 doi: 10.1111/j.1471-8286.2006.01547.x

PRIMER NOTE

Anonymous nuclear markers for the eastern fence lizard, *Sceloporus undulatus*

E. B. ROSENBLUM, N. M. BELFIORE and C. MORITZ
Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California at Berkeley, Berkeley, CA 94720, USA

Abstract

We present results from a screen for *de novo* variable nuclear loci using a genomic library approach in *Sceloporus undulatus*, the eastern fence lizard. We tested amplification success for 77 primer pairs in *S. undulatus*, *Sceloporus occidentalis* and *Sceloporus grammicus*. Many loci amplified in all three species suggesting that our primers will be useful for developing sequencing or single nucleotide polymorphism (SNP) genotyping markers in other sceloporine lizards. We also sequenced 19 loci, containing 158 variable sites, for 91 *S. undulatus* individuals. We report high levels of nucleotide variation in this species with an average of 38 SNPs per kilobase.

Keywords: nucleotide diversity, *Sceloporus*, single nucleotide polymorphism

Received 20 June 2006; revision accepted 25 July 2006

The eastern fence lizard, *Sceloporus undulatus*, is a broadly distributed taxon with extensive geographic variation and a complex evolutionary history. We developed multiple anonymous nuclear markers for studying the demography of *S. undulatus* in southern New Mexico where populations are adapted to novel substrate environments (Rosenblum 2006). Nuclear markers are also necessary to resolve persistent taxonomic and population genetics questions at multiple phylogenetic depths in sceloporine lizards (e.g. Sites *et al*. 1992; Leaché & Reeder 2002; Wiens & Penkrot 2002). Therefore, we tested *S. undulatus* primers on related species.

We constructed genomic libraries for two *S. undulatus* individuals from Otero County, NM (EBR186, EBR453, accessioned to the Museum of Vertebrate Zoology). Genomic DNA was extracted and sheared. Fragments of target length were isolated from an agarose gel, blunt-ended, ligated into a vector, transformed into *Escherichia coli* and plated on agar. To determine the frequency of repetitive elements in the *S. undulatus* genome, we probed several plates of colonies with the same sheared genomic DNA used in library construction. Hybridized colonies were extremely rare, indicating few high copy number regions. Subtractive hybridization was therefore not necessary. We sequenced 192 clone inserts, and performed BLAST and ENSEMBL searches to characterize the loci. Most loci remained anonymous; only a few exhibited significant matches to known sequences in lizards, other vertebrates, or pathogens (e.g. adenovirus).

Primers with target lengths of 20–28 bps were designed for 77 clones and were tested at three annealing temperatures (usually 57, 61 and 65 C, based on melting temperatures). The following polymerase chain reaction (PCR) parameters and chemistry apply to all reactions: 12- L reactions containing 10 ng DNA, 0.2 mM each dNTP, 0.25 mM each primer, 0.05 U/ L *Taq* Polymerase and 1× ThermoPol Buffer (New England Biolabs) were run for 5 min at 94 C, 30× (30 s at 94 C, 1 min at specified annealing temperature, 2 min at 72 C), 5 min at 72 C. Amplification was tested on a small panel (two *S. undulatus*, one *Sceloporus occidentalis* and one *Sceloporus grammicus*) to determine how *S. undulatus* primers would perform for other sceloporine species. We had high levels of cross-amplification success. Of the 50 loci that amplified for *S. undulatus*, 33 amplified cleanly for *S. occidentalis*, and 23 amplified cleanly for *S. grammicus*. As expected, amplification success decreased with increasing genetic distance. Compared to *S. undulatus*, *S. occidentalis* and *S. grammicus* are, respectively, 5% vs. 8% divergent for mtDNA 12S (Reeder 1995) and 11% vs. 31% divergent for 69 allozyme characters (Mindell *et al*. 1989). A small subset of primers was tested

Correspondence: Erica Bree Rosenblum, Department of Molecular and Cell Biology, University of California at Berkeley, Mailstop 84-171, One Cyclotron Road, Berkeley CA 94720, USA. Fax: 510-643-8238; E-mail: rosenblum@berkeley.edu

on a more divergent sand lizard, *Holbrookia maculata*, but many produced bands in this species as well (5/12 pairs). Primers and annealing temperatures are provided in Table 1. Of the loci that PCR amplified, a subset was optimized for sequencing in *S. undulatus*.

Nineteen loci that sequenced well for *S. undulatus* and contained nucleotide variation were chosen to screen for 91 *S. undulatus* individuals from southern New Mexico. These loci were sequenced directly using BigDye 3.1 cycle sequencing chemistry and visualized on an ABI PRISM 3730 (Applied Biosystems). ABI's KB base-calling software was used, but all sequences were checked by eye in SEQUENCHER (version 4.2, Gene Codes Corporation) to ensure variable sites and heterozygotes were scored correctly. Sequence data were obtained in one direction and truncated at the first insertion/deletion polymorphism. We considered nucleotide positions variable if base pair differences were observed in at least two chromosomes across the dataset. No fixed heterozygotes were observed, so variation was not due to co-amplification of duplicated regions. We resolved gametic phase computationally (PHASE, Stephens & Donnelly 2003), and found that results of analyses were robust to alternative phase calls at positions below the confidence probability threshold of 90%.

A total of 158 variable sites, or single-nucleotide polymorphisms (SNPs), were recorded in 4732 bp with an average of 3.8 SNPs/100 bp. There were many low frequency SNPs (e.g. 88 SNPs with a frequency less than 0.1) and increasingly few high frequency SNPs (e.g. 32 SNPs with a frequency greater than 0.25). We evaluated the minimum number of recombination events within loci (Hudson & Kaplan 1985) using DNASP (version 4.00, Rozas *et al*. 2003) and detected recombination at 11 of 19 loci. No significant linkage disequilibrium (LD) was observed among loci after correcting for multiple comparisons using ARLEQUIN (version 3.0.1, Excoffier *et al*. 2005), indicating that the 19 loci are effectively unlinked. Table 2 provides summary data at both nucleotide and haplotype levels for the 19 loci.

Levels of nucleotide variation observed for *S. undulatus* were high and consistent with other studies of nuclear variation in reptiles (e.g. Hughes & Mouchiroud 2001). The across-locus average of 38 SNPs per kilobase (kb) is an order of magnitude higher than SNP frequencies recorded in many model species [1–5 SNPs/kb in human, chimp, chicken, rat and mouse (e.g. International Chicken Map Consortium 2004 and references therein)]. SNP frequency in *S. undulatus* was also higher than reported in diverse nonmodel species; Brumfield *et al*. (2003) suggest 1 SNP/

Table 1 PCR primers (listed 5′–3′) and preferred annealing temperatures for 39 anonymous nuclear loci. Nineteen loci (Sun_001 through Sun_019) were optimized for sequencing in *Sceloporus undulatus* and comprised the final dataset presented here. Thirty-three loci (Sun_006 through Sun_012 and Sun_014 through Sun_039) cross-amplified in a PCR test with *S. undulatus* (*S.u.*), *Sceloporus occidentalis* (*S.o.*) and *Sceloporus grammicus* (*S.g.*). GenBank Accession nos refer to sequenced clones from which primers were developed

Locus ID	GenBank Accession nos.	Primer sequence (5′–3′)	PCR annealing temperature		
			S.u.	*S.o.*	*S.g.*
Sun_001	DQ784606	F: GTACCCAGTGCTTCCCAAAA R: CAACTCAGGGTTCCCAAAAG	60	n/a	n/a
Sun_002	DQ784607	F: TTTGCTGCGTAGGCTTATCC R: GCCTGTATGCTATGCTCCTTT	60	n/a	n/a
Sun_003	DQ784608	F: ATTTCCTCATAGACACTCCCATTTC R: GTATGTCTTTGAGTCATTTCTGAGCTAT	60	n/a	n/a
Sun_004	DQ784610	F: GGTCTTCTCTCCAAATACTAACAAGACT R: TTATGATATTGTACATGGAGATGTTGTC	60	n/a	n/a
Sun_005	DQ784611	F: GTTGGTTATTACCTTTAAAGCCCTACAT R: ATTTCGCTTTTGCCAATAACATACTT	60	n/a	n/a
Sun_006	DQ784614	F: CTTTTCTAGACCACATTTTACGAATTG R: TGCAAACATAAAAACCAATATTAAAACA	60	54	n/a
Sun_007	DQ784621	F: TTTCTGTCACGATGAAAATTGTAAACTA R: TAAACACAATGCTCACATTAGGAAAAAT	60	60	60
Sun_008	DQ784622	F: CTCTTGAAGTTCACAGGGTTTTCTTAG R: TAGCCTAGCTTCCTTACAGTTTGATAC	58	58	n/a
Sun_009	DQ784623	F: CAATCTCCCTCCCACCCTAAAATA R: ATGTCCACTTGTTGGACTGTCTTAT	61	61	61
Sun_010	DQ784624	F: CAGAAAGTAAATCCACTGTAGCTAGGA R: CTAATAATGGCATAGCAAGGAGTGTAG	61	61	61
Sun_011	DQ784627	F: CAGGAATTACGAGGTGTTTTACCTTAC R: CCTGTGCTTATTTCCTATCCAAAAC	58	58	58
Sun_012	DQ784629	F: TACAGAGTCTCCTCTTGACTGGATATT R: TTGGTACACTAACTCAAGCAAACCT	62	62	62

Table 1 *Continued*

Locus ID	GenBank Accession nos.	Primer sequence (5′–3′)	PCR annealing temperature		
			S.u.	*S.o.*	*S.g.*
Sun_013	DQ784637	F: CAAATAACCCCTAACTCTAGCCTTC R: CATACTTCCTAATAGCATTCTCTCCAC	62	n/a	n/a
Sun_014	DQ784639	F: GATTGCAACACACACTGGTTTATATC R: GTGTAAAACTAGAACCTAGAGTTAAAAGAA	57	57	57
Sun_015	DQ784640	F: GTACCCTAATAAGTGGGTATCTTCAGTC R: AGATAGTCCAGATTTGAGAAGTTACGTT	57	57	57
Sun_016	DQ784641	F: GTACCCTTTCTTTTGTAATGCTCTTTTT R: CTCTTCTTCCTCTTCCTCCACTTTCT	57	57	57
Sun_017	DQ784642	F: CATGTGATTTAAGGTGGTTTTATTTTAG R: AGATAATCTACAGTGATGGGATTTCATT	57	57	57
Sun_018	DQ784643	F: ATGACAGAAGTTGTGGTTCAACAGTAT R: AGTGAGATAGAAGTGGCTTTCTGATTAC	57	57	57
Sun_019	DQ784644	F: TGCTCTCATAGCTTTCATACATTATCTT R: CAATAGAGGATTAATCAGAACTGTCTC	57	57	57
Sun_020	DQ784609	F: CCCATGATTCAGTTTAAAGCACTAC R: AATGGGTTAAGAAGAGGCATTAAATTAT	61	61	61
Sun_021	DQ784612	F: GAGAGTCTCTATCAAATAACACACACAA R: TTTGGCTAAAGTGATACCTATTTTGATA	60	60	60
Sun_022	DQ784613	F: CATTCTTAGTCTTCTTCTGATTTCTTGA R: ACAAAATGTTGCCTGAAGATGAAT	58	58	58
Sun_023	DQ784615	F: GTATAATGACTTCAGCTTGATCTTCTTG R: GTATGCAGACAGAAATATGTTGTAAAGA	60	54	n/a
Sun_024	DQ784616	F: TACCCTCCTTCACTCATTCATCTAAAGT R: TCATTTATATACCGCTTCATACTGAACT	61	61	61
Sun_025	DQ784617	F: CTAATGTTCAGGTGGAATCTCTTTTC R: CAGACTTATCAGATTTGAAGATGACAC	60	57	57
Sun_026	DQ784618	F: AGATCATCTTCACCATAAGGTTTCTAAT R: AGTAGCAGCAGTACAGGCATTTAACTA	61	61	58
Sun_027	DQ784619	F: GCAGTATAGGAAGAACAGAACAAGATAG R: TGAAATCAGTCTCCTTGTAAGATTTGT	54	54	54
Sun_028	DQ784620	F: AATCTTATTTCTGCAGTTGATGTACTTT R: ATAAATGCAATGCCACAAATATAATAAG	54	54	n/a
Sun_029	DQ784625	F: GGGTACACATACATAGCATTTAACCAC R: AAGAGTGGCACATACATTACAGAGAGT	61	61	58
Sun_030	DQ784626	F: AGACCATGCTACTAAAACTGTGCTACT R: GTGGAGGGGAAATGAATATTTCTG	61	61	n/a
Sun_031	DQ784628	F: CTCTCTCCACATCCTCACCAAACAT R: AGCAGCAGCATTCTCACTGTGCAATAAA	65	65	65
Sun_032	DQ784630	F: GGACACAGAGAGCTGACTGTATACTAAA R: GCTAGGGAATTCTAGGAACTAGTTTTG	65	65	65
Sun_033	DQ784631	F: GTGTTATATAATGCCAGGAGCTCTCA R: CCCCACTTAAAAGGAATAGCACATA	60	60	60
Sun_034	DQ784632	F: CTGCCATGCAGGAAATCCAGATCA R: CCAAGTAGACCACAAGCTGAACAT	65	65	65
Sun_035	DQ784633	F: CATGTCTCTGAACTGTCTCCCTTTTA R: CTGCTGAGTAAATTTTGCCAAGAGA	60	60	60
Sun_036	DQ784634	F: GATATGACTTGCAACAGACTGGTTT R: GCTTGTAGAAGCCAACCTAACTATATG	60	60	n/a
Sun_037	DQ784635	F: AACACAATTCAGACCTCAAACAGAC R: ACTGTTGTACAAACAAACACACCAC	62	62	62
Sun_038	DQ784636	F: CCCCCGGCCACAAAAATTTAGTTA R: GGAACAAAGAATTCTGTTAGACCAACC	65	65	n/a
Sun_039	DQ784638	F: ACTATACAGCCTTCTGGACAGGAC R: CGATGTATGTATAGTGGACTACATGAGC	60	60	n/a

Table 2 Summary data for the 19 anonymous nuclear loci screened for variation in *Sceloporus undulatus*. Sample size indicates the number of individuals sequenced at each locus. Heterozygosities were calculated based on haplotypes (Guo & Thompson 1992) in ARLEQUIN (version 3.01). Asterisks in the final column indicate nine loci for which the observed heterozygosity (H_O) deviated significantly from the expected heterozygosity (H_E) after a correction for multiple comparisons, unsurprising given that sequence data for multiple populations were pooled. All other summary statistics were calculated using DNASP (version 4.0)

Locus ID	Sequencing direction	Sample size	No. base pairs	No. SNPs/ 100 bp	No. variable sites	No. haplotypes	Nucleotide diversity	Haplotype diversity	H_O	H_E
Sun_001	Reverse	63	68	10.3	8	8	0.013	0.531	0.29	0.53*
Sun_002	Reverse	90	475	1.7	12	11	0.005	0.845	0.73	0.85*
Sun_003	Forward	61	128	1.6	3	3	0.003	0.325	0.26	0.34
Sun_004	Reverse	78	281	4.6	14	22	0.006	0.853	0.59	0.85*
Sun_005	Forward	90	269	3.3	14	8	0.003	0.484	0.40	0.49
Sun_006	Forward	88	492	3.0	18	19	0.007	0.839	0.80	0.84
Sun_007	Forward	91	317	0.9	4	4	0.001	0.155	0.15	0.17
Sun_008	Reverse	85	186	1.6	5	7	0.004	0.643	0.65	0.64
Sun_009	Forward	84	453	7.3	39	27	0.015	0.902	0.75	0.90*
Sun_010	Reverse	81	259	7.7	21	22	0.013	0.817	0.31	0.82*
Sun_011	Forward	88	256	1.6	4	6	0.004	0.702	0.53	0.70*
Sun_012	Forward	90	248	2.0	9	6	0.003	0.562	0.52	0.57
Sun_013	Reverse	68	540	0.9	6	7	0.002	0.588	0.40	0.59*
Sun_014	Reverse	90	25	8.0	2	3	0.023	0.540	0.41	0.54
Sun_015	Forward	83	152	2.0	3	6	0.005	0.622	0.46	0.63*
Sun_016	Forward	83	237	5.9	15	12	0.013	0.751	0.66	0.76
Sun_017	Forward	70	162	2.5	6	6	0.006	0.739	0.66	0.74
Sun_018	Forward	84	154	4.5	7	9	0.005	0.524	0.35	0.53*
Sun_019	Reverse	72	30	3.3	1	2	0.012	0.371	0.24	0.38

200–500 bp of noncoding DNA as a rule of thumb. The current study recorded an average of roughly 1 SNP/25 bp, which is particularly striking given that sequences were obtained from a geographically restricted area. Researchers using primers presented here for marker development in other sceloporine lizards will likely obtain many SNPs by screening only a modest number of loci.

Acknowledgements

This work was supported by C.M. and NSF grants to E.B.R. (#DEB-0309327) and N.M.B. (#DBI-0306091). J. Sites and A. Leaché provided helpful comments.

References

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Guo S, Thompson E (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Hughes S, Mouchiroud D (2001) High evolutionary rates in nuclear genes of squamates. *Journal of Molecular Evolution*, **53**, 70–76.

International Chicken Polymorphism Map Consortium (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, **432**, 717–722.

Leaché AD, Reeder TW (2002) Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Systematic Biology*, **51**, 44–68.

Mindell DP, Sites JW Jr, Graur D (1989) Speciational evolution: a phylogenetic test with allozymes in *Sceloporus* (Reptilia). *Cladistics*, **5**, 49–62.

Reeder TW (1995) Phylogenetic relationships among phrynosomatid lizards as inferred from mitochondrial ribosomal DNA sequences: substitutional bias and information content of transitions relative to transversions. *Molecular Phylogenetics and Evolution*, **4**, 203–222.

Rosenblum EB (2006) Convergent evolution and divergent selection: lizards at the White Sands ecotone. *American Naturalist*, **167**, 1–15.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Sites JW Jr, Archie JW, Cold CJ, Flores-Villela O (1992) A review of phylogenetic hypotheses for lizards of the genus *Sceloporus* (Phrynosomatidae): implications for ecological and evolutionary studies. *Bulletin of the American Museum of Natural History*, **213**, 1–110.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Wiens JJ, Penkrot TA (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Systematic Biology*, **51**, 69–91.