Development of genic-SSRs markers from soybean aphid sequences generated by high-throughput sequencing of cDNA library

T.-H. Jun1,2, M. A. R. Mian1,2,3, K. Freewalt3, O. Mittapalli1 & A. P. Michel1

1 Department of Entomology, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH, USA
2 Department of Horticulture and Crop Science, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH, USA
3 USDA-ARS, Wooster, OH, USA

Abstract

The soybean aphid (Aphis glycines Matsumura) is one of the most important insect pests of soybean [Glycine max (L.) Merr.] in North America, and three biotypes of the aphid have been confirmed. Genetic studies of the soybean aphid are needed to determine genetic diversity, movement pattern, biotype distribution and mapping of virulence genes for efficient control of the pest. Simple sequence repeats (SSR) markers are useful for population and classical genetic studies, but few are currently available for the soybean aphid. In this study, we designed primers for 342 genic-SSR markers from a dataset of more than 102,024 transcript reads generated by 454 GS FLX sequencing of a cDNA library of the soybean aphid. Two hundred forty-six markers generated PCR products of expected size and 26 were polymorphic among four pooled aphid DNA samples. An additional five markers that were fixed for two alleles among the pooled samples were found to be polymorphic when tested on 96 individual aphids. Sequencing of the PCR products generated by two polymorphic SSR markers revealed that the polymorphisms were strictly because of variations in the SSR repeats among the aphids tested. The genetic diversity among 96 soybean aphids, 24 each from two field collections (South Dakota and Michigan, biotype unknown) and two laboratory colonies [biotype 1 (B1) and biotype 2 (B2)], was assessed with 29 polymorphic SSR markers. These markers discriminated laboratory colonies from field collections and field collections from different states. The genic-SSR markers developed and validated in this study will be a significant addition to the limited number of SSR markers, mostly genomic-SSR, currently available for the soybean aphid. These markers will be useful for genetic studies, including population genetics, genetic and QTL mapping, migration and biotype differentiation of the soybean aphid.
and yellowing (Hill et al. 2004; Krupke et al. 2005; Mensah et al. 2005). Soybean aphids can also transmit several plant viruses [e.g. * Alfalfa mosaic virus*, *Soybean dwarf virus* and *Soybean mosaic virus*] to soybean and cause the development of sooty mould because of honeydew excreted by soybean aphid, which inhibits photosynthesis (Iwaki et al. 1980; Hartman et al. 2001; Hill et al. 2001; Mueller and Grau 2007).

Host-plant resistance is an effective method for controlling insect pests, which also reduces the use of chemical insecticide applications and is suitable for organic soybean production. Recently, at least three soybean aphid resistance genes (*Rag1*, *Rag2* and *Rag3*) have been identified and mapped on three different soybean chromosomes (7, 13 and 16) (Li et al. 2007; Mian et al. 2008; Zhang et al. 2010). However, biotypes of soybean aphid have been identified in North America that can overcome host-plant resistance provided by these genes. As *Rag* resistance was found from Asian varieties, and soybean aphid’s native range is Asia, these biotypes were likely part of the initial invasion. For example, biotype 2 of soybean aphid, capable of defeating the *Rag1* gene, has been found in Ohio and biotype 3, able to colonize plants with the *Rag2* resistance gene, was identified from the overwintering host * Frangula alnus* in Indiana (Kim et al. 2008a; Hill et al. 2010). As a result, the presence of such soybean aphid biotypes can reduce efficacy and sustainability of host-plant resistance. Therefore, successful deployment of soybean aphid resistance genes depends on using them in areas where certain biotypes are non-existent, less likely to appear or develop. However, the knowledge of overall genetic diversity, population structure, biotype frequencies or abundance among North American soybean aphid populations is very limited, mainly because of the lack of informative molecular markers (Michel et al. 2009a,b). Thus, development of PCR-based co-dominant molecular markers is needed to clarify relationships among the soybean aphid populations and monitor biotype migration.

Molecular genetic markers have been developed and utilized for the study of population genetic structure of many insect species. In particular, simple sequence repeats (SSRs) or microsatellites are widely used markers because of their high abundance in the genome, locus-specific co-dominant inheritance, multiallelic nature, high polymorphism and high rates of transferability across species (Aggarwal et al. 2007; Choudhary et al. 2009; Dutta et al. 2011). The SSR markers can be found either in non-coding or coding regions of the genome, including protein-coding genes and expressed sequence tags (ESTs) (Tóth et al. 2000; Li et al. 2004; Choudhary et al. 2009). Expressed sequence tags databases are very useful resources to identify SSRs by using various bioinformatic tools (Varshney et al. 2005). A large number of SSR markers from ESTs have been developed and utilized in diverse species such as cotton (*Gossypium* spp., Guo et al. 2007), western corn rootworm (* Diabrotica virgifera*, Kim et al. 2008b), apple maggot (*Rhagoletis pomonella*, Schwarz et al. 2009), chickpea (*Cicer arietinum* L., Choudhary et al. 2009) and coral (*Acropora millepora*, Wang et al. 2009). In comparison with genomic-SSRs having no relation to the transcribed regions, genic or EST-SSRs have an advantage of being directly associated with expressed regions of the genes, because the markers are designed from the coding regions of the genome that are subject to selection pressures (Li et al. 2004; Choudhary et al. 2009; Garg et al. 2011). In addition, genic-SSRs are useful for comparative analysis across related species because the coding regions of the genome are more conserved than non-coding genomic regions (Varshney et al. 2005; Jakse et al. 2010; Zeng et al. 2010).

Traditionally, generating a large number of genic-SSRs has been technically challenging, expensive and time-consuming (Eujayl et al. 2004; Wang et al. 2010a). Recently developed next-generation sequencing platforms such as Illumina’s Genome Analyzer (GA), ABI’s SOLID and Roche’s 454 GS FLX have been used in various studies, including large-scale resequencing in well-characterized species, transcriptional profiling and transcriptome sequencing (Ossowski et al. 2008; Varshney et al. 2009; Wang et al. 2010b; Cullum et al. 2011). Additionally, de novo transcriptome sequencing for species without reference sequences have succeeded in generating and assembling draft sequences using advanced programming, which enables the development of a large number of genetic markers (Varshney et al. 2009; Bai et al. 2010; Li et al. 2010; Wang et al. 2010a; Garg et al. 2011). The objectives of this study were (i) to identify and validate genic-SSR markers from soybean aphid transcript sequences generated by 454 GS FLX transcript sequencing and (ii) to assess their polymorphism and usefulness in determining genetic diversity within and among soybean aphid populations.

**Materials and Methods**

**Sequence data sources**

Information regarding sampling and 454 GS FLX sequencing is published elsewhere (Bai et al. 2010).
Briefly, total RNA was isolated from 50 biotype 1 aphids using Trizol Reagent (Invitrogen). Approximately 10–20 µg total RNA were sent to the Purdue University Genome Center for 454 sequencing following manufacturer’s instructions. A total of 7 366 599 bp forming 19 293 high-quality transcript sequences were obtained through two rounds of assembly using Newbler and Phrap program (Bai et al. 2010). Of the 19 293 transcript sequences, a total of 881 (4.6%) transcript sequences containing SSR regions were detected using the program msat finder (Thurston and Field 2005) with a minimal repeat number of 8 (dinucleotide repeats) and 5 (all other repeats). Of these, 501 (2.6%) transcript sequences (mean length 615, range 119–2771 bp) that had at least 40 bp sequences in both flanking regions of each SSR were used to develop genic-SSR markers in the present study.

Development of SSR markers

The 501 transcripts from Bai et al. (2010) were searched again using BatchPrimer3 software (http://probes.pw.usda.gov/batchprimer3/index.html) by changing the minimum number of penta- and hexa-nucleotide repeats to 4. An additional six loci (3 penta- and 3 hexa-nucleotide repeats) were found among the 501 contigs that contained microsatellites. The primary parameters for primer design were the following: PCR product size ranging from 100 to 400 bp for multiplex analysis, primer length from 18 to 25 bp with 20 bp as the optimum, optimum annealing temperature 60°C having difference within 1°C between forward and reverse primers, and GC content from 40% to 60% with 50% as optimum. All forward primers had the universal M13 sequence (CACGACGTTGTAAAACGAC) at their 5’ end for use in fluorescent primer labelling (Schuelke 2000). The primers were synthesized by Integrated DNA Technologies Inc (Coraville, IA).

Soybean aphid samples

The adult soybean aphids collected from biotype 1 (B1), biotype 2 (B2), an Ohio field in 2009 (OH-09) and a South Dakota field in 2009 (SD) were used as pooled DNA templates. Both biotype 1 and biotype 2 aphids were collected from separate and isolated laboratory colonies that were started with limited (<50) individuals (see Michel et al. 2010). The OH-09 aphids were collected from a soybean field near the OARDC campus during the summer of 2009 (GPS: 40.768, –81.906), and the SD aphids were collected from a soybean field in SD in 2009 (GPS: 44.324, –96.775). The four pooled DNA sets (B1, B2, OH-09 and SD) each consisted of about 20 adult soybean aphids. Each of the adult aphids was collected from a different soybean plant to limit chances of collecting clones. This pooling of DNA was performed only for the initial validation for SSR markers (e.g. PCR amplification condition, product size estimation and potential for polymorphism).

For the population-level genetic diversity and differentiation analyses, 24 individual soybean aphids were sampled from biotype 1 and biotype 2 colonies and from two field sites, Michigan (MI, GPS: 42.691, –84.486) and South Dakota (SD, same as above). Infested leaves were collected throughout the field but only using one aphid per leaf per plant to control for the possibility of collecting clones.

PCR conditions and fragment analysis of SSRs

DNA was extracted using the QuickExtract™ Seed DNA Extraction Solution (Epicentre, Madison, WI) according to the manufacturer’s protocol. PCR amplifications were performed in 20-µl reactions containing approximate 50 ng of template DNA, 1× PCR buffer, 2.5 mM Mg2+, 200 µM dNTP, 40 nM of forward primer with M13 tail, 160 nM reverse primer, 160 nM of dye labelled M13 primer and 1.0 unit of Taq DNA polymerase (GenScript USA Inc., Piscataway, NJ). The PCR cycles consisted of initial denaturation at 95°C for 5 min, followed by 32 cycles of 30-s denaturation at 94°C, 30-s annealing temperature between 48 and 65°C depending on the optimum annealing temperature for each primer pair (see Table S2), and 30-s extension at 72°C, followed by eight cycles of 30-s denaturation at 94°C, 45-s annealing at 53°C and 45-s extension at 72°C. The optimum annealing temperature for each primer pair was determined empirically, using the theoretical melting temperature as the starting point and adjustments to annealing temperatures (increase and decrease) as needed to obtain clean PCR products. The PCR was finished with a final 12-min extension at 72°C on a thermal cycler (PTC-225; MJ Research Inc., Waltham, MA, USA; model TC-512; Techne Inc., Burlington, NJ, USA). Multiplex fragment analysis and allele sizing for the SSR markers were performed using the CEQ 8800 genetic analyzer and software, respectively (Beckman Coulter, Fullerton, CA). Allele scoring was checked for the presences of null alleles and stutter peaks with
Microchecker (Van Oosterhout et al. 2004), and no significant evidence for either was found.

Sequencing analysis of PCR products
Two genic-SSR markers (Ae-SSR32 and Ae-SSR331) that revealed length polymorphism with a single clear and specific band in the initial screening of the four pooled aphid samples were selected to identify the source of length variation among PCR products. Direct sequencing of PCR products amplified by the two genic-SSR markers in four aphid DNA samples (one each from B1, B2, OH-09 and SD) were performed. For the BigDye terminator cycle sequencing, the PCR products were directly purified using the mixture of exonuclease I and shrimp alkaline phosphatase (ExoSAP), which were conducted in 12-μl reactions containing 10 μl of PCR product, 2.0 units of exonuclease I and 1.8 units of shrimp alkaline phosphatase. Incubation was performed at 37°C for 1 h and 72°C for 15 min. Sequencing was performed in both forward and reverse directions with PCR amplification primers using ABI Prism 3100 genetic analyzer (Functional Biosciences, Madison, WI), and the sequences from each aphid sample were aligned using CodonCode Aligner 3.5.1 demo version (http://www.codoncode.com/aligner/) and Clustal W multiple alignment of BioEdit Sequence Alignment Editor 7.0.9 software (Hall 1999).

Genetic diversity analysis
For each marker, allelic richness ($R_s$), observed ($H_o$) and expected heterozygosity ($H_e$) were calculated using the program FSTAT (Goudet 1995, 2001). Hardy–Weinberg disequilibrium was also calculated using FSTAT, using 10 000 random permutations, with Bonferroni corrected P-values. FSTAT was also used to calculate the genetic differentiation statistic, $F_{ST}$, with significance determined by 10 000 permutations (analyses were performed with and without repeated genotypes, as common for aphid population genetic studies). The SSR allelic data were converted into binary code as 1 or 0 for each aphid individual to evaluate the genetic relationship. The genetic similarity among the aphids was determined using Dice coefficient (Dice 1945). The UPGMA (Unweighted Pair-Group Method with Arithmetic mean) based dendrogram was constructed based on similarity coefficients using SAHN module of the NTSYS-pc software package ver 2.2. The dendrogram was subjected to 1000 bootstraps using the FreeTree software (Hampel et al. 2001).

Results
Frequency distribution and development of genic-SSRs
A total of 621 SSRs were identified in 501 transcript sequences, which are derived from the total of 19 293 transcript sequences containing 11 866 singletons and 7427 contigs (fig. 1). Among the sequences used for the discovery of SSRs, 409 (81.6%) had single SSR, while 92 (18.4%) sequences had 2 (13.8%), 3 (3.4%) or 4 (1.2%) SSRs each. The trinucleotide motifs were the most abundant repeat motif (79.5%), followed by di- (16.2%), tetra- (2.4%), penta-(1.4%) and hexanucleotide motifs (0.5%) (fig. 1). The most abundant dinucleotide repeat unit detected was the AT/AT representing (10.1%) of the total 621 SSRs, followed by AC/GT (4.8%) and CA/TG (0.8%). The AAT/ATT motif was the most abundant trinucleotide repeat type (20.8%) of the total SSRs followed by ATA/TAT (19.0%) and TAA/TTA (16.3%) (Table S1).

A total of 342 primer pairs were designed from the total of 621 SSRs (fig. 1). The primers for the remaining 279 SSRs could not be designed because of flanking sequence being too short or poor base composition for primer design.

Validation of genic-SSR markers
Of the 342 primer pairs that were tested for initial validation for SSR markers using the four pooled aphid DNA templates (B1, B2, OH-09 and SD), a total of 246 (71.9%) primer pairs generated PCR products with expected size, whereas 72 (21.0%) primers pairs did not amplify even after retests under multiple PCR conditions (table 1; Table S2). In addition, 18 (5.3%) primers pairs produced PCR fragments much larger than the expected size and 6 (1.8%) primers pairs amplified multiple products (table 1; Table S2). Of the 246 primers pairs validated in the initial test, 208 (84.6%) primers pairs each generated a single (allele) whereas 38 (15.4%) primers pairs produced two alleles. A total of 26 (7.6%) primer pairs were polymorphic among the four pooled aphid DNA templates (table 1; Table S2). The remaining 12 primer pairs each showed the presence of two identical-sized bands among the four DNA pools, indicating either fixation of the heterozygous state or the presence of two alleles in the pooled DNA. These 12 markers were further tested using 96 individual aphids from four aphid collections (see below). Some individual aphids showed a second fainter band for each marker and some did
not have the second band, indicating non-specific primer-binding sites. Under more stringent PCR conditions (higher annealing temperatures), five primer pairs produced only a single band (implying previous co-amplification), two produced the same two bands in all 96 aphids tested (i.e. all individuals fixed for the heterozygote state) and the remaining five each produced polymorphic results (Table S2).

Multiple alignments were performed using nucleotide sequences containing the SSR regions from PCR products amplified with two genic-SSR markers in four aphid samples (fig. 2). Both B2 and SD aphid samples possessed 10 TA repeats for the Ae-SSR32, whereas OH-09 and B1 had 8 TA repeats each. For Ae-SSR331, OH-09 and B1 samples had 13 ATA motif units indicating two more repeats than the other two aphids that differed by 6 bp in size based on genotyping data. The SSR flanking sequences in Ae-SSR32 and Ae-SSR331 each were identical across all aphids tested. Therefore, the results from the sequence alignments confirm that the differences in number of motif repeats resulted in allele size variations for the two SSR markers.

Genetic relationship among soybean aphid populations

Twenty-nine polymorphic genic-SSRs markers were used to assess the genetic diversity and differentiation in a set of 96 soybean aphids collected from B1, B2, MI and SD with 24 individual aphids tested from each location (fig. 3). A total of 58 alleles were amplified from 29 SSRs with an average of 2.0 per locus. Within population polymorphism, statistics are shown in table 2. In both biotype colonies, Hardy–Weinberg equilibrium could not be calculated for several loci, as fixation of either the homozygote or heterozygote genotype occurred, but represents deviations nonetheless. Fixation of genotypes in these populations is likely due to continued asexual
reproduction and clonal amplification, where favoured genotypes quickly outnumber and outcompete less fit genotypes. In the field samples, only Ae-SSR 46 and Ae-SSR 11, Ae-SSR 152 and Ae-SSR 167 showed significant deviation in Hardy–Weinberg equilibrium in both field samples. A null allele effect is likely not the cause given that the direction of deviation was heterozygote excess (\(H_o > H_e\)). As these samples were collected late during the year after many asexual generations occurred, deviations are likely due to clonal amplification, similar to laboratory populations. Nonetheless, the microsatellites provided good resolution, as genotypic diversity was high in field populations. Each individual aphid represented a unique genotype among the 48 total individuals, and no genotype was found in more than one population.

A dendrogram calculated by dice similarity coefficient placed the aphid individuals into six clusters at a cut-off point of 0.86 similarity (fig. 3). Cluster II included all 24 aphids from biotype 2 with no within-cluster differences, and cluster V included all individuals of biotype 1 with minor differences among individuals. Twenty-one of the 24 aphids from Michigan grouped in cluster I with minor differences, while the remaining three aphids (MI-5, 15 and 17) grouped under cluster VI revealing a very high genetic dissimilarity with other 21 aphids from MI (fig. 3). Similarly, 23 of the SD individuals grouped together under cluster III with minor differences, but 1 individual (SD-16) showed a different genotype and was placed separately under cluster IV. The overall clustering found in the dendrogram was also corroborated by \(F_{ST}\) values (table 3). All \(F_{ST}\) values were significant (\(P = 0.006\) after Bonferroni corrections).

### Discussion

Frequency and distribution of genic-SSRs

Mining of genic-SSRs in soybean aphid transcript sequences from the next generation sequencing (454 GS FLX) was successfully carried out in the present study. In an earlier study, a total of 308 genomic-SSRs were detected from the 1240 contigs through Illumina GAII sequencing of soybean aphid (Jun et al. 2011). These SSR markers from soybean aphid demonstrate that the Illumina GAII technology could be very useful for developing molecular markers for organisms with limited genomic resources. A total of 621 genic-SSRs were obtained in our study, and the
frequency of genic-SSRs was approximately 1 SSR/11.9 kb. Higher or lower frequencies of genic-SSRs have been reported in various studies, for instance, rice (1 SSR/3.4 kb), pigeonpea (1 SSR/8.4 kb) and cotton (1 SSR/20.0 kb). However, such different frequencies could be due to the application of different repeat unit thresholds and repeat length, the amount of database searched and SSR identification tools used (Varshney et al. 2005; Dutta et al. 2011).

Trinucleotide motifs were the most abundant repeat type in these genic-SSRs, representing 79.5% of the total SSRs detected. This result is in agreement with a number of previous reports that the trinucleotide motifs are the most frequent genic repeats in most species, which might be due to the tendency towards negative selection against frame-shift mutations in the coding regions (Metzgar et al. 2000; Cloutier et al. 2009). However, in some species such as coffee, pigeonpea, and pumpkin dinucleotide repeats were the most frequent (Aggarwal et al. 2007; Gong et al. 2008; Dutta et al. 2011). The frequency and distribution of genic-SSRs can vary depending on (i) various factors in the SSR search such as software, search criteria and size of data set and (ii) genomic regions (e.g. coding regions, 5′-UTRs and 3′-UTRs) used for the detection of genic-SSRs (Li et al. 2004; Varshney et al. 2005; Garg et al. 2011). Yu et al. (2004) reported different frequency of di- and trinucleotide repeats in coding regions and UTRs – 74% of the trinucleotides repeats were in coding regions and 26% in UTRs, whereas 19% of dinucleotides repeats were observed in coding regions, and 42% and 39% were in 5’ and 3’ UTRs, respectively.

Development and validation of genic-SSRs
Approximately 72% of the 342 SSR primers used for initial validation generated the PCR product with the expected product size, while 96 primers (28%) failed
to amplify the expected PCR products. Among the 96 primers, 18 (5.3%) produced much larger than expected band sizes. These results might be due to presence of a small intron within the flanking region of each of the primer pairs. In contrast, 72 (21%) of the 96 primers amplified no PCR fragments despite various tests under multiple PCR conditions. These failures could be caused because of the following: (i) flanking primers designed across a splice site, (ii) the existence of large introns within the PCR target region, (iii) the existence of null alleles because of mutation at the primer-binding sites and (iv) the use of inaccurate sequences because of assembly errors in de novo transcriptome sequencing for species without reference sequences (Pemberton et al. 1995; Varshney et al. 2005; Cloutier et al. 2009; Wang et al. 2010a).

In the present study, the genic-SSR markers showed a low level of polymorphism (8.5%), compared with the result (about 20%) of previous marker validation using genomic-SSR markers (Jun et al. 2011) and cross-species amplification of microsatellites from *Aphis gossypii* and *Aphis fabae* (Michel et al. 2009a). Our observation was in agreement with the hypothesis of varietal adaptation in *A. glycines* (Moodley et al. 2005; Varshney et al. 2005; Cloutier et al. 2009; Wang et al. 2010a).
with other studies, showing that genomic-SSRs were more polymorphic compared with genic-SSRs, which may be due to greater levels of sequence conservation within transcribed or gene-rich regions (Eujayl et al. 2001; Chabane et al. 2005; Varshney et al. 2005). Nevertheless, it is possible that the low polymorphism rate of genic-SSRs developed in the present study may be partly due to small number of samples tested. Cloutier et al. (2009) suggested that the relatedness or the number of genotypes tested can partly influence the polymorphism rates. Additionally, polymorphism could be decreased due to a bottleneck during the initial North American invasion, use of field samples from few locations collected in early in the season and clonally propagated laboratory colonies. Sequencing of the PCR products of two SSR colonies showed that the polymorphism occurred strictly because of variations in the repeat units in the aphid samples used. However, SSR polymorphism can also occur from variations in the flanking regions (Eujayl et al. 2004; Saha et al. 2006).

Genetic differentiation within/among different soybean aphid populations

The 29 polymorphic SSR markers tested proved very useful for genetic differentiation among the aphid samples used in the genetic diversity test (fig. 3, table 3). The aphids of the two laboratory colonies formed their separate clusters with little or no within-colony diversity. Michel et al. (2010) reported similar results using FIVE laboratory and 12 field populations of soybean aphids. They suggested that the loss of genotypic diversity because of the fixation of homozygous or heterozygous genotypes in laboratory populations resulted in higher genetic differentiation among laboratory than field populations. The observed lack of polymorphism among individual aphids within a laboratory colony is also expected (Michel et al. 2010; Jun et al. 2011). Unfortunately, the studies cannot be directly compared with the current study because different individuals were used. More detailed study will need to be conducted to determine whether this set of genic-SSR markers can actually differentiate between field collections and biotype 1 and 2 of the soybean aphids.

Most of the aphid samples from MI and SD were grouped into clusters of their own. Although both MI and SD had equal numbers of genotypes, based on the dendrogram, MI contained more divergent genotypes. Three of the 24 aphids in the MI collection were very diverse compared with the rest of the aphids in the collection while only one aphid showed divergence from the rest of the aphids in the SD collection. These results match with the comparative genetic diversity expected of the soybean aphids in these two states. Michigan is one of the major soybean producing states with high population of buckthorn (winter host of the soybean aphid), and the aphid has been present in this state since its first discovery in 2000. On buckthorn, aphids undergo sexual reproduction (and hence recombination) and overwinter as eggs; thus, the amount of buckthorn in a given area is potentially correlated with soybean aphid genetic diversity. South Dakota is at the western fringe of the US soybean belt with very few buckthorn present in the state. The initial aphid colonization of soybean fields in SD each year may require migration of the aphids from distant overwintering sites creating a genetic bottleneck. Thus, given a smaller resident population, smaller genetic diversity would be expected in a field collection of soybean aphids from SD compared with a collection from MI.

In conclusion, here we report 246 high-quality genic-SSR markers experimentally validated from a set of 342 primer pairs designed from EST sequences generated by 454 GS FLX sequencing of a cDNA library of the soybean aphid. Twenty-nine of the 246 markers were polymorphic among aphids used in this study; however, more markers have the potential to be polymorphic among more soybean aphid populations. The usefulness of these markers in genetic diversity and population genetics has also
been demonstrated in a limited number of soybean aphid collections. These markers should be useful for genetic studies, including population genetics, genetic mapping, migration and biotype differentiation of the soybean aphid and are significant additions to the limited number of SSR markers that are currently available for the soybean aphid from the published literature.

Acknowledgements

We thank Wei Zhang, Lucia Orantes, Jane Todd, Ronald Fioritto and Tim Mendiola for their technical help in this study. We also thank the employees at the MCIC for their help in sequencing and genotyping. We appreciate Dr. Kelley Tilmon and Dr. Chris DiFonzo for their help in soybean aphid collection from South Dakota and Michigan, respectively. Two anonymous reviews provided critical insight to improve the manuscript. This study was supported by USDA-ARS, the OSU/OARDC and a grant from the Ohio Soybean Council (10-2-03).

References


Supporting Information

Additional Supporting Information may be found in the online version of this article:
Table S1. Distribution and frequency of different motif types identified in the 501 transcript sequences from soybean aphid.
Table S2. The list of 342 soybean aphid genic-SSR primer pairs with marker name, forward and reverse primer sequences, expected product size (bp), motif and aphid samples used for PCR amplification [B1 = Biotype1, B2 = Biotype2, OH(09) = Ohio in 2009, and SD = South Dakota].
Table S3. Summary of top-level GO terms assigned to the three A. glycines B1 transcript sequences containing polymorphic SSR region; P: Biological Process, C: Cellular Component, and F: Molecular Function.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.