Conservation Genet Resour (2013) 5:541–544
DOI 10.1007/s12686-012-9847-7

TECHNICAL NOTE

Characterization of EST-based SSR loci in the spruce budworm, *Choristoneura fumiferana* (Lepidoptera: Tortricidae)

B. M. T. Brunet · D. Doucet · B. R. Sturtevant · F. A. H. Sperling

Received: 22 November 2012 / Accepted: 19 December 2012 / Published online: 11 January 2013
© Her Majesty the Queen in Right of Canada 2013

Abstract After identifying 114 microsatellite loci from *Choristoneura fumiferana* expressed sequence tags, 87 loci were assayed in a panel of 11 wild-caught individuals, giving 29 polymorphic loci. Further analysis of 20 of these loci on 31 individuals collected from a single population in northern Minnesota identified 14 in Hardy–Weinberg equilibrium.

Keywords Spruce budworm · Expressed sequence tag · Microsatellite · Simple sequence repeat · *Choristoneura fumiferana* · Tortricidae

Expressed sequence tags (ESTs) are valuable resources for rapid discovery of genetic variation, including microsatellites (simple sequence repeats, SSRs). Despite their conserved nature relative to genomic SSRs and lower levels of polymorphism (Kim et al. 2008), EST-SSRs make good candidates for use in population genetic studies due to a lower incidence of null alleles and ease of use. This is particularly beneficial for SSR development in Lepidoptera, insects known to have high levels of transposable elements that complicate marker development (Sinama et al. 2011; van't Hof et al. 2007; Zhang 2004).

The spruce budworm [*Choristoneura fumiferana* (Clemens)] is the most ecologically important defoliator of North American coniferous forests (Volney and Fleming 2007). Although the spruce budworm is not itself an endangered species, it is nonetheless of great conservation interest because of its role in determining structure within North America's boreal forest ecozone. Its geographical distribution covers the boreal forest across Canada ranging from the Atlantic coast to the Yukon and Alaska, an area of ~1 billion hectares of forested land (Volney and Fleming 2000). As one of the world's largest resources of sequestered carbon (and therefore a significant contributor to the global carbon cycle), the boreal forest is of global conservation importance. Ironically however, the boreal forest is increasingly threatened by industry (e.g. oil exploration and extraction, timber harvesting) and natural disturbances (e.g. climate change, wildfire, insect outbreak). Impacts caused by the spruce budworm alone have been known to account for more than 30 % of the total forested area depleted annually in Canada (Volney and Fleming 2000).

Although, dendrochronological reconstruction of spruce budworm population cycles and forest structure has been documented for the last 400 years (Boulanger et al. 2012), surprisingly little is known about the synchronization of population cycles and dispersal patterns of the spruce budworm throughout its range (Anderson and Sturtevant 2011). This gap is largely due to a paucity of genetic markers from which fine-scaled surveys of population structure can be attained. Here, we characterize SSRs derived from EST libraries of the spruce budworm that complement an existing SSR resource developed by

Electronic supplementary material The online version of this article (doi:10.1007/s12686-012-9847-7) contains supplementary material, which is available to authorized users.

B. M. T. Brunet (✉) · F. A. H. Sperling
Department of Biological Sciences, University of Alberta, 11455 Saskatchewan Drive, Edmonton, AB T6G 2E9, Canada
e-mail: bbrunet@ualberta.ca

D. Doucet
Natural Resources Canada, Great Lakes Forestry Centre, 1219 Queen St. East, Sault Ste. Marie, ON P6A 2E5, Canada

B. R. Sturtevant
USDA Forest Service, Northcentral Research Station, 5985 Highway K, Rhinelander, WI 54501, USA

Table 1 Summary data for 20 EST-SSRs tested on 31 individuals of *Choristoneura fumiferana* from the Border Lakes landscape of northern Minnesota

Locus	Label	Repeat Motif	Ta (°C)	$MgCl_2$ (mM)	n	N_a	Range	H_o (H_e)	HWE
Fum107	VIC	$(TAA)_4$	57	2.5	31	2	133–136	0.1612 (0.2522)	0.0929
Fum120	NED	$(CTG)_6$	55	1.5	31	4	*134–148*	0.5161 (0.5684)	0.7483
Fum122	6FAM	$(TCA)_4$	59	2.5	31	3	108–114	0.1935 (0.1803)	1.0000
Fum128	NED	$(TGTT)_4$	57	2.5	31	2	124–140	0.0322	–
Fum129	NED	$(ATAA)_4$	59	2.5	30	10	*139–160*	0.2000 (0.7316)	*0**
Fum130	PET	$(TACC)_4$	55	1.5	31	2	*135–145*	0 (0.0634)	0.0168*
Fum133[a]	NED	$(GAG)_4$	53	3.5	31	2	*101–103*	0.0967 (0.0936)	1.0000
Fum137	PET	$(GCA)_6$	59	2.5	31	2	106–109	0 (0.0634)	0.0166*
Fum145	PET	$(ACT)_6$	59	2.5	31	5	*140–153*	0.2580 (0.4077)	0.0134*
Fum147	VIC	$(AAT)_4$	53	3.5	31	4	*102–109*	0.8064 (0.6409)	0.0992
Fum149	6FAM	$(TA)_7$	55	1.5	29	4	138–144	0.0689 (0.1336)	0.0359*
Fum157	VIC	$(GA)_7$	55	1.5	31	3	*147–150*	0.1612 (0.2363)	0.0097
Fum160	VIC	$(CGT)_5$	57	3.5	31	3	133–139	0.3548 (0.3484)	1.0000
Fum164	PET	$(ATA)_5$	57	3.5	31	5	*120–127*	0.1612 (0.5949)	*0**
Fum169	NED	$(AAT)_6$	55	3.5	31	4	144–150	0.0645 (0.1253)	0.0332*
Fum170	6FAM	$(ATA)_5$	59	2.5	31	3	*143–150*	0.2903 (0.4965)	0.0086*
Fum171	PET	$(GAG)_4$	59	2.5	31	1	112	0	–
Fum177	NED	$(AG)_7$	59	3.5	24	10	133–155	0.3333 (0.8617)	*0**
Fum185	6FAM	$(TCATA)_4$	59	3.5	31	5	130–140	0.5161 (0.6430)	*0.0004*
Fum187	6FAM	$(AGG)_5$	59	2.5	31	3	101–110	0.4516 (0.4394)	1.0000

Locus is the EST-SSR name with italics indicating loci that had inconsistent or low amplification success using a validation panel of 11 individuals, *Label* is the fluorophore used in PCR amplification (primer sequences are identified in Online Resource 1), *Repeat Motif* is the SSR sequence (in parentheses) and number of repeats (subscript), T_a and $MgCl_2$ are the optimal annealing temperature and magnesium chloride concentrations used during PCR, *n* is the number of individuals assayed, N_a and *Range* are the number of alleles detected and their respective size ranges (ranges are italicized if allele sizes are not multiples of the repeat length suggesting the presence of indels), $H_o(H_e)$ is the observed and expected heterozygosity (only a single value if no difference), and *HWE* is the *P* value for a test of deviation from Hardy–Weinberg equilibrium (italics indicate significant deviation after Bonferroni correction, $P < 0.0025$). * Loci with excess homozygosity. [a] Peaks 105 bases in length were also present in all individuals at this locus as would be expected based on information from Online Resource 1, indicating that the size range reported here may be due to unspecific amplification, but nonetheless the peaks in the size range reported here behaved in a manner similar to SSRs

Lumley et al. (2009) using traditional genomic isolation techniques.

Trace files for 41,230 ESTs of *C. fumiferana*, were base-called, assigned quality values, and trimmed of low-quality sequence using Phred v.0.071220.b (Ewing et al. 2008). Sequencing artifacts were removed from the quality filtered ESTs using Figaro v.1.05 (White et al. 2008) and SeqClean (http://compbio.dfci.harvard.edu/tgi/software/). Sequences were then assembled with TGICL (Pertea et al. 2002) by assembling ESTs in one direction. Approximately 1,893 unigenes and 1,406 singletons resulted from this processing.

Using a custom Perl script (Beldade et al. 2009) 150 EST-SSRs were discovered with at least three repeats of 2–5 bases in length. These were reduced to 114 unique loci after accounting for redundancy among loci using Micro-Family v.1.2 (Meglecz 2007). Primer3 (Rozen and Skaletsky 2000) was used to find primer sequences between 20 and 25 bases in length that yielded amplicons between 100 and 150 bases for each locus. Primers for 87 loci were designed using these criteria (Online Resource 1).

A validation panel of 11 *C. fumiferana* individuals (four females and seven males) collected throughout Alberta was used to test loci for amplification success. Microsatellites were amplified in 10 μL reactions containing 1X PCR buffer (10 mM Tris pH8.8, 0.1 % Triton X-100, 50 mM KCl, 0.16 mg/ml BSA), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 μM M13-tagged forward primer, 0.4 μm reverse primer, 0.4 μm fluorophore-tagged M13 primer, 0.1 U/μL Taq polymerase and 1 μL of template DNA (5 ng/μL). PCR cycling was as follows: 95 °C (3 min); 10 cycles of 94 °C (35 s), annealing temperature (T_a) plus 10 °C (−1 °C/cycle) (35 s) and 72 °C (45 s); 30 cycles of 94 °C (35 s), 53 °C (35 s) 72 °C (45 s); and 72 °C (20 min).

A T_a of 53 °C was used for M13-tagged primers (Schuelke 2000). Amplicons were co-loaded and diluted 1:30 in water, and 1 μL was added to 8.7 μL of Hi-Dye formamide (ABI) and 0.3 μL of GeneScan 500-LIZ size standard (ABI) and analyzed using an ABI 3730 sequencer. GeneMapper v.4.0 (ABI) was used to visualize and score genotypes. Genetic diversity statistics were obtained using the Excel Microsatellite Toolkit (Park 2001) (Online Resource 1). Tests for Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were performed using GenePop v.1.2 web server (Raymond and Rousset 1995; Rousset 2008).

Of the 87 loci tested, 59 produced detectable peaks in at least six individuals. Monomorphism was observed in 23 loci, and seven loci had inconsistent peak patterns that could not be reliably genotyped. Of the remaining 29 polymorphic loci, five had low allele frequencies and could not be tested for HWE, and one locus deviated from HWE (Online Resource 1). Evidence for null alleles was found for six loci using MicroChecker v.2.2.3 (van Oosterhout et al. 2004) (Online Resource 1). Significant linkage disequilibrium was not detected. The number of alleles per locus ranged from 2 to 8 with an average of 3.52. Observed and expected heterozygosity ranged from 0–0.91 (averaging 0.336) and 0.09–0.9 (averaging 0.438), respectively.

Since M13-tagged forward primers are known to complicate PCR amplification (de Arruda et al. 2010) we also obtained 5′ fluorescently labeled primers for 20 loci (including all that seemed potentially useful in the test panel) and assayed these in 31 individuals from northern Minnesota. Optimization of T_a and $MgCl_2$ concentrations was performed for each locus to increase amplification success (Table 1). Of these 20 loci, two had low allele frequencies and four had allele frequencies that deviated from HWE (Table 1). The remaining 14 loci satisfied HWE (Bonferroni corrected $P < 0.0025$). Significant LD was detected between two loci, Fum145 and Fum160 ($P = 0.0006$). Observed and expected heterozygosity ranged from 0–0.806 (averaging 0.233) and 0–0.862 (averaging 0.346), respectively. In comparison with the eight SSR loci previously developed by Lumley et al. (2009), our reliance on an EST resource has nearly tripled the number of useful SSRs but required only a fraction of the time needed using traditional cloning techniques.

Acknowledgments We thank P. Stothard for assistance in post-processing and analysis of ESTs, and P. Beldade and A. Long for the use of their SSR-mining Perl script. This study was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to F. Sperling, a Genomics Research and Development Initiative Grant (NRCAN) to D. Doucet, and a United States Department of Agriculture, Cooperative State Research Education and Extension Service, National Research Initiative Grant (WISR-2005-02815) to Brian Sturtevant.

References

Anderson DP, Sturtevant BR (2011) Pattern analysis of eastern spruce budworm *Choristoneura fumiferana* dispersal. Ecography 34:488–497

de Arruda MP, Goncalves EC, Schneider MPC, da Costa da Silva AL, Morielle-Versute E (2010) An alternative genotyping method using dye labeled universal primer to reduce unspecific amplifications. Mol Biol Rep 37:2031–2036

Beldade P, van't Hoff A, Jeronimo MA, Long AD (2009) Microsatellite markers associated with genes expressed in developing wings of *Bicyclus anynana* butterflies. Mol Ecol Resour 9: 1487–1492

Boulanger Y, Arsenault D, Morin H, Jardon Y, Bertrand P, Dagneau C (2012) Dendrochronological reconstruction of spruce budworm (*Choristoneura fumiferana*) outbreaks in southern Quebec for the last 400 years. Can J Forest Res 42:1264–1276

Ewing B, Hillier L, Wendl M, Green P (2008) Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Res 8:175–185

Kim KS, Ratcliffe ST, French BW, Liu L, Sappington TW (2008) Utility of EST-derived SSRs as population genetic markers in a beetle. J Hered 99:112–124

Lumley LM, Davis CS, Sperling FAH (2009) Isolation and characterization of eight microsatellite loci in the spruce budworm species *Choristoneura fumiferana* and *Choristoneura occidentalis*, and cross-species amplification in related tortricid moths. Conserv Genet Resour 1:501–504

Meglecz E (2007) MicroFamily: a computer program for detecting flanking region similarities among different microsatellite loci. Mol Ecol Notes 7:18–20

Park SDE (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. Ph.D. thesis, University of Dublin

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2002) TIGR Gene Indices Clustering Tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rousset F (2008) GENEPOP'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resour 8:103–106

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234

Sinama M, Dubut V, Costedoat C, Gilles A, Junker M, Malausa T, Martin J-F, Neve G, Pech N, Schmitt T, Zimmermann M, Meglecz E (2011) Challenges of microsatellite development in Lepidoptera: *Euphydryas aurinia* (Nymphalidae) as a case study. Eur J Entomol 108:261–266

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4: 535–538

van't Hof AE, Brakefield PM, Saccheri IJ, Zwaan BJ (2007) Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. Heredity 98:320–328

Volney WJA, Fleming RA (2000) Climate change and impacts of boreal forest insects. Agr Ecosyst Environ 82:283–294

Volney WJA, Fleming RA (2007) Spruce budworm (*Choristoneura* spp.) biotype reactions to forest and climate characteristics. Glob Change Biol 13:1630–1643

White JR, Roberts M, Yorke JA, Pop M (2008) Figaro: a novel statistical method for vector sequence removal. Bioinformatics 15:462–467

Zhang DX (2004) Lepidopteran microsatellite DNA: redundant but promising. Trends Ecol Evol 19:507–509