Theor Appl Genet (2010) 120:1063–1071
DOI 10.1007/s00122-009-1234-8

ORIGINAL PAPER

Fine mapping the soybean aphid resistance gene *Rag1* in soybean

Ki-Seung Kim · Stephanie Bellendir · Karen A. Hudson · Curtis B. Hill · Glen L. Hartman · David L. Hyten · Matthew E. Hudson · Brian W. Diers

Received: 23 July 2009 / Accepted: 30 November 2009 / Published online: 25 December 2009

Abstract The soybean aphid (*Aphis glycines* Matsumura) is an important soybean [*Glycine max* (L.) Merr.] pest in North America. The dominant aphid resistance gene *Rag1* was previously mapped from the cultivar 'Dowling' to a 12 cM marker interval on soybean chromosome 7 (formerly linkage group M). The development of additional genetic markers mapping closer to *Rag1* was needed to accurately position the gene to improve the effectiveness of marker-assisted selection (MAS) and to eventually clone it. The objectives of this study were to identify single nucleotide polymorphisms (SNPs) near *Rag1* and to position these SNPs relative to *Rag1*. To generate a fine map of the *Rag1* interval, 824 BC_4F_2 and 1,000 BC_4F_3 plants segregating for the gene were screened with markers flanking *Rag1*. Plants with recombination events close to the gene were tested with SNPs identified in previous studies along with new SNPs identified from the preliminary Williams 82 draft soybean genome shotgun sequence using direct re-sequencing and gene-scanning melt-curve analysis. Progeny of these recombinant plants were evaluated for aphid resistance. These efforts resulted in the mapping of *Rag1* between the two SNP markers 46169.7 and 21A, which corresponds to a physical distance on the Williams 82 8× draft assembly (Glyma1.01) of 115 kilobase pair (kb). Several candidate genes for *Rag1* are present within the 115-kb interval. The markers identified in this study that are closely linked to *Rag1* will be a useful resource in MAS for this important aphid resistance gene.

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1234-8) contains supplementary material, which is available to authorized users.

K.-S. Kim · S. Bellendir · C. B. Hill · M. E. Hudson · B. W. Diers (✉)
Department of Crop Sciences,
University of Illinois, Urbana, IL 61801, USA
e-mail: bdiers@illinois.edu; bdiers@uiuc.edu

K. A. Hudson
USDA-ARS, Purdue University,
West Lafayette, IN 47907, USA

G. L. Hartman
USDA-ARS, University of Illinois, Urbana, IL 61801, USA

D. L. Hyten
Soybean Genomics and Improvement Laboratory,
USDA-ARS, Beltsville, MD 20705, USA

Abbreviations

bp	Base pair
kb	Kilobase pair
LG	Linkage group
MAS	Marker-assisted selection
MCA	Melting curve assay
PCR	Polymerase chain reaction
SSR	Simple sequence repeat
SNP	Single nucleotide polymorphism
STS	Sequence tagged site

Introduction

Soybean aphid is a relatively new soybean pest in North America and was first observed on the continent in 2000 (Hartman et al. 2001). The soybean aphid was initially found in northern soybean growing regions of the US where it causes the greatest economic losses. The aphid has spread to 23 soybean growing states reaching as far south as Mississippi and Georgia in 2005 and to three Canadian

provinces (Venette and Ragsdale 2004). Kim et al. (2008) showed that there is genetic diversity in soybean aphid when they identified two distinct soybean aphid biotypes in North America based on their interaction with aphid resistance sources. Michel et al. (2009) also recently showed diversity in soybean aphid using microsatellite markers.

Focused research on the genetics of resistance to soybean aphid was initiated after its discovery in North America. Hill et al. (2006a, b) identified a single dominant gene named *Rag1* in Dowling and the *Rag* gene in Jackson and both were mapped to the same position on soybean chromosome 7 [formerly linkage group (LG) M] (Li et al. 2007). Mensah et al. (2008) reported that two recessive genes controlled soybean aphid resistance in PI 567541B and PI 567598B. Zhang et al. (2009) identified that the resistance in PI 567541B was controlled by two quantitative trait loci (QTL) that mapped to soybean chromosomes 7 and 13 (LG F). Recently, a soybean aphid resistance gene named *Rag2* was mapped to chromosome 13 from PI 243540 (Mian et al. 2008) and a resistance gene from PI 200538 was mapped to the same region (Hill et al. 2009).

Marker-assisted selection (MAS) has many advantages compared to phenotypic selection in breeding programs. MAS can be performed in segregating populations during early generations as well as at early stages of plant development. This can allow breeders to conduct many cycles of selection in a year for resistance without the natural occurrence or the necessity of inoculum maintenance of the pest or pathogen (Mohan et al. 1997). MAS also allows breeders to pyramid resistance genes without having to inoculate plants with specific isolates that can differentiate these genes. However, the essential requirements for MAS in a crop-breeding program are that the markers should co-segregate with genes controlling the desired trait, a high-throughput means of testing breeding populations with markers is needed, and the marker screening technique should be economical to use (Gupta et al. 1999; Francia et al. 2005).

Single nucleotide polymorphism (SNP) markers are becoming widely used in soybean breeding and research (Hyten et al. 2009b) because of their high frequency, widespread distribution throughout the genome (Choi et al. 2007), and their suitability for high-throughput automated genotyping (Hyten et al. 2008; Lee et al. 2004; Schork et al. 2000). Zhu et al. (2003) found 3.68 SNPs per kilobase pair (kb) in 25 diverse soybean genotypes. Hyten et al. (2006) suggested that the frequency of sequence variants in soybean is lower than other plant species due to historic genetic bottlenecks, subsequent intensive selection, and low sequence diversity in *Glycine soja* Sieb. and Zucc., the wild ancestor of soybean.

The *Rag1* gene from Dowling was mapped to a 12 centi-Morgan (cM) region on soybean chromosome 7 between the simple sequence repeat (SSR) markers Satt435 and Satt463 by Li et al. (2007). Other available SSR markers in this interval were tested, however, none were polymorphic between the parents of the crosses used to map this gene. To find additional markers near *Rag1*, Kaczorowski et al. (2008) hybridized nuclear DNA of the recurrent parent 'Dwight', the donor parent Dowling, and a pair of backcross derived isolines that differed for the *Rag1* region, onto Affymetrix soybean GeneChip microarrays. These hybridizations revealed 15 single feature polymorphisms (SFPs) closely linked to *Rag1*, of which 12 were confirmed through sequence analysis. SNP genotyping assays were developed and four SNPs were mapped to the *Rag1* region.

The objective of this study was to fine map the location of *Rag1*. This fine mapping will be useful for MAS as markers identified during this process that are closely linked to *Rag1* will almost perfectly segregate with the gene. In addition, the fine mapping will aid in gene cloning efforts by positioning the gene into a small interval containing few candidate genes.

Materials and methods

Plant material

The fine mapping was initiated by first identifying recombinants near *Rag1* in populations of BC_4F_2 plants that were segregating for the soybean aphid resistance gene. Populations were developed through four backcrosses using the maturity group (MG) VIII cultivar Dowling (PI 548663) (Craigmiles et al. 1978) as the donor parent of *Rag1* (Hill et al. 2004; Li et al. 2007) and the MG II cultivar Dwight (PI 587386) (Nickell et al. 1998) as the aphid-susceptible recurrent parent.

The BC_4F_2 populations were developed by first crossing Dowling and the MG II cultivar Loda (Nickell et al. 2001). *Rag1* was initially mapped in this F_2 population (Li et al. 2007). An aphid resistant F_2 plant was selected and four backcrosses to Dwight were then performed. The marker Satt435 was used to select for *Rag1* during the backcrossing process. Several BC_4F_1 plants were grown and BC_4F_2 seed were planted in the field at Urbana, IL in 2006. A total of 824 BC_4F_2 plants were screened with the SSR markers Satt463 and Satt540, which flank *Rag1*. One hundred and eleven plants with recombination events between the markers were selected and harvested for retesting with additional markers.

After the first set of recombinants was analyzed, a set of 1,000 BC_4F_3 plants that had the same pedigree as the BC_4F_2 plants described above were screened with two SNP markers to identify new recombination events close to *Rag1*. Plants used in this second screening were derived from five

BC_4F_2 plants from the first set that were heterozygous for Satt463, Satt435, and Satt540. BC_4F_3 plants were initially screened with TaqMan markers developed for the SNPs ss107918249 and ss107913360 which flank *Rag1*. Selected recombinants were then genotyped with additional markers which mapped closer to *Rag1*. From these screenings, one plant with a key recombination event was selected and grown to produce seed.

DNA extraction and quantification

Genomic DNA from the 824 BC_4F_2 plants and additional 1,000 BC_4F_3 plants was extracted from leaves prior to full expansion by a quick DNA extraction method (Bell-Johnson et al. 1998). Genomic DNA from the 111 selected $BC_4F_{2:3}$ recombinant lines was extracted from young trifoliolate leaf tissue bulked from 12 BC_4F_3 progeny plants using the CTAB method described by Keim and Shoemaker (1988) with the following modifications: an incubation time of 90 min, re-suspension of the DNA pellet in 500 µl 1× TE, and no RNase A treatment. After the completion of aphid resistance bioassays for the selected $BC_4F_{3:4}$ recombinant line and 11 selected $BC_4F_{2:3}$ recombinant lines, genomic DNA from each of the 44 plants in the bioassay was extracted by the CTAB method as described above. All CTAB DNA was quantified and diluted as described by Kaczorowski et al. (2008).

Genetic mapping and development of marker assays in the *Rag1* interval

Mapping and development of markers near *Rag1* was done in three rounds. The first round of mapping was carried out with the SSR markers Satt463, Satt540, and Satt435, which were previously found closely linked to *Rag1* (Li et al. 2007). The primer sequences for the SSR markers were available from the Choi et al. (2007) Soybean Linkage Map (http://bfgl.anri.barc.usda.gov/cgi-bin/soybean/Linkage.pl; accessed on May 27, 2009). Polymerase chain reaction (PCR) was performed according to Cregan and Quigley (1997) and gel electrophoresis was conducted as described by Kaczorowski et al. (2008). The second round of linkage analysis was carried out with the seven SNP markers 46169.7, 65906.2, 7623, 86377, 442-1688, ss107918249, and ss107913360. The first four of these markers were developed through hybridization of nuclear DNA onto Affymetrix soybean GeneChip microarrays (Kaczorowski et al. 2008). The fifth marker, 442-1688, was developed by re-sequencing PCR products using primers designed from the sequence of an early draft of the soybean genome sequence. The remaining markers, ss107918249 and ss107913360, were developed by re-sequencing sequence tagged sites (STSs) (Hyten et al. 2009a).

The physical location of the SSR and SNP markers on chromosome 7 was determined from a BLAST search of the primer and consensus sequences of the markers onto the soybean genome sequence available from the Soybean Genome Project, Department of Energy's Joint Genome Institute (http://www.Phytozome.net). Early genomic SNP discovery was performed using pre-release versions of the soybean draft genome sequence at 4× and 7× coverage, which was kindly supplied by Jeremy Schmutz, Joint Genome Institute, Stanford University Genome Sequencing Center.

Single nucleotide polymorphisms were genotyped with TaqMan SNP assays and MCA using a SNP-specific melt-curve probe with the LightCycler® 480 System, Roche Diagnostics, Indianapolis, IN, USA at the University of Illinois Genetic Marker Center (Supplementary Tables 1, 2 in Electronic Supplementary Material). TaqMan assays and MCAs were performed as described by Kaczorowski et al. (2008).

Linkage analysis was conducted with JoinMap 3.0 software (Van Ooijen and Voorrips 2001) using the Kosambi mapping function. A logarithm (base 10) of the odds (LOD) score of 5.0 was used as the threshold to group markers into LG. All 824 BC_4F_2 plants were tested with the SSR markers Satt463 and Satt540 and a Chi-square (χ^2) test was used to evaluate segregation of both markers. The 111 BC_4F_2 plants with recombination events between the two SSR markers were screened with the SSR marker Satt435. Plants with recombination events detected between Satt463 and Satt435 or between Satt435 and Satt540 were then tested with all of the SNP markers in the recombinant intervals. For the plants without recombination events in an interval, the genotypes for the markers flanking the interval were used to predict SNP or SSR marker genotypes within the interval. This prediction of the SNP data across these non-recombinant regions may have resulted in missing double recombinants in the intervals, which would have resulted in incorrect map constriction, but this would likely be rare enough to ignore.

Each plant in the $BC_4F_{2:3}$ or $BC_4F_{3:4}$ lines that was evaluated for aphid resistance test was also screened with a segregating marker from the *Rag1* interval. Genetic associations between the markers and aphid resistance where analyzed by single-factor analysis of variance with the PROC GLM procedure of SAS (SAS Institute 2002).

Re-sequencing of the *Rag1* interval based on the draft soybean genome sequence

The third round of linkage analysis was carried out by identifying SNPs in selected intervals between SNP marker ss107918249 and Satt435 based on the draft soybean genome sequence. This region was targeted for re-sequencing

based on results from the second round of linkage analysis. SNPs were identified through direct re-sequencing or melt-curve analysis followed by re-sequencing. Primer pairs were designed using Perl scripts (available on request) developed to identify primer pairs at 10 kb spacing across large intervals, or the IDT SciTools PrimerQuest^{SM} software tool for single primer pairs, and were ordered from Integrated DNA Technologies (IA, USA). The uniqueness of each primer pair was checked by BLAST search against the soybean draft genome sequence available at the time of design.

For direct re-sequencing of the target region, gel electrophoresis of the PCR products was initially run to verify that a single PCR product was produced from each primer pair. If primer pairs produce no product or multiple products, they were re-amplified with either a lower (no product) or a higher (multiple products) annealing temperature (Choi et al. 2007). After gel electrophoresis on a 0.9% TAE gel, PCR products from the two parents were purified with the QIAquick Gel extraction kit (Qiagen, CA, USA). Purified PCR products were sequenced from both ends using the same primers as used for PCR amplification with the ABI BigDye Terminator v3.1 cycle sequencing kit on an ABI PRISM 3730 sequencer (Applied Biosystems, Foster City, CA, USA) at the University of Illinois Keck Center Core Facility. To detect SNPs between the two parents, ABI trace files were analyzed by Sequencher version 4.7 (Gene Codes Corporation, Ann Arbor, MI, USA). Sequences of confirmed SNPs were used to design target amplification primers and probes for TaqMan assays or MCAs.

SNP discovery with melt-curve analysis

Polymerase chain reaction primer pairs generating products <200 bp in size and with a 10 kb approximate spacing across the *Rag1* region of the draft soybean genome sequence were designed using a Perl script (available from the authors) and synthesized by Integrated DNA Technologies, IA, USA. PCR and post amplification melt analysis of samples was preformed in 10-μL reaction volumes in 384 well plates on the Roche LightCycler® 480 System (Roche Applied Science), with a pre-incubation at 95°C 10 min hold followed by 45 cycles of 95°C 10 s hold, 54°C 15 s hold, and 72°C 20 s hold, with a single fluorescent reading at each extension step. Reactions contained 3 mM $MgCl_2$ and a final concentration of 1× LightCycler® 480 High Resolution Master Mix (Roche Applied Science). DNA amount for all reactions was 62.5 ng. Heteroduplex samples were created by spiking with 12.5 ng reference DNA (20% of total DNA). The primers were at 0.25 μM final concentration each. Melting analysis was performed with the Roche LightCycler® 480 immediately following the PCR amplification with an additional denaturation at 95°C 1 min hold, cooling at a programmed rate of 2.5°C/s to 40°C with a 1 min hold, and a continuous melting curve fluorescent acquisition during a 1°C/s ramp to 95°C. A derivative melting curve plot was obtained with the use of the LightCycler® 480 Gene Scanning software (Roche Applied Science).

Soybean aphid resistance bioassays

Eleven $BC_4F_{2:3}$ recombinant lines with unique recombination events in the *Rag1* interval selected from the 111 BC_4F_2 recombinants in the second round of linkage analysis and one $BC_4F_{3:4}$ line selected in the third round analysis were evaluated for aphid resistance in choice tests. These tests were conducted in an environmental plant growth chamber with temperatures ranging from 22 to 25°C and 14 h daily illumination at 30 μmol m^{-2} s^{-1} photosynthetically active radiation. Individual plants were grown in 60 × 60 × 60 mm plastic 48-pot inserts (Hummert Intl., Earth City, MO, USA) contained inside plastic trays without holes (Hummert Intl., no. F1020) as described by Hill et al. (2004). Each 48-pot insert included 44 plants from the selected recombinant $BC_4F_{2:3}$ lines or $BC_4F_{3:4}$ line and two replicates of the parents, Dowling and Dwight. The 48 plants in an insert were arranged in a complete randomized design (CRD). Aphid inoculation was conducted by placing leaves of Williams 82 that were each infested with 50–200 aphids on top of V_E-stage soybean seedlings (Fehr et al. 1971). The aphids included all summer aphid stages of soybean aphid biotype 1 which was collected in Illinois (Kim et al. 2008; Hill et al. 2009). Soybean aphid colonization was evaluated at 10 and 15 days after aphid infestation by counting the total number of aphids on each plant (Kim et al. 2008).

Results

Analysis of recombination events in the *Rag1* region

The *Rag1* locus was previously mapped between the SSR markers Satt540 and Satt463 on soybean chromosome 7 (formerly LG M) (Li et al. 2007). To more finely map the location of *Rag1*, we first tested 824 BC_4F_2 plants with these two markers to identify genetic recombinants near the gene in the first round of linkage analysis. The BC_4F_2 plants were developed by backcrossing *Rag1* into Dwight and the BC_4F_1 plant used to develop the population was heterozygous for *Rag1*, resulting in segregation of the gene in the population. Results from a χ^2 test for these markers in the BC_4F_2 population show that Satt540 fit a 1:2:1 segregation ratio ($\alpha = 0.05$) for a co-dominant marker but Satt463 did not. The lack of fit for Satt463 was the result of identifying fewer heterozygous plants for this marker than expected,

with there being 230 homozygous resistant, 366 heterozygotes, and 222 homozygous susceptible plants in the population. Satt463 did fit a 3:1 segregation ratio when the homozygous resistant and heterozygotes are combined into a single class. The mechanism causing the segregating distortion for Satt463 is not known, however, it is unlikely that the distortion was caused by marker scoring errors because this marker gives a clear banding pattern.

One hundred and eleven BC_4F_2 plants with recombination events between Satt463 and Satt540 were selected and screened with markers from recombinant intervals in the second round of linkage analysis. The markers tested include the SNP markers 46169.7, 65906.2, 7623, 86377, 442-1688, ss107918249, and ss107913360, and the SSR marker Satt435. The distance between Satt463 and Satt540 in the population was 7.4 cM and the largest marker interval was 4.7 cM between Satt463 and ss107918249 (Fig. 1). The relative order of the markers and genetic distances between them was consistent with their physical position on the Williams 82 chromosome 7 draft genome sequence (http://www.phytozome.net/soybean).

Eleven of the 111 $BC_4F_{2:3}$ recombinant lines were selected for aphid resistance testing based on the presence of unique recombination events in the *Rag1* region. Results from evaluating these selected lines for resistance and with the three SSR and seven SNP markers indicates that *Rag1* is in a 435 kb interval between ss107918249 and Satt435 (Table 1). The right most boundary of the position of the gene was identified through the analysis of lines 12 and 72. Both lines segregate for the region left of the recombination point and there was a significant ($P < 0.0001$) association between aphid resistance and segregation of the SNP marker 46169.7, indicating that *Rag1* must be to the left of ss107918249. The left border of the position of *Rag1* was demonstrated to be between SNP marker 65906.2 and Satt435. This border is shown by lines 6 and 100. Both lines are segregating for 65906.2 and genetic regions to the right of this marker and there was a significant association between aphid resistance and 65906.2 in both lines, showing that *Rag1* was segregating in both and therefore right of Satt435.

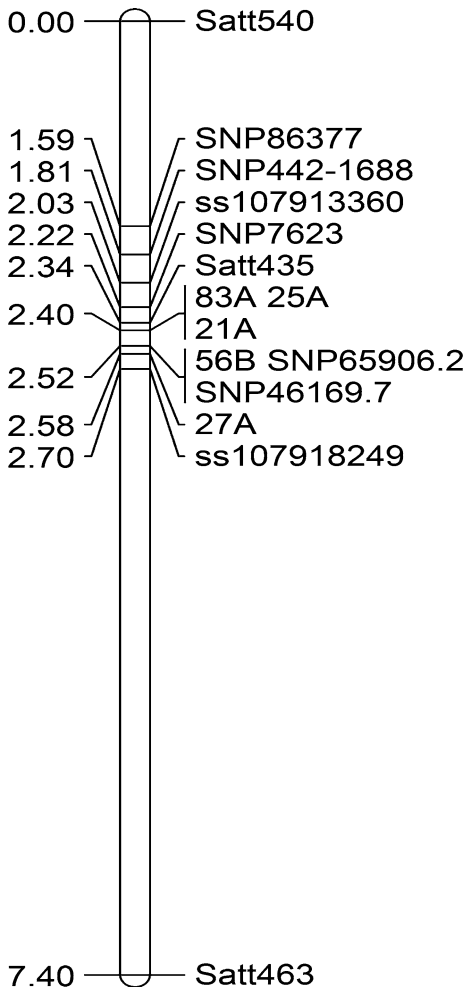

Fig. 1 Genetic linkage map for the interval between Satt463 and Satt540 on soybean chromosome 7 [formerly linkage group (*LG*) M] based on 824 BC_4F_2 plants

Saturation of the *Rag1* region with additional SNP markers

The region of the soybean draft genome sequence defined by the markers described above as containing the *Rag1* locus did not contain a sufficient number of known SNPs to narrow the locus to the smallest interval possible with the large mapping population available. Thus, in the third round of mapping additional SNP markers were developed that could be used to better define the *Rag1* region using one of two methods: dye-terminator re-sequencing or melt-curve SNP discovery. A total of 79 primer pairs were developed that produced products of approximately 1 kb in size, and these were directly sequenced using capillary dye-terminator chemistry. While PCR and sequencing from intergenic regions of the soybean genome was frequently not successful, this method produced 11 new SNP markers. To supplement the SNPs detected by direct sequencing, 20 primer pairs were designed which produced products ≤200 bp pairs in size. These primers were spaced at 10 kb intervals throughout a 200 kb region, and produced much more robust PCR products from intergenic DNA than the 1 kb products used for the direct sequencing approach. Using the Roche LightCycler® melt-curve SNP detection system, these products were screened for polymorphisms indicated by the formation of heteroduplexes in the presence of Dwight DNA. The LightCycler® Gene Scanning software predicted five of the 20 products contained polymorphisms, and four products were confirmed to contain polymorphisms by direct sequencing. The 15 products which were not predicted to contain polymorphisms were also sequenced, and of these one was found to contain a

Table 1 Marker genotypes and soybean aphid resistance reactions of 11 $BC_4F_{2:3}$ lines and one $BC_4F_{3:4}$ line selected based on locations of recombination events in each line

Physical position (Mb)[a]	4.963	5.055	5.141	5.227	5.421	5.464	5.468	5.479	5.493	5.546	5.578	5.608	5.645	5.899	8.244			Aphid number[b]				
First round markers	Satt540					Satt435									Satt463							
Second round markers		SNP 86377	SNP 442-1688	ss 107913360	SNP 7623						SNP 65906.2	SNP 46169.7		ss 107918249								
Third round markers							83A	25A	21A	56B			27A									
Line																Marker used in *F* test	Phenotype[c]	R	H	S	$P > F$[d]	R^2[e]
12	H[f]	H	H	H	H	H	H	H	H	H	H	H	*H*	*S*	S	SNP 46169.7	Segregating	90	117	629	<0.0001	0.94
72	R	H	H	H	H	H	H	H	H	H	H	H	*H*	*S*	S	SNP 46169.7	Segregating	80	68	510	<0.0001	0.96
4	R	R	R	R	R	R	R	R	R	R	R	*R*	*H*	H	H	ss107918249	Resistant	62	47	57	0.2043	0.08
82	H	H	H	H	H	H	H	H	*H*	*R*	R	R	R	R	R	21A	Resistant	49	58	55	0.2071	0.08
100	R	R	R	R	R	R	R	R	*R*	*H*	H	H	H	H	H	56B	Segregating	71	62	657	<0.0001	0.97
6	R	R	R	R	R	*R*	*H*	H	H	H	H	H	H	H	H	83A	Segregating	47	60	628	<0.0001	0.91
48	H	H	H	H	*H*	*S*	NT	NT	NT	NT	S	S	NT	S	S	SNP 7623	Susceptible	941	953	984	0.2080	0.17
7	H	H	H	*H*	*R*	R	NT	NT	NT	NT	R	R	NT	R	R	ss107913360	Resistant	233	308	291	0.3805	0.11
62	H	H	H	*H*	*R*	R	NT	NT	NT	NT	R	R	NT	R	R	ss107913360	Resistant	233	308	291	0.3805	0.11
73	H	H	*H*	*R*	R	R	NT	NT	NT	NT	R	R	NT	R	R	SNP 442-1688	Resistant	240	280	212	0.2155	0.13
92	H	*H*	*R*	R	R	R	NT	NT	NT	NT	R	R	NT	R	R	SNP 86377	Resistant	225	184	193	0.4319	0.09
K39[g]	S	S	S	S	S	S	S	S	S	S	*S*	H	H	H	H	SNP 46169.7	Susceptible	774	761	770	0.9010	0.01

[a] Physical position of the markers based on the Williams 82 chromosome 7 genome sequence of the 8× draft assembly (Glyma1.01) available at http://www.phytozome.net. The Mb positions of the SNP markers correspond to the SNP locations and the position of the SSR markers are the locations of the marker end sequence

[b] Mean number of aphids on each plant predicted to be homozygous resistant (R), heterozygous (H), and homozygous susceptible (S) for *Rag1* based on the segregation of the marker listed two columns to the left

[c] Phenotype of the line based on aphid numbers and the marker association test

[d] Significance level of the marker association

[e] R^2 value of the marker association

[f] Marker genotypes of the BC_4F_2 or BC_4F_3 plants that formed the recombinant lines; R designates that the plant was homozygous for the allele from Dowling, H designates heterozygous, S designates homozygous for the allele from Dwight, and NT designates not tested. Italicized values indicate genetic interval containing inferred recombination event

[g] The line K39 is a $BC_4F_{3:4}$ recombinant line selected from the 1,000 BC_4F_3 heterozygous plants. All other lines in the table are $BC_4F_{2:3}$ lines

SNP polymorphism not detected by the melt-curve screen and the other 14 were identical in Dowling and Dwight. Thus, the empirically determined false discovery rate was 20% and the false negative rate was 7% assuming there was no error in the marker validation through the direct sequencing.

The melt-curve method added a further five polymorphisms (four SNPs and one indel) to the 11 SNPs discovered by direct sequencing. Of these 16 polymorphisms within the interval, five primer pairs were chosen based on their location for development of genetic markers. Three of these pairs (56B, 21A, and 83A) contained SNPs that were appropriate for the development of MCA or TaqMan assays. Neither assay could be developed for the two additional primer pairs (27A and 25A) because of additional deletion polymorphisms near these SNPs. Therefore, the lines were genotyped for these two markers by sequencing bulked DNA from the 44 plants in the resistance assays. The six lines with recombination events between ss107918249 and Satt435 were screened with these SNP markers resulting in a more detailed positioning of the recombination events (Table 1). Data from these new markers adjusted the right most position of *Rag1* to 27A based on line 4 and the left most position to 21A based on lines 82 and 100. This analysis places *Rag1* into a 152 kb region.

To further narrow the genomic region containing the *Rag1*, a new set of 1,000 BC_4F_3 plants segregating for the gene were screened with SNP markers to identify additional recombination events, especially recombination events in the interval between SNP markers 27A and 21A. From the 1,000 plants screened, only one new recombination event in this interval was found. This plant had a recombination between 46169.7 and 65906.2 and genetic analysis of progeny from this plant positioned *Rag1* left of 46169.7

A BLAST analysis reveals that SNP marker 46169.7 and 21A, the two markers that flank *Rag1*, are at the nucleotide positions 5,608,084 and 5,492,694 on Soybean chromosome 7, respectively (http://www.phytozome.net/soybean). As before, the physical positions of all of the markers used on the Williams 82 draft genome sequence are consistent with their genetic position in the map of the *Rag1* locus, confirming that the assembly of the genome sequence in this region is correct as far as can be assessed using genetic linkage with these markers, and that the Dowling and Williams 82 genomes do not differ by any large-scale deletion, rearrangement or insertion events within this region. The cognate genomic region of the aphid-susceptible genotype Williams 82 to the region of the Dowling genome containing the *Rag1* locus, as defined by SNP markers 46169.7 and 21A, is approximately 115 kb (0.12 cM) in length (Table 1; Fig. 1).

Discussion

A fine map of the region containing the *Rag1* locus was developed. This mapping effort was greatly accelerated by the availability of the soybean genome sequence from the DoE Joint Genome Institute (http://www.phytozome.net/soybean). This sequence information was especially valuable for developing SNP markers through re-sequencing of targeted regions and identifying what genes underlay the interval where the gene is located.

Current gene annotation of the 115 kb region containing the *Rag1* locus on the Williams 82 genome sequence (http://www.phytozome.net/soybean) predicts the presence of 13 genes in this region, some of which have their expression supported by EST data. The predicted genes include two nucleotide binding leucine-rich repeat (NBS-LRR) genes and each is a potential candidates for *Rag1*. These NBS-LRR candidate genes are Glyma07g06890 and Glyma07 g06920 and they have homology with Arabidopsis genes encoding disease resistance proteins such as the resistance to *P. syringae*2 (*RPS2*) and TIR-NBS-LRR class (http://www.phytozome.net/soybean). The NBS-LRR genes are good candidates for the *Rag1*-encoding locus because the root knot nematode (*Meloidogyne incognita*) resistance gene in tomato, *Mi*, is an NBS-LRR gene (Milligan et al., 1998) and this gene also confers resistance to potato aphid (*Macrosiphum euphorbiae* Thomas) (Rossi et al. 1998). Another example of an aphid resistance gene that is an NBS-LRR is the *Vat* gene, which confers resistance to *Aphis gossypii* in melon (*Cucumis melo*). This gene was recently cloned and complimented through transformation (Dogimont et al. 2009). In addition, an aphid resistance gene was mapped in a NBS-LRR cluster in *Medicago truncatula* (Klingler et al. 2005).

Li et al. (2008) compared gene expression of the *Rag1* source Dowling with the aphid-susceptible cultivar Williams 82 using microarrays after infesting plants with soybean aphids. They found that changes in gene expression patterns suggesting that resistance in Dowling is mediated by an incompatible like interaction that is consistent with a gene for gene model. This supports the possibility that *Rag1* is one of the two NBS-LRR genes identified in the 115 kb interval and the gene activates this incompatible reaction. Further research is necessary to determine if one of the candidate loci within this interval corresponds to *Rag1*, since this analysis is based on the sequence of Williams 82, an aphid susceptible line without *Rag1* resistance.

We used TaqMan, and MCA to genotype SNPs in this study, and direct re-sequencing and melt-curve analysis to discover SNPs. The melt-curve analysis with the LightCycler® Gene Scanning system successfully predicted four of five PCR products that contained SNP polymorphisms, with a fifth detected by direct sequencing. However, direct

sequencing of the products is still necessary in order to develop a direct assay for the polymorphisms (whether MCA or TaqMan assays), and the reagent costs of the Gene Scanning system, while lower than the approximately $2 per reaction for direct sequencing, are still relatively high (around $0.40 per reaction). Thus, the use of Gene Scanning to discover new SNPs may be most cost-effective in regions or genomes where SNPs are rare, and thus many products must be screened in high throughput before the few that are predicted to be polymorphic are analyzed by direct sequencing.

There are many considerations in selecting a SNP genotyping assay, and the choice usually depends on cost per data point and simplicity of data acquisition (Lee et al. 2004). To obtain reliable and reproducible SNP genotyping results, we found that the MCA required relatively high quality and quantified DNA extracted using the CTAB extraction method, while the TaqMan assays produced high-quality genotypic data even with quick-extracted DNA (Bell-Johnson et al. 1998). For screening of the 1,000 BC_4F_3 plants for recombinants, we used two SNP markers designed for the TaqMan assay because the quick-extraction method required just 3 min to extract DNA from 96 samples. In addition, the quick-extraction method requires a relatively small amount of tissue, so the tissue sampling can be done much earlier during plant development than with the CTAB method. A minor drawback of the TaqMan assay compared to the MCA is that TaqMan assays take approximately 2 h while MCA takes 1 h. However, with a 384 well instrument, we were able to test a few thousand samples daily. On the basis of time and labor, the TaqMan assay with quick-extract DNA was the most efficient for MAS due to its simplicity and rapidity when a large number of samples need to be tested with a small number of markers. However, another drawback of the TaqMan assays is the reagent costs which were USD $0.40 per reaction compared to USD $0.13 for MCA in our study. For fine-scale mapping when markers are tested on a few key recombinants, MCA was also efficient because the lower cost to set up each marker compared to TaqMan assays.

Our high-resolution genetic map of the *Rag1* locus will facilitate MAS for *Rag1* in cultivar breeding programs. SNPs are becoming the marker of choice in MAS and the markers we identified within the 115 kb region that are closely linked to *Rag1* will be highly effective tools in MAS since only very rare recombinations are anticipated between these markers and the *Rag1* locus during selection. Our efforts to fine map the *Rag1* locus will also greatly facilitate the molecular identification and functional characterization of the *Rag1* locus. The cloning of *Rag1* will contribute to a better understanding of the mechanism of soybean aphid defense. This information could be used to identify other candidate soybean aphid resistance genes using the completed soybean genome sequence, for comparison to cloned insect resistance genes in other species or possibly to introduce aphid resistance into susceptible genotypes using biotechnology approaches.

The appearance of soybean aphid biotype variation in North America (Kim et al. 2008) increases the need to stack aphid resistance genes in cultivars. The breakdown of resistance genes by insect pests has frequently occurred, especially when resistance is conditioned by a single gene and cultivars carrying the gene are widely grown (Burd and Porter 2006; Haley et al. 2004). Stacking resistance genes may slow down or delay resistance gene adaptation. The close linkage between *Rag1* and the SNP markers we identified and the availability of efficient SNP marker detection assays will make these markers especially useful in MAS and the stacking of *Rag1* in combination with other aphid resistance genes.

Acknowledgments This work was supported by soybean check-off funding from the Illinois Soybean Association, the United Soybean Board, and the USDA CSREES grant #AG2006-34488-16915 to MEH and BD.

References

Bell-Johnson B, Garvey G, Johnson J, Lightfoot D, Meksem K (1998) Biotechnology approaches to improving resistance to SCN and SDS: methods for high-throughput marker-assisted selection. Soybean Genet Newsl 25:115–117

Burd JD, Porter DR (2006) Biotypic diversity in greenbug (Hemiptera: Aphididae): characterizing new virulence and host associations. J Econ Entomol 99:959–965

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685–696

Craigmiles JP, Hartwig EE, Sij JW (1978) Registration of 'Dowling' soybeans. Crop Sci 18:1094

Cregan PB, Quigley CV (1997) Simple sequence repeat DNA marker analysis. In: Caetano-Anolles G, Gresshoff PM (eds) DNA markers: protocols, applications and overview. Wiley, New York, pp 173–185

Dogimont C, Bendahmane A, Pitrat M, Burget-Bigeard E, Hagen L, Le Menn A, Pauquet J, Rousselle P, Caboche M, Chovelon V (2009) Gene resistant to *Aphis gossypii*. US Patent 7,576,264, 18 Aug 2009

Fehr WR, Caviness CE, Burmood DT, Pennington JS (1971) Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci 11:929–931

Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Valè G (2005) Marker assisted selection in crop plants. Plant Cell Tissue Organ Cult 82:317–342

Gupta P, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390

Haley SD, Peairs FB, Walker CB, Rudolph JB, Randolph TL (2004) Occurrence of a new Russian wheat aphid biotype in Colorado. Crop Sci 44:1589–1592

Hartman GL, Domier LL, Wax LM, Helm CG, Onstad DW, Shaw JT, Solter LF, Voegtlin DJ, D'Arcy CJ, Gray ME, Steffy KL, Orwick PL (2001) Occurrence and distribution of *Aphis glycines* on

soybean in Illinois in 2000 and its potential control. http://www.plantmanagementnetwork.org/pub/php/brief/aphisglycines/. Accessed 06 Mar 2009

Hill CB, Li Y, Hartman GL (2004) Resistance to the soybean aphid in soybean germplasm. Crop Sci 44:98–106

Hill CB, Li Y, Hartman GL (2006a) A single dominant gene for resistance to the soybean aphid in the soybean cultivar Dowling. Crop Sci 46:1601–1605

Hill CB, Li Y, Hartman GL (2006b) Soybean aphid resistance in soybean Jackson is controlled by a single dominant gene. Crop Sci 46:1606–1608

Hill CB, Kim KS, Crull L, Diers BW, Hartman GL (2009) Inheritance of resistance to the soybean aphid in soybean PI 200538. Crop Sci 49:1193–1200

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM et al (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103:16666–16671

Hyten DL, Song Q, Choi I-Y, Yoon M-S, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Hyten DL, Choi I-Y, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang E-Y, Matukumalli LK, Cregan PB (2009a) A high density integrated genetic linkage map of soybean and the development of a 1,536 Universal Soy Linkage Panel for QTL mapping. Crop Sci (in press)

Hyten DL, Smith J, Frederick RD, Tucker ML, Song Q, Cregan PB (2009b) Bulked segregant analysis using the GoldenGate assay to locate the *Rpp3* locus that confers resistance to soybean rust in soybean. Crop Sci 49:265–271

Kaczorowski KA, Kim KS, Diers BW, Hudson ME (2008) Microarray-based genetic mapping using soybean near-isogenic lines and generation of SNP markers in the *Rag1* aphid-resistance interval. Plant Genome 1:89–98

Keim P, Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. Soybean Genet Newsl 15:150–152

Kim KS, Hill CB, Hartman GL, Mian MAR, Diers BW (2008) Discovery of soybean aphid biotypes. Crop Sci 48:923–928

Klingler J, Creasy R, Gao LL, Nair RM, Calix AS, Jacob HS, Edwards OR, Singh KB (2005) Aphid resistance in *Medicago truncatula* involves antixenosis and phloem-specific, inducible antibiosis, and maps to a single locus flanked by NBS-LRR resistance gene analogs. Plant Physiol 137:1445–1455

Lee SH, Walker DR, Cregan PB, Boerma HR (2004) Comparison of four flow cytometric SNP detection assays and their use in plant improvement. Theor Appl Genet 110:167–174

Li Y, Hill CB, Carlson SR, Diers BW, Hartman GL (2007) Soybean aphid resistance in the soybean cultivars Dowling and Jackson map to linkage group M. Mol Breed 19:25–34

Li Y, Zou J, Li M, Bilgin DD, Vodkin LO, Hartman GL, Clough SJ (2008) Soybean defense responses to the soybean aphid. New Phytol 179(1):185–195

Mensah C, DiFonzo C, Wang D (2008) Inheritance of soybean aphid resistance in PI 567541B and PI 567598B. Crop Sci 48:1759–1763

Mian MAR, Kang ST, Beil SE, Hammond RB (2008) Genetic linkage mapping of the soybean aphid resistance gene in PI 243540. Theor Appl Genet 117:955–962

Michel AP, Zhang W, Jung JK, Kang ST, Mian MAR (2009) Population genetic structure of *Aphis glycines*. Environ Entomol 38:1301–1311

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM (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 10:1307–1320

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed 3:87–103

Nickell CD, Noel GR, Cary TR, Leitz RA (1998) Registration of 'Dwight' soybean. Crop Sci 38:1398

Nickell CD, Noel GR, Cary TR, Thomas DJ, Diers BW (2001) Registration of 'Loda' soybean. Crop Sci 41:589–590

Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM (1998) The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. Proc Natl Acad Sci USA 95:9750–9754

SAS Institute (2002) The SAS system for Windows. Release 9.00. SAS Institute, Cary

Schork NJ, Fallin D, Lanchbury JS (2000) Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet 58:250–264

Van Ooijen JW, Voorrips RE (2001) JoinMap 3.0 software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Venette RC, Ragsdale DW (2004) Assessing the invasion by soybean aphid (Homoptera: Aphididae) where will it end? Ann Entomol Soc Am 97:219–226

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

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