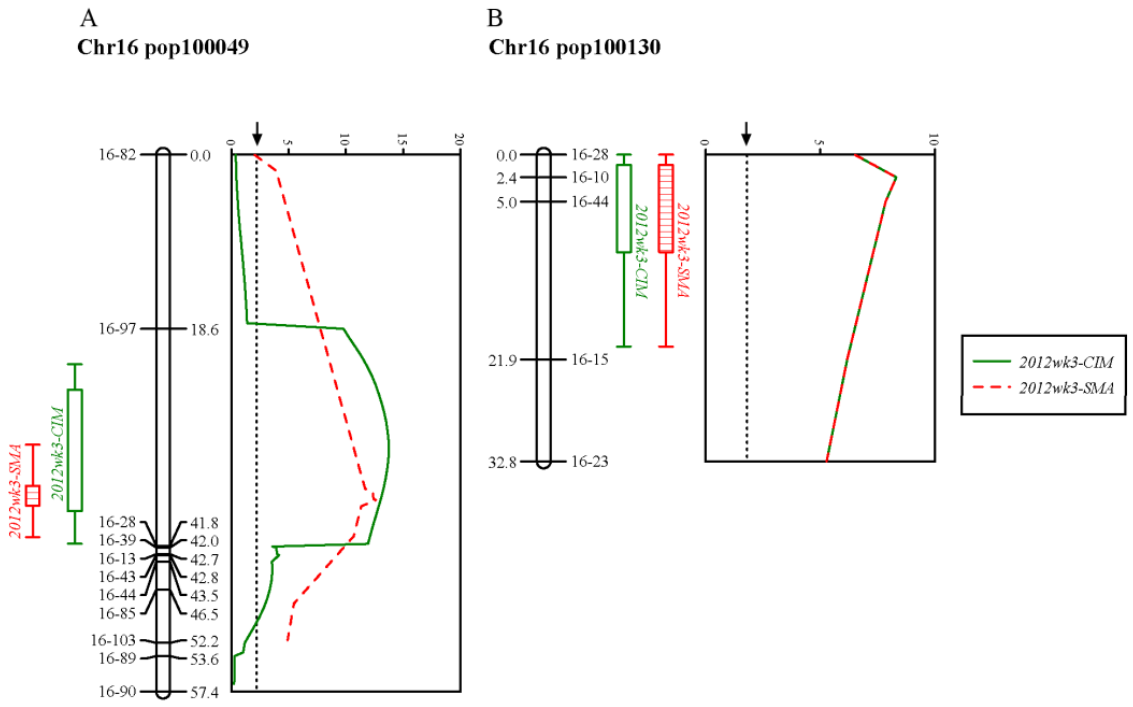
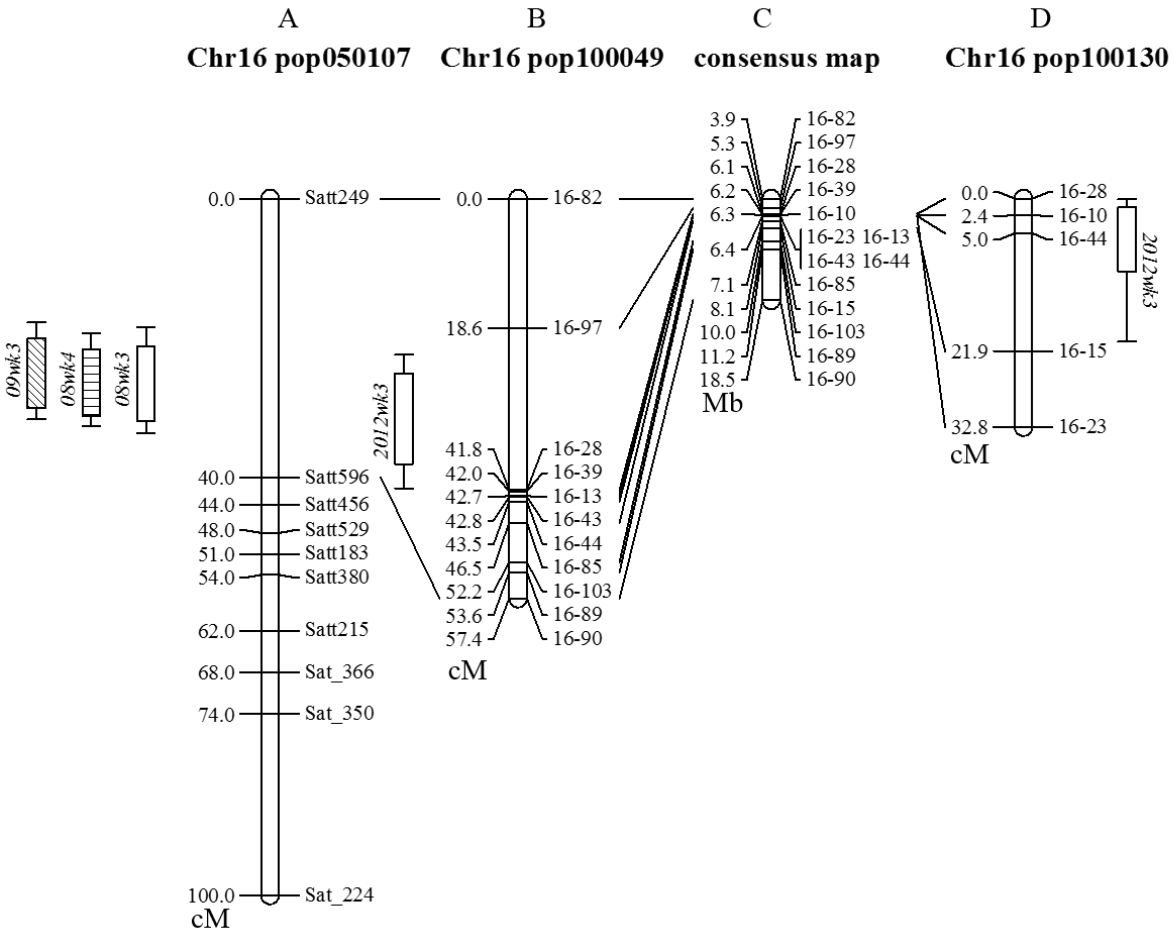


Figure 2.6: Locations of soybean aphid resistance loci, *Rag*3e, as determined using the composite interval mapping method. 1-LOD and 2-LOD support intervals of each locus are marked by *thick* and *thin* bars, respectively. *Unfilled bars* represent loci for the week-three rating in the field cage trial. *Bars filled with parallel hatch lines* represent loci for week-four rating in the field cage trial. *Bars filled with diagonal hatch lines* represent loci for week-three rating in the greenhouse trial. **a** Map of chromosome 16 from mapping population 050107 with the aphid resistance locus, *Rag*3e, shown on the *left*, unit centiMorgan; **b** Map of chromosome 16 from validation population 100049 with the aphid resistance locus, *Rag*3e, shown on the *left*, unit centiMorgan; **c** Map of chromosome 16 on the consensus map, unit Megabase; **d** Map of chromosome 16 from validation population 100130 with the aphid resistance locus, *Rag*3e, shown on the *right*



First round of fine mapping

Because *Rag*3e in PI 567597C was characterized as a partially dominant gene in this study, the progeny phenotypes of all the selected lines were categorized into resistant (R), segregating (H) and susceptible (S). The segregation ratio of each line was compared with the ideal ratio 1:2:1 by chi-square test. This information helped to distinguish whether the aphid resistance QTL was on the right or left side of the recombination breaking point. Detailed information was summarized (Table 2.6). The corresponding phenotype of each line determined by the progeny test was input in the second column with italic letters in Table 2.7. By comparing this with the phenotyping information, the side with candidate gene was identified by an arrow (Table 2.7).

Table 2.6: Chi-square test of progeny segregation ratio for recombinant lines in fine mapping study of PI 567597C

Generation [a]	Population [b]	line ID [c]	No. of progeny tested [d]	Aphid Phenotype [e]			1:2:1 ratio Chi-square Test *p* Value [i]	Segregation Pattern [j]
				R [f]	H [g]	S [h]		
F5:6	100049	28-6	12	12	0	0	<0.0001	F [k]
F5:6	100049	80-3	12	0	0	12	<0.0001	F
F5:6	100049	35-5	12	12	0	0	<0.0001	F
F5:6	100130	15-5	11	11	0	0	<0.0001	F
F5:6	100049	45-8	12	0	0	12	<0.0001	F
F5:6	100049	1-5	10	0	0	10	<0.0001	F
F5:6	100049	206-4	12	12	0	0	<0.0001	F
F5:6	100049	217-3	11	0	0	11	<0.0001	F
F5:6	100049	1-1	7	0	0	7	<0.0001	F
F5:6	100049	78-6	28	28	0	0	<0.0001	F
F5:6	100049	61-7	8	1	5	2	0.6873	S [l]
F5:6	100049	206-7	9	1	6	2	0.5427	S
F5:6	100130	101-2	46	46	0	0	<0.0001	F
F8:9	100049	147-2-5-79	10	10	0	0	<0.0001	F
F8:9	100130	52-1-30-13	10	10	0	0	<0.0001	F
F8:9	100130	52-1-30-17	9	9	0	0	<0.0001	F
F8:9	100049	45-4-5-1	8	8	0	0	<0.0001	F
F11:12	100049	63-1-5-30-5-3	8	0	0	8	<0.0001	F
F10:11	100049	38-5-3-11-4-2	10	0	0	10	<0.0001	F
F11:12	100049	45-4-4-5-3-2-18	7	7	0	0	<0.0001	F

[a] Generation of the recombinant lines

[b] The recombinant lines were selected from the listed population

[c] Line ID of the recombinant lines within their population

Table 2.6 (cont'd)

[d] The number of progeny rated for aphid resistance and used towards chi-square test
[e] Number of plants that has been categorized into R, H, S respectively
[f] Number of plants that were resistant to soybean aphids
[g] Number of plants that showed phenotype intermediate between resistant and susceptible to soybean aphids
[h] Number of plants that were susceptible to soybean aphids
[i] The *p* value of the chi-square test between the ideal segregation ratio 1:2:1 and the actual segregation for the progeny of each tested line. $\alpha = 0.05$
[j] The segregation pattern of each recombinant lines
[k] The progeny phenotype of the tested line is fixed (homozygous at the tested loci)
[l] The progeny phenotype of the tested line is segregating

Table 2.7: Recombination breakpoints among identified recombinants that mapped the position of *Rag*3e on Chromosome 16. Bold letters indicated the recombination breaking point. The arrow pointed to the side which the genotype agreed with the phenotype.

Line ID [a]	Pheno-type [b]	16-97 [g]	16-98	16-100	16-28	16-39	16-122	16-13	16-43	16-44*	16-124	16-128	16-132*	16-134	16-136*	16-110	16-112	16-85
Marker & Position [f] (bp)		5,259,121	5,555,122	5,809,541	6,079,769	6,214,642	6,260,278	6,424,067	6,431,101	6,438,676	6,484,276	6,517,204	6,571,636	6,624,879	6,721,743	6,774,822	6,868,110	7,070,805
28-6	*R* [c]	**S** [j] →	**R** [h]	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
80-3	*S* [e]	R	**R** →	**S**	S	S	S	S	S	S	S	S	S	S	S	S	S	S
35-5	*R*	S	S	**S** →	**R**	R	R	R	R	R	R	R	R	R	R	R	R	R
15-5	*R*		S	**S** →	**R**	R	R	R	R	R	R	R	R	R	R	R	R	
45-8	*S*	R	R	R	R	R	**R** →	**S**	S	S	S	S	S	S	S	S	S	S
1-5	*S*	S	S	S	S	S	S	S	S	S	S	**S** ←	**R**	R	R	R	R	R
206-4	*R*	R	R	R	R	R	R	R	R	R	R	R	R	**R** ←	**S**	S	S	S
217-3	*S*	S	S	S	S	S	S	S	S	S	S	S	S	**S** ←	**R**	R	R	R
1-1	*S*	**H** [i] →						**S**	S	S						S		S
78-6	*R*	**H** →						**R**	R	R						R		R
61-7	*H* [d]	H						H	H	**H**					←	**R**		R
206-7	*H*	H						H	H	**H**					←	**S**		S
101-2	*R*	R						R	R	R						**R**	←	**H**
147-2-5-79	*R*						H	**H** →				**R**		R				
52-1-30-13	*R*						R	**R**				← **S**		S				
52-1-30-17	*R*						R	**R**				← **S**		S				
45-4-5-1	*R*						R	**R**				← **H**		H				
63-1-5-30-5-3	*S*							S			**S**	← **R**						
38-5-3-11-4-2	*S*							**S**			← **R**	R						
45-4-4-5-3-2-18	*R*							**R**			← **S**	S						

Table 2.7 (cont'd)

*Markers developed through whole genome sequence SNP discovery pipeline
[a] Line ID of the recombinant lines within their population
[b] The corresponding phenotype (tested through progeny test, Table 2.6) of each recombinant lines
[c] The progenies of the recombinant line were resistant to soybean aphids.
[d] The progenies of the recombinant line were segregating for aphid resistance.
[e] The progenies of the recombinant line were susceptible to soybean aphids.
[f] The upper portion of the row displayed the marker name. The lower portion of the row presented the physical position of each marker on chromosome 16. The unit was in base pair (bp)
[g] For convenience and saving space, "MSUSNP" which was the first part of every marker's name, was not included in the table.
[h] Both SNP alleles of this loci was from the resistant parent.
[i] One SNP allele of this loci was from the resistant parent and another one was from the susceptible parent.
[j] Both SNP alleles of this loci was from the susceptible parent.

SNP marker MSUSNP16-97 (5,259,121bp) and MSUSNP16-28 (6,079,769bp) for population 100049, MSUSNP16-10 (6,262,227bp) and MSUSNP16-44 (6,438,676bp) for population 100130 were flanking markers next to the aphid resistance QTL peak, respectively. To cover a larger interval to reduce the probability of missing the candidate genes, MSUSNP16-97 (5,259,121bp) for population 100049 and MSUSNP16-28 (6,079,769bp) for population 100130 as the left border marker, MSUSNP16-85 (7,070,805bp) for population 100049 and MSUSNP16-15 (8,051,585bp) for population 100130 as the right border marker, were used for the first round of fine mapping. Three hundred and ninety-nine F5:6 plants from population 100049 and two hundred and eight F5:6 plants from population 100130 were genotyped with the two border makers and a middle marker, MSUSNP16-44 (6,438,676bp). Forty-four and thirty F5:6 recombinant lines, four and seven F5:6 heterozygous lines were selected from population 100049 and population 100130, respectively. The F6:7 progenies of these lines were both phenotyped and genotyped. Also, more markers in between the border markers and the middle marker were used to genotype the selected F5:6 recombinants to saturate the QTL region. (Table 2.7)

For lines 28-6, 35-5, 15-5 and 78-6, their progenies were all resistant to the soybean aphid. Correspondingly, genetic materials derived from the resistant parent were at the right side of the recombination breaking point, which were labelled with ‘R’. The aphid

resistance gene should then reside right of the breaking points within ‘R’ labelled region. Similarly, for line 80-3, 45-8 and 1-1, whose progenies were all susceptible to the soybean aphid, the aphid resistance gene would locate right of the breaking points with ‘S’ labelled region. From all these results combined, the left border of the QTL was narrowed from MSUSNP16-97 (5,259,121bp) to MSUSNP16-98 (5,555,122bp) to MSUSNP16-100 (5,809,541bp) to MSUSNP16-122 (6,260,278bp). The progenies of line 206-4 and 101-2, line 1-5 and 217-3, line 61-7 and 206-7 were resistant, susceptible and segregating to the soybean aphid, respectively. The matching genotypes of these lines were all at the left side of the breaking points. These results pushed the right border of the QTL from MSUSNP16-85 (7,070,805bp) to MSUSNP16-110 (6,774,822bp) to MSUSNP16-136 (6,721,743bp) to MSUSNP16-132 (6,571,636bp). To sum up, after the first round of the fine mapping, the QTL region was narrowed from a 1,811kb region between MSUSNP16-97 (5,259,121bp) and MSUSNP16-85 (7,070,805bp) to a 311kb region betweenMSUSNP16-122 (6,260,278bp) and MSUSNP16-132 (6,571,636bp).

Second round of fine mapping

After the first round of fine mapping, SNP marker MSUSNP16-122 (6,260,278bp) and MSUSNP16-132 (6,571,636bp) became the flanking markers for the second round of fine mapping. Marker MSUSNP16-122 (6,260,278bp), MSUSNP16-13 (6,424,067bp), MSUSNP16-128 (6,517,204bp) and MSUSNP16-134 (6,624,879bp) were used for

genotyping. Based on both genotypes and phenotypes, one hundred and thirty-one and twenty-four recombinant lines were selected from a total of one thousand and forty-six F8:9 plants from population 100049 and 100130, respectively. The progenies of these recombinant lines were planted in the field in the summer of 2014. They were both phenotyped and genotyped.

The progenies of line 147-2-5-79 were resistant to soybean aphids. For this line, the genotypes on the right side of the recombination breaking point, right side of marker MSUSNP16-13, matched with the phenotypes. The progenies of line 52-1-30-13, 52-1-30-17 and 45-4-5-1 were resistant to soybean aphids. For these three lines, the genotypes on the left side of the recombination breaking point, left side of marker MSUSNP16-128, matched with the phenotypes. Through this round of fine mapping, the boundary of the QTL was refined to a 93kb region between MSUSNP16-13 (6,424,067bp) and MSUSNP16-128 (6,517,204bp).

Third round of fine mapping

Besides the new flanking markers, MSUSNP16-13 and MSUSNP16-128, MSUSNP16-124 was added as a middle marker for the third round of fine mapping. One hundred and twenty-one F9:10, F10:11 and F11:12 plants from both populations were genotyped and phenotyped. Fifty-two recombinant lines were selected for progeny tests conducted in

summer of 2015.

Line 63-1-5-30-5-3 had progenies that were susceptible for soybean aphids. The matching genotypes of this line, ‘S’, confirmed the left and right QTL border refined by the second round of fine mapping. Progenies of line 38-5-3-11-4-2 and 45-4-4-5-3-2-18 were susceptible and resistant to soybean aphids, respectively. The corresponding ‘S’ and ‘R’ genotypes for these two lines were all on the left side of marker MSUSNP16-124. This helped to delimit the QTL into a 60kb region between MSUSNP16-13(6,424,067bp) and MSUSNP16-124 (6,484,676bp).

All the SNP marker positions, including those from the SoySNP50K chip, were designed based on the Williams 82 soybean genome version Glyma.Wm82.a1. Since the Williams 82 genome has been updated into Glyma.Wm82.a2, the sequence of marker MSUSNP16-13 and MSUSNP16-124 were used to BLAST against Glyma.Wm82.a2. The QTL candidate region shifted 147 kb towards the centromere with no inflation in between (60 kb). Nine genes with predicted gene model or functions were located in the candidate region (Table 2.8).

Table 2.8: Genes with predicted gene model in *Rag*3e QTL interval

Locus Name	**Physical Position** (bp, Wm82.a2)	**Database ID**	**Annotation Description**
Glyma.16g066000	6570689-6574950	AT5G40660.1	ATP12 protein-related
		GO:0043461	proton-transporting ATP synthase complex assembly
		KOG3015	F1-ATP synthase assembly protein
		PTHR21013	ATP SYNTHASE MITOCHONDRIAL F1 COMPLEX ASSEMBLY FACTOR 2/ATP12 PROTEIN, MITOCHONDRIAL PRECURSOR
		PF07542	ATP12 chaperone protein
Glyma.16G066100	6575317-6575700	N. A.	N. A.
Glyma.16G066200	6576967-6580036	AT5G40650.1	succinate dehydrogenase 2-2
		GO:0009055	electron carrier activity
		GO:0051536	iron-sulfur cluster binding
		KOG3049	Succinate dehydrogenase, Fe-S protein subunit
		PTHR11921	SUCCINATE DEHYDROGENASE IRON-SULFUR PROTEIN
Glyma.16G066300	6587408-6590067	AT3G46610.1	Pentatricopeptide repeat (PPR-like) superfamily protein
		PTHR24015	FAMILY NOT NAMED
		PF01535	PPR repeat
Glyma.16G066400	6592258-6601109	AT5G40640.1	N. A.
Glyma.16G066500	6600655-6600822	N. A.	No Annotation Available
Glyma.16G066600	6606039-6612286	AT2G25940.1	alpha-vacuolar processing enzyme
		GO:0004197	cysteine-type endopeptidase activity
		GO:0006508	proteolysis
		KOG1348	Asparaginyl peptidases
		PTHR12000	HEMOGLOBINASE FAMILY MEMBER

Table 2.8 (cont'd)

		PTHR12000:SF2	
		PF01650	Peptidase C13 family
Glyma.16G066700	6620330-6621988	AT5G14360.1	Ubiquitin-like superfamily protein
		GO:0005515	protein binding
		PTHR10666	UBIQUITIN
		PTHR10666:SF85	UBIQUITIN-LIKE PROTEIN 4A
		PF00240	Ubiquitin family
Glyma.16G066800	6627025-6628243	AT1G58190.2	receptor like protein 9
		GO:0005515	protein binding
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF474	PROTEIN KINASE DOMAIN-CONTAINING PROTEIN, CYTOPLASMIC
		PF00560	Leucine Rich Repeat

Discussion

In this study, the aphid resistance gene from PI 567597C was characterized as partially dominant through an inheritance pattern study. A major QTL, *Rag*3e, on chromosome 16 was detected across different years and genetic backgrounds. Then, fine mapping using residual heterozygous lines and SNP markers delimited the QTL region from 29cM (approximately 6Mb) to 60kb. *rag*3 from PI 567598B and *Rag*3b from PI 567537 localized up stream (left side) of *Rag*3e. *rag*3 was a recessive gene. *Rag*3b was a dominant gene. Both *rag*3 and *Rag*3b carried antibiosis resistance to aphid biotype 1, 2 and 3. Comparing *Rag*3e with *rag*3 and *Rag*3b, the QTL interval, the inheritance pattern, the resistance type and the resistance to aphid biotypes were distinctively different. Very likely, *Rag*3e is a unique aphid resistance gene (Bales 2013; Zhang et al. 2013, Table 1.1, Chapter III). It is also possible that this gene is allelic with other *Rag*3*/rag*3 genes. The variations in different alleles caused the phenotypic differences. Interestingly, fine mapped *Rag*3e interval overlapped with the fine mapping QTL region of *Rag*3d from PI 567585A (Chapter III). The same as *Rag*3e, *Rag*3d was also a partially dominant gene that resistant to all four biotypes of soybean aphids. Its resistance to the soybean aphid has been confirmed in different genetic background as well. However, *Rag*3d (PI 567585A) confers antixenosis and antibiosis resistance to the soybean aphid whereas *Rag*3e (PI 567597C) only controlls antixenosis resistance. It is unknown if *Rag*3d and *Rag*3e are allelic with

each other or if they are different genes closely linked. (Liu 2010, Table 1.1)

If comparing the power of detecting QTL among all the populations, for a trait with relatively low heritability, such as aphid resistance, a population with advanced generations (mapping population 050107 and 050018) was more helpful in identifying the major QTL and its effect. The genetic map conducted through a higher generation population had less inflation and was more close to the consensus map. Even though, less powerful, F3:4 generation population (validation population 100049 and 100130) could be used in confirmation of the identified QTL. Also, a population size around two hundred would be sufficient for QTL mapping studies of aphid resistance in soybean.

We used SSR and SNP markers that were operating on totally different platforms in this study. The SSR markers were cheaper and provided a wide range of polymorphism. However, it was time consuming. The SNP markers and its platform may cost more and the markers are di-allelic. But it was high through-put, easy to quantify and had fewer missing data. As the technology quickly advances, there may be cheaper and more user-friendly platforms (Yuan and Wen 2014; Du and Wang 2015). Compared with genotyping, phenotyping is more time consuming and more subjective to quantify, especially for disease and insect resistance. For traits like plant height and leaf area, there are already automatic quantifying method available (Fanourakis et al. 2014). For aphid resistance,

high through-put phenotyping technologies are relatively difficult to develop and would be desirable in the future.

SNP markers used in this study were either developed from the SoySNP50K chip or the whole genome next generation sequencing (NGS) data. There was a significant difference when using data from these two data sets. Even though the NGS data was pre-filtered, when performing polymorphic test, only 30% of SNP makers designed through the NGS data would be polymorphic. This may be due to the low coverage and high error rate of the NGS data. On the other hand, because the SoySNP50K chip was carefully selected, filtered and evaluated, 80% of the SNP markers designed through the SoySNP50K chip were polymorphic (Song et al. 2013). However, since the SoySNP50K chip was aimed to cover the whole genome for a diverse genetic pool, there may not be enough polymorphic markers between two designated parents within a specific region. The NGS data would be a great source to cover the gap. In fact, for fine mapping studies, the narrower the QTL region, less polymorphic markers can be found from the SoySNP50K chip. It would be more important to develop SNP markers from the NGS data.

Aphid resistance genes have been identified in other species rather than soybean. There are tomato gene *Mi* that is resistant to potato aphid (Milligan et al. 1998; Rossi et al. 1998),

the melon gene *Vat* that is resistant to melon aphid (Chen et al. 1997; Pauquet et al. 2004) and the *AKR* gene that is resistant to blue-green aphids in *Medicago truncatula* (Guo et al. 2009). They were all nucleotide-binding site leucine-rich repeat (NBS-LRR) genes (Milligan et al. 1998; Ohnishi et al. 2012; Bales 2013; Dogimont et al. 2014). In soybean research, soybean cyst nematode (SCN) resistance gene *Rhg1* has been well characterized. The resistance in *Rhg1* was conferred by copy number variation of three tandem genes. After studying forty-one diverse soybean accessions, the resistant accessions were grouped into low and high copy number variation groups that actually come from two distinct ancestors. DNA methylation was also found correlated with the soybean cyst nematode resistance (Cook et al. 2012; Cook et al. 2014).

There are nine annotated genes in the *Rag*3e QTL interval. Seven of them could be candidate genes for *Rag*3e. Glyma.16g066000 was annotated as ATP12 protein related. ATP12 protein is a group of ATPase F1F0-assembly proteins that locate in mitochondria. They are essential for the assembly of the mitochondrial F1-F0 complex for phosphorylation. There were studies showing ATPase or protein involved with phosphorylation participating plant response to biotic stress. One of the three tandem repeat genes in *Rhg1* for soybean cyst nematode (SCN) was an ATPase (Cook et al. 2012). Early phosphorylation served as signals in signaling pathways during plant response to biotic stress (Peck 2003). Thus, Glyma.16g066000 could be a candidate

gene for *Rag*3e.

There was no annotation for Glyma.16g066100. Glyma.16g066200 was annotated as succinate dehydrogenase (SDH). Succinate dehydrogenase was a crucial parameter in plant response to abiotic stress. Jardim-Messeder et al. (2015) demonstrated that SDH was a direct source of Reactive oxygen species (ROS) in *Arabidopsis thaliana* and *Oryza sativa*, and the induction of ROS production by specific SDH inhibitors impaired plant growth. No direct evidence has been shown for succinate dehydrogenase involved in plant response to biotic stress such as insect resistance. But the possibility cannot be ruled out.

Glyma.16G066300 is predicted to be a pentatricopeptide repeat (PPR-like) superfamily protein that binds one or several organellar transcripts in mitochondria or chloroplasts. They target a specific RNA sequence to alter the sequence, the turnover, the processing, or the translation of the transcripts. Because these transcripts are translated into proteins of different functions, each PPR protein may be involved in totally different biological pathways. There were studies showing PPR proteins involved in photosynthesis, leaf development, leaf pigmentation, seed or embryo development, growth of the plant, pollen fertility and ABA response. Zsigmond et al. (2012) reported that the overepxression of the mitochondrial *PPR40* gene improves salt tolerance in Arabidopsis (Barkan and Small

2014). Thus, Glyma.16G066300 could be a very good candidate gene for *Rag*3e.

Glyma.16G066400 is annotated using the Arabidopsis gene AT5G40640.1. AT5G40640.1 has an unknown function. It is expressed in multiple developmental stages and multiple tissues. The interesting aspect of AT5G40640.1 is the subcellular localization. It localizes in chloroplast and plasma membranes. It could be a trafficking protein working in the secretory pathways or a protein in the hormone or disease signaling transduction pathways.

Glyma.16G066600 is characterized as an alpha-vacuolar processing enzyme. It's homolog in Arabidopsis, AT2G25940.1 was up-regulated by wounding treatment, ethylene and salicylic acid (Kinoshita et al. 1999). This suggests that Glyma.16G066600 may be one of the pathogen responsive genes downstream of the pathogen response signaling pathways.

Glyma.16G066700 is related to a ubiquitin-like superfamily protein, AT5G14360.1, in Arabidopsis. Ubiquitin has a very important role in protein degradation. Through E1 to E2 to E3 cascade, the length of the ubiquitin poly tail determines the fate of the protein. Undoubtedly, ubiquitin is involved in many biological processes, like plant growth and development, signaling of hormones, the cell cycle, and circadian clock etc. Fascinating

enough, during pathogen response, E3s from the host plant can ubiquitylate and remove the pathogenic factors inside the plant. Correspondingly, pathogen would inject inhibitors to disrupt plant E3s's functions. Hence, ubiquitin could be key factors in defense response (Schwartz and Hochstrasser 2003; Vierstra 2012).

Glyma.16G066800 is a leucine-rich repeat receptor-like kinase (LRR-RLK) that would be possibly involved in defense response. LRR-RLK acts at the front line of disease resistance by interacting with the pathogen with leucine-rich repeat at the surface of plant cell. The signal then would be transmitted through the kinase domain down to the disease response cascade (Sekhwal et al. 2015).

To sum up, *Rag*3e in PI 567597C was a precious source of soybean aphid resistance. This study discovered a major aphid resistance gene, *Rag*3e, from PI 567597C and fine mapped the QTL region to a 60kb interval with seven candidate genes. Flanking markers from fine mapping have already been used in a marker assisted selection breeding procedure at Michigan State University. This would help incorporate *Rag*3e into elite cultivars together with different resistant sources to gain a durable and long lasting aphid resistance. Further study of R*ag*3e*,* for example molecular cloning, would help to characterize aphid resistance mechanisms in soybean.

APPENDIX

APPENDIX

Table 2.9 SNP markers used in PI 567597C QTL mapping study. SNPs in corresponding wild-type and mutant-alleles are in brackets [ ].

SNP assay ID	SoySNP50K chip ID	Chromosome	Physical position (bp) [a]	Target sequence [b]
MSUSNP16-10	Gm16_6262227_C_T	16	6,262,227	5'GTTGGGCTATGTCCAAAATAGTATCCCCATTAGT TAGTATCCCATGATGTCATGAGGTGTAAACTTGTTA AGACATATCAAACTTAGGGTTTAAGTTAAC[C/T]A GATCCGAAAAAGCTGCCACTATAGTGCCTTCTCTTT GAGTATGTGGTAATTATTGATTGAAGGCTTGATTGA AGGATCATCCTCATAGCTTAGGTTTTG -3'
MSUSNP16-13	Gm16_6424067_A_G	16	6,424,067	5'CTCACTCGTTGATAAGAAAATGCATAAAACCTGC AACCCTCAACTTCCTGACACCACTCGCAGTCCCTGA GATTCGGCGGCGGCTAGCGTCGGTGGCGGC[A/G]G CGGCGGACGAGGACCCTCCGCAATCGCCGTCGTCGT TCACTTTCTCGTCGGAGGGGGAGAAGGTGTACGTGA AGCCGCCGGCGGGGAAGTGGAAGCCGT -3'
MSUSNP16-15	Gm16_8051585_T_C	16	8,051,585	5'CATATTTAACATTATTCCTCAATCATGAACGGTA CTTATCTCCGTTTCATGTGTTTCACAATATCCTTAT ACTTAGAGCTATCAAAATGGGTCAGCCCGG[T/C]C

Table 2.9 (cont'd)

MSUSNP16-23	*	16	6,355,235	TACATGGGCTGACCCGCAACGGGTTGAGCTAAAAGT GGGCTAGTCCAGCTCGGCTCACTTTTGTGTGGGCTA ATAAAATGCCAGCCTGACCCAGTCCAC -3' 5'AGGATCTTACCTTTCTGATTCAGATCCAACCCCT AAAATAACTTTTGCATACAAATACTACTCGTGAATT ATGCAATACCCACGGTCTTACACTTATTTT[T/C]A AAACACATTTAACCCAATGCATTACAATTTAACTCC TCAGGTTCTTAACTTGGAACCGTACACTCTTCCTTT AACACTTCTCGCATTGCACTACAATTT -3'
MSUSNP16-28	Gm16_6079769_A_G	16	6,079,769	5'CATAGAGGGCTTGAGCGATGTCTTGATCGTTGAC CGAAAGTTGTGATCAAGTGTGGTAGTGTACGTCATC TCTCTCAGTTCCCCCACGATTCCTAATAAC[A/G]C ATCAATATTCTTCTCTTTTGAAACACACCAATTATA TATATTTTCTTTTCTTTTTATTCATTTTCTTTCATC GTCATAATTTTAATTTTTTTATCTAAA -3'
MSUSNP16-39	Gm16_6214642_C_T	16	6,214,642	5'ACACGATTGAAGAAAATTGAAAAAGAAATACTAC TACTATTGAAGAAAGTTGAAAAAGAAATACCAGTAC CTTGTCCATTTCTGCTTCTTTTGCGGGTTG[C/T]G GAAAGTGTTCCAATTTAATGCTGGTCCTAAGTCCTA ACCAAGTTATCAAGATTCAAGCCGTGGCTTGCAGGT AATATTTAACTCTCTGTTTAGACTTTA -3'
MSUSNP16-43	Gm16_6431101_A_C	16	6,431,101	5'GCATCAAAGAATGTATTAACAATACAATGACAAA CATAGAATTCAGCAAAGAGCTCTCTTATACTAGCTG AAGTACAAAGCATAGCACCAAGAACAGCAG[A/C]T GAGATATTATGCTAGTATTAACAAGTAAATCAACAA AACACCATGCATATATCATGGCAGGTGTGAACTATA TAATCCTTTCTCAACCCAAGCTTCAAA -3'

Table 2.9 (cont'd)

MSUSNP16-44	*	16	6,438,676	5'CCTTAAGAAGGATTCTCAAAAGTTTACTTTTAGC TCCAACAAGACATGTTCTTACATCTAAGCCCAACCA AACAAAAATAGAAAAACCAAATTTTAAATT[T/C]T TTATTATCAACCTCATGATCACCATGTCTACCACGA TTTATCCATGGTTGTGTTTGGTTATCAATTTTAGCT TTTTCATCAATTTTGGTTAATAATTTT -3'
MSUSNP16-82	Gm16_3912739_T_C	16	3,912,739	5'TCAATGAGACTTACCTTCTCTTCTGGTGTTATAT GTCTTATAATGGTCGTAGCTATCAAGTTGGTAAATC CCGAGCAATCACCCACTACGAACATGTGAG[T/C]A TGGTGGGCATTGTACAGTACACAAGCCCTCCAAACT ATGAGTCTGATTTGGATGAAAGAGCTTCATGATGAT GGATGTTTCAACAAAGATGGCCGAGTT -3'
MSUSNP16-85	Gm16_7070805_G_A	16	7,070,805	5'ATGCAAGGGAAGCAGCTGCAAGAGATGCAAGGGA TGCAAAGGTGGAGGCGAGAGATGTAAAGAGAACAAC AGTGACAGCAACAACCGCAACCGCATGAAC[G/A]T GATGAGTATTAATGTGTTGTTATGAACTTATGATGT TGGTTTATGTGGGGAAATAAATGATGTATGTACCTC TTCTTGCCTATGTAGTAGGTTTGGGTG -3'
MSUSNP16-89	Gm16_11194677_C_T	16	11,194,677	5'TATGCTCATTGGCTTGTTGTTCCTTGGGGGTGTG GTGCTGAGTTCGTGGTCGTGTTTCCTAAATGTCGAC TTGGCAACTCCGACAGAATGTGAGGCGTAA[C/T]A AAGGTTGGTGAGGTGAAGAGGGTGTACTCCTAATCG TTATCCACTTTAAAATGTTGTAAGACCATTAAGTTC TAGAAGTTCCAGTGGCAACATACCCAA -3'
MSUSNP16-90	Gm16_ 18529171_A_G	16	18,529,171	5'CGAGTGATGTGACATTAAAGTACACACGTCAACC CTCCTCGTCAGCCCTAAATCCAGTATGTTGATGCTC CTCGTATAATATGAACCCTAGCTCCCGACA[A/G]C CAAATGGTTGGGCCATTATGATTTTGTTATATGAGA

Table 2.9 (cont'd)

MSUSNP16-97	Gm16_5259121_A_G	16	5,259,121	ATCTTGGTGATAAGGTTGTTACTTTATTATCAAGTA TCAACAAAATTGTTATGTTGTTACAAC -3' 5'TTGAGAATTCAACAATCACGACTATTAAAAGCTC CATGAGATCCACAACAAGTGGTGGTGCAGTTTTGGG TCCTGGGGTGGTGACGCAATTTCTGGTGGT[A/G]C AGGGTGACGACACAACGATGATGGTGGTCAGGGAGG TGGGTGGTGCAATGATGTGGCAGCACGGTTCTAGTG GTGCAAGGTGGGCCTAATGCACAAATT -3'
MSUSNP16-103	Gm16_9958372_G_A	16	9,958,372	5'ACACATATAGAGAGTTGATGGTATGTTTATGTCG CACGAGTAAGAATTTTTATCTGTCTCACTCTCAAAT CTTCTTCTCTAACAGTGGTACATTCATTTC[G/A]T CCCATGTTTATTGCTTATAGTTCGTTTTTCTTTACT ATGTTGTCAAATTTAATAACATTTGAAATCTTTGTC ATTTGATTTTTATTTCAATGTACACAA -3'

[a] Genomic position of single nucleotide polymorphism on the Williams 82 genome assembly, Glyma1 (Schmutz et al., 2010).

[b] Target sequence for KASP custom design with 100-bp upstream and downstream of the single nucleotide polymorphism.

*Markers developed through whole genome sequence SNP discovery pipeline

Table 2.10: SNP markers used in fine mapping study of *Rag*3e in PI 567597C. SNPs in corresponding wild-type and mutant-alleles are in brackets [ ].

SNP assay ID	SoySNP50K chip ID	Chrom osome	Physical position (bp) [a]	Target sequence [b]
MSUSNP16-13	Gm16_6424067_A_G	16	6,424,067	5'CTCACTCGTTGATAAGAAAATGCATAAAACCTGCAA CCCTCAACTTCCTGACACCACTCGCAGTCCCTGAGATT CGGCGGCGGCTAGCGTCGGTGGCGGC[A/G]GCGGCGG ACGAGGACCCTCCGCAATCGCCGTCGTCGTTCACTTTC TCGTCGGAGGGGGAGAAGGTGTACGTGAAGCCGCCGGC GGGGAAGTGGAAGCCGT -3'
MSUSNP16-28	Gm16_6079769_A_G	16	6,079,769	5'CATAGAGGGCTTGAGCGATGTCTTGATCGTTGACCG AAAGTTGTGATCAAGTGTGGTAGTGTACGTCATCTCTC TCAGTTCCCCCACGATTCCTAATAAC[A/G]CATCAAT ATTCTTCTCTTTTGAAACACACCAATTATATATATTTT CTTTTCTTTTTATTCATTTTCTTTCATCGTCATAATTT TAATTTTTTTATCTAAA -3'
MSUSNP16-39	Gm16_6214642_C_T	16	6,214,642	5'ACACGATTGAAGAAAATTGAAAAAGAAATACTACTA CTATTGAAGAAAGTTGAAAAAGAAATACCAGTACCTTG TCCATTTCTGCTTCTTTTGCGGGTTG[C/T]GGAAAGT GTTCCAATTTAATGCTGGTCCTAAGTCCTAACCAAGTT ATCAAGATTCAAGCCGTGGCTTGCAGGTAATATTTAAC TCTCTGTTTAGACTTTA -3'
MSUSNP16-43	Gm16_6431101_A_C	16	6,431,101	5'GCATCAAAGAATGTATTAACAATACAATGACAAACA TAGAATTCAGCAAAGAGCTCTCTTATACTAGCTGAAGT ACAAAGCATAGCACCAAGAACAGCAG[A/C]TGAGATA TTATGCTAGTATTAACAAGTAAATCAACAAAACACCAT GCATATATCATGGCAGGTGTGAACTATATAATCCTTTC TCAACCCAAGCTTCAAA -3'

Table 2.10 (cont'd)

MSUSNP16-44	*	16	6,438,676	5'CCTTAAGAAGGATTCTCAAAAGTTTACTTTTAGCTC CAACAAGACATGTTCTTACATCTAAGCCCAACCAAACA AAAATAGAAAAACCAAATTTTAAATT[T/C]TTTATTA TCAACCTCATGATCACCATGTCTACCACGATTTATCCA TGGTTGTGTTTGGTTATCAATTTTAGCTTTTTCATCAA TTTTGGTTAATAATTTT -3'
MSUSNP16-85	Gm16_7070805_G_A	16	7,070,805	5'ATGCAAGGGAAGCAGCTGCAAGAGATGCAAGGGATG CAAAGGTGGAGGCGAGAGATGTAAAGAGAACAACAGTG ACAGCAACAACCGCAACCGCATGAAC[G/A]TGATGAG TATTAATGTGTTGTTATGAACTTATGATGTTGGTTTAT GTGGGGAAATAAATGATGTATGTACCTCTTCTTGCCTA TGTAGTAGGTTTGGGTG -3'
MSUSNP16-97	Gm16_5259121_A_G	16	5,259,121	5'TTGAGAATTCAACAATCACGACTATTAAAAGCTCCA TGAGATCCACAACAAGTGGTGGTGCAGTTTTGGGTCCT GGGGTGGTGACGCAATTTCTGGTGGT[A/G]CAGGGTG ACGACACAACGATGATGGTGGTCAGGGAGGTGGGTGGT GCAATGATGTGGCAGCACGGTTCTAGTGGTGCAAGGTG GGCCTAATGCACAAATT -3'
MSUSNP16-98	Gm16_5555122_T_C	16	5,555,122	5'ACTGGAAGACCTAAAGATTGAAAACATCTCTATGCC TGACGAGTTCGATTCTGAACTCCTGATTGAGAAGCTAC CAAAGTCCTAGATAGATTATAAACAA[T/C]AATTAAA GCACAAACACAAATAGATGTTACGACCAGACCTTATCA TCCACATTATCATTGAAGATACAAGCAGGAAGGAAAAT GTTATTGCAAGGACCAA -3'
MSUSNP16-100	Gm16_5809541_C_T	16	5,809,541	5'GAGGAAGACGATGGCATTTGCGGTGGTTGCTGATTA GCCTCCAGCTGCGTCATACGCTGGAGGAGGTCGTCAAT TTTCGTGTTCATTGAGAGCTGAGACG[C/T]CGATAGT TTCGCGATTGCGTCCTCGAGACAATATGCCAAAACCCT

Table 2.10 (cont'd)

MSUSNP16-110	Gm16_6774822_T_C	16	6,774,822	GGAACGAGTCGCATCCGCCATTGTTGTTCAATGAAAGC ACCAATGTTATGCTTGG -3' 5'CAAACATATTAAAGCAAACAATCGGGTTCAGATACA ACGTGCAAAAAAATTCAAATTCCCACCCAATCGTGTAA ACCCATATGAAACCATTTTTTGCAAT[T/C]TGATAAC ACGTGGTCCATCTATATCTGTCCGATACAACTAAACTG GTCTTCATTCAGATCAGTCGGTTTCAGGCGGGGGTGTT GCTAGGTGCACCCAGCA -3'
MSUSNP16-112	Gm16_6868110_C_A	16	6,868,110	5'TTAAATTTATTGAACCTCGACAAAAAAAAATATTGG GTAATAAACTTCCAAGCAACCTAAAAAACATGGTTTAG AGAGTAACCTACGCTGCCCTTCTGTA[C/A]AGTTTGT AAGCCACAACAACAACTGCTACAGCTCCAAGCACACCA ACAGCTGTGCGAATATCACGCTGCTGCAGTTGAGGCCA ATTCGAACAGATTTTGC -3'
MSUSNP16-122	Gm16_6260278_T_C	16	6,260,278	5'ATCCGACAAATTAATTTTTATTTTAAACTTTTTCAT ATCCACATGTTCCTTAAAAAAATGAATACTGTTATGGA AACAAATGAAAGTTCTGAACCAGGGG[T/C]GGACCCA GGATCTATTGACTGGGGGAACCAGGTTTAAAAATTAAA ATTAATACAATAAATATACTAACATAAAAATTTATACA AAGTATTTATAATTATA -3'
MSUSNP16-124	Gm16_6484276_A_G	16	6,484,276	5'ATCAAATTATCATGATTAATTATTAAACTTACGACA AAATTAATATATATACTTATATTACTTATACCAATTTG TGCGGGTCAAATACTTTGACTATAAT[A/G]ACCGATC CAATTCTTTAAATTATTTATAAGTGAAAAAAAAATAAC CTCCTATATTATATCAGATCATTTAATTTATATATTTG GTTTCACATTTCAAATA -3'
MSUSNP16-128	Gm16_6517204_T_C	16	6,517,204	5'CAATTCTTGTGGCGATCCTCTAGGCTTATCAAACCT CGTTGTTTGTGGTGCCTATTTGATTTTCTATCTTACAC

Table 2.10 (cont'd)

MSUSNP16-132	*	16	6,571,636	ATTTTATGTTCGTGTCGGCATTTCTG[T/C]GCTCTCC AAATTGTGCCTTCAGCCCGATTTTTCTCAGGTTCGTTC TGCACGTTTTTACGAATCTGTTTTTAGTGTTTCTTGAC CCTAAATTACAGTTCAA -3'
MSUSNP16-134	Gm16_6624879_C_T	16	6,624,879	5'ACAAATTAAGGAACCATATAACTGACTGTATAGTTA TCATAGATTAGCCCATCAGCTAAGACACTAGACCCCCC TTCTTACAAATCCCTCCCTTATGTTA[C/T]TTGATCA GTGGTGGAAAGGCACTTCAAATCCTTTTTCGAGATTGG TTAGCCATGTCCAACATACAAATATGGCATCCTCCATC AGTTTGTTAGCGTTGAA -3'
MSUSNP16-136	*	16	6,721,743	5'GATGTATCTTGTGTGGTGGCGGTGGTGGCCCAAGGC CGCGGTGTGTCGCGTGACTGCGTGAGTCGTGTCCACGG TGAGGAGAAGAAGATGAGAAGAAATG[C/T]TGTAAGA GGAGAAGAATAAAGCAAGGTACTAGTCCTTAAAGTGGT ACTAGTCCAATGGTTCTTAAAGTGAAAAAGAAAAAATC CAACGTACTAAAACTAA -3'
				5'ATTCACCATGAGATCCCTTTAACAATTTCCTCTAAG GTCTTTCTTCGTCTAACCCCGCCCCTTTTCACAGGACT AGAAACCATGCAATCTATTCTTCTAA[C/T]TAGTACT TCTTTGCCTTTCACAATTTATGACTTTATTCATTAACC ACAAAAAAAGCAAAAAACACTACCTTTCAACAATATGC ATATTGGAAAGAATGAG -3'

[a] Genomic position of single nucleotide polymorphism on the Williams 82 genome assembly, Glyma1 (Schmutz et al., 2010).

[b] Target sequence for KASP custom design with 100-bp upstream and downstream of the single nucleotide polymorphism.

*Markers developed through whole genome sequence SNP discovery pipeline

REFERENCES

REFERENCES

Alt J, Ryan-Mahmutagic M (2013) Soybean Aphid Biotype 4 Identified. Crop Sci 53:1491

Bales C, Zhang G, Liu M, et al (2013) Mapping soybean aphid resistance genes in PI 567598B. Theor Appl Genet 126:2081–2091

Bales CJC (2013) Fine mapping of aphid resistance genes in soybean plant introduction (PI) 567598B. Ph.D. dissertation, Michigan State University

Bansal R, Mian MAR, Michel AP (2013) Identification of Novel Sources of Host Plant Resistance to Known Soybean Aphid Biotypes. J Econ Entomol 106:1479–1485

Barkan A, Small I (2014) Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol 65:415–442

Beckendorf EA, Catangui MA, Riedell WE (2008) Soybean Aphid Feeding Injury and Soybean Yield, Yield Components, and Seed Composition. Agron J 100:237

Bhusal SJ, Jiang G-L, Hesler LS, Orf JH (2014) Soybean Aphid Resistance in Soybean Germplasm Accessions of Maturity Group I. Crop Sci 54:2093

Bhusal SJ, Jiang G-L, Tilmon KJ, Hesler LS (2013) Identification of Soybean Aphid Resistance in Early Maturing Genotypes of Soybean. Crop Sci 53:491

Catangui MA, Beckendorf EA, Riedell WE (2009) Soybean Aphid Population Dynamics, Soybean Yield Loss, and Development of Stage-Specific Economic Injury Levels. Agron J 101:1080

Chen JQ, Rahbé Y, Delobel B, et al (1997) Melon resistance to the aphid Aphis gossypii: behavioural analysis and chemical correlations with nitrogenous compounds. Entomol Exp Appl 85:33–44

Cook D, Bayless A, Wang K, et al (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

Cook DE, Lee TG, Guo X, et al (2012) Copy Number Variation of Multiple Genes at Rhg1 Mediates Nematode Resistance in Soybean. Science 338:1206–1209

Diehl SR, Bush GL (1984) An evolutionary and applied perspective of insect biotypes. Annu Rev Entomol 29:471–504

Dogimont C, Chovelon V, Pauquet J, et al (2014) The Vat Locus Encodes for a CC-NBS-LRR Protein that Confers Resistance to Aphis gossypii Infestation and A. gossypii-Mediated Virus Resistance. Plant J 80(6):993-1004

Du W, Wang D (2015) Perspectives from Soybean Breeding: A Combination of Classic and New Technologies to Map and Fine Map Aphid Resistance in Soybean and Its Application in Horticulture. ASHS Search and Access Archived Conference Presentations. https://ashs.confex.com/ashs/2015/videogateway.cgi/id/1983?recordingid=1983. Accessed 5 Sep 2015

Eddy SR (1998) Profile hidden Markov models. Bioinforma Oxf Engl 14:755–763

Fanourakis D, Briese C, Max JF, et al (2014) Rapid determination of leaf area and plant height by using light curtain arrays in four species with contrasting shoot architecture. Plant Methods 10:9

Grant D (2015) What's New at SoyBase? Plant and Animal Genome. https://pag.confex.com/pag/xxiii/webprogram/Paper16170.html. Accessed 7 Apr 2016

Guo S, Kamphuis LG, Gao L, et al (2009) Two independent resistance genes in the *Medicago truncatula* cultivar jester confer resistance to two different aphid species of the genus Acyrthosiphon. Plant Signal Behav 4:328–331

Hanson AA, Orf JH, Koch RL (2016) Sources of Soybean Aphid Resistance in Early-Maturing Soybean Germplasm. Crop Sci 56:154

Hesler LS (2013) Resistance to soybean aphid among wild soybean lines under controlled conditions. Crop Prot 53:139–146

Hesler LS, Dashiell KE, Lundgren JG (2007) Characterization of resistance to Aphis glycines in soybean accessions. Euphytica 154:91–99

Hill CB, Kim K-S, Crull L, et al (2009) Inheritance of Resistance to the Soybean Aphid in Soybean PI 200538. Crop Sci 49:1193

Hill CB, Li Y, Hartman GL (2006a) A Single Dominant Gene for Resistance to the Soybean Aphid in the Soybean Cultivar Dowling. Crop Sci 46:1601

Hill CB, Li Y, Hartman GL (2006b) Soybean Aphid Resistance in Soybean Jackson Is Controlled by a Single Dominant Gene. Crop Sci 46:1606

Hill JH, Alleman R, Hogg DB, Grau CR (2001) First Report of Transmission of *Soybean mosaic virus* and *Alfalfa mosaic virus* by *Aphis glycines* in the New World. Plant Dis 85:561–561

Hyten DL, Choi IY, Song Q, Specht JE, Carter TE, Shoemaker RC, et al. (2010) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for QTL mapping. Crop Sci 960–968

Jardim-Messeder D, Caverzan A, Rauber R, et al (2015) Succinate dehydrogenase (mitochondrial complex II) is a source of reactive oxygen species in plants and regulates development and stress responses. New Phytol 208:776–789

Jun T-H, Mian MAR, Michel AP (2012b) Genetic mapping revealed two loci for soybean aphid resistance in PI 567301B. Theor Appl Genet 124:13–22

Kang S-T, Rouf Mian MA, Hammond RB (2008) Soybean Aphid Resistance in PI 243540 Is Controlled by a Single Dominant Gene. Crop Sci 48:1744

Kim K-S, Bellendir S, Hudson KA, et al (2010a) Fine mapping the soybean aphid resistance gene Rag1 in soybean. Theor Appl Genet 120:1063–1071

Kim K-S, Hill CB, Hartman GL, et al (2010b) Fine mapping of the soybean aphid-resistance gene Rag2 in soybean PI 200538. Theor Appl Genet 121:599–610

Kinoshita T, Yamada K, Hiraiwa N, et al (1999) Vacuolar processing enzyme is up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions. Plant J 19:43–53

Li Y, Hill CB, Carlson SR, et al (2006) Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. Mol Breed 19:25–34

Liu G, Xing H, Diao Y, et al (2014) Identification of Resistance to Soybean Aphids in Early Germplasm. Crop Sci 54:2707

Liu M (2010) Inheritance of resistance and molecular mapping of soybean aphid resistance genes in soybean PI 567585A; identification of aphid resistance genes in soybean using modified nested association mapping (MNAM). Ph.D. dissertation, Michigan State University

McCarville MT, Kanobe C, MacIntosh GC, O'Neal M (2011) What is the economic

threshold of soybean aphids (Hemiptera: Aphididae) in enemy-free space? J Econ Entomol 104:845–852

Mensah C, DiFonzo C, Wang D (2008) Inheritance of soybean aphid resistance in PI 567541B and PI 567598B. Crop Sci 48:1759

Mensah C, DiFonzo C, Nelson RL, Wang D (2005) Resistance to Soybean Aphid in Early Maturing Soybean Germplasm. Crop Sci 45:2228

Mian MAR, Hammond RB, St. Martin SK (2008) New Plant Introductions with Resistance to the Soybean Aphid. Crop Sci 48:1055

Milligan SB, Bodeau J, Yaghoobi J, et al (1998) The Root Knot Nematode Resistance Gene Mi from Tomato Is a Member of the Leucine Zipper, Nucleotide Binding, Leucine-Rich Repeat Family of Plant Genes. Plant Cell 10:1307–1319

Ohnishi S, Miyake N, Takeuchi T, et al (2012) Fine mapping of foxglove aphid (Aulacorthum solani) resistance gene Raso1 in soybean and its effect on tolerance to Soybean dwarf virus transmitted by foxglove aphid. Breed Sci 61:618–624

Painter RH (1951) Insect resistance in crop plants. Macmillan Publishing Co, New York

Pauquet J, Burget E, Hagen L, et al (2004) Map-based cloning of the Vat gene from melon conferring resistance to both aphid colonization and aphid transmission of several viruses. In: Progress in cucurbit genetics and breeding research. Proceedings of Cucurbitaceae 2004, the 8th EUCARPIA Meeting on Cucurbit Genetics and Breeding. Palacký University, Olomouc, pp 325–329

Peck SC (2003) Early phosphorylation events in biotic stress. Curr Opin Plant Biol 6:334–338

Ragsdale DW, Landis DA, Brodeur J, et al (2011) Ecology and Management of the Soybean Aphid in North America. Annu Rev Entomol 56:375–399

Ragsdale DW, McCornack BP, Venette RC, et al (2007) Economic Threshold for Soybean Aphid (Hemiptera: Aphididae). J Econ Entomol 100:1258–1267

R. L. Blackman, V. F. Eastop (2000) Aphids on the World's Crops: An Identification and Information Guide, 2nd Edition. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471851914.html. Accessed 26 Apr 2016

Rossi M, Goggin FL, Milligan SB, et al (1998) The nematode resistance gene Mi of tomato

confers resistance against the potato aphid. Proc Natl Acad Sci 95:9750–9754. doi: 10.1073/pnas.95.17.9750

Schmutz J, Cannon SB, Schlueter J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schwartz DC, Hochstrasser M (2003) A superfamily of protein tags: ubiquitin, SUMO and related modifiers. Trends Biochem Sci 28:321–328

Sekhwal MK, Li P, Lam I, et al (2015) Disease Resistance Gene Analogs (RGAs) in Plants. Int J Mol Sci 16:19248–19290

Semagn K, Babu R, Hearne S, Olsen M (2013) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 33:1–14

Song F, Swinton S, DiFonzo C, et al (2006) Profitability Analysis of Soybean Aphid Control Treatments in Three North-central States. Staff paper at Michigan State University. http://msuent.com/assets/Song06SBAProfitablity.pdf. Accessed 11 Apr 2016

Song Q, Hyten DL, Jia G, et al (2013) Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. PLoS ONE 8(1):e54985

Song Q, Jia G, Zhu Y, Grant D, Nelson RT, Hwang EY, et al. (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci 50:1950

Van Ooijen (2006) JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B. V., Wageningen, Netherlands

Vierstra RD (2012) The Expanding Universe of Ubiquitin and Ubiquitin-Like Modifiers. Plant Physiol 160:2–14

Voorrips RE (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. J Hered 93:77–78

Wang S, Basten CJ, Zeng ZB (2012) Windows QTL Cartographer 2.5. Dep. Stat. N. C. State Univ. Raleigh NC http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

Xiao L, Hu Y, Wang B, Wu T (2013) Genetic mapping of a novel gene for soybean aphid resistance in soybean (Glycine max [L.] Merr.) line P203 from China. Theor Appl Genet 126:2279–2287

Yan Li CBH (2012) Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. Mol Breed 19:25–34

Yuan J (2014) Fine Mapping Soybean Aphid Resistance Gene *rag1c* in Soybean (*Glycine max* L.). Plant and Animal Genome XXII. https://pag.confex.com/pag/xxii/webprogram/Paper13372.html. Accessed 9 Dec 2015

Yuan J, ZW (2014) Introduction of High Throughput and Cost Effective SNP Geno- typing Platforms in Soybean. Plant Genet Genomics Biotechnol 2:90–94

Zhang G, Gu C, Wang D (2013) Mapping and validation of a gene for soybean aphid resistance in PI 567537. Mol Breed 32:131–138

Zhang G, Gu C, Wang D (2009) Molecular mapping of soybean aphid resistance genes in PI 567541B. Theor Appl Genet 118:473–482

Zhang G, Gu C, Wang D (2010) A novel locus for soybean aphid resistance. Theor Appl Genet 120:1183–1191

Zsigmond L, Szepesi A, Tari I, et al (2012) Overexpression of the mitochondrial PPR40 gene improves salt tolerance in Arabidopsis. Plant Sci Int J Exp Plant Biol 182:87–93

CHAPTER III: FINE MAPPING SOYBEAN APHID RESISTANCE GENES IN PI 567585A AND PI 567537

Abstract

The soybean aphid is a major pest of soybean [*Glycine max* (L.) Merr] in North America. PI 567585A confers both antixenosis and antibiosis resistance to all four biotypes of soybean aphids. It would be a great source of aphid resistance for soybean breeding program. Aphid resistance from PI 567585A was controlled by a major partially dominant gene which was mapped to a 28 cM / 8.5 kb QTL interval on chromosome 16. This QTL explained 93.1% of the total phenotypic variation and was designated as *Rag*3d (Liu 2010). In order to utilize aphid resistance from PI 567585A, further study on *Rag*3d was needed. The objectives of this study were to use residual heterozygous lines (RHL) to fine map aphid resistance QTL in PI 567585A, to develop SNP markers for marker assisted selection and to find candidate genes for functional study. RHLs were selected from two segregating populations,100040 and 100041. Important RHLs were genotyped with the SoySNP50K chip to help identifying recombination breaking points. *Rag*3d was fine mapped to a 46kb interval on chromosome 16 with five candidate genes.

PI 567537 possess antibiosis resistance to soybean aphid biotype 1, 2 and 3 (no data for biotype 4) (Table 1.1). It was a valuable addition to soybean aphid resistance germplasm

pool. The objectives of this study were to use SNP markers to fine map the *Rag*3b aphid resistance QTL region, to identify candidate genes for *Rag*3b and to develop genetic markers for marker assisted selection. Three F2 populations derived from the cross between the resistant and the susceptible parents were used to validate *Rag*3b and screening for recombination breaking points. The *Rag*3b QTL was delimited to a 199kb region with 12 candidate genes.

Key word Soybean, Aphid Resistance, Fine Mapping, *Rag*3d, Partial dominance, *Rag*3b, Dominance, Residual Heterozygous Lines, SoySNP50K

Introduction

The soybean aphid (*Aphis glycines* Matsumura) was first discovered in the Midwest of the U. S. in 2000. It was an invasive species from Asia. The soybean aphid can damage plants severely by sap-feeding and virus transmission. Affecting 30 states in the U. S. and three provinces in Canada, the soybean aphid damage has resulted in a huge economic impact on soybean production in North America (Ragsdale et al. 2011).

Foliar insecticide application to treat soybean aphids has been well adopted in the U. S. The application decision was made based on scouting and economic threshold. When

the number of aphids reached the economic threshold of 250 per plant during scouting, insecticide treatment can effectively suppress aphid outbreak (Ragsdale et al. 2007; Johnson et al. 2009). However, insecticides may have negative effect on aphid natural enemies and are not friendly to the environment (Johnson et al. 2008). Cultivars with host plant resistance would be a great alternative to reduce aphid damage.

There are two types of host plant resistance to insects: antixenosis and antibosis (Painter 1951). Antixenosis are mechanisms that deter insects from colonization, oviposition and feeding. Antibiosis influences insect biology. Among the 43 identified soybean aphid resistant germplasm (Li et al. 2006; Hesler et al. 2007; Mian et al. 2008; Hill et al. 2009; Zhang et al. 2009; Zhang et al. 2010; Liu 2010; Jun et al. 2012; Bhusal et al. 2013; Bansal et al. 2013; Zhang et al. 2013; Bales et al. 2013; Xiao et al. 2013; Hesler 2013; Bhusal et al. 2014; Liu et al. 2014; Hanson et al. 2016), 12 of them have been characterized with genetic studies. Both Dowling and Jackson have a single dominant gene on chromosome 7 to control antibiosis, named as *Rag*1 (Resistance to *Aphis glycines*) and *Rag*, respectively (Li et al. 2006); PI 567541B has two recessive genes, *rag*1c (chromosome 7) and *rag*4 (chromosome 13), conferring antibiosis (Zhang et al. 2009); PI 567301B has two genes on chromosome 8 and 13 controlling antixenosis (Jun et al. 2012); P203 has a single dominant gene *[Rag*6*]_P203* with antixenosis (Xiao et al. 2013); both PI 243540 and PI 200538 have *Rag*2, a single dominant gene, contributing antibiosis (Kang et al.

2008; Hill et al. 2009); both PI 567543C and PI 567537 have a single dominant gene, *Rag*3 and *Rag*3b, on chromosome 16 responsible for antixenosis and antibiosis, respectively (Zhang et al. 2010; Zhang et al. 2013); both PI 567585A and PI 567597C have a single partially dominant gene, *Rag*3d and *Rag*3e, on chromosome 16 in charge of antixenosis (Liu 2010, chapter II); PI 567598B has two genes, *rag*1b and *rag*3, on chromosome 7 and 16 for antibiosis (Bales et al. 2013).

*Rag*1, *Rag*2, *rag*1c, *[Rag*6*]_P203* and *Rag*3e have been fine mapped to a 115 kb region on chromosome 7, a 54 kb region on chromosome 13, a 96kb region on chromosome 7 (different region from *Rag*1), a 192-kb interval on chromosome 8, and a 60kb region on chromosome 16, respectively (Kim et al. 2010a; Kim et al. 2010b; Xiao et al. 2013; Yuan 2014, chapter II).

Intriguingly, out of the twelve germplasm with *Rag*/*rag* genes, five of them have aphid resistance gene located on chromosome 16. Their QTL intervals are very close to each other. In order to further utilize these germplasm, to fine map these QTLs is the key. The objective of this study is to fine map aphid resistance genes *Rag*3b from PI 567537 and *Rag*3d from PI 567585A for a deeper understanding of *Rag*3/*rag*3 region on chromosome 16.

Material and Method

Plant materials for fine mapping study of *Rag*3d in PI567585A

Fine mapping populations 100040 and 100041 were developed by single seed descent (SSD) to F3:4 [derived from the cross of a resistant parent (E09914 for population 100040, E09915 for population 100041) and a susceptible parent (Skylla)]. SSD F3:4 plants were planted in the field during summer, 2012. Fifteen F3:4 seeds were planted for each line. The resistant parent E09914 and E09915 are resistant lines selected from the F3:4 of two rounds of backcross between Skylla (recurrent parent) and PI 567585A (donor parent).

F3:4 and F4:5 were selected phenotypically based on an aphid rating. The seeds of segregating lines were kept for planting in the greenhouse in fall 2012 and spring 2013, respectively. F5:6 plants were planted in the field in the summer of 2013. F7:8, F8:9, F9:10 and F10:11 were planted in greenhouse spring 2014, in field summer 2014 and in greenhouse in fall 2014 and in spring 2015. The F3:4 plants were evaluated for aphid severity individually. The leaf sample of each line (fifteen plants) were bulked for genotyping (see detailed description in next session). The F5:6 plants and higher generations were both phenotyped (aphid rating) and genotyped individually with SNP markers. Table 3.1 is a summary of the fine mapping populations.

Table 3.1: Fine mapping population information for PI 567585A. R stands for resistant parent; S stands for susceptible parent.

Population	Female Parent	Male Parent	Generation	# of Lines
100040	E09914^{R} (Skylla x PI 567585A)	Skylla^{S}	F3:4	187
100041	E09915^{R} (Skylla x PI 567585A)	Skylla^{S}	F3:4	177

Plant materials for fine mapping study of *Rag*3b in PI567537

Population100047-3, 100047-4 and 100048-5 with one hundred and seventy-three, two hundred and ten and a hundred and twenty-five F2 seeds, respectively, were used in the study. Population 100047-3 and 100047-4 were derived from the crosses of E09928 and E00003. Population 100048-3 was derived from the cross of E09929 and E00003. E09928 and E09929 are breeding lines selected from the progenies of the cross E00003xPI 567537 and carries the aphid resistance from PI 567537. E00003 is an elite cultivar susceptible to soybean aphids.

F2, F2:3, F3:4 and F4:5 were planted in the greenhouse spring 2014, in field summer 2014, in spring 2015 and in the field summer 2015, respectively. Each F2 plant and the selected plants from other generations were both phenotyped (aphid rating) and genotyped individually with SNP markers. Table 3.2 is a summary of the fine mapping populations.

Table 3.2: Fine mapping population information for PI 567537. R stands for resistant parent; S stands for susceptible parent.

Population	Female Parent	Male Parent	Generation	# of Lines
100047-3	E09928^{R} (E00003xPI 567537)	E00003^{S}	F2	173
100047-4	E09928^{R} (E00003xPI 567537)	E00003^{S}	F2	210
100048-5	E09929^{R} (E00003xPI 567537)	E00003^{S}	F2	125

DNA extraction and SNP genotyping

For each F3:4 line, a 1cm^2 leaf tissue from all the fifteen plants were collected in a 15 ml tube (centrifuge tubes, Corning Inc.). In each tube, all the leaf tissues together was called a bulk collection. Leaf tissue of individual plants for the other generations were collected in 96 well-plates. After one day in the -80 C° freezer, the samples were freeze-dried, then ground. The DNA extraction process was conducted following the CTAB method (Bales et al. 2013). For PCR reactions, the original DNA were diluted 50 and 5 times for tube and plate extraction, respectively. Kompetitive Allele Specific PCR (KASP) assays were developed (Semagn et al. 2013) by comparing parent information either from the SoySNP50K iSelect SNP beadchip (Song et al. 2013) or the whole genome re-sequencing data. The complete list of SNP markers used in this study were listed in table 3.10 and table 3.11. KASP assays were run with 3 µL reaction system including 1.5 µL KASP master mix (KBiosciences, Herts, England), 0.045 µL of primer mix and 1.5 µL of 10-25 ng/µL genomic DNA. The PCR conditions for KASP marker assay were 95°C for 15 min, followed by 10 cycles of 95°C for 20 seconds and 65°C for 1 minute, then followed

by 32 cycles of 95°C for 20 seconds and 58°C for 1 minute. For the KASP markers, the PCR reactions are either running in the Bio-Rad PCR machine (model C1000 touch, Bio-Rad Laboratories, Inc., USA) or Roche 480 light cycler (Roche Diagnostics, Germany). The fluorescent level of the final PCR products was measured and analyzed with the Roche 480 light cycler.

Parent DNAs and the selected F3:4 bulk DNAs were sent for whole genome SNP genotyping analysis with the SoySNP50K iSelect SNP beadchip (Song et al., 2013). DNA concentration was determined by the Quant-iT™ Picogreen® dsDNA Assay Kit (Invitrogen, USA) and quantified using BioTek Multi-Detection Microplate Reader (Biotek, USA). After normalized to 50 ng/ul. Each DNA sample was prepared for Infinium assay following manufacture's protocol. GenomeStudio Genotyping module was used for data analysis.

Aphid infestation and rating

The planting conditions followed these described by Bales et al. (2013). In greenhouse trials, eight seeds per line were planted in a plastic pot with 105mm in diameter and 125 mm deep. The greenhouse temperature was maintained at 26/15^{0}C day/night with sodium vapor lights supplementing light intensity during the day (14h). In summer field

trials, with 10 to 15 seeds, each line was planted in a single-row plot, 60 cm long with a row spacing of 60 cm. Aphid resistance was evaluated either in an aphid cage [a 12.2 x 18.3 m aphid and predator-proof polypropylene cage with 0.49-mm size mesh (Redwood Empire Awning Co., Santa Rosa, CA, USA)] in the summer or in the greenhouse without the cage in the fall and spring. Summer aphid colonies were collected by Dr. Christina DiFonzo from state-wide scouting every year. The colonies were kept and propagated in a small field cage. The greenhouse aphid colonies were kept in the greenhouse aphid room all year around. By the end of every summer season, aphid colonies in the field will also be moved to the greenhouse aphid room. For inoculation, small paint brushes were used to transfer the aphid from the original colonies to the testing soybean plants. Each plant was infested at the V2 stage with two wingless soybean aphids. Aphid resistance evaluation took place three or four weeks after the infestation. Each plant was rated with a zero to four scale (zero is resistant, four is susceptible) developed by Mensah et al.(2005, 2008). The phenotypes of plants within each line were converted into damage index (DI): the sum of the scale value times number of plants in each scale category, divided by 4 times the total number of plants and then times 100 [DI=Σ (Scale value × No. of plants in the category)/ (4× Total No. of plants) ×100] (Mensah et al. 2005). Rating value and damage index corresponding to number of aphid per plants is summarized in Table 2.2.

Statistics

The phenotypic distributions of recombinants were compared with the ideal distributions by Chi-square test using Excel 2013 (Microsoft Office Professional Plus 2013).

QTL mapping for validation

Genotyping data of population 100040 and population 100041 were used to construct the linkage map to validate *Rag*3d in PI 567585A. Genotyping data of population 100047-3, 100047-4 and 100048-5 F2 plants were used to construct the linkage group to confirm *Rag*3b in PI 567537. All the linkage maps were constructed with an LOD score of 3.0 and Kosambi function using the software JoinMap 4.0 (Van Ooijen 2006). The damage index of F3:4 and F2:3 lines were used as corresponding phenotyping data for PI 567585A and PI 567537, respectively. Single marker analysis (SMA) and composite interval mapping (CIM) were conducted using QTL cartographer V2.5 (Wang et al. 2012). For the CIM analysis, a 1000-permutation test was used to determine the LOD threshold at the 5% probability level. The forward and backward regression method with a walking speed of 1cM was used when running the method. Results displayed were output from QTL cartographer V2.5.

Candidate gene search and annotation

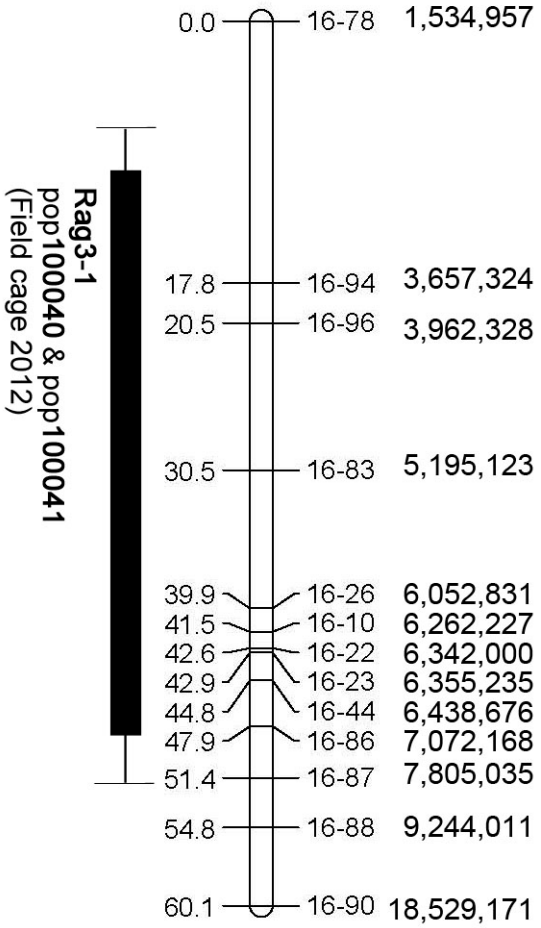
A candidate gene search was conducted using both soybean genome assemblies Glyma.Wm82.a1 and Glyma.Wm82.a2 (Schmutz et al. 2010; Grant 2015). Because the SNP position for Glyma.Wm82.a1 was kept throughout this study, the corresponding position of the two flanking markers can be used directly to search in Glyma.Wm82.a1 from soybase (http://www.soybase.org/gb2/gbrowse/gmax1.01/). To match the fine mapping interval into Glyma.Wm82.a2, two methods were used. For SNP developed from the SoySNP50K chip, the corresponding position in Glyma.Wm82.a2 is already labelled by Song et al. (2013). After the corresponding new name of the SNP was found, the new position in Glyma.Wm82.a2 is searched in soybase (http://www.soybase.org/gb2/gbrowse/gmax2.0/). For the SNP marker developed from SNP discovery, the primer sequence of the SNP was used to BLAST against the Glyma.Wm82.a2 version genome to locate the marker position (http://www.soybase.org/ NCBI Blast report, blastn). Candidate genes and their annotations from Glyma.Wm82.a1 and Glyma.Wm82.a2 were compared.

Results

QTL validation

Aphid resistance gene *Rag*3d from PI 567585A was mapped to *Rag3/rag*3 region by Menghan Liu (2010). It was a major gene with partially dominant inheritance. Simple sequence repeat markers Satt674 - Sct_065 and Satt654 - Sct_065 were flanking markers of the QTL identified in Menghan's study. These markers spanned an interval of 28 cM / 8.5 Mb at the beginning of chromosome 16 (Liu 2010). To verify the aphid resistance QTL in fine mapping populations100040 and 100041, one hundred and eighty-seven and one hundred and seventy-seven F3:4 lines of these two populations were phenotyped and genotyped for QTL mapping, respectively. The fine mapping populations were genotyped with thirteen single nucleotide polymorphism markers. These markers spanned a 60.1cM/18Mb region at the beginning of chromosome 16 covering the QTL mapping interval. A 6.3 Mb QTL interval between marker MSUSNP16-78 and MSUSNP16-87 was identified in both populations through CIM method (Figure 3.1). This confirmed the aphid resistance QTL, *Rag*3d, in the fine mapping populations. Also, the new flanking markers labelled a narrower region than the initial QTL mapping.

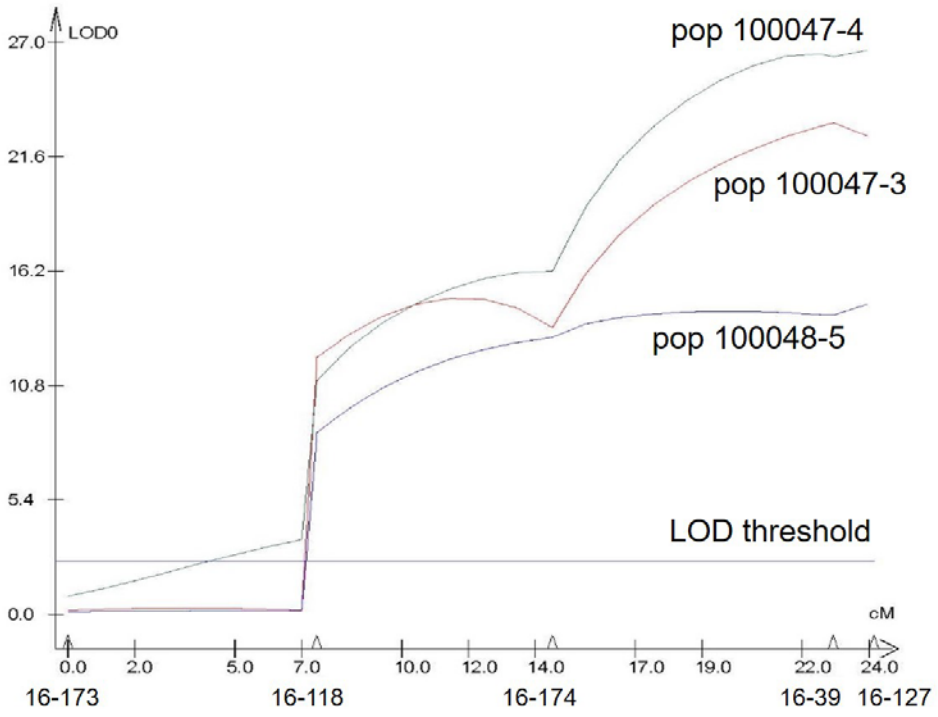
Figure 3.1: Validation of Initial mapping results by the fine mapping population 100040 and 100041. Both populations have been SSD to F3:4 from the cross of Skylla and PI 567585A. The phenotyping was conducted in the field cage summer 2012. Fifteen seeds were planted from each line and DNA are bulked for genotyping. The black bar on the left shows CIM results from both populations. This indicates the aphid resistant QTL was located between SNP marker MSUSNP16-78 and MSUSNP16-87. Numbers on the right shows the physical distance of these markers on chromosome 16 (bp).



Soybean aphid resistance gene *Rag*3b from plant introduction (PI) 567537 was characterized by Zhang (2013) using a F4 population as the mapping population and a F2 population as the validation population. Both populations were segregating for soybean aphid resistance. *Rag*3b was a major dominant gene controlling aphid resistance in PI 567537, explaining 80-90% of the phenotypic variation. The QTL has been detected in an interval between Sat_339 (located approximately at 3.0Mb on Chromosome 16) and Satt654 (7.8Mb on Chromosome 16) in the mapping population and an interval between Sat_339 and Sct_065 (10.5Mb on Chromosome 16) in the validation population.

To verify the existence of *Rag*3b in the fine mapping populations 100047-3, 100047-4 and 100048-5, QTL mapping was conducted. In order to construct a linkage map, five markers covering the QTL region (20cM/3.0Mb) were used to genotype the three F2 populations. The aphid ratings of the F2 progenies were converted into damage index for each line. As shown in figure 3.2, for each population, there is a QTL peak between marker MSUSNP16-118 (4,399,162bp) and MSUSNP16-127 (6,510,537bp). LOD scores of three peaks are from 15 to 27, exceeding the LOD threshold 3.0. The QTL mapping results validated a strong QTL in all three fine mapping populations.

Figure 3.2: QTL validation in fine mapping population 100047-3, 100047-4 and 100048-5. For convenience and space, “MSUSNP” which was the first part of every marker’s name, was not included in the graph. *pop* stands for population. LOD threshold was set to 3.0.



Fine mapping *Rag*3d in PI567585A

Recombinants identified by the SoySNP50K iSelect SNP beadchip (SoySNP50K chip)

For the first round of fine mapping, seventy-six aphid resistant lines (forty-two lines from population 100040 and thirty-four lines from population 100041), seventy-four susceptible lines (forty-one lines from population 100040 and thirty-three lines from population 100041) and a hundred and twenty-three segregating lines (sixty-four lines from population 100040 and fifty-nine lines from population 100041) were selected from F3:4 populations. These lines were genotyped with thirteen SNP markers. Fifteen lines with possible recombination breaking points were selected. The bulk DNA of each line and the parents of the two populations were genotyped with the SoySNP50K chip. For the resistant lines, regions with alleles from the susceptible parents can be excluded for aphid resistance QTL. For the susceptible lines, the region with alleles from the resistant parents can be excluded. For the heterozygous lines, the ones (line 85, 101, 65 and 39) with a recombination breaking point in the QTL region, were planted for further fine mapping. Table 3.3 summarizes the phenotypes and genotypes of four informative recombinants and the resistant and susceptible parents. For region 1.4 to 1.9 Mb, parent E09915, line 37, 116 and 100 were resistant to aphid phenotypically. However, the genotypes of these lines were the same as the susceptible parent (Skylla) which means this region was not responsible for aphid resistance. For region 2.2 to 3.7 Mb, line 145 was phenotypically

susceptible while its genotype in this region was the same as the resistant parent indicating this region was not related to aphid resistance. Through the first round of fine mapping, the QTL was narrowed to a 4.1 Mb (from 3.7 Mb to 7.8 Mb) region.

Table 3.3: Genotyping parents (E09914, E09915 and Skylla) and bulk DNA of F3:4 lines with SoySNP50K chip excluded the possibility of *Rag*3d QTL locating between 1.4 Mb and 3.7 Mb on Chromosome 16. Genotyping was conducted with the SoySNP50K chip. Column 1, the physical position on chromosome 16 (bp); Column 2, 3 and 4 are genotypes of E09914, E09915 and Skylla respectively; Column 5, 6, 7 and 8 are genotypes of bulk DNA from resistant (column 5, 6 and 7) and susceptible (column 8) lines respectively. The cage rating on top of column 5-8 indicates the damage index calculated from rating of fifteen plants for each line (0 or R means resistant, 100 or S indicates susceptible).

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Aphid Rating (Damage Index)	19.8 *R*	13 *R*	87.5 *S*	22.12 *R*	22.22 *R*	25 *R*	93.2 *S*
Position (bp)	E09914	E09915	skylla	Line 37	Line 116	Line 100	Line 145
1,421,831	GG	AA	AA	AA	AA	AA	
1,422,983	GG	AA	AA	AA	AA	AA	
1,434,889	TT	CC	CC	CC	CC	CC	
1,534,957	TT	CC	CC	CC	CC	CC	
1,844,621	AA	GG	GG	GG	GG	GG	
1,846,998	GG	AA	AA	AA	AA	AA	
1,867,614	AA	GG	GG	GG	GG	GG	
2,299,577	CC	CC	TT				CC
2,314,458	CC	CC	TT				CC
2,331,985	TT	TT	CC				TT
2,332,760	AA	AA	GG				AA
2,333,905	TT	TT	CC				TT
2,345,330	CC	CC	AA				CC
2,385,162	AA	AA	GG				AA
2,388,421	GG	GG	AA				GG
2,398,326	TT	TT	GG				TT
2,405,914	AA	AA	GG				AA
2,462,464	AA	AA	GG				AA
2,467,542	TT	TT	GG				TT

Table 3.3 (cont'd)

2,487,058	AA	AA	GG	AA
2,504,489	CC	CC	TT	CC
2,506,509	GG	GG	AA	GG
2,513,346	CC	CC	TT	CC
2,515,540	GG	GG	AA	GG
2,520,441	GG	GG	AA	GG
2,521,327	AA	AA	GG	AA
2,522,576	CC	CC	TT	CC
2,524,076	GG	GG	AA	GG
2,556,969	GG	GG	TT	GG
3,657,324	AA	AA	GG	AA
3,701,652	CC	CC	TT	CC

Progeny test of residual heterozygous line to further narrow down

For the second round of fine mapping, nine SNP markers spanning the narrowed QTL region from the first round were used to genotype five hundred and ninety F5:6 lines. Among these F5:6 lines, seventy-six lines were progenies of F3:4 lines that had been genotyped with the SoySNP50K chip. The other five hundred and fourteen F5:6 lines were residual heterozygous lines that were phenotypically segregating at F3:4 and F4:5.

For line 8-2, 8-6 and 101-7-17, the right side genotypes of the breaking point were homozygous resistant. The progenies of these lines were all resistant to soybean aphids. This matched with the genotypes on the right side of the breaking point. Likewise, the genotype of line 169-3 on the right of the breaking point was homozygous susceptible, corresponding with susceptible progenies. For lines 101-7-5, 101-8-6 and 65-2-1, the progeny phenotypes were segregating at a 1:2:1 ratio by Chi-square test (Table 3.4). This fitted the heterozygous genotypes at the right side of the breaking point. Also, the 1:2:1 ratio agreed with the inheritance pattern of *Rag*3d as a partially dominant gene. As indicated by arrows in Table 3.5, the results of these lines pushed the left border of the QTL region from 3.6Mb (MSUSNP 16-94) to 3.9Mb, from 3.9 Mb (MSUSNP16-96) to 5.3 Mb, and from 5.3 Mb (MSUSNP16-97) to 6.1 Mb (MSUSNP16-26) on chromosome 16. On the other side, the recombination breaking point in lines 65-2-1, 101-7-17 and 29-1 helped to move the right border from 9.2 Mb (MSUSNP16-88) to 7.1 Mb (MSUSNP16-

86). Thus, after the second round of fine mapping, the QTL region was narrowed to a 1020 kb region, between 6.1 Mb (MSUSNP16-26) and 7.1 (MSUSNP16-86) on chromosome 16.

For the third round of fine mapping, two thousand seven hundred and seven phenotypically and genotypically segregating residual heterozygous lines were used (two thousand two hundred and fifty-five F7:8, one hundred and thirty-four F8:9 and three hundred and eighteen F9:10). They were derived from the original fine mapping population 100040 and 100041. Three new markers in between 6.4 Mb (MSUSNP16-44) and 7.1 Mb (MSUSNP16-86), MSUSNP16-124, MSUSNP16-128 and MSUSNP16-134, were added. Marker MSUSNP16-10, MSUSNP16-44, MSUSNP16-128 and MSUSNP16-134 spanning a 400 kb region were used to genotype F7:8 lines. Marker MSUSNP16-44, MSUSNP16-124 and MSUSNP16-128 spanning an 80 kb region were used to genotype F8:9 and F9:10. F7:8 line 39-6-10-4-8, 15-4-1-4 and 39-2-1-12-31 indicated that the QTL was between MSUSNP16-44 and MSUSNP16-128. F8:9 and F9:10 lines 101-7-5-1-1-2, 39-2-3-14-6-1, 101-8-2-23-5-8-2, 15-4-1-6-3-17, 39-2-2-16-3-1-14 and 15-4-1-10-3-1 located the QTL between MSUSNP16-44 (6,438,676 bp) and MSUSNP16-124 (6,484,276 bp), spanning a 46 kb region. (Table 3.5)

Table 3.4: Chi-square test of progeny segregation ratio for recombinant lines in fine mapping study of PI 567585A

Generation [a]	Population [b]	line ID [c]	No. of progeny tested [d]	Aphid Phenotype [e]			1:2:1 ratio Chi-square Test *p* Value [i]	Segregation Pattern [j]
				R [f]	H [g]	S [h]		
F5:6	100040	8-2	26	26	0	0	<0.0001	F [k]
F5:6	100040	169-3	4	0	0	4	<0.0001	F
F5:6	100040	101-7-5	9	4	3	2	0.3889	S [l]
F5:6	100040	8-6	16	16	0	0	<0.0001	F
F5:6	100040	101-8-6	27	9	13	5	0.5427	S
F5:6	100041	65-2-1	8	2	5	1	0.6873	S
F5:6	100040	101-7-17	5	5	0	0	<0.0001	F
F5:6	100040	29-1	8	8	0	0	<0.0001	F
F7:8	100041	39-6-10-4-8	15	15	0	0	<0.0001	F
F7:8	100041	15-4-1-4	14	0	0	14	<0.0001	F
F7:8	100041	39-2-1-12-31	14	0	0	14	<0.0001	F
F8:9	100040	101-7-5-1-1-2	11	0	0	11	<0.0001	F
F8:9	100041	39-2-3-14-6-1	12	12	0	0	<0.0001	F
F9:10	100040	101-8-2-23-5-8-2	12	12	0	0	<0.0001	F
F9:10	100041	15-4-1-6-3-17	10	2	5	3	0.9048	S
F9:10	100041	39-2-2-16-3-1-14	6	3	2	1	0.3679	S
F9:10	100041	15-4-1-10-3-1	12	6	4	2	0.1353	S

[a] Generation of the recombinant lines

[b] The recombinant lines were selected from the listed population

[c] Line ID of the recombinant lines within their population

[d] The number of progeny rated for aphid resistance and used towards chi-square test

[e] Number of plants that has been categorized into R, H, S respectively

[f] Number of plants that were resistant to soybean aphids

Table 3.4 (cont'd)

[g] Number of plants that showed phenotype intermediate between resistant and susceptible to soybean aphids
[h] Number of plants that were susceptible to soybean aphids
[i] The *p* value of the chi-square test between the ideal segregation ratio 1:2:1 and the actual segregation for the progeny of each tested line. $\alpha = 0.05$
[j] The segregation pattern of each recombinant lines
[k] The progeny phenotype of the tested line is fixed (homozygous at the tested loci)
[l] The progeny phenotype of the tested line is segregating

Table 3.5: Recombination breakpoints among identified recombinants that mapped the position of *Rag*3d on Chromosome 16. Bold letters indicated the recombination breaking point. The arrow pointed to the side which the genotype agreed with the phenotype.

Line ID [a]	Pheno-type [b]	16-94 [g] 3,657,324	16-96 3,962,328	16-97 5,259,121	16-98 5,555,122	16-26 6,052,831	16-10 6,262,227	16-44* 6,438,676	16-124 6,484,276	16-128 6,517,204	16-134 6,624,879	16-86 7,072,168	16-88 9,244,011
8-2	*R* [c]		**H** [i] →	**R** [h]		R	R	R				R	
169-3	*S* [e]		**H** →	**S** [j]		S	S	S				S	
101-7-5	*H* [d]	S	S	**S** →	**H**	H	H	H				H	H
8-6	*R*		H	**H** →		**R**	R	R				R	
101-8-6	*H*	S	S	S	S	**S** →	**H**	H				H	H
65-2-1	*H*	R	R	R	R	**R** →	**H**	H				**H** ←	**S**
101-7-17	*R*	S	S	S	S	**S** →	**R**	R				**R** ←	**H**
29-1	*R*		R	R		R	R	**R**				← **H**	
39-6-10-4-8	*R*						R	**R**		← **S**	S		
15-4-1-4	*S*						S	**S**		← **R**	R		
39-2-1-12-31	*S*						H	**H** →		**S**	S		
101-7-5-1-1-2	*S*							S	**S**	← **H**			
39-2-3-14-6-1	*R*							R	**R**	← **H**			
101-8-2-23-5-8-2	*R*							**R**	← **H**	H			
15-4-1-6-3-17	*H*								**H**	← **R**			
39-2-2-16-3-1-14	*S*							**H**	← **S**	S			
15-4-1-10-3-1	*S*							H	**H**	← **R**			

*Markers developed through whole genome sequence SNP discovery pipeline

[a] Line ID of the recombinant lines within their population

Table 3.5 (cont'd)

[b] The corresponding phenotype (tested through progeny test, Table 3.4) of each recombinant lines
[c] The progenies of the recombinant line were resistant to soybean aphids.
[d] The progenies of the recombinant line were segregating for aphid resistance.
[e] The progenies of the recombinant line were susceptible to soybean aphids.
[f] The upper portion of the row displayed the marker name. The lower portion of the row presented the physical position of each marker on chromosome 16. The unit was in base pair (bp)
[g] For convenience and saving space, "MSUSNP" which was the first part of every marker's name, was not included in the table.
[h] Both SNP alleles of this loci was from the resistant parent.
[i] One SNP allele of this loci was from the resistant parent and another one was from the susceptible parent.
[j] Both SNP alleles of this loci was from the susceptible parent.

All the SNP marker positions, including those from the SoySNP50K chip, were designed based on the Williams 82 soybean genome version Glyma.Wm82.a1. The Williams 82 genome has been updated into Glyma.Wm82.a2. Sequence of flanking marker MSUSNP16-44 and MSUSNP16-124 were used to match the new genome and identify candidate genes for aphid resistance. The QTL candidate region shifted 147 kb towards the centromere with no inflation (46 kb). Six genes with predicted gene model or functions were located in the candidate region (Table 3.6). Among these six genes, only Glyma.16G066500 has no annotation information. The rest five genes could be candidate gene for aphid resistance from PI 567585A. These five genes were described in details in discussion of Chapter II.

Table 3.6: Genes with predicted gene model in *Rag*3d QTL interval

Locus Name	**Physical Position** (bp, Wm82.a2)	**Database ID**	**Annotation Description**
Glyma.16G066300	6587408-6590067	AT3G46610.1	Pentatricopeptide repeat (PPR-like) superfamily protein
		PTHR24015	FAMILY NOT NAMED
		PF01535	PPR repeat
Glyma.16G066400	6592258-6601109	AT5G40640.1	N. A.
Glyma.16G066500	6600655-6600822	N. A.	No Annotation Available
Glyma.16G066600	6606039-6612286	AT2G25940.1	alpha-vacuolar processing enzyme
		GO:0004197	cysteine-type endopeptidase activity
		GO:0006508	Proteolysis
		KOG1348	Asparaginyl peptidases
		PTHR12000	HEMOGLOBINASE FAMILY MEMBER
		PTHR12000:SF2	
		PF01650	Peptidase C13 family
Glyma.16G066700	6620330-6621988	AT5G14360.1	Ubiquitin-like superfamily protein
		GO:0005515	protein binding
		PTHR10666	UBIQUITIN
		PTHR10666:SF85	UBIQUITIN-LIKE PROTEIN 4A
		PF00240	Ubiquitin family
Glyma.16G066800	6627025-6628243	AT1G58190.2	receptor like protein 9
		GO:0005515	protein binding
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF474	PROTEIN KINASE DOMAIN-CONTAINING PROTEIN, CYTOPLASMIC
		PF00560	Leucine Rich Repeat

Fine mapping *Rag*3b in PI 567537

First round of fine mapping

Because *Rag*3b in PI 567537 was characterized as a dominant gene (Zhang et al. 2013), the progeny phenotypes of all the selected lines were categorized into resistant (R) and susceptible (S). The segregating ratio of each line was compared with the ideal ratio 3:1 by chi-square test. This information helped to distinguish whether the aphid resistance QTL was on the right or left side of the recombination breaking point. Detailed information was summarized (Table 3.7). The corresponding phenotype of each line determined by the progeny test was input in the second column with italic letters in Table 3.8. By comparing this with the phenotyping information, the side with candidate gene was identified by an arrow (Table 3.8).

Table 3.7: Chi-square test of progeny segregation ratio for recombinant lines in fine mapping study of PI567537

Generation [a]	Population [b]	line ID [c]	No. of progeny tested [d]	Aphid Phenotype [e]		3:1 ratio	Segregation
				R [f]	S [g]	Chi-square Test *p* Value [h]	Pattern [i]
F2	100048-5	87	8	6	2	1.0000	S [j]
F2	100047-4	210	11	8	3	0.8618	S
F2	100047-4	22	11	0	11	0.0000	F [k]
F2	100048-5	25	10	10	0	0.0679	F
F2:3	100047-4	56-4	12	0	12	0.0000	F
F2:3	100047-3	148-11	11	11	0	0.0555	F
F2	100047-4	116	16	13	3	0.5637	S
F2	100047-3	41	13	13	0	0.0374	F
F2	100047-4	199	9	0	9	0.0000	F
F2	100047-3	148	16	12	4	1.0000	S
F2	100047-4	93	12	9	3	1.0000	S
F2:3	100047-4	46-6	10	0	10	0.0000	F
F2:3	100047-4	157-8	9	9	0	0.0833	F
F2	100047-3	136	11	0	11	0.0000	F
F2	100047-4	139	12	12	0	0.0455	F
F2	100047-4	157	15	11	4	0.8815	S
F2	100047-4	54	14	10	4	0.7576	S
F2:3	100047-4	54-4	8	0	8	0.0000	F
F3:4	100047-4	54-2-3	8	12	3	0.6547	S
F3:4	100047-4	54-3-6	9	7	2	0.8474	S
F3:4	100047-4	157-3-5	13	13	0	0.0374	F
F3:4	100047-4	54-3-8	11	0	11	0.0000	F

[a] Generation of the recombinant lines

Table 3.7 (cont'd)

[b] The recombinant lines were selected from the listed population
[c] Line ID of the recombinant lines within their population
[d] The number of progeny rated for aphid resistance and used towards chi-square test
[e] Number of plants that has been categorized into R and S respectively
[f] Number of plants that were resistant to soybean aphids
[g] Number of plants that were susceptible to soybean aphids
[h] The *p* value of the chi-square test between the ideal segregation ratio 3:1 and the actual segregation for the progeny of each tested line. $\alpha = 0.1$
[i] The segregation pattern of each recombinant lines
[j] The progeny phenotype of the tested line is segregating
[k] The progeny phenotype of the tested line is fixed (homozygous at the tested loci)

Table 3.8: Recombination breakpoints among identified recombinants that mapped the position of *Rag*3b on Chromosome 16. Bold letters indicated the recombination breaking point. The arrow pointed to the side which the genotype agreed with the phenotype.

Marker & Position[f] (bp) / Line ID [a]	Pheno-type [b]	16-173 [g] 3,530,651		16-118 4,399,162		16-174 5,101,636		16-39 6,214,642		16-40 6,413,214		16-127 6,510,537
87	*H* [d]	**R** [h]	➡	**H** [i]		H		H				H
210	*H*	**S** [j]	➡	**H**		H		H				H
22	*S* [e]	**H**	➡	**S**		S		S				S
25	*R* [c]	**H**	➡	**R**		R		R				R
56-4	*S*	**R**	➡	**S**		S				S		S
148-11	*R*	**S**	➡	**R**		R				R		R
116	*H*	S		**S**	➡	**H**		H				H
41	*R*	H		**H**	➡	**R**		R				R
199	*S*	H		**H**	➡	**S**		S				S
148	*H*	S		S		**S**	➡	**H**				H
93	*H*	R		R		**R**	➡	**H**				H
46-6	*S*	R		R		**R**	➡			**S**		S
157-8	*R*	S		S		**S**	➡			**R**		R
136	*S*	H		H		**H**	➡	**S**				S
139	*R*	H		H		**H**	➡	**R**				R
157	*H*	S		S		S		**S**	➡			**H**
54	*H*	H		H		H		**H**			⬅	**R**
54-4	*S*	S		S		S		**S**			⬅	**R**
54-2-3	*H*							**H**		⬅ **R**		R
54-3-6	*H*							**H**		⬅ **R**		R
157-3-5	*R*							**S**	➡	**R**		R
54-3-8	*S*							S	➡	**S**	⬅	**R**

[a] Line ID of the recombinant lines within their population

[b] The corresponding phenotype (tested through progeny test, Table 3.7) of each recombinant lines

[c] The progenies of the recombinant line were resistant to soybean aphids.

[d] The progenies of the recombinant line were segregating for aphid resistance.

[e] The progenies of the recombinant line were susceptible to soybean aphids.

Table 3.8 (cont'd)

[f] The upper portion of the row displayed the marker name. The lower portion of the row presented the physical position of each marker on chromosome 16. The unit was in base pair (bp)
[g] For convenience and saving space, "MSUSNP" which was the first part of every marker's name, was not included in the table.
[h] Both SNP alleles of this loci was from the resistant parent.
[i] One SNP allele of this loci was from the resistant parent and another one was from the susceptible parent.

[j] Both SNP alleles of this loci was from the susceptible parent.

All the lines without dash are F2 lines with progenies tested for aphid resistance in the summer of 2014. SNP marker MSUSNP16-173 (3,530,651bp), MSUSNP16-118 (4,399,162bp), MSUSNP16-174 (5,101,636bp), MSUSNP16-39 (6,214,642bp) and MSUSNP16-127 (6,510,537bp) were used to genotype one hundred and seventy-three, two hundred and ten and a hundred and twenty-five F2 individuals from population 100047-3, 100047-4 and 100048-5, respectively. From table 3.8, the progenies of line 87 and 210 were segregating for soybean aphid resistance level. The progenies of line 22 and 25 were susceptible and resistant to the soybean aphid, respectively. For the above four lines, the corresponding genotypes to the phenotypes ('H', 'S' and 'R') were all on the right side of the breaking point, the right side of marker MSUSNP16-173. Similarly, for line 116, 41 and 199, the matching genotypes to the phenotypes were on the right side of the breaking point, the right side of marker MSUSNP16-118. For line 148, 93, 136 and 139, the right side of marker MSUSNP16-174 matched with the phenotypes. For line 157, the genotype right side of marker MSUSNP16-39, 'H', agreed with the segregating phenotype. Lastly, line 54 had a segregating phenotype. This corresponded with the left side genotype of marker MSUSNP 16-127. Through the first round of fine mapping, the left border of the aphid resistance QTL has been refined from MSUSNP16-173 to MSUSNP16-118 to MSUSNP16-174 to MSUSNP16-39. The right border remained at 16-127. The size of the candidate region was narrowed from 3Mb to 296kb. (Table 3.8)

Second round of fine mapping

Lines with one dash were selected from two hundred and eighty-nine F2:3 recombinant lines harvested in the summer of 2014. They were genotyped with marker MSUSNP16-173, MSUSNP16-118, MSUSNP16-174, MSUSNP16-40 (6,413,214bp) and MSUSNP16-127. The progeny of twenty-eight lines were tested for aphid resistance in the spring of 2015. Line 56-4 and 148-11 confirmed the QTL at the right side of MSUSNP16-173. Line 46-6 and 157-8 indicated the QTL at the right side of MSUSNP16-174. The progeny of line 54-4 were all susceptible to soybean aphids. This matched with the 'S' genotype on the left side of MSUSNP16-127. By the second round of fine mapping, the QTL candidate region was refined to the right side of MSUSNP16-39 (6,214,642bp) and the left side of MSUSNP16-127 (6,510,537bp). (Table 3.8)

Third round of fine mapping

Four hundred and forty-five F3:4 lines (line ID with two dashes) were genotyped with marker MSUSNP16-39, MSUSNP16-40 and MSUSNP16-127. The progeny of forty-seven lines were planted in the summer 2015 to test for aphid resistance. Progenies of line 157-3-5 were resistant to the soybean aphid. The genotypes 'R' at the right side of marker MSUSNP16-39 agreed with the phenotypes, verifying the QTL at the right side of marker MSUSNP16-39. For line 54-3-8, the progenies were susceptible to the soybean

aphid. This line validated the QTL at the left side of MSUSNP16-127. Progenies for line 54-2-3 and 54-3-6 were segregating for aphid resistance. The genotype ‘H’, on the left side of marker MSUSNP16-40 matched with the phenotype. To conclude, after three rounds of fine mapping, the *Rag*3b QTL candidate region was narrowed to a 199kb region between marker MSUSNP16-39 (6,214,642bp) and MSUSNP16-40 (6,413,214bp) (Table 3.8).

All the SNP marker positions, including those from the SNP chip, were designed based on the Williams 82 soybean genome version Glyma.Wm82.a1. Since the Williams 82 genome has been updated into Glyma.Wm82.a2, the sequence of marker MSUSNP16-39 and MSUSNP16-40 were used to BLAST against Glyma.Wm82.a2. Marker MSUSNP16-39 and marker MSUSNP16-40 shifted towards the centromere 43 kb and 147kb respectively, resulting in the QTL candidate region shifting towards the centromere with inflation (303 kb). Twenty-five genes with predicted gene model or functions were located in the candidate region (Table 3.9). Twenty-three of them had annotations. Twelve of them were annotated as leucine-rich repeat (LRR) receptor-like kinase family protein. As discussed in a previous chapter (Discussion, chapter II), aphid resistance gene identified so far were all nucleotide-binding site leucine-rich repeat (NBS-LRR) genes (Ohnishi et al. 2012; Bales 2013). Thus, these twelve LRR genes could be the candidate genes for *Rag*3b in PI 567537.

Table 3.9: Genes with predicted gene model in *Rag3b* QTL interval. Genes annotated with leucine rich repeat were labelled in bold.

Locus Name	**Physical Position** (bp, Wm82.a2)	**Database ID**	**Annotation Description**
Glyma.16g063500	6244922-6246453	AT4G12560.1	F-box and associated interaction domains-containing protein
		GO:0005515	protein binding
		PF00646	F-box domain
Glyma.16g063600	6255512-6260421	AT4G12560.1	F-box and associated interaction domains-containing protein
		GO:0005515	protein binding
		PF00646	F-box domain
		PF07734	F-box associated
Glyma.16g063700	6283066-6284827	AT5G14340.1	myb domain protein 40
		GO:0003682	chromatin binding
		KOG0048	Transcription factor, Myb superfamily
		PTHR10641	MYB-LIKE DNA-BINDING PROTEIN MYB
		PF00249	Myb-like DNA-binding domain
Glyma.16g063800	6285954-6286301	N. A.	N. A.
Glyma.16g063900	6293218-6299916	AT3G01720.1	N. A.
Glyma.16g064000	6307477-6311935	AT3G27350.2	N. A.
		PF06886	Targeting protein for Xklp2 (TPX2)
Glyma.16g064100	6322034-6325343	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase

Table 3.9 (cont'd)

			LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN
		PTHR24420	KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain
Glyma.16g064200	6333543-6337085	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase
			LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN
		PTHR24420	KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain
			P-loop containing nucleoside triphosphate hydrolases superfamily
Glyma.16g064300	6351138-6351481	AT5G60930.1	protein
Glyma.16g064400	6355500-6365869	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase

Table 3.9 (cont'd)

			LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN
		PTHR24420	KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
Glyma.16g064500	6365591-6366186	N. A.	N. A.
Glyma.16g064600	6368644-6368931	ATCG01280.1	Chloroplast Ycf2;ATPase, AAA type, core
		GO:0005524	ATP binding
		GO:0009507	chloroplast
		PF05695	Plant protein of unknown function (DUF825)
Glyma.16g064700	6370916-6373389	AT1G60600.1	UbiA prenyltransferase family protein
		AT1G60600.2	UbiA prenyltransferase family protein
		GO:0004659	prenyltransferase activity
		GO:0016021	integral component of membrane
			1,4-DIHYDROXY-2-NAPHTHOATE
		PTHR13929	OCTAPRENYLTRANSFERASE
		PF01040	UbiA prenyltransferase family
Glyma.16g064800	6379732-6383232	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase
			LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN
		PTHR24420	KINASE
		PTHR24420:SF472	

Table 3.9 (cont'd)

		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
Glyma.16g064900	6400179-6404141	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain
Glyma.16g065000	6424303-6427965	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain

Table 3.9 (cont'd)

Glyma.16g065100	6441260-6441580	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF472	
		PF07714	Protein tyrosine kinase
Glyma.16g065200	6441736-6445069	AT1G35710.1	Protein kinase family protein with leucine-rich repeat domain
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG0472	Leucine-rich repeat protein
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
Glyma.16g065300	6445882-6447544	ATCG01280.1	Chloroplast Ycf2;ATPase, AAA type, core
		PF05758	Ycf1
Glyma.16g065400	6461077-6466135	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation

Table 3.9 (cont'd)

		KOG1187	Serine/threonine protein kinase
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain
Glyma.16g065500	6484260-6487299	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG0472	Leucine-rich repeat protein
		KOG1187	Serine/threonine protein kinase
		PTHR24420	LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain
Glyma.16g065600	6501502-6505551	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase

Table 3.9 (cont'd)

			LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN
		PTHR24420	KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain
Glyma.16g065700	6527718-6531948	AT4G08850.1	Leucine-rich repeat receptor-like protein kinase family protein
		GO:0004672	protein kinase activity
		GO:0005515	protein binding
		GO:0005524	ATP binding
		GO:0006468	protein phosphorylation
		KOG1187	Serine/threonine protein kinase
			LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN
		PTHR24420	KINASE
		PTHR24420:SF472	
		PF00069	Protein kinase domain
		PF00560	Leucine Rich Repeat
		PF08263	Leucine rich repeat N-terminal domain
Glyma.16g065800	6533053-6535758	AT5G40670.1	PQ-loop repeat family protein / transmembrane family protein
		PTHR13131	CYSTINOSIN
		PF04193	PQ loop repeat
Glyma.16g065900	6555138-6559286	AT5G40670.1	PQ-loop repeat family protein / transmembrane family protein
		KOG2913	Predicted membrane protein
		PTHR13131	CYSTINOSIN
		PF04193	PQ loop repeat

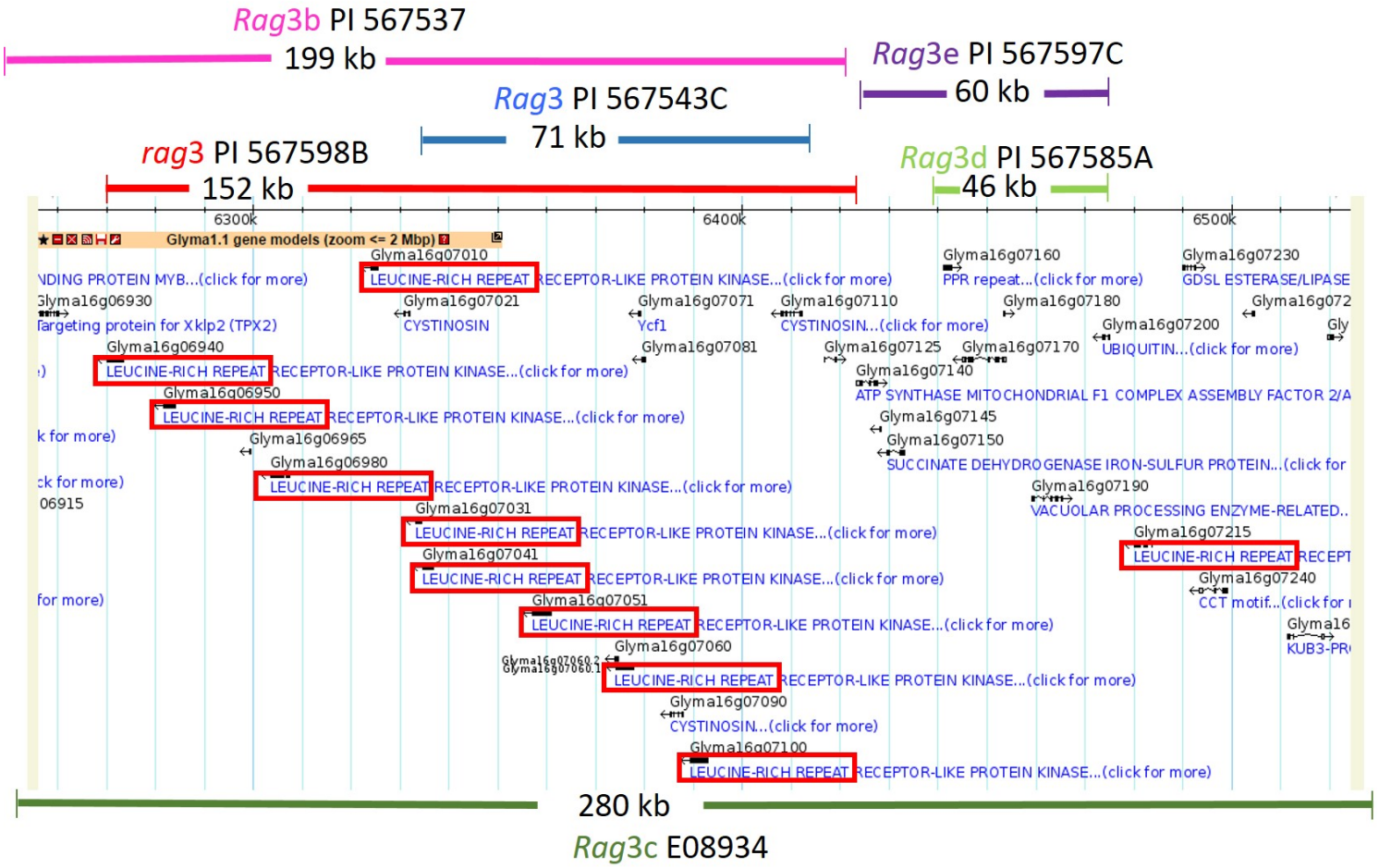
Discussion

As first verified in this study, the aphid resistance gene, *Rag*3d, in PI 567585A was located within the QTL mapping interval identified by Liu (2010). Utilizing the SoySNP50K chip and residual heterozygous lines, this study fine mapped *Rag*3d to a 46 kb (6,438,676 bp - 6,484,276 bp) region on chromosome 16. Fine mapped *Rag*3d QTL fell in the *Rag*3*/rag*3 region where several aphid resistance genes have been characterized from different resistance sources (Zhang et al. 2010; Zhang et al. 2013; Bales et al. 2013). Compared with *Rag*3 from PI 567543C, *rag*3 from PI 567598B, *Rag*3e from PI 567597C and *Rag*3b from PI 567537, *Rag*3d had a different inheritance pattern, a different resistance type and was resistant to different biotypes of soybean aphids (Table 1.1). The *Rag*3d QTL did not overlap with the first fine mapped *rag*3 gene in PI 567598B (6,270,557 bp to 6,423,098 bp), but on the right side (downstream) of *rag*3 (Bales 2013). For the other *Rag*3*/rag*3 genes, further information is needed to distinguish if *Rag*3d is controlled by a different gene or if *Rag*3d is allelic with other genes.

Following the QTL study of *Rag*3b by Zhang et al. (2013), the single dominant gene controlling aphid resistance from PI 567537 was fine mapped to a 199kb region on chromosome 16. As discussed earlier (Discussion, chapter II & III), *Rag*3b was a different gene from *Rag*3d and *Rag*3e. Interestingly, *rag*3 fell into the *Rag*3b QTL interval. Both

*rag*3 and *Rag*3b confer antibiosis resistance to aphid biotype 1,2 and 3. However, their inheritance pattern was significantly distinct. *rag*3 was a recessive gene whereas *Rag*3b was a dominant gene. *rag*3 and *Rag*3b may carry different alleles from the same gene. Further study is needed to confirm this assumption. Figure 3.3 is a summary of all *Rag*3/*rag*3 genes on chromosome 16. *Rag*3 from PI 567543C (Zhang 2012), *rag*3 from PI567598B and *Rag*3b from PI567537 were located in the same interval. QTL region of *Rag*3d from PI 567585A and *Rag*3e from PI 567597C overlapped with each other.

Figure 3.3: A summary of *Rag3*/*rag3* genes with LRR proteins highlighted on the Williams 82 genome assembly, Glyma1 (Schmutz et al., 2010)



Soybean genome went through two rounds of ancient duplications and diploidized eventually with 2n=40 (Roulin et al. 2013). It is not rare to find synteny within the soybean genome. The *Rag*3d QTL interval has a syntenic region on chromosome 19. This region includes QTLs responsible for white mold, flood tolerance and aluminum tolerance. Similarly, The *Rag*3b QTL interval has a syntenic region on chromosome 5 that contains QTL of SCN resistance (Grant et al. 2010). The different but related function of *Rag*3d and *Rag*3b syntenic regions may be due to sub-functionalization after polyploidization (Roulin et al. 2013). However, the *Rag*3/*rag*3 region enriched with predicted NBS-LRR genes (a 223kb region) on chromosome 16 do not have a syntenic region on other chromosomes (Grant et al. 2010), showing the uniqueness of this region. In fact, out of the 314 NBS-LRR genes in soybean genome, 40 of them are on chromosome 16. There were 19 disease resistance QTLs reported within 2Mb flanking region of these NBS-LRR genes. Chromosome 16 possessed the most number of NBS-LRR genes with most number of disease resistance QTLs (Kang et al. 2012). Besides soybean aphid resistance QTLs, there were whitefly resistance QTL and *Phytophthora sojae* resistance QTL in this Rag3/rag3 region (Grant et al. 2010). Thus, this *Rag*3/*rag*3 region on chromosome 16 is very important for disease and insect resistance in soybean.

When conducting the QTL mapping and fine mapping procedures, markers designed from the SoySNP50K chip and the NGS data were based on Williams 82 soybean genome

version Glyma.Wm82.a1. For consistency, version Glyma.Wm82.a1 SNP position was kept throughout the study. When matching the relative position to Glyma.Wm82.a2, the marker sequence was used to BLAST against the new assembly. The SoySNP50K Nomenclature Conversion Tool on Soybase was also a good reference. The positions of most SNP markers used in this study shifted 147kb towards the centromere when comparing Glyma.Wm82.a2 with Glyma.Wm82.a1. Importantly, their relative position did not change, which means the relative position of the fine mapping results were not influenced (Table 3.10). The candidate genes for *Rag*3d were exactly the same between the two versions, Glyma.Wm82.a1 and Glyma.Wm82.a2. However, the QTL interval of *Rag*3b was inflated in Glyma.Wm82.a2. It is possible that some repeat sequences were inserted or more genes were assembled and annotated in the new assembly. In general, with new nomenclature, the annotations were more precise and informative in Glyma.Wm82.a2 than in Glyma.Wm82.a1.

Comparing with fine mapping study for a partially dominant gene (*Rag*3d), fine mapping for a dominant gene (*Rag*3b) was more difficult. For a co-dominant gene, the “R” (homozygous alleles derived from the resistant parent), “H” (one allele from the resistant parent, one allele from the susceptible parent) and “S” alleles (homozygous alleles derived from the susceptible parent) expressed three different phenotypes, resistant to soybean aphids, intermediate and susceptible, respectively. Before a progeny test, a

proper recombinant line can be selected based on both the phenotype and the genotype. However, for a dominant gene, the “R” and “H” genotypes would express the same phenotype, resistant to soybean aphids. Recombinant lines can only be selected based on genotypes, adding more work for the next generation. The standard of categorizing the phenotypes influences the results as well. Soybean aphid resistance was measured quantitatively in this study. The rating distribution would normally be continuous (partially dominant) or scattered (dominant). Only a suitable cutoff (partially dominant: <2 resistant, 2-3 intermediate, >3 susceptible; dominant: <2.5 resistant, >2.5 susceptible) could help distinguish the phenotypic classes clearly.

In summary, *Rag*3d from PI 567585A and *Rag*3b from PI 567537 have been fine mapped to a 46kb interval on chromosome 16 with five candidate genes and a 199kb region on chromosome 16 with twelve NBS-LRR genes as candidate genes, respectively. The fine mapped *Rag*3d from PI 567585A and *Rag*3b from PI 567537 have been utilized in the soybean breeding program at Michigan State University. The flanking markers from the fine mapping study were used for marker assisted selection, saving time, space, labor and money (Du and Wang 2014). These markers could also help to identify the resistance gene and characterize the *Rag*3/*rag*3 region on chromosome 16 to understand the mechanisms of aphid resistance in soybean.

APPENDIX

APPENDIX

Table 3.10: SNP markers used in fine mapping study of *Rag*3d in PI 567585A. SNPs in corresponding wild-type and mutant-alleles are in brackets [ ]

SNP assay ID	SoySNP50K chip ID	Chrom osome	Physical position (bp) [a]	Target sequence [b]
MSUSNP16-10	Gm16_6262227_C_T	16	6,262,227	5'GTTGGGCTATGTCCAAAATAGTATCCCCATTAGTTAGT ATCCCATGATGTCATGAGGTGTAAACTTGTTAAGACATAT CAAACTTAGGGTTTAAGTTAAC[C/T]AGATCCGAAAAAG CTGCCACTATAGTGCCTTCTCTTTGAGTATGTGGTAATTA TTGATTGAAGGCTTGATTGAAGGATCATCCTCATAGCTTA GGTTTTG -3'
MSUSNP16-26	Gm16_6052831_T_C	16	6,052,831	5'CTGGCAGGCTACCACTAGTGGTCGCGCCTGGGGCCCAC CACTAGTGGTCGCGCCTGGCAGGCCACCACTTTCACCTCT GTCCCATCGTCCTGTCAAGTCA[T/C]GACATGTGTCGCG TTCTGGTGGAATGCGCCCCTCAGAAAAGCGCTTTGTAGTA AAATAACAGACCCCCTTGATAAATAAAAATGAAACAGACC CATTTTA -3'
MSUSNP16-44	*	16	6,438,676	5'CCTTAAGAAGGATTCTCAAAAGTTTACTTTTAGCTCCA ACAAGACATGTTCTTACATCTAAGCCCAACCAAACAAAAA TAGAAAAACCAAATTTTAAATT[T/C]TTTATTATCAACC TCATGATCACCATGTCTACCACGATTTATCCATGGTTGTG

Table 3.10 (cont'd)

MSUSNP16-86	Gm16_7072168_A_C	16	7,072,168	TTTGGTTATCAATTTTAGCTTTTTCATCAATTTTGGTTAA TAATTTT -3' 5'TATTTAGCTTTGAATCATTATTGGGTATCCAAAACGTG AGAACACTTGAAGCTGCAGCGAACATGGCCTTGTGAGGGG CCCACTTCAAGCACGAAGGAAC[A/C]CCAATATGGCTGG TCCAACATGCCACCTGCAGGGAAGAAAACCAGAGGCTCAT CATCGCAAAAATTTTGCGCTATTACATTGACAAAATATAA AATATTG -3'
MSUSNP16-88	Gm16_9244011_A_G	16	9,244,011	5'ATTCCCACTCTCCTCACTGAATTTCATGCTACCCCCAC TGGGGGCCACTCCGGTGTCGCCAAGACCATTGCTCGTGTT TCTGAGAATTTTTATTGGCCCA[A/G]CCTTCGCGATGAT GTTGCTACTTTTGTGGTCAATTGCCTTGATTGCCAGTCCA CCAAATATGAGGCAAAGAAGCTTGCTGGTCTCCTATGCTC ACTTCCA -3'
MSUSNP16-94	Gm16_3657324_G_A	16	3,657,324	5'ACGCTTGCATTTAAATTTAATAATTTTATTACGATAAA ACGAGCAATAATAGACATAGTACAGCAGGATCTTGCTAAA TGCAATGAATTGGGAGGAAGTT[G/A]CTTGCTTCTCCCA ACTTGCTTCCTAGAGGCCATGATGCATTTGGATAATTGTT GAACACGTTACGACGCAGAAAAAAGAAAAATTAGAAACCT CATGATC -3'
MSUSNP16-96	Gm16_3962328_C_A	16	3,962,328	5'ACTGACCACATCATCTCCCAGAAGTGGTCAAGGTTCTC TCATTAGAGGCCACAGCATTGAGGGTCATTGAAGGTGAAG GAAGAGATCCTGAACCAAGAGC[C/A]GGAGCAGAAGCAG TTTCGGTTGCCAACTCTTCGTTTTCATGTGAAGGTTGTAA TGGTTTTGGTGAGTAGGTGAAGGAAAGATCTTTGTTGGCA GCAAGAG -3'
MSUSNP16-97	Gm16_5259121_A_G	16	5,259,121	5'TTGAGAATTCAACAATCACGACTATTAAAAGCTCCATG AGATCCACAACAAGTGGTGGTGCAGTTTTGGGTCCTGGGG

Table 3.10 (cont'd)

MSUSNP16-98	Gm16_5555122_T_C	16	5,555,122	TGGTGACGCAATTTCTGGTGGT[A/G]CAGGGTGACGACA CAACGATGATGGTGGTCAGGGAGGTGGGTGGTGCAATGAT GTGGCAGCACGGTTCTAGTGGTGCAAGGTGGGCCTAATGC ACAAATT -3'
MSUSNP16-124	Gm16_6484276_A_G	16	6,484,276	5'ACTGGAAGACCTAAAGATTGAAAACATCTCTATGCCTG ACGAGTTCGATTCTGAACTCCTGATTGAGAAGCTACCAAA GTCCTAGATAGATTATAAACAA[T/C]AATTAAAGCACAA ACACAAATAGATGTTACGACCAGACCTTATCATCCACATT ATCATTGAAGATACAAGCAGGAAGGAAAATGTTATTGCAA GGACCAA -3'
MSUSNP16-128	Gm16_6517204_T_C	16	6,517,204	5'ATCAAATTATCATGATTAATTATTAAACTTACGACAAA ATTAATATATATACTTATATTACTTATACCAATTTGTGCG GGTCAAATACTTTGACTATAAT[A/G]ACCGATCCAATTC TTTAAATTATTTATAAGTGAAAAAAAAATAACCTCCTATA TTATATCAGATCATTTAATTTATATATTTGGTTTCACATT TCAAATA -3'
MSUSNP16-134	Gm16_6624879_C_T	16	6,624,879	5'CAATTCTTGTGGCGATCCTCTAGGCTTATCAAACCTCG TTGTTTGTGGTGCCTATTTGATTTTCTATCTTACACATTT TATGTTCGTGTCGGCATTTCTG[T/C]GCTCTCCAAATTG TGCCTTCAGCCCGATTTTTCTCAGGTTCGTTCTGCACGTT TTTACGAATCTGTTTTTAGTGTTTCTTGACCCTAAATTAC AGTTCAA -3'
				5'GATGTATCTTGTGTGGTGGCGGTGGTGGCCCAAGGCCG CGGTGTGTCGCGTGACTGCGTGAGTCGTGTCCACGGTGAG GAGAAGAAGATGAGAAGAAATG[C/T]TGTAAGAGGAGAA GAATAAAGCAAGGTACTAGTCCTTAAAGTGGTACTAGTCC AATGGTTCTTAAAGTGAAAAAGAAAAAATCCAACGTACTA AAACTAA -3'

Table 3.10 (cont'd)

[a] Genomic position of single nucleotide polymorphism on the Williams 82 genome assembly, Glyma1(Schmutz et al. 2010).
[b] Target sequence for KASP custom design with 100-bp upstream and downstream of the single nucleotide polymorphism.
*Markers developed through whole genome sequence SNP discovery pipeline

Table 3.11: SNP markers used in fine mapping study of *Rag3b* in PI 567537. SNPs in corresponding wild-type and mutant-alleles are in brackets [ ].

SNP assay ID	SoySNP50K chip ID	Chromosome	Physical position (bp) [a]	Target sequence [b]
MSUSNP16-39	Gm16_6214642_C_T	16	6,214,642	5'ACACGATTGAAGAAAATTGAAAAAGAAATAC TACTACTATTGAAGAAAGTTGAAAAAGAAATAC CAGTACCTTGTCCATTTCTGCTTCTTTTGCGGG TTG[C/T]GGAAAGTGTTCCAATTTAATGCTGG TCCTAAGTCCTAACCAAGTTATCAAGATTCAAG CCGTGGCTTGCAGGTAATATTTAACTCTCTGTT TAGACTTTA -3'
MSUSNP16-40	Gm16_6413214_A_G	16	6,413,214	5'TGACATTTTTATAAATATATTATCTTTTTAT ATGCCATGTGCATGGCGCGTGACACATTCAACA ATGTTCATTGGGTAGCCCGTCTTAGTAGGTTAC GCA[A/G]CAGGTAAGTTAAGACGATGTATTTG AAAACACTAGAAATTTTGAATGTTAACGACGTT TTGTTGAACCACCGTCGTTAACATTGAAGTTTA TAAGTTTTT -3'
MSUSNP16-118	Gm16_4399162_A_G	16	4,399,162	5'GAGCATTACAAATAAACTTCACTTGTTTGGT GAGTGTTAGAATAAAGCACGAAAGTGTGAAGCT AAGTATTTTTAGCTGAAGAGAATGGATGTCAGG AAA[A/G]CATGGAACGTGAAAACAAAAACACT TAACCACCCAAACAAATGAAAAGAAAGAACCTA AAACTGAGCTTCTTTAACAAGAGGGCCTAGAGA TACAACTAC -3'
MSUSNP16-127	Gm16_6510537_A_G	16	6,510,537	5'CTCCAAGACTAGACGAACTCTTCAAGCTTTT CTCCAACTCCAAAACTCACTAAAAAACCTCACA

Table 3.11 (cont'd)

MSUSNP16-173	Gm16_3530651_G_A	16	3,530,651	AAATCAATAACTTTTCTCTACTTGGTACTAGTA GCT[A/G]GTGTGAAATGAGCAATGGTTGAGGC TCTATTTGCAGGGGCAGATGAAGGTCCTAGAAG GTGTTGCCTGAAGCTTGGTCTAGGGAAGATGGC AAGGATGGC -3'
MSUSNP16-174	Gm16_5101636_A_G	16	5,101,636	5'TTTCAACCTACCATCCTGGTTGGACGTCACC TCTACCATGTGCATCTTCCCTTGACCTTCATTT TCCTGCCATTGTGACTCCATGTGATGTTCAGAC ACT[G/A]TCACACTTGCACGCCCATGAGAATA TCTCCCTTCTTCCTCTTGTTACTCTGCACATGC ACCTGAGTGATTAAAGGAGCAATAACCTCTTTG TTAATAGTC -3'
				5'TTTCATTATAAAAAATGGAGAAATAAGTAAG AAAGAAAATGATCTAGAATAACATCAAGTTGTG AATTTTTTTTTGTGTGTGTACTAAGATTGATAA TAA[A/G]GATAAAACTAAACTAAGTCCTCATG CATTTATTTCAACCCAACCATTAGACCGACTCT AGTAGGTCTTAGGAATCCAATTAGAATGTATTA TAAAAAAAA -3'

[a] Genomic position of single nucleotide polymorphism on the Williams 82 genome assembly, Glyma1 (Schmutz et al., 2010).

[b] Target sequence for KASP custom design with 100-bp upstream and downstream of the single nucleotide polymorphism.

REFERENCES

REFERENCES

Bales C, Zhang G, Liu M, et al (2013) Mapping soybean aphid resistance genes in PI 567598B. Theor Appl Genet 126:2081–2091

Bales CJC (2013) Fine mapping of aphid resistance genes in soybean plant introduction (PI) 567598B. Ph.D. dissertation, Michigan State University

Bansal R, Mian MAR, Michel AP (2013) Identification of Novel Sources of Host Plant Resistance to Known Soybean Aphid Biotypes. J Econ Entomol 106:1479–1485

Bhusal SJ, Jiang G-L, Hesler LS, Orf JH (2014) Soybean Aphid Resistance in Soybean Germplasm Accessions of Maturity Group I. Crop Sci 54:2093

Bhusal SJ, Jiang G-L, Tilmon KJ, Hesler LS (2013) Identification of Soybean Aphid Resistance in Early Maturing Genotypes of Soybean. Crop Sci 53:491

Du W, Wang D (2014) Marker Assisted Selection as a Powerful Strategy for Gene Pyramiding in Soybean Breeding. WENYANs Teach. WORLD. http://wdmsu.weebly.com/marker-assisted-selection.html. Accessed 14 Mar 2016

Grant D (2015) What's New at SoyBase? Plant and Animal Genome. https://pag.confex.com/pag/xxiii/webprogram/Paper16170.html. Accessed 7 Apr 2016

Grant D, Nelson RT, Cannon SB, Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 38:D843–D846

Hanson AA, Orf JH, Koch RL (2016) Sources of Soybean Aphid Resistance in Early-Maturing Soybean Germplasm. Crop Sci 56:154

Hesler LS (2013) Resistance to soybean aphid among wild soybean lines under controlled conditions. Crop Prot 53:139–146

Hesler LS, Dashiell KE, Lundgren JG (2007) Characterization of resistance to Aphis glycines in soybean accessions. Euphytica 154:91–99

Hill CB, Kim K-S, Crull L, et al (2009) Inheritance of Resistance to the Soybean Aphid in Soybean PI 200538. Crop Sci 49:1193

Johnson KD, O'Neal ME, Bradshaw JD, Rice ME (2008) Is preventative, concurrent management of the soybean aphid (Hemiptera: Aphididae) and bean leaf beetle (Coleoptera: Chrysomelidae) possible? J Econ Entomol 101:801–809

Johnson KD, O'Neal ME, Ragsdale DW, et al (2009) Probability of cost-effective management of soybean aphid (Hemiptera: Aphididae) in North America. J Econ Entomol 102:2101–2108

Jun T-H, Mian MAR, Michel AP (2012b) Genetic mapping revealed two loci for soybean aphid resistance in PI 567301B. Theor Appl Genet 124:13–22

Kang S-T, Rouf Mian MA, Hammond RB (2008) Soybean Aphid Resistance in PI 243540 Is Controlled by a Single Dominant Gene. Crop Sci 48:1744

Kang YJ, Kim KH, Shim S, et al (2012) Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol 12:139

Kim K-S, Bellendir S, Hudson KA, et al (2010a) Fine mapping the soybean aphid resistance gene Rag1 in soybean. Theor Appl Genet 120:1063–1071

Kim K-S, Hill CB, Hartman GL, et al (2010b) Fine mapping of the soybean aphid-resistance gene Rag2 in soybean PI 200538. Theor Appl Genet 121:599–610

Li Y, Hill CB, Carlson SR, et al (2006) Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. Mol Breed 19:25–34

Liu G, Xing H, Diao Y, et al (2014) Identification of Resistance to Soybean Aphids in Early Germplasm. Crop Sci 54:2707

Liu M (2010) Inheritance of resistance and molecular mapping of soybean aphid resistance genes in soybean PI 567585A; identification of aphid resistance genes in soybean using modified nested association mapping (MNAM). Ph.D. dissertation, Michigan State University

Mensah C, DiFonzo C, Wang D (2008) Inheritance of soybean aphid resistance in PI 567541B and PI 567598B. Crop Sci 48:1759

Mensah C, DiFonzo C, Nelson RL, Wang D (2005) Resistance to Soybean Aphid in Early Maturing Soybean Germplasm. Crop Sci 45:2228

Mian MAR, Hammond RB, St. Martin SK (2008) New Plant Introductions with Resistance to the Soybean Aphid. Crop Sci 48:1055

Ohnishi S, Miyake N, Takeuchi T, et al (2012) Fine mapping of foxglove aphid (Aulacorthum solani) resistance gene Raso1 in soybean and its effect on tolerance to Soybean dwarf virus transmitted by foxglove aphid. Breed Sci 61:618–624

Painter RH (1951) Insect resistance in crop plants. Macmillan Publishing Co, New York

Ragsdale DW, Landis DA, Brodeur J, et al (2011) Ecology and Management of the Soybean Aphid in North America. Annu Rev Entomol 56:375–399

Ragsdale DW, McCornack BP, Venette RC, et al (2007) Economic Threshold for Soybean Aphid (Hemiptera: Aphididae). J Econ Entomol 100:1258–1267

Roulin A, Auer PL, Libault M, et al (2013) The fate of duplicated genes in a polyploid plant genome. Plant J 73:143–153

Schmutz J, Cannon SB, Schlueter J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Semagn K, Babu R, Hearne S, Olsen M (2013) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 33:1–14

Song Q, Hyten DL, Jia G, et al (2013) Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. PLoS ONE 8(1):e54985

Van Ooijen (2006) JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B. V., Wageningen, Netherlands

Wang S, Basten CJ, Zeng ZB (2012) Windows QTL Cartographer 2.5. Dep. Stat. N. C. State Univ. Raleigh NC http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

Xiao L, Hu Y, Wang B, Wu T (2013) Genetic mapping of a novel gene for soybean aphid resistance in soybean (Glycine max [L.] Merr.) line P203 from China. Theor Appl Genet 126:2279–2287

Zhang G, Gu C, Wang D (2013) Mapping and validation of a gene for soybean aphid resistance in PI 567537. Mol Breed 32:131–138

Zhang G, Gu C, Wang D (2009) Molecular mapping of soybean aphid resistance genes in PI 567541B. Theor Appl Genet 118:473–482

Zhang G, Gu C, Wang D (2010) A novel locus for soybean aphid resistance. Theor Appl Genet 120:1183–1191

Zhang Z (2012) QTL mapping of Phytophthora root rot resistance and aphid resistance in soybean. Ph.D. dissertation, Michigan State University