9-2009

Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis

Brad S. Coates
Iowa State University, brad.coates@ars.usda.gov

Douglas V. Sumerford
Iowa State University

Nicholas J. Miller
Iowa State University

Kyung S. Kim
Iowa State University

Thomas W. Sappington
Iowa State University, tsapping@iastate.edu

Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis

Abstract

Microsatellite loci are standard genetic markers for population genetic analysis, whereas single nucleotide polymorphisms (SNPs) are more recent tools that require assessment of neutrality and appropriate use in population genetics. Twelve SNP markers were used to describe the genetic structure of *Diabrotica virgifera virgifera* (LeConte; Coleoptera: Chrysomelidae) in the United States of America and revealed a high mean observed heterozygosity (0.40 ± 0.059) and low global F_{ST} (0.029). Pairwise F_{ST} estimates ranged from 0.007 to 0.045, and all but 2 populations showed significant levels of genetic differentiation ($P \leq 0.008$). Population parameters and conclusions based on SNP markers were analogous to that obtained by use of microsatellite markers from the identical population samples. SNP-based F_{ST} estimates were 3-fold higher than corresponding estimates from microsatellites, wherein lower microsatellite F_{ST} estimates likely resulted from an overestimate of migration rates between subpopulations due to convergence of allele size (homoplasy). No significant difference was observed in the proportion of SNP or microsatellite markers loci that were nonneutral within populations. SNP markers provided estimates of population genetic parameters consistent with those from microsatellite data, and their low back mutation rates may result in reduced propensity for error in estimation of population parameters.

Keywords

neutrality, population structure, single nucleotide polymorphism

Disciplines

Entomology | Genetics | Population Biology

Comments

This article is from *Journal of Heredity* 100 (2009): 556, doi:10.1093/jhered/esp028

Authors

Brad S. Coates, Douglas V. Sumerford, Nicholas J. Miller, Kyung S. Kim, Thomas W. Sappington, Blair D. Siegfried, and Leslie C. Lewis

Journal of Heredity 2009:100(5):556–564
doi:10.1093/jhered/esp028
Advance Access publication June 12, 2009

Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis

Brad S. Coates, Douglas V. Sumerford, Nicholas J. Miller, Kyung S. Kim, Thomas W. Sappington, Blair D. Siegfried, and Leslie C. Lewis

From the USDA-ARS, Corn Insects and Crop Genetics Research Unit, Genetics Laboratory, Iowa State University, Ames, IA 50011 (Coates, Sumerford, Miller, Kim, Sappington, and Lewis); the Department of Entomology, Iowa State University, Ames, IA 50011 (Sumerford, Sappington, and Lewis); and the Department of Entomology, University of Nebraska, Lincoln, NE 68588 (Siegfried). Kyung S. Kim is now at the College of Veterinary Medicine, Seoul National University, Sillim-dong San 56-1, Gwanak-gu, Seoul 151-742, South Korea.

Address correspondence to Brad S. Coates at the address above, or e-mail: brad.coates@usda.ars.gov.

Abstract

Microsatellite loci are standard genetic markers for population genetic analysis, whereas single nucleotide polymorphisms (SNPs) are more recent tools that require assessment of neutrality and appropriate use in population genetics. Twelve SNP markers were used to describe the genetic structure of *Diabrotica virgifera virgifera* (LeConte; Coleoptera: Chrysomelidae) in the United States of America and revealed a high mean observed heterozygosity (0.40 ± 0.059) and low global F_{ST} (0.029). Pairwise F_{ST} estimates ranged from 0.007 to 0.045, and all but 2 populations showed significant levels of genetic differentiation ($P \leq 0.008$). Population parameters and conclusions based on SNP markers were analogous to that obtained by use of microsatellite markers from the identical population samples. SNP-based F_{ST} estimates were 3-fold higher than corresponding estimates from microsatellites, wherein lower microsatellite F_{ST} estimates likely resulted from an overestimate of migration rates between subpopulations due to convergence of allele size (homoplasy). No significant difference was observed in the proportion of SNP or microsatellite markers loci that were nonneutral within populations. SNP markers provided estimates of population genetic parameters consistent with those from microsatellite data, and their low back mutation rates may result in reduced propensity for error in estimation of population parameters.

Key words: *neutrality, population structure, single nucleotide polymorphism*

Single nucleotide polymorphisms (SNPs) are single base substitutions found at a single genomic locus. Individual biallelic SNPs have lower allele diversity and provide less statistical power to discriminate unique genotypes compared with microsatellite loci but have a more dense and uniform distribution within genomes which make them useful for population and mapping studies (Xing et al. 2005). Due to low mutation rate, forward and backward mutation is generally ignored such that SNP-based lineages are more stable over microevolutionary time (Morin et al. 2004). SNPs that are highly frequent within or between species may be neutral or subject to balancing selection, whereas low-frequency SNPs may be deleterious and undergo negative selection (Zhu et al. 2004). Additionally, synonymous SNPs (nonamino acid changing) mutations are more likely to be selectively neutral than SNPs that lead to an amino acid change (Zhao et al. 2003).

SNPs can provide tools for population and genetic mapping studies (Glaubitz et al. 2003; Morin et al. 2004; Seddon et al. 2005). Benefits of SNP-based markers include universal applicability of nucleotide sequence calls between laboratories and flexibility in detection protocols. In contrast, the migration of microsatellite fragments during electrophoresis and comparison to known standards are required for size-based allele determination, but migration rate can differ between electrophoresis methods that may cause genotype comparison between collaborating laboratories to be difficult (Kim et al. 2008a). SNP genotypes

based on single nucleotide changes are universally comparable and do not require standardization across detection platforms. Studies that use SNP markers can be replicated and performed in parallel among different laboratories and the results directly compared without adjustment of raw data. Despite the abundance of SNPs within expressed sequence tag (EST) collections, their application in population genetics has been limited largely due to the question of neutrality of the mutations that are used in genotyping (Zhang and Li 2005; Moen et al. 2008).

The western corn rootworm, *Diabrotica virgifera virgifera*, is a major pest of cultivated maize in the United States of America and causes greater than $1 billion in crop damage each year (Levine et al. 2002; Sappington et al. 2006). In the early 1990s, *D. v. virgifera* was reported in Eastern Europe (Baca 1994; Sivcev et al. 1994), wherein spread is aided by continued trans-Atlantic and intra-European introductions (Miller et al. 2005; Ciosi et al. 2008). The species has one generation per growing season and individuals overwinter in the soil as diapausing eggs (Chiang 1973). Larvae emerge in the spring and feed on maize roots, which lead to decreased yield due to a reduction in nutrient uptake and plant instability (Spike and Tellefson 1991). *Diabrotica virgifera virgifera* has evolved resistance to insecticide treatments (Meinke et al. 1998), and laboratory populations respond rapidly to selection against transgenic maize that expressed the *Bacillus thuringiensis* CryBb1 toxin and have the potential to evolve field resistance to this management tool (Lefko et al. 2008). A behavioral adaptation where females oviposit into soybean fields (Sammons et al. 1997) has resulted in the presence of larva in the soil of first year maize planted the subsequent growing season and circumvents crop rotation practices for insect control. Range expansion of rotation resistance has raised concerns among producers, both in North American and European countries where rotation is often a mandatory component of the response to new outbreaks of *D. v. virgifera* (European Commission Decision 2003/733/EC).

Diabrotica virgifera virgifera microsatellite loci have been isolated from anonymous genomic DNA (Kim and Sappington 2005b; Kim et al. 2008a) and from EST library sequences (Kim et al. 2008b). Substantial genetic variation was detected in North American populations using anonymous microsatellite loci but showed no population structure (Kim and Sappington 2005a; Miller et al. 2006). In contrast, isolated European sites show significant differentiation due to founder effects (Miller et al. 2005; Ciosi et al. 2008). Additional molecular genetic markers are needed to investigate *D. v. virgifera* genetics and inheritance of traits of economic importance. Herein we report 20 *D. v. virgifera* SNP loci that were tested for Mendelian inheritance and selective neutrality within 4 North American sample sites. SNP markers were compared with prior microsatellite data to determine performance in detecting population structure. The neutrality of SNP marker loci was tested and compared with anonymous microsatellite markers. *Diabrotica virgifera virgifera* midgut and head ESTs are available and were assembled into hundreds of unique contigs. Our results suggest that putative SNPs identified among redundant EST reads can be developed into single locus markers for population analyses and linkage mapping for this species.

Materials and Methods

Sample Collection and DNA Extraction

Single pair matings were established among *D. v. virgifera* from a nondiapausing (ND) colony at the North Central Agricultural Research Laboratory (NCARL) in Brooking, SD. Three pedigrees designated 3ND, 5ND, and 6ND were selected for analysis because >1 SNP allele was segregating. Each pedigree consisted of the 2 parents and 22 F_2 offspring. Population genetic analyses used SNP markers that were to be directly compared with prior microsatellite marker-based analyses. To accomplish this, SNP genotyping was conducted on the same *D. v. virgifera* samples previously collected and used for microsatellite genotyping by Kim and Sappington (2005a) and Kim et al. (2008a; Ankeny, IA, Champaign, IL, Cobleskill, NY, and New Deal, TX). All DNAs were extracted and stored as described by Kim and Sappington (2005a).

SNP Validation

SNPs were identified from contiguous sequences (contigs; read depth ≥2) assembled from *D. v. virgifera* third instar midgut ESTs (Siegfried et al. 2005; accession numbers CN497248–CN498776). SNPs were detected visually from ClustalW alignments or using PolyPhred (Stephens et al. 2006) that examined aligned sequence traces assembled by Consed (Gordon et al. 1998). Polymerase chain reaction (PCR) primer pairs were designed for 54 contigs containing putative SNPs within restriction enzyme recognition sites and synthesized by Integrated DNA Technologies (Coralville, IA). Loci were PCR amplified in 10 μl reactions with 2.5 mM $MgCl_2$, 150 μM dNTPs, 10 ng DNA, 2.5 pmol each primer, 2 μl 5× PCR buffer, and 0.3125U Go*Taq* DNA polymerase (Promega, Madison, WI). The thermocycler program had a 96 °C denaturation for 2 min, followed by a touchdown (TD) phase of 65 °C at −2 °C/cycle for 7 cycles. Subsequent cycles were at 96 °C for 20 s, 52 °C (TD2) or 50 °C (TD5) for 30 s, and 72 °C for 60 s for 30 (TD2) or 35 (TD5) cycles. PCR products were digested after addition of 0.5 units of restriction enzyme (Table 1), 2 μl 10× reaction buffer, and 8 μl nuclease-free water and incubated at 37 °C (60 °C for *Taq*I and *Tfi*I) for 14 h. Products were separated on 1 mm × 20 cm 6% acrylamide:bisacrylamide (19:1) gels at 140 V for 4.5 h and visualized under UV after staining in 0.5 μg/ml ethidium bromide. Twenty SNP markers were tested for Mendelian inheritance in *D. v. virgifera* F_1 pedigrees 3ND, 5ND, and 6ND. DNA from parents and 22 progeny per pedigree were PCR amplified and SNPs detected by restriction enzyme digestion as described previously. Mendelian inheritance of SNP markers among progeny was assessed by a chi-square (χ^2) goodness-of-fit test using SAS (v. 9.1.3) based on expectations from parental genotypes.

Table 1. *Diabrotica virgifera virgifera* EST contigs, SNP marker loci, and pedigree results

Locus/Contig	Oligonucleotide primers	PCR	RE	Pedigrees Parent: GT	P^a
1125	F: GGAACAGGGTGTTCTTGCAT R: TGGCGAAATCAAGTGGGTAT	TD2	*Afl*II	♂ A379/G379 ♀ A379/G379	0.6698
1154	F: TGCAAACCWTCCTGCTCATGG R: GCTAAGTAAGAAACAAGCGCAACA	TD5	*Msp*I	♂ G179/C179 ♀ C179/ C179	0.2008
1185	F: GGTTCAGCTCATTATTGTACAGCAT R: GCTTTGCGATCTTCATTCCA	TD2	*Nla*III	♂ G207/A207 ♀ A207/A207	0.3937
1200	F: CAAAGAAGATGCCGCCAATAA R: TTATCTACCATAAAGGCACAGTCTGA	TD2	*Sau*3AI	♂ T507/C507 ♀ T507/C507	1.0000
1203	F: CGTTGCGCCCAATACAAAAA R: CACCGTCAGGTAAGATTTCAGG	TD2	*Cfo*I	♂ T564/C564 ♀ T564/T564	1.0000
1207	F: ATGAAGTGTCTAGTTCTTCTTTCCCT F: TTGGTTGTTGACACCTTGGAG	TD5	*Scr*FI	♂ A236/G236 ♀ A236/G236	0.5254
1214	F: GGAGACTCTGACCTCCAACTTGA R: TAATGTCCCTTGGCCCAGTT	TD2	*Msp*I	♂ T283/C283 ♀ C283/C283	0.0881
1224	F: GGCCGTATTGGYCGTCTTGT R: TGTGGTGAACACTCCGGTAGA	TD2	*Tfi*I	♂ C194/C194 ♀ C194/T194	0.6698
1245	F: CAGGTCGCCATGGATATTCT R: GCAGCAGCACATGTAGCAAT	TD2	*Alu*I	♂ A578/A578 ♀ G578/G578	1.0000
1276	F: TGGTAACTGGTATTAAGGTAGTAGACC R: TCCAGATTCAATCATTTCATGGTA	TD5	*Hae*III	♂ C569/C569 ♀ G569/G569	1.0000
1280	F: TTACCAAACAGGCTGATGTTGA R: TTCAATAGTGGTTTAGGGCGATT	TD5	*Hin*fI	♂ C396/C396 ♀ C396/C396	NA
1283	F: AGAACTAGGCTACCGCCACA R: CAGCCAACGATTTTGTCAGA	TD2	*Ear*I	♂ T390/C390 ♀ C390/C390	0.0881
1300	F: GGATCGATGGGATTCTACCC R: CAAACTGTGGCTAAAGCCCTAAG	TD5	*Sau*3AI	♂ A210/T210 ♀ A210/T210	0.0956
1312	F: TTTAACAGGAAAATTCGTCAATATGT R: GACCACACCATGGAGCATAAAA	TD2	*Taq*I	♂ A217/null ♀ G217/G217	0.2008
1327	F: ATTTCGCACAATTGGGAAGC R: TGTCGCCCTTTTTCGTTTG	TD2	*Bfa*I	♂ A550/A550 ♀ A550/G550	0.6698
1341	F: GGTGAAATAACCAAGCAAATGAA R: TTGCCCAATGAACCGAATAA	TD2	*Mse*I	♂ C115/C115 ♀ T115/T115	NA
1345	F: GTGGAAGAAACAAGCGCTGA R: GCATTTTCAGGCCAATCACA	TD2	*Nla*III	♂ A385/C385 ♀ A385/A385	0.6698
1376	F: CAGACAGTGTGGACGTTGAGA R: CACAGGTTTGCCCCTACATT	TD2	*Aci*I	♂ A377/G377 ♀ A377/G377	1.0000
1397	F: CATCGGCAAAGTTCTCAACA R: CTTCCAAACAGCCAGATGGT	TD2	*Mnl*I	♂ T220/C220 ♀ T220/C220	NA
1411	F: CTGAAGGAGCGAAAGGTGAC R: TGTTCCAGTCCTGATGCGTA	TD2	*Cfo*I	♂ A044/G044 ♀ A044/G044	0.0053

NA, not applied.

P values are associated with χ^2 goodness-of-fit to Mendelian expectation among F_1 progeny observed from the given parent genotype (GT).

[a] *P* value testing significance of χ^2 values between observed and expected genotypic frequencies.

SNP-Based Population Differentiation

Samples collected from Ankeny, Champaign, Cobleskill, and New Deal (previously screened with microsatellite markers by Kim and Sappington [2005a] and Kim et al. [2008a]) were SNP genotypes as described previously. The observed heterozygosity (H_O) and expected heterozygosity (H_E) and significance of deviations from Hardy–Weinberg equilibrium (HWE) were tested using Markov chain exact tests in the Arlequin software package (v. 3.1; Excoffier et al. 2005). Pairwise F_{ST} estimates of locus-by-locus F_{ST}, F_{IS}, and F_{IT} estimates and exact tests for sample differentiation also were performed with Arlequin (v. 3.1; Excoffier et al. 2005; see references therein for all tests). An isolation-by-distance model was tested by the relationship between $F_{ST}/(1 - F_{ST})$ employing SNP-based F_{ST} estimates and geographic distance between sample sites ($\log_{10}$ km) using the IBD Web service v.3.15 (http://ibdws.sdsu.edu/~ibdws/; Jensen et al. 2005), and significance was estimated by 1000 jackknife permutation steps. A principal component (PC) analysis was applied to a covariance matrix of SNP allele frequencies across all loci by using the program XLSTAT (Addinsolt, NY). The geometric relationships among populations were visualized on a scattergram of factor scores along PC axes that accounted for the most variation.

STRUCTURE 2.0 (Prichard et al. 2000) was used to estimate the number of genetically distinct populations (K). For each category of marker (SNPs, anonymous, and EST microsatellites), 10 STRUCTURE runs were carried out for each value of K from 1 to 4. Each run consisted of 5×10^7 iterations, preceded by a burn-in of 10^5 iterations that used an admixture model of individual ancestry. The median value of the estimated log probability of the data, conditional on K, (ln $\Pr(X|K)$), was taken for each set of 10 runs and used to compute the posterior probability of K, $\Pr(K|X)$, assuming a uniform prior distribution for K.

SNP and Microsatellite Marker Comparisons

HWE, H_E and H_O, and global and pairwise F_{ST} estimates, as well as isolation by distance, PC analyses, and STRUCTURE analyses were performed for EST microsatellite and anonymous microsatellite loci as described above. Neutrality for each SNP, EST microsatellite, and anonymous microsatellite marker was tested using the Ewens–Watterson–Slatkin exact test and implemented by the use of the Arlequin software package (v. 3.1; Excoffier et al. 2005; see references therein for all tests). The proportions of loci that were neutral for each marker type were evaluated and the significance determined by a Kruskal–Wallis test (Kruskal and Wallis 1962) using the program STATISTIX 8 (Analytical Software. Tallahassee, FL).

Table 2. *Diabrotica virgifera virgifera* EST-based SNP marker loci

	Allele		Ames, IA		
Locus/Contig	Major	Minor	H_O	H_E	P value
1125	G379	A379	0.41	0.39	0.8200
1154	G179	C179	0.41	0.32	0.1220
1185	G207	A207	0.53	0.46	0.2819
1200	T507	C507	0.82	0.48	0.0001
1203	T564	C564	0.55	0.50	0.5366
1207	G236	A236	0.37	0.36	0.7809
1214	C283	T283	0.34	0.43	0.1884
1224	C194	T194	0.39	0.48	0.1992
1245	G579	A579	0.45	0.43	0.7521
1276	C569	G569	0.58	0.46	0.3437
1280	C396	T396	0.10	0.10	0.7034
1283	T390	C390	0.58	0.48	0.1800
1300	A210	T210	0.44	0.50	0.4533
1312	G217	A217	0.11	0.11	0.7199
1327	A550	G550	0.16	0.15	0.5665
1341	C115	T115	0.47	0.49	0.7385
1345	A385	C385	0.30	0.32	0.7585
1376	A377	G377	0.72	0.46	0.0012
1397	C220	T220	0.12	0.16	0.1608
1411	A044	G044	0.32	0.42	0.1179

Contig assignment of SNP marker loci ascribed to those defined by Siegfried et al. (2005). Each allele consists of nucleotide present and position with corresponding contigs. Estimates observed (H_O) and associated significance (P value) of deviation from the expected heterozygosity (H_E) from the Ames population are given.

Results

SNP Validation

A total of 54 primer pairs were designed to PCR amplify regions of *D. v. virgifera* midgut EST contigs (genes) that contained putative SNPs, of which 20 of 50 SNPs (37%) were validated for use in population or pedigree analysis. Of the loci that failed, 14 (26%) primer pairs did not generate a PCR product, 8 (15%) had products with $\geq$3 PCR coamplifying fragments, and 4 (7.5 %) showed poor or inconsistent PCR amplification. Furthermore, 4 putative SNPs (3.7% of total) did not exhibit polymorphism and were considered to be due to sequencing errors.

PCR product sizes for the 20 validated SNPs corresponded to those predicted from ESTs, except for those of markers 1312, 1327, and 1345, where introns likely were included in the amplicon. Two alleles were observed at all 20 validated SNP loci (Table 1). The F_1 pedigree data further indicated that 18 of the 20 validated SNPs were inherited according to Mendelian expectations based on parental genotypes (χ^2 $P > 0.05$; Table 1). Only SNP markers 1312 and 1411 showed significant deviations. The SNP assay for marker 1312 indicated homozygous parents and a Mendelian expectation of 100% heterozygous progeny (SNP genotype A217/G217). In contrast, putative progeny SNP genotypic proportions were approximately 50% A217/G217 and 50% G217/G217. A correction that assumed a null allele was present in the male parent (A217/null) and resultant progeny would show G217/null or A217/G217 genotypes resulted in no significant deviation from expected ratios ($P \geq 0.2008$; Table 1). An analogous correction for marker 1411 could not be made.

Significant deviation of observed heterozygosity (H_O) from expected (H_E) was not detected in the preliminary screen of the Ames, IA, population using χ^2 tests ($\alpha = 0.05$), except at loci 1200 and 1376 (Table 2). Because markers 1200 and 1376 were Mendelian in all pedigrees where initial parents showed polymorphism, these results suggest that null alleles may exist within populations even though they were not detected by pedigree analysis (Table 1). Conversely, markers 1312 and 1411 showed no significant deviation from HWE in wild populations but showed null alleles with F_1 pedigrees 3ND, 5ND, and 6ND. Population screening indicated that markers 1280, 1312, 1327, and 1397 harbored low-frequency SNPs (minor allele frequency $\leq$0.16; data not shown). These markers were less informative than the remaining SNP markers and were removed from the marker set. The SNP locus from marker 1411 was retained for population analysis due to adherence to HWE within populations and assuming that the frequency of null alleles was low but sampled by chance during formation of F_1 pedigrees 3ND, 5ND, and 6ND.

SNP-Based Population Differentiation

After marker testing by pedigree and preliminary Ames, IA, population screening, 13 SNP markers were chosen to genotype population samples from Ankeny, IA, Champaign,

Table 3. North American *Diabrotica virgifera virgifera* population statistics

	SNP marker loci											
	1125	1203	1207	1214	1224	1245	1276	1283	1300	1341	1345	1411
IA	0.262	0.477	0.233	0.318	0.378	0.341	0.364	0.453	0.463	0.375	0.170	0.330
IL	0.319	0.430	0.235	0.361	0.304	0.167	0.355	0.417	0.468	0.552	0.115	0.281
NY	0.404	0.438	0.215	0.337	0.189	0.413	0.198	0.500	0.415	0.143	0.200	0.372
TX	0.365	0.236	0.266	0.266	0.245	0.418	0.478	0.447	0.543	0.351	0.163	0.146
μ	0.338	0.395	0.237	0.321	0.279	0.335	0.349	0.454	0.472	0.355	0.362	0.282
σ	0.061	0.108	0.021	0.040	0.081	0.117	0.115	0.034	0.053	0.168	0.035	0.098
H_E	0.453	0.467	0.379	0.442	0.370	0.430	0.437	0.498	0.497	0.424	0.275	0.387
H_O	0.436	0.401	0.345	0.424	0.300	0.442	0.450	0.467	0.461	0.420	0.307	0.352
F_{ST}	0.006	0.053	−0.008	0.003	0.036	0.057	0.043	−0.005	0.002	0.135	0.003	0.029
F_{IS}	−0.005	0.031	−0.007	0.004	0.034	0.020	0.085	−0.006	0.008	0.053	−0.002	0.061
F_{IT}	0.011	0.022	−0.001	−0.001	0.002	0.038	0.046	−0.001	−0.007	0.086	0.005	−0.034

Minor SNP allele frequencies by locus across Ankeny, IA, Champaign, IL, Cobleskill, NY, and New Deal, TX, population samples. The overall North American population mean (μ) and variance (σ) as well as observed (H_O) and expected heterozygosity (H_E) are given for each locus. Lastly, locus-specific F_{ST}, F_{IS}, and F_{IT} estimates among populations are given.

IL, Cobleskill, NY, and New Deal, TX. A total of 26 alleles were scored from 13 biallelic loci among 190 individuals. No private alleles were detected. Minor SNP allele frequencies ranged from 0.500 (marker 1283) to 0.115 (marker 1345) across all populations and all loci (mean = 0.332 ± 0.114; Table 3). Inbreeding coefficient (F_{IS}) and F_{IT} estimates across all populations and all loci ranged from −0.007 to 0.085 and −0.034 to 0.086, respectively. The global estimates of F_{IS} and F_{IT} averaged across loci also were not significant ($P \geq 0.05$; Table 3).

All loci were in HWE except for marker 1185 in the Ankeny ($H_E = 0.304$ $H_O = 0.516$, $P = 0.0099$), Champaign ($H_E = 0.508$ $H_O = 0.240$, $P = 0.0004$), Cobleskill ($H_E = 0.447$ $H_O = 0.156$, $P = 0.0001$), and New Deal ($H_E = 0.529$ $H_O = 0.262$, $P = 0.0021$) populations. Although SNP marker 1185 performed as expected in pedigree analysis and the preliminary population screen, it showed a significant departure from HWE in all populations that were tested and thus not listed in Table 3 and was not included in any calculations of population statistics. A null allele may be present in natural populations that was not sampled within pedigrees 3ND, 5ND, and 6ND or was at a low frequency in the Ames, IA, test sample. Departure of SNP marker 1185 from HWE also may be attributed to the Wahlund effect, selection at or near the genomic locus, nonrandom mating, inbreeding, or genetic drift. Marker 1185 was hence omitted from all further population genetic calculations. Of the remaining 12 loci, only 2 instances of significant departure from HWE were observed at marker 1245 in Ankeny ($H_E = 0.483$ $H_O = 0.319$, $P = 0.03101$) and marker 1207 in Champaign ($H_E = 0.265$ $H_O = 0.403$, $P = 0.02167$). The remaining loci did not exhibit significant deviation between observed and expected heterozygosity levels (Table 3) or any departure from HWE.

The locus-by-locus F_{ST} estimates for the 12 SNP markers were low and ranged from −0.008 to 0.135 among subpopulations and showed a global estimate of 0.029 ± 0.040 (Table 3). Although pairwise F_{ST} estimates between subpopulations were significant for all comparisons except Ankeny versus Champaign, all values were low (≤0.0445; Table 4). All subpopulation comparisons remained significant when a Bonferroni-adjusted P value was used ($\alpha = 0.0083$). Regardless, the global estimate of $F_{ST} = 0.029$ indicated that very little genetic differentiation was accounted for by differences between subpopulations. Exact test of population differentiation gave a single significant P value between the Ankeny and Champaign samples ($P = 0.01940 \pm 0.0080$), but this was not significant after Bonferroni correction. Analysis of molecular variance results indicated that 96.07% of variance detected by SNP markers was accounted among individuals within population samples, and only 2.50% could be attributed to differences between population samples (results not shown).

SNP and Microsatellite Marker Comparisons

Mean factor scores for SNP and microsatellite markers were obtained independently from PC analysis. Component 1 accounted for 52%, 51%, and 50%, and component 2 accounted for 35%, 25%, and 30% of the total variance in the covariance matrix for SNP, EST microsatellite, and anonymous microsatellite data sets, respectively (Figure 1). PC analyses indicated that Ankeny and Champaign sample sites were the most similar using the SNP or either microsatellite data set. Relative divergence among subpopulations as a function of increased geographic distance using an isolation-by-distance model was tested for

Table 4. Pairwise F_{ST} estimates (below diagonal) and corresponding P values (above diagonal; significance set at a Bonferroni-adjusted P value $\alpha = 0.0083$)

	IA	IL	NY	TX
IA	—	0.1084	<0.0001	0.0049
IL	0.0067	—	<0.0001	<0.0001
NY	0.0290	0.0435	—	<0.0001
TX	0.0163	0.0285	0.0445	—

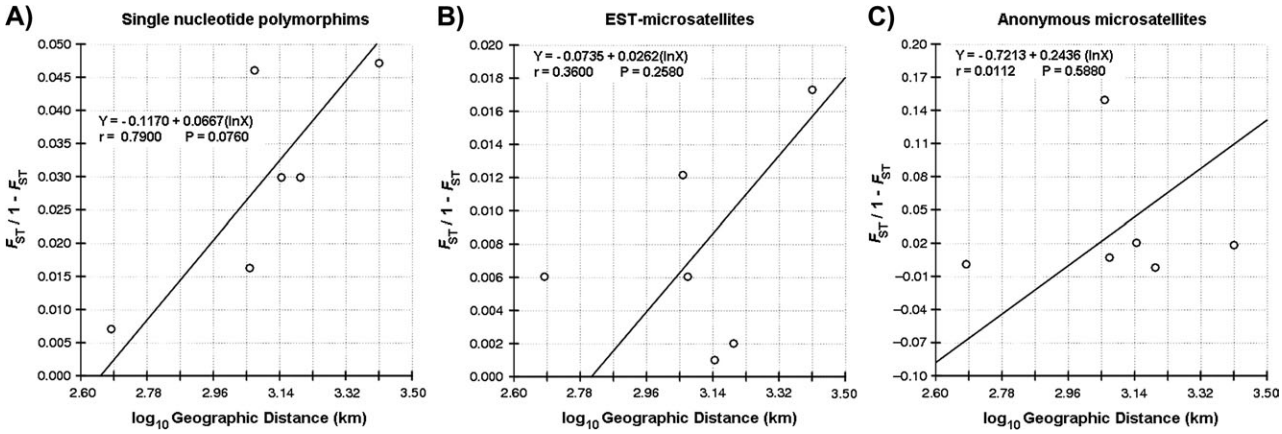

Figure 1. Isolation by distance assessed by plotting $F_{ST}/(1 - F_{ST})$ versus log-transformed geographic distance between North American *Diabrotica virgifera virgifera* sample sites for of SNP (**A**), EST microsatellite (**B**), and anonymous genomic microsatellite data (**C**).

SNP- and microsatellite-based $F_{ST}/(1 - F_{ST})$ estimates. The $F_{ST}/(1 - F_{ST})$ statistic among the 4 *D. v. virgifera* subpopulations tested indicated an analogous positive, but nonsignificant, correlation with $\log_{10}$(km) for both SNP and both microsatellite marker data sets (Figure 2).

Analyses using the STRUCTURE program did not reveal any evidence of population subdivision. The posterior probabilities for a value of $K > 1$ appeared negligible ($<10^{-10}$) for the SNP and anonymous microsatellite markers. A slightly different result was obtained for EST-derived microsatellites, the posterior probability was predominantly supportive of $K = 3$ ($\Pr(3|X) \approx 1$). However, the increase in ln $\Pr(X|K)$ between $K = 1$ and $K = 3$ was small relative to the decrease between $K = 3$ and $K = 4$ and for $K = 3$, the ancestry of each individual was divided evenly between the 3 putative populations. Both of these observations suggest that there was no genuine signal of population structure in the data from EST-derived microsatellites either.

One-third of SNP marker loci showed nonneutrality in 1 of the 4 natural populations tested in this study (Table 5). The proportion of SNP marker loci showing nonneutrality in at least one population was comparable to that estimated for microsatellite markers by the Ewens–Watterson–Slatkin exact tests (Kruskal–Wallis statistic = 0.9885; $P = 0.6100$).

Discussion

SNP Validation

EST collections often show redundancy among reads, where duplicate sequences are derived from independent sampling of cloned inserts transcribed from the same genomic locus. Orthologous sequences (derived from the same transcription locus) allow DNA sequence variation among alleles at a locus to be identified (Hayes et al. 2007). SNPs are identified by computational methods or by manual alignment of orthologous sequences within EST collections and are suitable for genetic marker development (Huntley et al. 2006). Although individual SNP loci have fewer alleles per locus than most microsatellites, SNPs have a higher density and more uniform distribution in genomes (Xing et al. 2005), are less expensive, produce lower error rates during

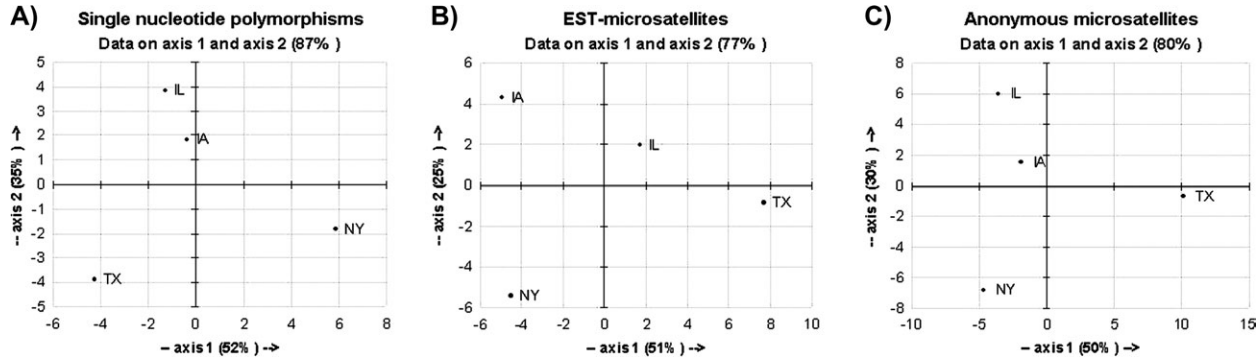

Figure 2. Plot of factor scores among 4 North American *Diabrotica virgifera virgifera* subpopulations generated from covariance matrices of SNP (**A**), EST microsatellite (**B**), and anonymous genomic microsatellite allele frequencies (**C**).

Table 5. Statistics for Ewens–Watterson–Slatkin (E–W–S) exact tests for each EST-based SNP marker locus in 4 North American *Diabrotica virgifera virgifera* populations

Locus	$F_{observed}/F_{expected}$	E–W–S
1125	0.720	0
1203	0.693	1
1207	0.826	0
1214	0.725	0
1224	0.778	0
1245	0.742	0
1276	0.726	1
1283	0.652	1
1300	0.648	1
1341	0.701	0
1345	0.922	0
1411	0.785	0

Number of populations that differ significantly from neutrality under each test indicated for each SNP marker locus ($P \leq 0.05$).

genotyping (Montgomery et al. 2005), and have universal comparability and transportability across detection platforms (Morin et al. 2004).

The current study suggests that EST collections from midgut tissues (Siegfried et al. 2005) contain readily available SNPs that may be developed into additional validated marker loci. Preliminary bioinformatics data indicate an additional 3964 putative SNPs are located within 3224 head tissues EST contigs (Ratcliffe ST and Liu L, unpublished data; GenBank dbEST accessions EW761110–777362), but none of these were evaluated in the current study. An overall 37% success rate of SNP marker development from a set of putative substitutions within 54 *D. v. virgifera* midgut EST contigs is typical for a nonmodel organism. Previously, Coates et al. (2008) successfully amplified 36 of 48 (75%) EST-based markers from genomic DNA by PCR and further validated the presence of SNPs at 28 of these 36 marker loci (77.8%; overall success rate of 58%). Of the 34 *D. v. virgifera* SNP markers that failed, a majority (14 or 41%) did not generate a PCR product. This highlights a common difficulty observed during the development of genomic marker loci from sequence data based on fully processed (intron spliced) complementary DNA (cDNA). In the absence of corresponding genomic sequence, the position and size of intervening intron regions are unknown. Oligonucleotide primers designed from cDNA sequence often result in no amplification by PCR due to disruption of primer annealing sites by intervening intron sequences, differential splicing, or presence of introns of unexpectedly large size (Cox et al. 1998). These factors are a prevalent source of PCR amplification failure in nonmodel organisms and likely were responsible for most failures in *D. v. virgifera* SNP marker development.

SNP-Based Population Differentiation

Validated SNPs from *D. v. virgifera* midgut EST contigs were informative for population genetic analyses providing results equivalent to those obtained from microsatellite markers. Similar to previous studies that used microsatellite markers (Kim and Sappington 2005a; Kim et al. 2008a), the *D. v. virgifera* populations sampled from North America in the current study showed little genetic structuring based on SNP marker loci. Low global and pairwise F_{ST} estimates from SNP genotype data indicated that subpopulations sampled from across North America are genetically similar, and no significant level of differentiation could be detected in 4 populations by use of 12 marker loci. Homogeneity of the population could be maintained by a high migration rate between widely spaced sample sites located in a largely open landscape, but this seems unlikely given the apparent flight capacity of adult *D. v. virgifera* and documented estimates of recent range expansions (for references, see Kim and Sappington 2005b). Range expansion that has occurred since the mid 1900s and associated effects on population dynamics also could explain genetic similarity among subpopulations. Insufficient time may have elapsed since the initial range expansion for subpopulation divergence to have arisen, such that population homogeneity and nondifferentiation of subpopulations prevails.

Comparison between SNP- and Microsatellite-Based Markers

The results from analyses were consistent across all marker types. Population genetic data from SNP and both microsatellite marker types produced low F_{ST} estimates indicative of little or no differentiation between subpopulations. Marker types also performed similarly in the detection of population differences within isolation by distance, PC, and STRUCTURE 2.0 analyses. Although all marker types indicated a trend toward a positive correlation between F_{ST} and geographic distance in isolation-by-distance analysis, the SNP marker data provided the best fit of data points to the regression line but still remained nonsignificant. Lack of isolation by distance based on the microsatellite-derived estimates of F_{ST} was expected given previous results (Kim and Sappington 2005b). However, increasing the number of populations for SNP-based marker estimates seems likely to reveal isolation-by-distance pattern consistent with their increased sensitivity in detecting pairwise differences in F_{ST}. The difference in sensitivity for both F_{ST} and isolation-by-distance estimates may reflect the different mutation processes that give rise to alleles in microsatellites and SNPs.

A high mutation rate among polymorphic microsatellites often can lead to underestimation of subpopulation divergence. The large number of alleles typical of microsatellite markers has the effect of reducing the maximum value that F_{ST} can attain, even under circumstances when there is no gene flow between populations (Hedrick 1999; Balloux et al. 2000; Hedrick 2005). In addition, microsatellite allele homoplasy that results from reversion mutation that can cause gene flow and homogeneity levels among subpopulations to be overestimated. *Diabrotica virgifera virgifera* has a moderate dispersal rate estimated at

10–30 km/year to 33.3 km/year (for references, see Kim et al. 2005b). Overall, the F_{ST} estimate for SNP-based markers across all loci in the 4 populations tested (global estimate = 0.029) was higher than that estimated from anonymous genomic microsatellite (F_{ST} = 0.009; Kim and Sappington 2005a) and EST microsatellites (F_{ST} = 0.010; Kim et al. 2008b). Kim and Sappington (2005a) showed that microsatellite-based F_{ST} estimates did not fit isolation-by-distance models when New Deal samples were removed due to suspected effects from hybridization with Mexican corn rootworm, *Diabrotica virgifera zeae*. Kim and Sappington (2005a) suggested that a lack of a significant correlation between $F_{ST}/(1 - F_{ST})$ and geographic distance also may result from a population that has not reached equilibrium due to recent range expansion. Global F_{ST} estimates from SNP-based markers are 3-fold higher than for microsatellite markers, which likely result because of the biallelic nature of SNPs and the marginal effects of homoplasy at SNP loci (Morin et al. 2004). *Diabrotica virgifera virgifera* population homogeneity likely was influenced by a recent range expansion, but affects may have been overestimated by homoplasy among microsatellite alleles. Ryynänen et al. (2007) indicated that profound population differentiation can be detected using few biallelic SNP markers. Ryynänen et al. (2007) also indicated that >7 SNP marker loci are needed to estimate genetic parameters when population differentiation is low or when multilocus analyses are used, which suggested that our inclusion of 12 *D. v. virgifera* SNP markers was sufficient and appropriate for population genetic analyses.

The proportion of SNP markers that were nonneutral within at least one of the populations tested was not significantly different from the proportion of microsatellite markers likewise showing occasional nonneutrality. SNPs that are observed frequently within or between species likely are evolutionarily neutral, whereas low-frequency SNPs are likely deleterious and undergo negative selection (Zhu et al. 2004). SNP frequencies also are influenced by random genetic drift and other genome processes (Miller and Kwok 2001), such that all those at low frequency cannot be assumed deleterious or under selection. Data from 17 synonymous third codon position *D. v. virgifera* SNP loci indicated that the mean minor allele frequency was 0.324. In contrast, 2 nonsynonymous SNP marker loci located at second codon positions, markers 1280 and 1327, had significantly lower minor allele frequencies (0.05 and 0.08, respectively; mean 0.065) when compared with synonymous SNPs (nonparametric *t*-test statistic = 5.48×10^{-6}, 2-tailed $P < 0.0001$, degree of freedom = 1). Human SNPs that are deleterious or acted on by selection are found at lower frequency compared with SNPs that give rise to synonymous mutations, such that SNPs below a population frequency threshold of 6% is an indication that they may be harmful to gene function (Wong et al. 2003). From *D. v. virgifera* SNP data, we cannot directly observe any effects brought about by selection but have used high SNP frequencies as an indicator of selective neutrality or balancing selection.

Conclusions

We showed that SNP marker loci are viable tools for characterization of natural populations and tracking segregation in pedigrees. Analyses of SNP and microsatellite marker data resulted in similar conclusions with respect to population structure. SNP marker loci provided a higher estimate of F_{ST} that may reflect the lack of systematic downward bias due to numerous alleles and possibly the lower reversion rate of substitution mutations compared with microsatellite repeats. Because SNP marker loci are less susceptible to these effects, they may provide better estimates of F_{ST}. Available EST resources can provide a large number of putative SNPs that can be validated for population and genetic mapping studies and will be a focus of future research.

Funding

United States Department of Agriculture Current Research Information System (CRIS) project 016.

Acknowledgments

Mention of proprietary products does not constitute an endorsement or a recommendation by United States Department of Agriculture, Iowa State University, or University of Nebraska for its use. We thank B. Wade French, USDA-ARS, NCARL, Brookings, South Dakota, USA, for providing individuals used in pedigree analyses.

References

Baca F. 1994. New member of the harmful entomofauna of Yugoslavia, *Diabrotica virgifera virgifera* LeConte (Coleoptera, Chrysomelidae). Zastita Bilja. 45:125–131.

Balloux F, Brunner H, Lugon-Moulin N, Hausser J, Goudet J. 2000. Microsatellites can be misleading: an empirical and simulation study. Evolution. 54:1414–1422.

Chiang HC. 1973. Bionomics of the northern and western corn rootworms. Annu Rev Entomol. 18:47–72.

Ciosi M, Miller NJ, Giordano R, Estoup A, Guillemaud T. 2008. Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. Mol Ecol. 17:3614–3627.

Coates BS, Sumerford DV, Hellmich RL, Lewis LC. 2008. Mining an *Ostrinia nubilalis* midgut expressed sequence tag (EST) library for candidate genes and single nucleotide polymorphisms (SNPs). Insect Mol Biol. 17:607–620.

Cox LA, Jett C, Hixson JE. 1998. Molecular basis of an apolipoprotein[a] null allele: a splice site mutation is associated with deletion of a single exon. J Lipid Res. 39:1319–1326.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online. 1:47–50.

Glaubitz JC, Rhodes OE, Dewoody JA. 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphism. Mol Ecol. 12:1039–1047.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Hayes BJ, Nilsen K, Berg PR, Grindflek E, Lien S. 2007. SNP detection exploiting multiple sources of redundancy in large EST collections improves validation rates. Bioinformatics. 23:1692–1693.

Hedrick PW. 1999. Perspective: highly variable Loci and their interpretation in evolution and conservation. Evolution. 53:313–318.

Hedrick PW. 2005. A standardized genetic differentiation measure. Evolution. 59:1633–1638.

Huntley D, Baldo A, Hohri S, Sergot M. 2006. SEAN: SNP prediction and display utilizing EST sequence clusters. Bioinformatics. 22:495–496.

Jensen JL, Bohonak AJ, Kelley ST. 2005. Isolation by distance, web service v.3.15. BMC Genet. 6:13.

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

Kim KS, Sappington TW. 2005a. Genetic structuring of western corn rootworm (Coleoptera: Chrysomelidae) populations in the U.S. based on microsatellite loci analysis. Environ Entomol. 34:494–503.

Kim KS, Sappington TW. 2005b. Polymorphic microsatellite loci from the western corn rootworm (Insecta: Coleoptera: Chrysomelidae) and cross-amplification with other *Diabrotica* spp. Mol Ecol Notes. 5:115–117.

Kim KS, Stolz U, Miller NJ, Waits ER, Guillemaud T, Sumerford DV, Sappington TW. 2008b. Core set of microsatellite markers for western corn rootworm (Coleoptera: Chrysomelidae) population genetic studies. Environ Entomol. 37:293–300.

Kruskal WH, Wallis WA. 1962. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 47:583–621.

Lefko SA, Nowatzki TM, Thompson SD, Binning RR, Pascual MA, Peters ML, Simbro EJ, Stanley BH. 2008. Characterizing laboratory colonies of western corn rootworm (Coleoptera: Chrysomelidae) selected for survival on maize containing event DAS-59122-7. J Appl Entomol. 132:189–204.

Levine E, Spencer JL, Isard SA, Onstad DW, Gray ME. 2002. Adaptation of the western corn rootworm to crop rotation: evolution of a new strain in response to management practice. Am Entomol. 48:94–107.

Meinke JL, Siegfried BD, Wright RJ, Chandler LD. 1998. Adult susceptibility of Nebraska western corn rootworm populations to selected insecticides. J Econ Entomol. 91:594–600.

Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, Derridj S, Kim KS, Reynaud P, Furlan L, Guillemaud T. 2005. Multiple transatlantic introductions of the western corn rootworm. Science. 310:992.

Miller N, Kim KS, Ratcliffe ST, Estoup A, Bourguet D, Guillemaud T. 2006. Absence of genetic divergence between western corn rootworms resistant and susceptible to control by crop rotation. J Econ Entomol. 99:685–690.

Miller RD, Kwok PY. 2001. The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine. Hum Mol Genet. 10:2195–2198.

Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT, Fevolden SE, Berg PR, Lien S. 2008. Identification and characterization of novel SNP markers in Atlantic cod: evidence for directional selection. BMC Genet. 9:18.

Montgomery GW, Campbell MJ, Dickson P, Herbert S, Siemering K, Ewen-White KR, Visscher PM, Martin NG. 2005. Estimation of the rate of SNP genotyping errors from DNA extracted from different tissues. Twin Res Hum Genet. 8:346–352.

Morin PA, Luikart G, Wayne RK. 2004. The SNP Workshop Group SNPs in ecology, evolution, and conservation. Trends Ecol Evol. 19: 208–215.

Prichard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR. 2007. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). J Hered. 98:692–704.

Sammons AE, Edwards CR, Bledsoe LW, Boeve PJ, Stuart JJ. 1997. Behavioral and feeding assays reveal a western corn rootworm (Coleoptera: Chrysomelidae) variant that is attracted to soybeans. Environ Entomol. 26:1336–1342.

Sappington TW, Siegfried BD, Guillemaud T. 2006. Coordinated *Diabrotica* genetics research: accelerating progress on an urgent insect pest problem. Am Entomol. 52:90–97.

Seddon JM, Parker HG, Ostrander EA, Ellegren H. 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Mol Ecol. 14:503–511.

Siegfried BD, Waterfield N, ffrench-Constant RH. 2005. Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathespins. Insect Mol Biol. 14:137–143.

Sivcev I, Manojlovic B, Krnjajic S, Dimic N, Draganic M, Baca F, Kaitovic Z, Sekulic R, Keresi T. 1994. Distribution and harmful effect of *Diabrotica virgifera* LeConte (Coleoptera: Chrysomelidae), a new maize pest in Yugoslavia. Zastita Bilja. 45:19–26.

Spike BP, Tellefson JJ. 1991. Yield response of corn subjected to western corn rootworm (Coleoptera: Chrysomelidae) infestation and lodging. J Econ Entomol. 84:1584–1590.

Stephens M, Sloan JS, Robertson PD, Scheet P, Nickerson DA. 2006. Automating sequence-based detection and genotyping of SNPs from diploid samples. Nature Genet. 38:475–481.

Xing C, Schumacher FR, Xing G, Lu Q, Wang T, Elston RC. 2005. Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis. BMC Genetics. 6(Suppl 1):S29.

Wong GK, Yang Z, Passey DA, Kibukawa M, Paddock M, Liu CR, Bolund L, Yu J. 2003. A population threshold for functional polymorphisms. Genome Res. 13:1873–1879.

Zhang L, Li WH. 2005. Human SNPs reveal no evidence of frequent positive selection. Mol Biol Evol. 22:2504–2507.

Zhao Z, Fu YX, Hewett-Emmett D, Boerwinkle E. 2003. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene. 17:207–213.

Zhu Y, Spitz MR, Amos CI, Lin J, Schabath MB, Wu X. 2004. An evolutionary perspective on single nucleotide polymorphism screening in molecular cancer epidemiology. Cancer Res. 64:2251–2257.

Received November 3, 2008; Revised April 2, 2009; Accepted April 17, 2009

Corresponding Editor: Lacey Knowles