Use of expressed sequence tag microsatellite markers for population genetic research of *Helicoverpa armigera* (Lepidoptera: Noctuidae) from India

Thiruvengadam Venkatesan,[2] Vaddi Sridhar, Yan R. Tomason, Sushil Kumar Jalali, Gajanan T. Behere, Rajagopalan M. Shanthi, Ravinder Kumar, Venkata Gopinath Vajja, Padma Nimmakayala, Umesh K. Reddy[1]

Abstract—Cotton bollworm, *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae), is a serious pest of several crops throughout the world, representing millions of United States of America dollars worth of damage. This pest can adapt to various cropping systems in a wide geographical range and has high migratory potential. It features high fecundity and can develop resistance to almost all insecticides used for its management. Several investigations to develop microsatellite markers for *H. armigera* have not been successful because of the paucity of microsatellites in the lepidopteran genome. As well, collections of *H. armigera* from cotton fields of southern and western India were not yet studied for molecular genetic diversity. The current study aimed to screen publicly available expressed sequence tag resources for simple sequence repeats and assess their potential as DNA markers for assessment of gene flow between collections of southern and western India. We identified 30 polymorphic microsatellites for potential use in diversity analysis of *H. armigera* collections. Genetic diversity analysis revealed that the collections were widely diverse with population differentiation index (F_{st}) of 0.17. Furthermore, gene flow analysis revealed a mean frequency of private alleles of 11% within the collections. The microsatellite resources we developed could be widely used for molecular diversity or population genetic research involving this important pest of cotton and food crops.

Introduction

Cotton bollworm, *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae), is a serious pest of several crops including maize, sorghum, tobacco, pigeon pea, chickpea, tomato, pepper, cotton, and sunflower and is widely distributed throughout the world. Damage to crops represents millions of United States of America dollars (Tan *et al.* 2001; Ji *et al.* 2003; Scott *et al.* 2004). This pest is found on more than 182 plant species belonging to 47 families: 56 species are heavily damaged and 126 are severely affected (Pawar *et al.* 1986). It can adapt to various cropping systems in a wide geographical range and has high migratory potential, travelling up to 1000 km.

Received 26 November 2014. Accepted 18 March 2015. First published online 14 July 2015.

T. Venkatesan,[2] S.K. Jalali, ICAR-National Bureau of Agricultural Insect Resources, Bangalore 560 024, India
V. Sridhar, ICAR-Indian Institute of Horticultural Research, Bangalore 560 089, India
Y.R. Tomason, V.G. Vajja, P. Nimmakayala, U.K. Reddy,[1] Gus R. Douglass Institute and Department of Biology, West Virginia State University, West Virginia 25112-1000, United States of America
G.T. Behere, Indian Council of Agricultural Research-Research Complex-North East Region, Umiam 793 103, Meghalaya, India
R.M. Shanthi, R. Kumar, Indian Council of Agricultural Research-Sugarcane Breeding Institute, Coimbatore 641 007, India

[1]Contributed equally to the first author.

[2]Corresponding author (e-mail: tvenkat12@gmail.com).
Subject editor: Amanda Roe
doi:10.4039/tce.2015.47

© 2015 Entomological Society of Canada

It has high fecundity (Fitt 1989), with 3000 eggs/female, and can develop resistance to almost all insecticides used for its management (Tang *et al.* 1988; Forrester *et al.* 1993; Ahmad *et al.* 2001).

Microsatellite markers or simple sequence repeats (SSRs) are the most widely used tools to infer ecological and evolutionary hypotheses among natural or experimental populations. Simple sequence repeats are highly polymorphic, co-dominant, and reliable for population genetic studies and can be used with multiplexing (Tautz and Renz 1984; Tautz *et al.* 1986; Zhang and Hewitt 2003; Selkoe and Toonen 2006). Some microsatellites from expressed sequence tags (ESTs) and unigenes have potential to alter gene function by affecting phenotype and development. Microsatellites and other markers have been used in population genetic studies of *H. armigera* across the world. Microsatellites are the markers of choice to study genetic diversity and gene flow among the populations, but readily available polymorphic microsatellites are < 50 (McKenchnie *et al.* 1993; Scott *et al.* 2003; Subramanian and Mohankumar 2006; Behere *et al.* 2007). Several investigations have attempted to develop microsatellite markers for *H. armigera* but have not been successful (Nève and Meglécz 2000; Ji *et al.* 2003; Meglecz *et al.* 2004; Zhang 2004).

Helicoverpa armigera feed on cotton from August to October in western India and September to December in southern India before switching to other food crops (Behere *et al.* 2013). Genetic variation among various geographic populations depends on several factors, including gene flow among the populations, host range and time since separation (Templeton *et al.* 1990; Behere *et al.* 2013). Behere *et al.* (2013) hypothesised increased gene flow between central and southern Indian *H. armigera* populations.

Collections of *H. armigera* from cotton fields of southern and western India have not been analysed for genetic diversity or gene flow. In this study, we screened publicly available EST resources for SSRs to develop polymorphic markers for use in diversity analysis of *H. armigera* populations from southern and western India. The microsatellite resources we developed can be used widely for molecular diversity or population genetic research involving this important pest of cotton and food crops.

Results

Frequency, distribution, and organisation of microsatellites in *Helicoverpa* gene assemblies

In total, 3082 repeat motifs were identified from 57 248 contigs: 559 were dinucleotide repeats (DNRs), 819 trinucleotide repeats (TNRs), 1455 tetranucleotide repeats (TTNRs), and 249 pentanucleotide repeats (PNRs) (Table 2). The most common motif type of DNRs was AT/TA (34%) followed by AC/TG (21%) and CA/GT (18%). We identified 27% TNR repeats from all repeats; the most prevalent was AAT/TTA (32%) followed by ATC/TAG (16%). We identified 559 loci among the GC-rich, frequent TNR motifs (ACA/TGT, ATC/TAG, ACC/TGG, CAA/GGT, CGA/GCT, and CTA/GAT). AT-rich TTNR motifs had the highest frequency, representing 64% of the total tetra class. In total, 139 AAAAC/TTTTG loci in the penta class represented 56% of the total PNR class. Of all 69% were simple imperfect repeats, followed by 17% simple perfect repeats and the rest compound motifs.

Dinucleotide repeats tended to be longer than TNRs (Table 1S). Many of the DNRs (474) and TNRs (69) contained five units. None of the motifs were longer than 11 units. In total, 1338 repeat units were located in open reading frames (ORFs). Among the rest, 892 and 841 were in 5' and 3' UTRs, respectively. Overall, 211 could not be localised to any geneic region. Sequences for forward and reverse primers and expected product sizes for 993 SSR motives are in Table 2S.

Amplification pattern of ORF-specific microsatellites and their use in analysis of genetic divergence among the populations

We used 30 microsatellite primer pairs to amplify 20 DNA samples of *H. armigera* from southern and western regions of India (Table 3). Ha1, 8, 36, 38, 51, 54, 56, 69, and 74 significantly deviated from Hardy–Weinberg equlibrium (Table 4). The mean observed heterozygosity for all microsatellite alleles was 0.59 (range 0.1–0.84). The Shannon index (SI) can be used to determine how a locus resolves genetic divergence in ecologically diverse populations. As well, to confirm gene sharing across populations, we used number of migrants (Nm) testing with private alleles (Barton and Slatkin 1986). A mean Nm

© 2015 Entomological Society of Canada

Table 1. Sampling of *Helicoverpa armigera* populations used in the study.

Sample	Location	Latitude	Longitude
1	Aurangabad, Maharashtra	19°53'N	75°23'N
2	Rajkot, Gujarat	22°18'N	70°56'N
3	Gulbarga, Karnataka	17°19'N	76°54'N
4	Coimbatore, TamilNadu	11°00'N	77°00'N
5	Junagadh, Gujarat	21°31'N	70°36'N
6	Guntur, Andhra Pradesh	16°18'N	80°29'N
7	Warangal, Andhra Pradesh	17°58'N	79°40'N
8	Davanagiri, Karnataka	14°31'N	75°58'N
9	Dharwad, Karnataka	15°27'N	75°05'N
10	Anand, Gujarat	22°32'N	73°00'N
11	Kolar, Karnataka	13°09'N	78°11'N
12	Jamnagar, Gujarat	22°27'N	70°07'N
13	Mysore, Karnataka	12°18'N	76°42'N
14	Akola, Maharastra	21°18'N	77°33'N
15	Ahmedabad, Gujarat	20°03'N	72°40'N
16	Bellary, Karnataka	15°09'N	76°55'N
17	Ongole, Andhra Pradesh	15°30'N	80°03'N
18	Vadadora, Gujarat	22°00'N	73°16'N
19	Tumkur, Karnataka	13°20'N	77°08'N
20	Bangalore, Karnataka	12°58'N	77°38'N

Table 2. Distribution of dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide microsatellites by repeat number.

Repeat number	Di	Tri	Tetra	Penta	Total
3	0	0	1411	223	1634
4	0	744	39	25	808
5	474	63	3	1	541
6	69	7	1	0	77
7	14	3	0	0	17
8	0	1	1	0	2
11	2	1	0	0	3
Total	559	819	1455	249	3082

frequency > 1.0 implies significant gene flow among the populations. The mean frequency of private alleles across groups was 0.11 and the Nm after correction for population size was 1.18688. Locus wise Nm estimates (locus wise gene flow) were calculated based on $F_{st} = 0.25(1 - F_{st})/F_{st}$. Nm indices of individual SSR loci estimated by ranged from 0 to 0.39 (Table 3S). Principal component analysis (PCA) resolved genetic relationships of various collections from southern and western regions of India, which indicates gene flow and migration among the pest populations (Fig. 2). The first two eigen vectors cumulatively accounted for 23.1% of the variation (vector I = 13.5 and vector II = 9.6). To corroborate the results of PCA, genetic diversity analysis of the collections ranged from 0% to 32%.

Pairwise genetic diversity estimates varied between 0.25 (between collection from Warangal, Andhra Pradesh, and Ongole, Andhra Pradesh) to 0.77 (between Junagarh, Gujrat collection and Tumkur, Karnataka) (Table 5). A phenogram created by the unweighted pair group method with arithmetic mean (UPGMA) contained groups representing mixtures of collections from various states (Fig. 3). To further partition genetic diversity, analysis of molecular variance (AMOVA) of two collections, southern and western, revealed 90.93% intragroup variance as compared with 9.07% intergroup variance, which corroborates the results of the PCA and UPGMA.

Population structure analysis based on model-based assumptions was used to estimate K-2 to K-10 clusters each with three iterations. The results were analysed for mean ± SD LnP(K) and ΔK values as estimated by Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/), which implements the Evanno method for visualising structure output (Earl and vonHoldt 2012).

© 2015 Entomological Society of Canada

Table 3. Gene-specific (ORF) microsatellites used in the current investigation.

Name of the primer	Forward primers	Reverse primers	Motif	Repeat	Blast score	Match	Gene	Organism
Ha1	AACACGTTACCGGAAACACC	TCTGTGCCACTCCACAAGTC	aag	5	1e-35	Q5R4T9	Eukaryotic translation initiation factor 2 subunit 2	*Pongo abelii* (Lesson) (Primates: Hominidae)
Ha5	CGGAGAAATCGTTGAGGAAG	AAGAGTGTTTTCGAGTCGGG	aag	5				
Ha8	GAACCTGAACCTGAACCTGC	GGGGCTTTTCTTCAATTTCC	aag	7				
Ha24	AACGAAGCCATCCGAGTCTA	GCACCCCTTGCAATTCAGTA	atg	5	5e-08	Q8K039	Uncharacterised protein KIAA1143 homolog	*Mus musculus* (Linnaeus) (Rodentia: Muridae)
Ha25	CCGCCTTCTGAAGATGGATA	TCGTTTTTAGTGGCTTCATCA	atg	5	5e-22	O46072	Probable ATP-dependent RNA helicase kurz	*Drosophila melanogaster* Meigen (Diptera: Drosophilidae)
Ha28	CTACGAGAAGGACTGCTGGG	ATCGGTCCCAGTAGGAAAGG	ca	11	0.006	Q9Z1K5	Protein ariadne-1 homolog	*M. musculus*
Ha29	GAATGGGGTCATTCCTGAAA	TATTCCTCCGTGCGACTACC	ca	7				
Ha32	CAACCCAGGAAACGTGAACT	TTGTAGGGTCTTGTTGGCCT	caa	5				
Ha36	GCAGCTGGCTGATCTCCTTA	CAAGCAATTTCGGCATCTTT	cat	5				
Ha37	AATGTTGTCGGTTCACATCG	ACAAAGGCTTCATCGACACC	cct	5	5e-64	Q1HPS0	Myosin regulatory light chain 2	*Bombyx mori* (Linnaeus) (Lepidoptera: Bombycidae)
Ha38	TGAGGTCCAGCCTTCTGACT	ACATGCCATTGTCAAGCAAA	cga	5	5e-64	Q1HPS0	Myosin regulatory light chain2	*B. mori*
Ha40	GTGCTTCTTCCTGGGCG	CACAGCAACAGAGGAGCAGA	ctg	5	2e-37	P56941	Niemann-Pick C1 protein	*Sus scrofa* (Linnaeus) (Artiodactyla: Suidae)
Ha41	TGGCTTTTTCTGGTTCTGCT	GCCAGTGACCGTAGTTGTGA	ctt	7				
Ha42	AAGAAAAAGCCGCTAAAGCC	TAAGTGGGGAGATCATTCGC	gaa	6	3e-15	Q05783	High mobility group protein D	*D. melanogaster*
Ha44	CCGTGATATCTTCCGTGGTT	CATACGCGTCGTTTACGAAT	gaa	5	2e-18	Q8BKU8	Trans membrane protein 87B	*M. musculus*
Ha49	GAAGCAGAAGCACAAGTCCC	CTTTCTTCCGTCGAGCATTC	gac	5	5e-26	Q7ZY81	Nucleosome assembly protein1-like1-B	*Xenopus laevis* (Daudin) (Anura: Pipidae)

© 2015 Entomological Society of Canada

Table 3. *Continued*

Name of the primer	Forward primers	Reverse primers	Motif	Repeat	Blast score	Match	Gene	Organism
Ha51	AGAGGACGACGATCTCGAAG	AACTGCACCCGTTTTAGGTG	gag	5	0	Q9URV0	Uncharacterised RNA-binding protein C106.12c	*Schizosaccharomyces pombe* Lindner (Saccharomycetaceae)
Ha52	CCTAAAAGGAGCCAAAAGGG	TAAAGAGGCACGAGGCGTAT	gag	5	0	Q70IV5	Synemin	*M. musculus*
Ha54	TGGAAGAGGAGCTACTCGGA	AACTGCACCCGTTTTAGGTG	gag	5				
Ha55	CAGATTTGCTCTGGAAGCCT	CTCTCCCTTCAACTGCCAAG	gat	5	1e-55	O00566	U3 small nucleolar ribonucleo protein MPP10	*Homo sapiens* Linnaeus (Primates: Hominidae)
Ha56	CTTAGCCGAACAGACCGAAG	GAGGCTACCACCATCAAGGA	gcg	5	7e-54	Q9U3U0	60S acidic ribosomal protein P0	*Ceratitis capitata* (Wiedemann) (Diptera: Tephrididae)
Ha57	GTCAAGAAGAACCAAGGCCA	ACCACAGCCCTGACGAGA	gcg	5	3e-05	Q3T0T5	Hematological and neurological expressed 1 protein	*Bos taurus* Linnaeus (Artiodactyla: Bovidae)
Ha59	GGAGGAGGATGAGGAGGAAC	CAGTCGGGGGCAGCAT	gct	6				
Ha60	AAACTCAAGTCAAGTGGCGG	GGGTAGGGGTAGGGTTGGTA	ggc	5				
Ha67	AGTACAATCCGACCCGTGAG	GAGGTTGAAGAGGACGACGA	tcc	5	7e-08	P19351	Troponin T, skeletal muscle	*D. melanogaster*
Ha69	AAACAGTACAGTCCGACCCG	GAGGTTGAAGAGGACGACGA	tct	15	7e-11	P19351	TroponinT, skeletal muscle	*D. melanogaster*
Ha74	GTGGTCCGGTGAAGAAGAGA	ATTACAAACAGTGGCCGAGG	tg	7				
Ha80	GAGCGTGCCAAGATACACAA	TAAGGAGGTGGTGGTTCGAC	tgc	5				
Ha82	TCATCCACCACGCAAGTTTA	CACGGTTAATGAGATACAGCC	tgtc	5				

© 2015 Entomological Society of Canada

Table 4. Amplification pattern and heterozygosity levels of individual microsatellites.

Name of the primer	Repeat length	Allele number	PIC	Ho	He	*P*-value
Ha1	5	2	0.5	0	0.52	0.00023
Ha5	5	10	0.72	0.9	0.74	0.73336
Ha8	7	3	0.28	0.06	0.29	0.01027
Ha24	5	2	0.5	0	0.53	0.0727
Ha25	5	3	0.51	0.05	0.54	0.13081
Ha28	11	3	0.49	0.52	0.5	0.04488
Ha29	7	2	0.09	0	0.1	0.10207
Ha32	5	3	0.61	0.05	0.64	0.64602
Ha36	5	10	0.79	0.68	0.81	0.02156
Ha37	5	2	0.46	0	0.48	0.17825
Ha38	5	3	0.59	0	0.62	0.00665
Ha40	5	7	0.69	0.5	0.71	0.09447
Ha41	7	2	0.36	0	0.38	0.45607
Ha42	6	5	0.48	0.33	0.5	0.11367
Ha44	5	7	0.75	0.25	0.78	0.34532
Ha49	5	5	0.65	0.38	0.68	0.30285
Ha51	5	4	0.66	0	0.69	0.00176
Ha52	5	5	0.6	0.32	0.62	0.79801
Ha53	5	8	0.58	0.1	0.6	0.13775
Ha54	5	2	0.38	0	0.39	0.01946
Ha55	5	2	0.27	0	0.28	0.79285
Ha56	5	4	0.55	0.11	0.57	0.01161
Ha57	5	8	0.79	0.6	0.82	0.44159
Ha59	6	5	0.67	0.25	0.71	0.10606
Ha60	5	10	0.81	0.53	0.84	0.08633
Ha67	5	6	0.76	0.25	0.79	0.04528
Ha69	15	6	0.72	0.2	0.75	0.01293
Ha74	7	3	0.62	0	0.66	0
Ha80	5	3	0.64	0	0.67	0.49918

K-4 was the most appropriate cluster size for this population, with the highest $\Delta K = 312$ (Fig. S1). Population structure analysis revealed information pertaining to lineage sharing among the collections in addition to corroborating the results of a neighbour joining analysis.

Discussion

Microsatellites are located all over genome and are the hot spots because of their ability to resolve diversity and gene flow (Katti *et al.* 2001; Archak *et al.* 2007). In this study of collections of cotton bollworm, *H. armigera*, from southern and western India, we have made available 3082 EST specific microsatellite markers and provide information for the motif type, frequency, distribution, location (ORF or UTR), nature (perfect, imperfect, and compound) as well as abundance of the markers. Similar to our findings of the abundance of short microsatellites and imperfect repeat units in the *Helicoverpa* genome, Archak *et al.* (2007) noted increased numbers of short microsatellites with imperfect units in *Anopheles* Meigen (Diptera: Culicidae), *Bombyx* Linnaeus (Lepidoptera: Bombycidae), *Drosophila* Fallén (Diptera: Drosophilidae), and *Tribolium* MacLeay (Coleoptera: Tenebrionidae). Furthermore, Archak *et al.* (2007) observed $>70\%$ TNR microsatellites in ORFs which is in confirmity with our findings.

With the ORF-specific microsatellites, the *Helicoverpa* collections from various locations exhibited wide genetic diversity and also provided evidence of allele sharing and significant gene flow across the populations. Researchers across the world have investigated genetic differences among various *Helicoverpa* collections from

© 2015 Entomological Society of Canada

Table 5. Pair-wise genetic distances of various collections in the study. Serial number refers to the sample number in Table 1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0																			
2	0.7193	0																		
3	0.66667	0.76471	0																	
4	0.64286	0.66667	0.65957	0																
5	0.7377	0.67568	0.73333	0.56757	0															
6	0.41176	0.67347	0.65957	0.66667	0.7193	0														
7	0.56098	0.63158	0.67442	0.65385	0.68889	0.45455	0													
8	0.63265	0.71429	0.63158	0.53846	0.65854	0.57143	0.63636	0												
9	0.48718	0.5	0.55556	0.48718	0.51515	0.48718	0.4	0.55	0											
10	0.4	0.59184	0.59184	0.41463	0.53488	0.45455	0.47619	0.53191	0.35135	0										
11	0.65	0.80392	0.7619	0.75439	0.80769	0.51724	0.66667	0.69231	0.71429	0.55	0									
12	0.59184	0.69231	0.68627	0.63636	0.70909	0.59184	0.52632	0.64706	0.64286	0.54717	0.65854	0								
13	0.63265	0.61111	0.74545	0.66667	0.70213	0.59091	0.65957	0.66667	0.64	0.51111	0.68421	0.65385	0							
14	0.41463	0.41176	0.53488	0.45455	0.55556	0.35135	0.42105	0.35294	0.14286	0.31707	0.6	0.52941	0.54167	0						
15	0.51351	0.72	0.72549	0.67857	0.76271	0.63265	0.6	0.63636	0.59091	0.56	0.76923	0.64	0.65957	0.45	0					
16	0.59091	0.76667	0.74074	0.70968	0.70833	0.65385	0.69231	0.66667	0.66667	0.56863	0.7	0.64	0.65217	0.59259	0.55556	0				
17	0.31579	0.6	0.58491	0.57377	0.62712	0.31579	0.25	0.55556	0.42222	0.375	0.47368	0.45833	0.52941	0.43396	0.42857	0.48936	0			
18	0.56757	0.75	0.79661	0.71429	0.77358	0.65217	0.62162	0.70833	0.66667	0.52381	0.70588	0.62791	0.64103	0.64286	0.57576	0.57576	0.35294	0		
19	0.62791	0.76923	0.72093	0.65217	0.7551	0.62791	0.58824	0.5	0.61905	0.53488	0.66667	0.61905	0.70213	0.48718	0.65	0.65	0.47619	0.72727	0	
20	0.59184	0.60976	0.64444	0.60784	0.62791	0.63636	0.6	0.51351	0.52381	0.52	0.63158	0.55556	0.48571	0.44186	0.6	0.5	0.54386	0.61905	0.44828	0.00000

© 2015 Entomological Society of Canada

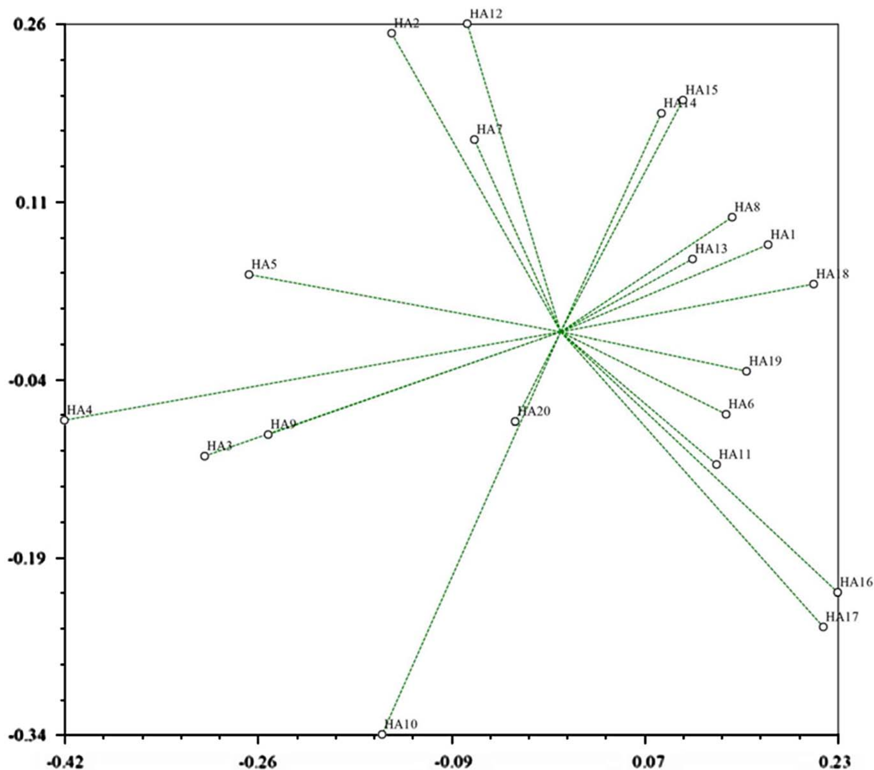
Fig. 1. Sampling locations of *Helicoverpa armigera* in India.

diverse topological barriers and found that this pest population is panmictic and highly mobile, which facilitates strong evolutionary potential to cope with pest management practices (Zhou *et al.* 2000; Grasela and McIntosh 2005; Ji *et al.* 2005; Subramanian and Mohankumar 2006; Behere *et al*. 2007). Behere *et al.* (2013), using exon-primed, intron-crossing (EPIC) markers, noted seasonal and geographical variation in Indian *H. armigera* populations from various crops. Frequent migration among the natural cotton bollworm populations was previously noted (Scott *et al.* 2006; Vassal *et al.* 2008; Behere *et al*. 2013). In addition, Scott *et al.* (2003) and Selkoeand Toonen (2006) reported that the adult moth movement varies among seasons, thereby affecting genetic variation. *Helicoverpa armigera* populations differentially responded to pheromones (Tamhankar *et al.* 2003) and parasitoids (Manjunath *et al.* 1970) in India. Kranthi *et al.* (1997) reported diverse metabolic mechanisms mediating pyrethroid resistance among collections of *H. armigera* in India. Various reports suggested that *H. armigera* in India could not be categorised into races, from their host feeding preferences (Reed and Pawar 1982).

Ji *et al.* (2003) noted that the expected heterozygosity of various genomic microsatellites ranged from 0.62 to 0.91 among *H. armigera* collections and the observed allele numbers ranged from four to 12. Of the 30 microsatellite loci we tested, nine deviated significantly from Hardy–Weinberg equilibrium (Table 4), as indicated by heterozygotes deficit. This phenomenon is not uncommon with pest species such as *H. armigera*, which are under extreme selection

© 2015 Entomological Society of Canada

Fig. 2. Principal component analysis of genetic relationships of *Helicoverpa armigera* collections.



pressure for every generation (Scott *et al.* 2005). The assumption of absence of migration and selection pressure under Hardy–Weinberg equilibrium is inappropriate in pest species such as *H. armigera* that undergo extreme selection pressure because of the application of insecticides (Scott *et al.* 2005). Nevertheless, we used microsatellites specific to ORFs to obtain a range of heterozygosity of 0.1 to 0.85 with allele numbers from two to 10, which is almost comparable to values for non-genic microsatellites used in previous studies (Tan *et al.* 2001; Yajie and Dexing 2003; Scott *et al.* 2004; Perera and Blanco 2011). Endersby *et al.* (2006) reported a high frequency of null alleles typically found in Lepidoptera. Populations of *H. armigera* are not strongly structured among regions in south-eastern Australia (Endersby *et al.* 2006; Weeks *et al.* 2010).

Our study identified wide genetic differentiation within *Helicoverpa* populations collected from cotton in southern and western India. However, the inclusion of a larger number of collections from various crops would further resolve the extent of gene flow across the pest populations. Our study was conducted with a limited number of collections from various locations, and larger samples from various collections need to be investigated to precisely understand the trend and mechanisms of dynamic evolution in this important pest population. Nevertheless, we have identified several useable microsatellites and various population genetic tools never previously used to analyse genetic diversity among *H. armigera* collections. The results we obtained with the newly developed microsatellites could be compared with those from investigations of larger sample involving multiple locations, cropping seasons and different host crops from India.

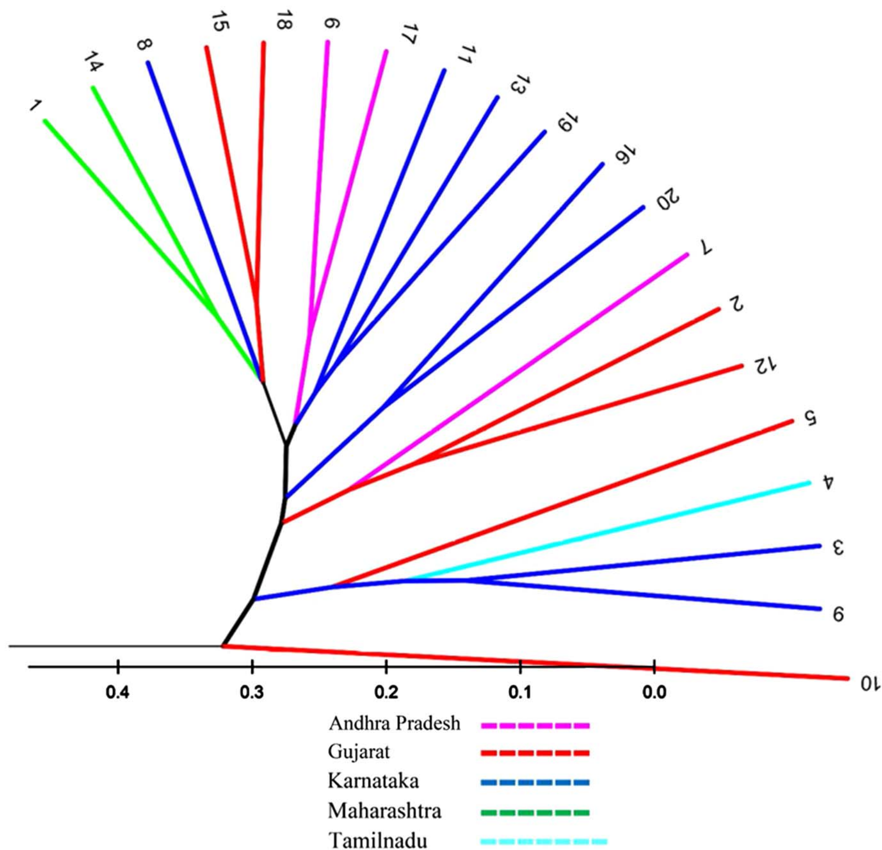
Methods

Insect collections and DNA isolation

The larvae of *H. armigera* were collected during the peak incidence from non-*Bt*-cotton

© 2015 Entomological Society of Canada

Fig. 3. Unweighted pair group method with arithmetic mean (UPGMA) phenogram of the association of *Helicoverpa armigera* collections from southern and western states of India.



fields of southern and western regions of India during 2008–2009 and 2009–2010. The collections were from 20 locations in the three southern states (Andhra Pradesh, Karnataka, and Tamil Nadu) and two western states (Maharashtra and Gujarat) (Fig. 1 and Table 1). In total, 50 larvae (4th–5th instar) per location were collected. A total of 1000 larvae belonging to 20 locations were transferred immediately into glass vials for rearing with an artificial diet (Nagarkatti and Prakash 1974). The larvae were transferred to fresh diet on arrival in the laboratory and allowed to pupate. The pupae were surface-sterilised in 0.1% sodium hypochlorite and allowed to emerge into moths. Newly emerged adults were identified as *H. armigera* (Hübner) by taxonomic keys provided by Hardwick (1965). Adults belonging to a particular location were separately preserved in 95% ethanol at −20 °C and used for DNA extraction. All larvae were thoroughly washed with formaldehyde and alcohol and the gut contents were removed to avoid contamination of any other DNA. The genomic DNA isolation involved use of an animal kit (catalogue number 69504, Qiagen (Valentia, California, United States of America) by standard procedures.

Marker development

A set of 57 248 EST assemblies were used for SSR searching with a high-performance computer cluster version developed with a Perl script available as SSRIT (SSR tool) at cotton microsatellite database (CMD; www.cottonmarker.org) (Blenda *et al.* 2006). The CMD SSR tool

© 2015 Entomological Society of Canada

parameters were set to detect dinucleotides to pentanucleotides with minimum repeat size of five, four, and three for dinucleotide, trinucleotide, and tetranucleotide and pentanucleotide motifs, respectively. To examine the location of SSRs in the sequences relative to the putative coding region, the SSR server is equipped with FLIP, a UNIX C program that identifies relative SSR location (Brossard 1997). The location of the SSRs within the contigs was further confirmed by use of prot4EST (Wasmuth and Blaxter 2004) and the fasty35 module of the FASTA package (Iseli *et al.* 1999). Potential primers were designed by using the Primer3 module with the SSR tool (Rozen and Skaletsky 2000). The individual properties included sequence name, repeat(s) motif and number and SSR location relative to the ORF. A set of alternative primers was designed with optimum primer length 20 nucleotide (range 18–26 nt), optimum melting temperature 50 °C (range 45–55 °C), and optimum product size range 100–350 base pairs. Simple sequence repeat containing sequences were identified by use of BLAST with the tblastx option (E-value cut-off 1e-6) against the Swiss-Prot database.

Polymerase chain reaction conditions

Polymerase chain reactions involved the use of 200 ng genomic DNA, 0.20 µM mixed forward and reverse primers, 1X Buffer (10 mM de Tris-HCl, pH 8.2, 50 mMKCl, Triton 0.1%, BSA 1 mg/mL), 1.5 mM MgCl2, 0.2 mMdNTPs, and 1 U Taq polymerase in a 10-µL reaction volume. Amplification involved a GeneAmp PCR 9700 System thermal cycler (Applied Biosystems Inc., Waltham, Massachusetts, United States of America) programmed to 94 °C for two minutes, followed by 35 cycles of 94 °C for 30 seconds, 50–65 °C for 30 seconds, 72 °C for one minute, and a final extension step at 72 °C for 10 minutes. Amplified products were separated on 3% SFR agarose gels. For scoring, stutters were avoided and discernible bands were scored as alleles. Estimation of allele sizes for the survey panel was by comparison with a 50 base pair molecular weight ladder that was loaded twice on each SFR grade agarose gel.

Data analysis

Estimation of molecular genetic diversity of ecological populations by the Shannon diversity index, F_{st} and heterozygosity based on Wright's F-statistic (Wright 1978) involved use of Popgene 1.31 (Yeh and Boyle 1997). Analysis of gene flow (number of migrants; Nm) by the private allele method (Slatkin 1985) involved use of Genepop 4.1 (Rousset 2008). Genetic distance and principal component analysis involved use of NTSYS-pc (Rohlf 2000). Structure 2.2 (Pritchard *et al.* 2000) was used to cluster individuals into K groups by estimating the posterior probability of the data for a given K, $\Pr(X|K)$. The number of populations (K) was determined by use of an admixture model with correlated alleles and $K = 2$–10. Five independent runs of 100 000 Markov Chain Monte Carlo generations, after 100 000 generation burn-in periods, were used to estimate each value of K. The optimal K value was determined by use of the ad-hoc statistic ΔK (Evanno *et al.* 2005). The number of Ks in each data set was evaluated by use of Structure Harvester (http://taylor0.biology.ucla.edu/structure Harvester), which implements the Evanno method for visualising structure output (Earl and vonHoldt 2012).

Acknowledgements

The authors are grateful to the Human Resource Development programme of the National Agricultural Innovation Project (NAIP/O&M/Hort/AM/4/1683), Indian Council of Agriculture Research, New Delhi, India for funding support for T.V. and V.S. to conduct research at West Virginia State University. We acknowledge the professional editing of Laura Smalls.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.4039/tce.2015.47.

References

Ahmad, M., Iqbal Arif, M., and Ahmad, Z. 2001. Resistance to carbamate insecticides in *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Pakistan. Crop Protection, **20**: 427–432.

Archak, S., Meduri, E., Kumar, P.S., and Nagaraju, J. 2007. InSatDb: a microsatellite database of fully sequenced insect genomes. Nucleic Acids Research, **35**: D36–D39.

© 2015 Entomological Society of Canada

Barton, N. and Slatkin, M. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity, **56**: 409–415.

Behere, G.T., Tay, W.T., and Russell, D.A. 2007. Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. BMC Evolutionary Biology, **7**: 117–126.

Behere, G.T., Tay, W.T., Russell, D.A., Kranthi, K.R., and Batterham, P. 2013. Population genetic structure of the cotton bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) in India as inferred from EPIC-PCR DNA Markers. Public Library of Science One, **8**: e53448.

Blenda, A., Scheffler, J., and Scheffler, B. 2006. CMD: a cotton microsatellite database resource for *Gossypium* genomics. BMC Genomics, 7: 132. Available from http://www.biomedcentral.com/1471-2164/7/132 [accessed 18 May 2015].

Brossard, N. 1997. FLIP: a Unix program used to find/translate ORFs. Bionet Software, Montréal, Québec, Canada. Available from http://megasun.bch.umontreal.ca/ogmp/ogmpid.html [accessed 25 May 2015].

Earl, D. and vonHoldt, B. 2012. Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. Conservation Genetics Resources, **4**: 359–361.

Endersby, N.M., McKechnie, S.W., Ridl, P.M., and Weeks, A.R. 2006. Microsatellites reveal a lack of structure in Australian populations of the diamondback moth, *Plutella xylostella* (L.). Molecular Ecology, **15**: 107–118.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology, **14**: 2611–2620.

Fitt, G.P. 1989. The ecology of *Heliothis* species in relation to agroecosystems. Annual Review of Entomology, **34**: 17–53.

Forrester, N.W., Cahill, M., Bird, L.J., and Layland, J.K. 1993. Management of pyrethroid and endosulfan resistance in *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Australia. Bulletin of Entomological Research: Supplement Series, **1**: 1–132.

Grasela, J.J. and McIntosh, A.H. 2005. Cross-species investigation of *Helicoverpa armigera* microsatellites as potential markers for other related species in the *Helicoverpa-Heliothis* complex. Journal of Insect Science, **5**: 1–13.

Hardwick, D.F. 1965. The corn earworm complex. Memoirs of the Entomological Society of Canada, **97**: 5–247.

Iseli, C., Jongeneel, C.V., and Bucher, P. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proceedings of the International Conference on Intelligent Systems for Molecular Biology, **1999**: 138–148.

Ji, Y.J., Hewitt, G., Kang, L., and Li, D.M. 2003. Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. Molecular Ecology Notes, **3**: 102–104.

Ji, Y.J., Wu, Y.C., and Zhang, D.X. 2005. Novel polymorphic microsatellite markers developed in the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae). Insect Science, **12**: 331–334.

Katti, M.V., Ranjekar, P.K., and Gupta, V.S. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Molecular Biology and Evolution, **18**: 1161–1167.

Kranthi, K.R., Armes, N.J., Rao, N.G., Raj, S., and Sundaramurthy, V.T. 1997. Seasonal dynamics of metabolic mechanisms mediating pyrethroid resistance in *Helicoverpa armigera* in central India. Pesticide Science, **50**: 91–98.

Manjunath, T., Phalak, V., and Subramanian, S. 1970. First records of egg parasites of *Heliothis armigera* (Hubn.) (Lep.: Noctuidae) in India. Technical Bulletin of the Commonwealth Institute of Biological Control, **13**: 111–115.

McKenchnie, S.W., Spackman, M.E., and Naughton, N.E. 1993. Assessing budworm population structure in Australia using the AT-rich region of mitochondrial DNA. Beltwide Cotton Insect Research and Control Conference, New Orleans, Louisiana, United States of America.

Meglecz, E., Petenian, F., and Danchin, E. 2004. High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. Molecular Ecology, **13**: 1693–1700.

Nagarkatti, S. and Prakash, S. 1974. Rearing *Heliothis armigera* (Hubn.) on an artificial diet. Technical Bulletin Commonwealth Institute of Biological Control, **17**: 169–173.

Nève, G. and Meglécz, E. 2000. Microsatellite frequencies in different taxa. Trends in Ecology and Evolution, **15**: 376–377.

Pawar, C., Bhatnagar, V., and Jadhav, D. 1986. *Heliothis* species and their natural enemies, with their potential for biological control. Proceedings: Animal Sciences, **95**: 695–703.

Perera, O. and Blanco, C. 2011. Microsatellite variation in *Helicoverpa zea* (Boddie) populations in the southern United States. Southwestern Entomologist, **36**: 271–286.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**: 945–959.

Reed, W. and Pawar, C. 1982. *Heliothis*: a global problem. *In* Proceedings of international workshop on *Heliothis* management problem. *Edited by* W. Reed and V. Kumble. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India. Pp. 9–14.

© 2015 Entomological Society of Canada

Rohlf, F.J. 2000. NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software, Setauket, New York, United States of America.

Rousset, F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources, **8**: 103–106.

Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology, **132**: 365–386.

Scott, K., Lange, C., Scott, L., and Graham, G. 2004. Isolation and characterization of microsatellite loci from *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae). Molecular Ecology Notes, **4**: 204–205.

Scott, K., Lawrence, N., Lange, C.L., Scott, L.J., Wilkinson, K.S., Merritt, M.A., *et al*. 2005. Assessing moth migration and population strcturing in *Helicoverpa armigera* (Lepdoptera: Noctuidae) at the regional scale: example from the Darling Downs, Australia. Journal of Economic Entomology, **98**: 2210–2219.

Scott, K. Wilkinson, K., and Merritt, M. 2003. Genetic shifts in *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae) over a year in the Dawson/Callide Valleys. Crop and Pasture Science, **54**: 739–744.

Scott, L.J., Lawrence, N., and Lange, C.L. 2006. Population dynamics and gene flow of *Helicoverpa armigera* (Lepidoptera: Noctuidae) on cotton and grain crops in the Murrumbidgee Valley, Australia. Journal of Economic Entomology, **99**: 155–163.

Selkoe, K.A. and Toonen, R.J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters, **9**: 615–629.

Slatkin, M. 1985. Gene flow in natural populations. Annual Review of Ecology and Systematics, **16**: 393–430.

Subramanian, S. and Mohankumar, S. 2006. Genetic variability of the bollworm, *Helicoverpa armigera*, occurring on different host plants. Journal of Insect Science, **6**: 1–28.

Tamhankar, A., Rajendran, T., and Hariprasad Rao, N. 2003. Variability in response of *Helicoverpa armigera* males from different locations in India to varying blends of female sex pheromone suggests male sex pheromone response polymorphism. Current Science, **84**: 448–450.

Tan, S., Chen, X., Zhang, A., and Li, D. 2001. Isolation and characterization of DNA microsatellite from cotton bollworm (*Helicoverpa armigera*, Hübner). Molecular Ecology Notes, **1**: 243–244.

Tang, Z., Gong, K., and You, Z. 1988. Present status and countermeasures of insecticide resistance in agricultural pests in China. Pesticide Science, **23**: 189–198.

Tautz, D. and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research, **12**: 4127–4138.

Tautz, D., Trick, M., and Dover, G.A. 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature, **322**: 652–656.

Templeton, A.R., Shaw, K., Routman, E., and Davis, S.K. 1990. The genetic consequences of habitat fragmentation. Annals of the Missouri Botanical Garden, **77**: 13–27.

Vassal, J., Brevault, T., Achaleke, J., and Menozzi, P. 2008. Genetic structure of the polyphagous pest *Helicoverpa armigera* (Lepidoptera: Noctuidae) across the Sub-Saharan cotton belt. Communications in Agricultural and Applied Biological Sciences, **73**: 433–437.

Wasmuth, J.D. and Blaxter, M.L. 2004. prot4EST: translating expressed sequence tags from neglected genomes. BMC Bioinformatics, 5: 187. Available from http://www.biomedcentral.com/1471-2105/5/187 [accessed 18 May 2015].

Weeks, A.R., Endersby, N.M., Lange, C.L., Lowe, A., Zalucki, M.P., and Hoffmann, A.A. 2010. Genetic variation among *Helicoverpa armigera* populations as assessed by microsatellites: a cautionary tale about accurate allele scoring. Bulletin of Entomological Research, **100**: 445–450. doi: 10.1017/S0007485309990460.

Wright, S. 1978. Evolution and the genetics of population, variability within and among natural populations. The University of Chicago Press, Chicago, Illinois, United States of America.

Yajie, J. and Dexing, Z. 2003. Characteristics of microsatellite DNA in lepidopteran genomes and implications for their isolation. Acta Zoologica Sinica, **50**: 608–614.

Yeh, F.C. and Boyle, T. 1997. POPGENE, Version 1.21: software Microsoft window-based freeware for population genetic analysis. University of Alberta, Edmonton, Alberta, Canada.

Zhang, D.X. 2004. Lepidopteran microsatellite DNA: redundant but promising. Trends in Ecology and Evolution, **19**: 507–509.

Zhang, D.X. and Hewitt, G.M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology, **12**: 563–584.

Zhou, X., Faktor, O., Applebaum, S.W., and Coll, M. 2000. Population structure of the pestiferous moth *Helicoverpa armigera* in the eastern Mediterranean using RAPD analysis. Heredity, **85**: 251–256.

© 2015 Entomological Society of Canada