Paired-End Illumina Shotgun Sequencing Used to Develop the First Microsatellite Primers for *Megacopta cribraria* (F.) (Hemiptera: Heteroptera: Plataspidae)[1]

Tracie M. Jenkins[2], Tyler D. Eaton, Schyler O. Nunziata[3], Stacey L. Lance[3], and Kenneth L. Jones[4]

Department of Entomology, University of Georgia, Griffin Campus, Griffin, Georgia 30223-1797 USA

J. Entomol. Sci. 48(4): 345-351 (October 2013)

Abstract Microsatellite or simple sequence repeats (SSRs) are useful markers for testing hypotheses related to intraspecific genetic diversity and phylogeographic dispersal patterns of invasive species. *Megacopta cribraria* (F.), a bean plataspid from Asia, is an invasive insect pest first discovered in the Western Hemisphere in 2009 in 9 counties in northeast Georgia, USA. By the end of 2012, *M. cribraria* had been confirmed in 392 counties in 8 U.S. states. To study the genetic diversity over time of this dispersing invasive population we developed and characterized the first codominant markers from *M. cribraria* genomic sequences. Seventeen genomic microsatellite loci in *M. cribraria* were characterized. These loci were screened in 23 individuals from 11 counties in Georgia, USA, collected in 2009 and 2010, and from 6 individuals from Kanagawa Province in Japan collected in 2010. The number of alleles per locus ranged from 2 - 7, observed heterozygosity ranged from 0.00 - 0.60, and the probability of identity values ranged from 0.12 - 0.92. These new loci will be used to examine the spatial and temporal genetic diversity as well as the genetic structure of *M. cribraria* as it stochastically disperses and survives transitions into new geographic areas. They also will provide broader insights into how genes and epistatic interactions facilitate adaptive radiations of *M. cribraria* specifically and exotic insect species in general.

Key words *Megacopta cribraria*, microsatellite, PCR primers, SSR, STR

Megacopta cribraria (F.) (Heteroptera: Plataspidae), a known pest of legumes in Asia, was first discovered in the Western Hemisphere in October 2009 in Jackson Co., GA, USA (Suiter et al. 2010). Identified from morphological characters and DNA sequence (Eger et al. 2010, Jenkins et al. 2010, Jenkins and Eaton 2011), the insect was found in kudzu fields, *Pueraria montana* var. *lobata* (Willd.) Ohwi (Fabaceae), near residential areas and on the sides of houses by the thousands apparently seeking overwintering sites (Suiter et al. 2010, Ruberson et al. 2013). By the end of 2009 *M. cribraria*, generally identified as the kudzu bug, had been confirmed in 9 contiguous counties in northeast Georgia accounting for a land area of 7,050 km^2 (Gardner et al. 2013). By the end of 2012, *M. cribraria* had increased its dispersal range by 70-fold

[1]Received 24 May 2013; accepted for publication 12 July 2013.
[2]Corresponding author (email: jenkinst@uga.edu).
[3]Savannah River Ecology Laboratory, University of Georgia, Aiken, SC 298023.
[4]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045.

to 497,819 km^2 to include 392 counties in 8 states: Georgia, AL, SC, North Carolina, VA, Tennessee, MS, and Florida (Gardner et al. 2013). Since that report, the insect has been confirmed in 4 additional states (Louisiana, MD, Delaware, and Kentucky) and the District of Columbia (W.A. Gardner, University of Georgia, personal communication).

Megacopta cribraria feeds not only on kudzu but also on soybean plants, a commodity valued in 2010, according to the USDA National Agricultural Statistics Service, at $38,915,328,000 (Jenkins and Eaton 2011). It is, therefore, imperative to develop genomic, markers which provide insights into *M. cribraria* population structure and intraspecific dispersal patterns to predict and control dispersal as well as develop effective control strategies. Genomic DNA markers will also provide the opportunity to test hypotheses on the evolutionary process of the adaption of a recent, genetically depauperate and aggressively expanding invasive insect. Furthermore, since discovery in the Western Hemisphere temporal DNA samples from individuals throughout the expanded range have consistently been collected (Jenkins and Eaton 2011), which provides the opportunity to examine changes in genetic variation over the course of the expansion.

Microsatellites or simple sequence repeats (SSRs) are powerful multilocus genomic markers for testing population genetic questions. They have been used to track the inheritance of alleles from progenitor to progeny, to illuminate intraspecific population diversity and structure (Wang et al. 2009), and to delineate phylogenetic relationships (Jenkins et al. 2012). The purpose of this work, therefore, was to characterize for the first time microsatellite markers which can be used to study the population structure, dispersal patterns, and evolution of *M. cribraria* in the Western Hemisphere.

Materials and Methods

Total DNA was extracted from 2 individuals of *M. cribraria* following the protocol of Jenkins and Eaton (2011) for use in isolation of microsatellite loci. An Illumina paired-end shotgun library was prepared by shearing 1 µg of DNA using a Covaris S220 and following the standard protocol of the Illumina TruSeq DNA Library Kit and using a multiplex identifier adaptor index. This library was pooled with those from other species. Illumina sequencing was then done on the HiSeq with 100 bp paired-end reads. Five million of the resulting reads were analyzed with the program *PAL_FINDER_v0.02.03* (Castoe et al. 2012) to extract those reads that contained di-, tri, tetra-, penta-, and hexanucleotide microsatellites. Once positive reads were identified in *PAL_FINDER_v0.02.03,* they were batched to a local installation of the program Primer3 (version 2.0.0) for primer design. The loci for which the primer sequences occurred only 1 or 2 times in the 5 million reads were selected to avoid issues with copy number of the primer sequence in the genome. Forty-eight loci of the 5,149 that met this criterion were chosen. One primer from each pair was modified on the 5′ end with an engineered sequence (CAG tag 5′-CAGTCGGGCGTCATCA-3′) to enable use of a third primer in the PCR (identical to the CAG tag) that was fluorescently labeled. The sequence GTTT was added to primers without the universal CAG tag addition.

Forty-eight primer pairs were tested for amplification and polymorphism using DNA obtained from 8 individuals. PCR amplifications were performed in a 12.5 µL volume (10 mM Tris pH 8.4, 50 mM KCl, 25.0 µg/ml BSA, 0.4 µM unlabeled primer,

0.04μM tag labeled primer, 0.36μM universal dye-labeled primer, 3.0 mM $MgCl_2$, 0.8 mM dNTPs, 0.5 units AmpliTaq Gold® Polymerase (Applied Biosystems), and 20 ng DNA template) using an Applied Biosystems GeneAmp 9,700. Touchdown thermal cycling programs (Don et al. 1991) encompassing a 10°C span of annealing temperatures ranging between 65 - 55°C (TD65) or 58 - 48°C (TD58) were used for all loci. Touchdown cycling parameters consisted of an initial denaturation step of 5 min at 95°C followed by 20 cycles of 95°C for 30 sec, highest annealing temperature (decreased 0.5°C per cycle) for 30 sec, and 72°C for 30 s; and 20 cycles of 95°C for 30 sec, lowest annealing temperature for 30 sec, and 72°C for 30 sec. PCR products were run on an ABI-3130xl sequencer and sized with Naurox size standard prepared as described in DeWoody et al. (2004), except that unlabeled primers started with GTTT. Results were analyzed using GeneMapper version 3.7 (Applied Biosystems). Seventeen of the tested primer pairs amplified high quality PCR product that exhibited polymorphism.

GenAlEx v6.4 (Peakall and Smouse 2006) was used to estimate the number of alleles per locus (*k*), observed and expected heterozygosity (H_o and H_e), and probability of identity (PI). Tests for deviations from Hardy-Weinberg equilibrium (HWE) and for linkage disequilibrium were conducted using GENEPOP v4.0 (Rousset 2008).

Results and Discussion

We assessed the variability of the 17 polymorphic loci in 23 specimens collected from Georgia and Japan. Conditions and characteristics of the loci are provided in Table 1. Three loci - Mecr3, Mecr39, and Mecr45 (Table 1) - showed significant deviation from expectations under HWE after Bonferroni correction for multiple comparisons. No linkage disequilibrium was detected for any of 136 paired loci comparisons. These are the first microsatellite loci to provide the scientific community with a powerful molecular tool capable of illuminating the spatial and temporal genetic structure of *M. cribraria*. DNA samples collected from 2009 through 2013 and into the future can now be analyzed in light of changes in population genetic structure. Hypotheses can be tested which will provide insights into the sources of *M. cribraria* phenotypic variation encoded and not encoded (epigenetically inherited) in DNA sequence. Understanding how environmental variance and genomic imprinting contribute to the insects' capacity to successfully tolerate and adapt to biotic and abiotic (Gardner et al. 2013) parameters will afford the opportunity to test hypotheses specific to directional dispersal of genetically depauperate, peripheral populations.

Since its discovery in 2009 *M. cribraria*'s population genetics (Jenkins et al. 2010, Jenkins and Eaton 2011), taxonomy (Eger et al. 2010), biology (Ruberson et al. 2013), and dispersal (Suiter et al. 2010, Gardner et al. 2013) in the Western Hemisphere has been and continues to be well documented (T.M. Jenkins, unpubl data, W.A. Gardner, personal communication). Microsatellite markers (Table 1) will not only provide insights into the adaptive correlation between genetics and environmentally-induced epigenetic phenomena, but also the opportunity to create a genomic baseline to complement the mitochondrial baseline established by Jenkins and Eaton (2011). The collaborative and integrative studies done since 2009 in combination with genomic studies generated by these microsatellite markers will collectively serve as a template for the study of other invasive systems.

Table 1. Details for 17 polymorphic microsatellite loci developed for *Megacopta cribraria.* The size indicates the range of observed alleles in base pairs and includes the length of the CAG tag; number of individuals genotyped is *N*; *k* is number of alleles observed; H_o and H_e are observed and expected heterozygosity, respectively; PI is the probability of identity for each locus, and TD.

Locus	Primer Sequence 5′ → 3′	Repeat motif	Size (bp)	N	*K*	H_o	H_e	PI	TD
Mecr3	F: *TCGAACTCAGGAACCTCTATACG	ATCT	165 - 177	9	2	0	0.444	0.41	65
	R: AAATGTTCCAATCGCCTCG		165 - 165	3	1	0	0	1	
Mecr4	F: *ATAGCTTCCTGGTAACTGTCACG	TCTG	157 - 165	15	3	0.467	0.487	0.35	65
	R: AACACGTCAGGATGTCACGC		149 - 161	8	3	0.625	0.508	0.32	
Mecr6	F: *ACTTAAATTCACCCGCTCCC	AAAT	220 - 224	15	2	0.267	0.231	0.62	65
	R:AAGCAACTGTTAACATCAAGTATCTGG		212 - 224	8	2	0.25	0.5	0.38	
Mecr7	F: *GAATAGAGAGGAACGGGAAAGC	ATT	171 - 174	14	2	0.429	0.459	0.4	65
	R: AATCGGAGGATTAGACTGCG		162 - 174	7	3	0.429	0.357	0.44	
Mecr8	F: *GTAGGTCCTGTATTTCTTACTCATGG	AATT	166 - 170	12	2	0.333	0.278	0.56	65
	R: ACGGTGATGCAATAGCTGG		166 - 166	8	1	0	0	1	
Mecr11	F: *TGTTGTTTAATGATGGATATGAGC	AGGT	249 - 253	15	2	0.133	0.124	0.77	65
	R: AGAGGTTATCAGAGGAGCGG		249 - 253	8	2	0.25	0.375	0.46	
Mecr17	F: *GGAAAGCGAAGTTAGGACCC	ATT	118 - 118	15	1	0	0	1	65
	R: CCAAACTACTAAGCCATTGGGC		115 - 118	8	2	0.125	0.117	0.79	
Mecr18	F: *CTAGAATGACACGATTAAGACGGC	AGT	196 - 209	15	4	0.4	0.471	0.32	65
	R: CCGGTGCAGTTAAAGGCG		193 - 208	6	3	0.667	0.611	0.23	
Mecr28	F: *TAGCGCAAGCATCTTCAAGC	AAAT	217 - 221	15	2	0.533	0.48	0.39	65
	R: GTGCCGGATTAGCACATCG		217 - 221	4	2	0.5	0.375	0.46	

Table 1. Continued.

Locus	Primer Sequence 5′ → 3′	Repeat motif	Size (bp)	N	*K*	H_o	H_e	PI	TD
Mecr30	F: *TGAAGATATTGCATTCCATCAGG	ATT	183 - 195	15	3	0.533	0.58	0.24	65
	R: GTGTCAGAGGATGGACAGCG		183 - 204	8	7	0.5	0.813	−0.06	
Mecr35	F: *TGTTCATCACTGTTAACAAACGC	ATAC	208 - 216	12	4	0.5	0.566	0.25	65
	R: TCAGCTAACATTGGGTACACG		206 - 216	8	6	0.75	0.773	0.08	
Mecr39	F: *GGTTAATTTGGTCTTATGACCTTCG	ATCT	340 - 356	14	5	0.143	0.707	−0.14	58
	R: TGAAAGAACGGAATACATGCG		340 - 344	8	2	0.25	0.5	0.38	
Mecr40	F: *TTTGAAATGCCCGAAGAGG	TTC	120 - 126	15	2	0.133	0.124	0.77	65
	R: TGGAGATAGGATCTGTCGGC		126 - 126	8	1	0	0	1	
Mecr43	F: *TGGAGTTAAGCTGCCAGGG	TTGG	288 - 288	15	1	0	0	1	58
	R: TTAACTCACCTTCGTCGGGC		276 - 288	8	2	0.125	0.117	0.79	
Mecr45†	F: *CTAGGCGATCCAATTGCTGC	ATT	278 - 290	14	2	0	0.459	0.4	65
	R: TTCTTGTCTCTTGGGACGGC		278 - 290	6	2	0	0.444	0.41	
Mecr46	F: *TGAGTAATCGAATAAATTGAGGAGC	ATC	180 - 189	11	2	0.455	0.483	0.38	65
	R: TTGAGTCTACAGGAGTTGTTGCG		180 - 189	8	2	0.25	0.219	0.63	
Mecr48	F: *TTCCAATGCATCTTCTAATCGC	ATAC	237 - 249	15	3	0.533	0.54	0.3	65
	R: TTTCAAAGAGGAAGTCGAGACC		233 - 247	8	5	0.5	0.5	0.28	

Acknowledgments

This research was supported by an USDA-NIFA-AFRI Foundational Grant (A1111) under Award Number 043661-01. Manuscript preparation was also partially supported by the DOE under Award Number DE-FC09-07SR22506 to the University of Georgia Research Foundation.

References Cited

Castoe, T. A., A. W. Poole, A. P. J. de Koning, K. L. Jones, D. F. Tomback, S. J. Oyler-McCance, J. A. Fike, S. L. Lance, J. W. Streicher, E. N. Smith and D. D. Pollack. 2012. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. PLoS ONE 7: e30953.

DeWoody, A. J., J. Schupp, L. Kenefic, J. Busch, L. Murfitt and P. Keim. 2004. Universal method for producing ROX-labeled size standards suitable for automated genotyping. Biotechniques 37: 348-350.

Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker and J. S. Mattick. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19: 4008.

Eger, Jr., J. E., L. Ames, D. R. Suiter, T. M. Jenkins, D. A. Rider and S. E. Halbert. 2010. Occurrence of the old world bug *Megacopta cribraria* (Fabricius) (Heteroptera: Plataspidae) in Georgia: A serious home invader and potential legume pest. Insecta Mundi 87: 0121.

Gardner, W. A., H. B. Peeler, J. LaForest, P. M. Roberts, A. N. Sparks Jr., J. K. Greene, D. Reisig, D. R. Suiter, J. S. Bacheler, K. Kidd, C. H. Ray, X. P. Hu, R. C. Kemerait, E. A. Scocco, J. E. Eger Jr., J. R. Ruberson, E. J. Sikora, D. A. Herbert Jr., C. Campana, S. Halbert, S. D. Stewart, G. D. Buntin, M. D. Toews and C. T. Bargeron. 2013. Confirmed distribution and occurrence of *Megacopta cribraria* (F.) (Hemiptera: Heteroptera: P)lataspidae) in the southeastern United States. J. Entomol. Sci. 48: 118-127.

Jenkins, T. M., T. D. Eaton, D. R. Suiter, J. E. Eger, L. M. Ames and G. D. Buntin. 2010. Preliminary genetic analysis of a recently-discovered invasive true bug (Hemiptera: Heteroptera: Plataspidae). J. Entomol. Sci. 45: 1-2.

Jenkins, T. M. and T. D. Eaton. 2011. Population genetic baseline of the first plataspid stink bug symbiosis (Hemiptera: Heteroptera: Plataspidae) reported in North America. Insects 2: 264-272.

Jenkins, T. M., N. A. Barkley and M. L. Wang. 2012. Microsatellite markers in plants and insects, Part II: Microsatellite databases and *in silico* tools for analyses of population genetic structure. Genes, Genomes. Genomics 6: 60-75.

Peakall, R. and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288-295.

Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Res 8: 103-106.

Ruberson, J. R., K. Takasu, G. D. Buntin, J. E. Eger, W. A. Gardner, J. K. Greene, T. M. Jenkins and W. A. Jones. 2013. From Asian curiosity to eruptive American pest: *Megacopta cribraria*

(Hemiptera: Plataspidae) and prospects for its biological control. Appl. Entomol. Zool. (Jpn.) 48: 3-13.

Suiter, D. R., J. E. Eger Jr., W. A. Gardner, R. C. Kemerait, J. N. All, P. M. Roberts, J. K. Green, L. M. Ames, G. D. Buntin, T. M. Jenkins and G. K. Douce. 2010. Discovery and distribution of *Megacopta cribraria* (Hemiptera: Heteroptera: Plataspidae) in Northeast Georgia. J. Integrat. Pest Manag. 1(1): 2010., doi: 10.1603/IPM10009.

Wang, M. L., N. A. Barkley and T. M. Jenkins. 2009. Microsatellite markers in plants and insects, Part I: Applications and Biotechnology. Genes, Genomes. Genomics 3: 54-67.