Isolation and Characterization of Microsatellite Markers for Soybean Looper (Lepidoptera: Noctuidae)

C. S. Silva,1 E. M. G. Cordeiro,1 and A. S. Corrêa1,2,*

1Laboratory of Arthropod Molecular Ecology, Department of Entomology and Acarology, University of Sao Paulo, Luiz de Queiroz College of Agriculture, Piracicaba, Sao Paulo 13418900, Brazil and 2Corresponding author, e-mail: ascorrea@usp.br

Subject Editor: Margaret Allen

Received 14 July 2019; Editorial decision 16 September 2019

Abstract

We constructed 13 microsatellite markers for Chrysodeixis includens (Walker), a serious crop pest in the Americas. All SSR (Simple Sequence Repeat) markers were polymorphic, with no evidence of linkage disequilibrium between any loci in any population. The total number of alleles per locus ranged from 5 for L3, L9, and L11 to 26 for L6; the mean number of alleles per locus in three populations of C. includens ranged from 2.33 for L3 to 14.67 for L6. Hardy–Weinberg equilibrium (HWE) deviation was not observed in four loci for at least one population (L3, L5, L9, L10). Markers L6, L7, L8, L10, L11, L12, and L13 showed a frequency of null alleles > 0.2 for at least one population. STRUCTURE and F-statistics revealed low population structure among the populations (FST = 0.013) and a high degree of inbreeding (FIS = 0.658). The SSR markers developed here will be useful in future studies on the ecology, demography, host dynamics, and gene flow of C. includens. This information is essential to understand the recent status of C. includens as a key pest in South America.

Key words: SSR markers, agricultural pest, genetic structure, gene flow, pest outbreak

The soybean looper Chrysodeixis includens (Walker, 1858) is one of the most serious lepidopteran agricultural pests, attacking a wide range of crops including soybean, bean, cotton, and tomato (Baldin et al. 2014, Specht et al. 2015). This polyphagous species causes extensive defoliation and damage to reproductive structures, including maturing fruits (Herzog 1980). Chrysodeixis includens is widely distributed in the western hemisphere, from the northern United States to Argentina (Kogan and Turnpseed 1987). Historically, C. includens has been a pest of secondary importance; however, in the last two decades, numerous population outbreaks have occurred in soybean crops throughout South America, raising C. includens to the status of key pest for this crop (Specht et al. 2015).

Population outbreaks of C. includens have been associated with excessive use of fungicides to control Asian soybean rust (Phakopsora pachyrhizi Sydow & Sydow [Uredinales: Phakopsoraceae]), which can negatively affect entomopathogenic fungi such as Metarhizium rileyi (Farl.) Samson (Ascomycota: Clavicipitaceae) and beneficial predatory and parasitoid arthropods (Sosa-Goméz et al. 2003; Pereira et al. 2018). Chrysodeixis includens outbreaks may also be fostered by the advancement of agricultural frontiers into central and northern Brazil, together with the intensive cropping systems that provide abundant food, high selection pressure for pesticides and cultivated hosts, and increase connections among distinct populations (Santos et al. 2017, Soares et al. 2018, Corrêa et al. 2019, Stacke et al. 2019). These hypotheses are untested, and the lack of information on the historical and contemporary dynamics of C. includens populations increases the difficulty of answering these questions.

A population study of C. includens, using ISSR (Inter Simple Sequence Repeat) markers, indicated low genetic diversity and high genetic similarity among populations from different Brazilian regions (Palma et al. 2015). However, ISSRs are dominant markers and have problems of reproducibility, which may affect measurement of genetic structure and gene flow among populations (McGregor et al. 2000). An alternative is microsatellite markers (SSRs), which are widely used in population studies. These markers are distributed throughout the genome and have features such as codominance, multi-allelic, high reproducibility, and transferability among related species (Mason 2015).

SSR markers have been used to determine the genetic structure, gene flow, and hybridization events for several lepidopteran pests (A’Hara and Cottrell 2013, Wei et al. 2013, Jing et al. 2016, Leite et al. 2016, Leitão et al. 2016, Seymor et al. 2016, Arias et al. 2019). To develop markers for C. includens, we 1) constructed 13 microsatellite markers, 2) evaluated the polymorphism of these markers, and 3) validated them for future population studies of C. includens.

Materials and Methods

Insect Collection and DNA Extraction

We used individuals of C. includens obtained from the Laboratory of Insect Biology of USP/Esalq to construct the SSR
library and standardize the amplification reactions of SSR markers. Subsequently, to validate the SSR markers and estimate the genetic indexes, we used 72 insects from three locations in Brazil: Coxilha, Rio Grande do Sul (28°10′54″S; 52°44′46″W); Campo Verde, Mato Grosso (15°25′34″S; 54°48′05″W); and Russas, Ceará (04°44′28″S; 38°00′14″W).

Total genomic DNA (gDNA) was extracted from the thorax of adult insects, using the CTAB protocol adapted from Black and Dutek (1997). The gDNA quality and quantity were checked on agarose gel (0.8% w/v) and Nanodrop (Thermo Fisher Scientific, Madison, WI), respectively.

Microsatellite-Enriched Library Protocol
The microsatellite-enriched library was constructed according to the protocol described by Billotte et al. (1999). Initially, the total gDNA was digested with the Afa I/Rsa I restriction endonucleases. We then performed the linkages of Rsa21 and Rsa25 adapters, followed by fragment amplification. The SSR library was enriched using the biotinylated probes Biotin-III (CT) and Biotin-III (GT) (Promega, Madison, WI). The enriched fragments were cloned into a pGEM-T vector (Promega) and then transformed via electroporation into Escherichia coli XL1-Blue competent cells (Stratagene, San Diego, CA). Transformed bacteria were selected using the β-galactosidase gene and grown overnight in 2YT-HMFM medium with ampicillin and tetracycline. Finally, 96 cloned bacteria were amplified and sequenced using the SP6 and T7 primers in a Genetic Analyzer 3500XL (Applied Biosystems, Waltham, MA) automated sequencer. The presence of nucleotide tandem repeats was verified using WebSat software (Martins et al. 2009).

Primer Design, PCR, and Genotype
We found 16 sequences with tandem-repeat regions among 96 fragments sequenced. The primers were designed using WebSat and Primer3 (Untergasser et al. 2012) for the following conditions: amplicon size from 100 to 400 bp and annealing temperature of 55–60°C.

PCRs were performed in a total volume of 15 μl containing 50 ng DNA, 10× (100 mM Tris-HCl [pH 8.8], 500 mM KCl, 1% Triton X-100), dNTP (100 μM each), 0.5 mM of each primer (R and F), 5 U of Taq, and MgCl2 (25 mM); The MgCl2 was adjusted according to the protocol described by Billotte et al. (1999). The reaction mixture and PCR conditions were as follows: initial denaturation at 94°C for 3 min, followed by 35 cycles of 94°C for 45 s, 55–60°C for 45 s, and 72°C for 1 min. After 35 cycles, a final extension was performed at 72°C for 20 min. The amplification products were inspected using agarose gel (0.8% w/v).

Our preliminary results indicated that 13 loci showed satisfactory amplicons. These 13 forward primers were synthesized with fluorescent-labeled dyes (FAM, NED, VIC, and PET) (Applied Biosystems). The primers were grouped into four multiplex reactions based on the size and type of the fluorescent, three of them performed with three primer sets (L1-L4-L7; L3-L8-L12; L4-L5-L11) and one with four primer sets (L2-L9-L10-L13), based on the size and type of the fluorescent labeling. Reactions were performed in a total volume of 15 μl, containing 50 ng of gDNA, 10 mM buffer (100 mM Tris-HCl [pH 8.8], 500 mM KCl, 1% Triton X-100, 1.5), dNTP (100 μM each), 0.5 mM of each primer (R and F), 5 U/μl of Taq, and 25 mM MgCl2. The genotype reaction was carried out with 1 μl of PCR product of the multiplex reactions with 9.2 μl formamide and 0.2 μl of the GeneScan 500 Liz dye size standard (Applied Biosystems). Genotyping was carried out in a sequencer ABI 3500xL Genetic Analyzer. The fragments were visualized and tagged using the GeneMarker v.2.6.3 program (SoftGenetics, State College, PA).

Data Analyses
The number of alleles per locus, the observed heterozygosity (H0), the expected heterozygosity (H†), and the inbreeding coefficient (Fis) were calculated using the software GDA v. 1.1 (Lewis and Zaykin 2001). The number of private alleles, linkage disequilibrium test, Hardy–Weinberg equilibrium (HWE), and F-statistics (Weir and Cockerman 1984) were estimated by the software Fstat (Goudet 2001). The frequency of null alleles and pairwise FIST were determined using the FreeNA software (Chapuis and Estoup 2007).

We also estimated the genetic relationships among the C. includens subpopulations, using a Bayesian approach with the software STRUCTURE v.2.3.4 (Pritchard et al. 2000). We performed 20 independent runs with 250,000 burn-in and 750,000 MCMC steps. The number of clusters (K) tested ranged from 1 to 6, as suggested by Evanno et al. (2005). The criterion for establishing the best K was the AK method estimated in the Structure Harvester Program (Earl and vonHoldt 2012).

Results
We constructed and genotyped 13 microsatellite markers for C. includens. All SSR markers were polymorphic, with fragment amplitudes ranging from 99 to 392 bp (Table 1). The L1 and L4 SSR markers, although polymorphic, were successfully genotyped in only 10% and 25% of the individuals tested, respectively. These two SSR markers were not used in the subsequent analyses because of the large amount of missing data.

The total number of alleles per locus ranged from 5 for L3, L9, and L11 to 26 for L6 (Table 2). The mean number of alleles per locus, including all three populations of C. includens, ranged from 2.33 for L3 to 14.67 for L6 (Table 2). The number of private alleles including the three populations and all 11 loci was 51, consisting of 20 for Coxilha (RSCO), 19 for Campo Verde (MTCV), and 12 for Russas (CERU). The observed and expected heterozygosity indicated a high proportion of homozygotes for all loci and populations of C. includens (Table 2).

There was no evidence of linkage disequilibrium between any of the loci in any population, confirming that they can be used as independent genetic markers. The HWE test indicated that four loci were in equilibrium for at least one population (L3, L5, L9, L10), while the others were not in HWE (Table 2). The markers L6, L7, L8, L9, L10, L11, L12, and L13 showed a frequency of null alleles P > 0.20 (Table 2).

The F-statistic revealed a low population structure among the populations of C. includens (FST = 0.013) and a high degree of inbreeding (Fis = 0.658 and FST = 0.662). The STRUCTURE analysis grouped the individuals of C. includens into four clusters (K = 4); however, comparison of the clusters considering each individual identified no genetic structuring among the sites (Fig. 1).

Discussion
Microsatellite markers are employed extensively in genetic studies (Varshney et al. 2005, Dutech et al. 2007, Chapman et al. 2009). We developed the first microsatellite markers for the moth C. includens, a pest of crops in the Americas. The SSR loci segregated independently, were polymorphic, and generated a high number of alleles.
even in a relatively small number of individuals, confirming their usefulness for genetic studies of *C. includens*. Two markers (L1 and L4), despite the low amplification success, are polymorphic and can be used in future studies with better optimization of amplification reactions (PCR). Another interesting trait in these molecular markers was the high number of private alleles, 51 in the 11 SSR markers. Private alleles are useful to characterize the genetic structure and gene flow of a species (Slatkin 1985).

### Table 1. Characteristic of the 13 microsatellite loci isolated for *Chrysodeixis includens* (Lepidoptera: Noctuidae)

| Locus | Size range | Motif$^a$ | Primers | Annealing temperature | GenBank accession number |
|-------|------------|----------|---------|-----------------------|-------------------------|
| L1    | 236 bp$^b$ | (TG)$_7$ + (GT)$_7$ | F – TGGCAACCAATGAAATGAAA R – ACTAGCTGTTGCGCGTACT | 58°C | MN204626 |
| L2    | 384 bp (358–392) | (GT)$_4$ | F – TTTGTAATGCGACCACATTG R – AAGGCAGTGGACTGAAAGCA | 62°C | MN204627 |
| L3    | 209 bp (189–221) | (TG)$_6$ | F – CCTGAGGATAAGGTGACGACT R – GCAATCAGGATATACGGGAAC | 57°C | MN204628 |
| L4    | 339 bp$^b$ | (AC)$_9$ | F – GAGTTCGAGGCCTGTTTGAG R – GCGGTTAGTGGGACTGTTGATA | 60°C | MN204629 |
| L5    | 109 bp (99–137) | (AC)$_9$ | F – ATTTACGGCTTCACGACAGG R – TGCAACGAAGCGAATTTACA | 55°C | MN204630 |
| L6    | 336 bp (304–392) | (GT)$_8$ | F – TTTGTATGGCACCCATTGTG R – AAGCAGTGGACTGCAACAGA | 62°C | MN204631 |
| L7    | 209 bp (189–221) | (TG)$_6$ | F – CCTGAGGATAAGGTGACGACT R – GCAATCAGGATATACGGGAAC | 57°C | MN204632 |
| L8    | 273 bp (239–295) | (CA)$_8$ | F – GGCAATGGTGTAAGCAGGTT R – TTACGTAGATACCTCCTTTTAGT | 62°C | MN204633 |
| L9    | 222 bp (220–236) | (GT)$_4$ | F – AAATCCTGGAGCCTGTTTGAG R – GCGGTTAGTGGGACTGTTGATA | 55°C | MN204634 |
| L10   | 156 bp (142–173) | (ATC)$_5$ | F – AAGGCAGAAGCAACCATCG R – GAAATCCTGGAGCCTGTTTGAG | 62°C | MN204635 |
| L11   | 217 bp (213–225) | (TG)$_5$ + (CA)$_6$ | F – GCTCGTGACCCTATTTCGAG R – GCGCCTATTCGCATCACTAT | 58°C | MN204636 |
| L12   | 359 bp (311–359) | (ACAT)$_7$ | F – TCCCCGAGAACCACTCTATG R – GGAAAGAGGCGTGACGTTTA | 55°C | MN204637 |
| L13   | 287 bp (287–335) | (GT)$_9$ | F – ATTTACGGCTTCACGACAGG R – TGGTCTTCTCCGACGTTTTT | 62°C | MN204638 |

$^a$Nucleotide tandem repeats.

$^b$Low frequency of amplification.

### Table 2. Genetic diversity estimates for each microsatellite loci and three populations of *Chrysoideixis includens*

| Population | A | 1 | 2 | 3 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | Mean pop. |
|------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---------|
| RSCO       | 11 | 0.811 | 0.921 | 0.924 | 0.796 | 0.923 | 0.268 | 0.818 | 0.543 | 0.836 | 0.921 | 0.706 | 8.363 |
|            | 0.400 | 0.666 | 0.200 | 0.000 | 0.222 | 0.125 | 0.272 | 0.166 | 0.000 | 0.187 | 0.203 | |
|            | 0.513* | 0.281* | 0.788* | 1.000* | 0.764* | 0.540 | 0.677* | 0.697* | 1.000* | 0.801* | 0.715* | |
|            | 0.220 | 0.001 | 0.120 | 0.369 | 0.432 | 0.357 | 0.154 | 0.288 | 0.231 | 0.441 | 0.372 | |
| MTCV       | 9 | 0.853 | 0.483 | 0.895 | 0.926 | 0.887 | 0.921 | 0.119 | 0.923 | 0.547 | 0.869 | 0.910 | 0.758 |
|            | 0.473 | 0.076 | 0.590 | 0.300 | 0.210 | 0.157 | 0.041 | 0.428 | 0.125 | 0.111 | 0.357 | 0.261 | |
|            | 0.451* | 0.846* | 0.345* | 0.682* | 0.767* | 0.832* | 0.656 | 0.555 | 0.775* | 0.878* | 0.616* | 0.663* | |
|            | 0.282 | 0.267 | 0.145 | 0.317 | 0.350 | 0.390 | 0.119 | 0.235 | 0.247 | 0.391 | 0.279 | |
| CERU       | 9 | 0.879 | 0.198 | 0.882 | 0.899 | 0.865 | 0.903 | 0.167 | 0.836 | 0.427 | 0.848 | 0.924 | 0.712 |
|            | 0.333 | 0.214 | 0.727 | 0.315 | 0.181 | 0.350 | 0.086 | 0.300 | 0.090 | 0.133 | 0.450 | 0.289 | |
|            | 0.627* | 0.083 | 0.179 | 0.654* | 0.797* | 0.618* | 0.485 | 0.653* | 0.791* | 0.847* | 0.519* | 0.606* | |
|            | 0.192 | 0.000 | 0.082 | 0.299 | 0.354 | 0.282 | 0.115 | 0.274 | 0.247 | 0.380 | 0.242 | |
| Mean/locus | 14 | 9.666 | 2.333 | 14.000 | 14.666 | 8.000 | 12.333 | 3.333 | 6.666 | 3.666 | 7.333 | 12.666 | 8.066 |
|            | 0.848 | 0.227 | 0.900 | 0.916 | 0.849 | 0.916 | 0.185 | 0.859 | 0.506 | 0.851 | 0.918 | 0.725 | |
|            | 0.402 | 0.097 | 0.661 | 0.271 | 0.130 | 0.240 | 0.084 | 0.333 | 0.127 | 0.081 | 0.331 | 0.251 | |
|            | 0.532 | 0.581 | 0.269 | 0.708 | 0.850 | 0.739 | 0.548 | 0.627 | 0.752 | 0.907 | 0.646 | 0.659 | |

$^A$: number of alleles; $H_e$: expected heterozygosity; $H_o$: observed heterozygosity; $f$: inbreeding coefficient; $N_a$: frequency of null alleles; $A_t$: total number of alleles per locus.

*Depart significantly from HWE ($P < 0.05$).
The analyses of heterozygosity revealed a high degree of inbreeding in these *C. includens* populations, where the observed heterozygosity (*H*<sub>O</sub>) was significantly lower than the expected heterozygosity (*H*<sub>E</sub>). This extensive inbreeding agrees with the significant deviation of the HWE and with the estimate of null alleles > 0.20 at some loci (Chakraborty et al. 1992). A deviation of HWE and a high frequency of null alleles are recurrent in studies with SSRs in natural populations of lepidopteran species, and may lead to over-estimation of population differentiation (Chapuis and Estoup 2007, Chapuis et al. 2008).

Huang et al. (2016) suggested that the frequency of a null allele should be considered individually for each locus, and that loci with a frequency of null alleles > 0.50 should be excluded. In some population studies with noctuid moths and other agricultural pests, few distortions were observed in the genetic structure and gene flow when comparing analyses using SSR loci with a frequency of null alleles > 0.2, or excluding them (see Endersby et al. 2007, Song et al. 2017, Leite et al. 2017). This may be explained by high natural homozygosity/interbreeding and low genetic structure of these insect pests, due to their high dispersal capacity and successive events of (re)colonization of annual crops.

The *F*<sub>ST</sub> index and STRUCTURE analysis indicated an absence of genetic structure and high genetic similarity among populations of *C. includens*, despite the wide geographical distances among the sites sampled. Two hypotheses can be proposed to account for this genetic similarity among the populations: high gene flow (Palma et al. 2015) and a recent population expansion. Both hypotheses should be more thoroughly tested for *C. includens*, using a larger number of subpopulations collected in different regions of Brazil and supported by different molecular markers (e.g., mitochondrial and single-nucleotide polymorphism [SNP] markers).

These SSR markers will be useful in future studies of the ecology, demography, host dynamics, and gene flow of *C. includens*. This information is essential to understanding the recent population outbreaks of *C. includens* in South America.

Studies of historical and current population dynamics may guide integrated and sustainable management strategies for this pest, as well as provide information for more-efficient pesticide applications and the use of genetically modified crops in order to contain or even prevent the evolution of resistance of *C. includens* to these control tactics.

**Acknowledgments**

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, fellowship 311465/2018-8 to A.S.C. and fellowship 140844/2016-3 to C.S.S.), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, fellowship 2017/02393-0 to E.M.G.C.). Authors were also supported by other grants given by FAPESP.

**References Cited**

A’Hara, S., and J. Cottrell. 2013. Development and characterisation of ten polymorphic microsatellite markers for the pine-tree lappet moth *Dendrolimus pini* (Lepidoptera: Lasiocampidae). Conserv. Genet. Resour. 5: 1135–1137.

Arias, O., Cordeiro, E., Corrêa, A. S., Domingues, F. A., Guidolin, A. S., and C. Omoto. 2019. Population genetic structure and demographic history of *Spodoptera frugiperda* (Lepidoptera: Noctuidae): implications for insect resistance management programs. Pest Manag. Sci. 75: 2948–2957.

Baldin, E. L. L., Lourenço, A. L., and E. C. Schlick-Souza. 2014. Outbreaks of *Chrysodeixis includens* (Walker) (Lepidoptera: Noctuidae) in common bean and castor bean in São Paulo State, Brazil. Bragantia. 73: 458–465.

Billotte, N., Lagoda, P. J. L., Risterucci, A. M., and F. C. Baurens 1999. Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits. 54: 277–288.

Black, W. C., and N. M. Duteau. 1997. RAPD-PCR and SSCP analysis for insect population genetic studies, pp 361–373. In J. M. Crampton, C. B. Beard, and C. Louis (eds.), *The molecular biology of insect disease vectors: a methods manual*. Chapman & Hall, London, United Kingdom.
