Published in final edited form as:

Conserv Genet. 2008 December ; 9(6): 1657–1660. doi:10.1007/s10592-008-9511-7.

Isolation and characterization of SNP variation at 90 anonymous loci in the banded wren (*Thryothorus pleurostictus*)

EMILY R. A. CRAMER^{*,†,‡}, L. STENZLER^{*}, A. L. TALABA^{*}, C. A. MAKAREWICH^{*}, S. L. VEHRENCAMP^{†,‡}, and I. J. LOVETTE^{*}

^{*}Fuller Evolutionary Biology Program, Laboratory of Ornithology, Cornell University, Ithaca, New York 14850, USA

[†]Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA

[‡]Bioacoustics Research Program, Laboratory of Ornithology, Cornell University, Ithaca, New York 14850, USA

Abstract

Single nucleotide polymorphisms (SNPs) are becoming more commonly used as molecular markers in conservation studies. However, relatively few studies have employed SNPs for species with little or no existing sequence data, partly due to the practical challenge of locating appropriate SNP loci in these species. Here we describe an application of SNP discovery via shotgun cloning that requires no pre-existing sequence data and is readily applied to all taxa. Using this method, we isolated, cloned and screened for SNP variation at 90 anonymous sequence loci (51kb total) from the banded wren (*Thryothorus pleurostictus*), a Central American species with minimal pre-existing sequence data. We identified 168 SNPs (a mean of one SNP/305 bp, with SNPs unevenly distributed across loci). Further characterization of variation at 41 of these SNP loci among 256 individuals including 37 parent-offspring families suggests that they provide substantial information for defining the genetic mating system of this species, and that SNPs may be generally useful for this purpose when other markers are problematic.

Keywords

Thryothorus pleurostictus; SNP; single-nucleotide polymorphism; paternity analysis

Conservation efforts depend on molecular tools both in practical applications and in research. The availability of diverse types of molecular markers encourages the choice of the marker best-suited to a particular question, and it provides alternatives if one type of marker fails. Recent theoretical work (Morin et al. 2004, Anderson et al. 2006) and empirical studies (Belfiore et al. 2003, Herreaz et al. 2005, Seddon et al. 2005) suggest that a relatively novel molecular marker, the single nucleotide polymorphism (SNP), will become particularly useful in conservation. SNPs are broadly distributed across the genome, they have a well-characterized model of molecular evolution, and they are readily compared among laboratories using efficient, cost-effective scoring techniques (Vignal et al 2002, Morin et al. 2004). Moreover, SNPs can be easily isolated and characterized using existing genomic data, which are becoming increasingly available, or using basic cloning and sequencing techniques (Morin et al. 2004). Moreover, SNPs can be easily isolated and characterized using existing genomic data, which are becoming increasingly available, or using basic cloning and sequencing

techniques (Morin et al. 2004). Most SNP discovery to date has used existing sequence information from closely related species or from genomic regions flanking microsatellites, which may limit the breadth of taxa and/or the genomic distribution of SNP loci used. Less restrictive methods of SNP discovery, such as the one presented here, will allow conservation geneticists to apply these powerful markers more broadly.

We describe an application of SNP discovery via shotgun cloning that requires no pre-existing sequence data and is readily applied to all taxa. Using this technique, we rapidly isolated and characterized 90 anonymous genomic loci from which we developed a panel of 41 SNP markers for the banded wren (*Thryothorus pleurostictus*). These SNPs will be used to determine the genetic mating system of this species for which the dozens of microsatellite loci screened to date are insufficiently variable to determine paternity with adequate statistical confidence (Brar et al. 2007; unpublished data). Although banded wrens are not threatened or endangered (IUCN 2007), their habitat—tropical dry forest—is of great conservation concern (Janzen 1988), and the species has atypically low microsatellite allelic variation (Brar et al. 2007; unpublished data). Informed management of dry forest tracts will preserve the existing genetic variation. Banded wrens may offer insight into recovery from population bottlenecks and fragmentation, as glaciation events apparently caused such demographic events in the species (Brar et al. 2007).

To generate an initial genomic library of anonymous sequence loci, we extracted genomic DNA from the blood of a single individual using Perfect gDNA Blood Mini Kits (Eppendorf). Separate DNA aliquots were digested with restriction enzymes EcoRV or HincII (Invitrogen). We cloned size-selected fragments 600–2000 bp long into pSMART-HC Amp plasmids and heat-shock transformed plasmids into *E. coli*® (Lucigen) chemically competent cells using the CloneSmart Blunt Cloning Kit (Lucigen). We then plated bacteria on agar/ampicillin media and grew hundreds of bacterial colonies containing wren DNA inserts. For sequencing, we transferred cells with a sterile toothpick to 50 µL 5 mM Tris, pH 7.5 (Fisher) in 96-well polymerase chain reaction (PCR) plates. We lysed cells by heating the mixture to 98°C for 10 minutes, then centrifuged the solution and used 1 µL supernatant as template DNA for PCR. We amplified clones using 2 pmol of plasmid primers SL1 (5'-CAGTCCAGTTACGCTGGAGTC-3') and SR2 (5'-GGTCAGGTATGATTTAAATGGTCAGT-3'), 0.5 units JumpStart™ Taq polymerase (Sigma), 10 uM TrisHCl pH 8.3, 50 mM KCl, 2.0 mM MgCl₂, and 200 uM dNTPs (Invitrogen) in 10 µL reaction mixes. Thermal cycling conditions were: initial denaturing at 95°C for 3 min; 35 cycles of 95°C for 50 s, 50 °C for 50 s, and 72 °C for 50 s; and a final extension at 72°C for 5 min. We checked product quality on a gel and prepared those with single, robust bands > 300 bp long for sequencing by adding 5 units Exonuclease I and 0.5 units Shrimp Alkaline Phosphatase (USB), then incubating the mixture at 37 °C for 30 min and 90 °C for 10 min. We sequenced loci on both strands using Applied Biosystems (ABI) BigDye chemistry version 3.1 with the amplification primers. Products were cleaned using Sephadex G-50 columns, and sequence analysis was performed with an ABI automated 3730xl DNA analyzer. We amplified DNA from 783 colonies, 255 of which we sequenced.

We aligned and edited sequences in Sequencher 4.5 (GeneCodes). For 104 sequences of sufficient length and quality, we designed wren-specific primers using PrimerSelect 5.07 (DNASTAR). Ninety of these loci were successfully optimized for PCR by varying the annealing temperature, MgCl₂ concentration, or the number of PCR cycles; all other conditions remained as described above. We forward-sequenced these loci from a panel of 6–12 individual wrens and reverse-sequenced when necessary to obtain a high-quality complete sequence. Sequences were aligned and edited in Sequencher 4.5. We considered base-pair differences among individuals to be SNPs only when flanking sequence was high quality and/or the alternative bases were each present in several individuals.

Across all 51,250 bp of high-quality sequence, we found 168 SNPs, equivalent to a mean of 1 SNP per 305 bp, although SNPs were unevenly distributed. Twenty-six fragments lacked SNPs entirely (GenBank dbGSS accession numbers ER936130-ER936155), and the remaining 64 loci each contained 1-13 SNPs (ER936066- ER936129). We designed an assay including 48 SNPs, each from a different DNA fragment, for cost-effective throughput (Table 1). Many technologies are available for genotyping SNPs (Vignal et al. 2002); we used SNPlex (ABI), an oligoligation assay, to genotype 37 putative banded wren families and 43 candidate extra-pair parents (256 individuals total). The SNPlex assay was performed at the Cornell University Core Laboratory Center. Genotyping results were analyzed with GeneMapper 3.7 software (ABI). Seven of the 48 SNPs (14.6%) failed for unknown reasons, whereas the remaining 41 gave robust separation of homozygous and heterozygous genotypes.

Using the genotypes of 101 presumably unrelated adult individuals, we found no evidence for linkage disequilibrium or for deviation from Hardy-Weinberg equilibrium using Genepop v3.4 (Raymond and Rousset 1995) and controlling for multiple testing with false discovery rate (Pounds and Cheng 2006). There was no evidence of sex linkage, as observed heterozygosity for 34 known females was non-zero for all loci (Kalinowski et al. 2007). Preliminary analyses using Cervus (Kalinowski et al. 2007) suggest the probability of correctly excluding unrelated males as fathers is substantially higher using this panel of 41 SNPs (exclusion probability of 0.998) than using the 7 most informative microsatellite loci tested for this species (0.877). The presence of readily scored SNP variation in this species with unusually low microsatellite allele variation suggests that SNPs will be broadly useful for relatedness assays in taxa for which other markers are problematic.

Acknowledgements

We thank the many field-crew participants who collected banded wren DNAs and the Cornell Core Laboratory Center staff for assistance and advice with SNPlex. Funding was provided by NIH grant R01-MH60461 to SLV.

References

- Anderson EC, Garza JC. The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics* 2006;172:2567–2582. [PubMed: 16387880]
- Belfiore NM, Hoffman FG, Baker RJ, et al. The use of nuclear and mitochondrial single nucleotide polymorphisms to identify cryptic species. *Mol Ecol* 2003;12:2011–2017. [PubMed: 12803650]
- Brar RK, Schoenle A, Stenzler LM, et al. Eleven microsatellite loci isolated from the banded wren (*Thryothorus pleurostictus*). *Mol Ecol Notes* 2007;7:69–71. [PubMed: 18392115]
- Herraez DL, Schafer H, Mosner J, et al. Comparison of microsatellite and single nucleotide polymorphism markers for the genetic analysis of a Galloway cattle population. *Z Naturforsch C* 2005;60:637–643. [PubMed: 16163842]
- IUCN. 2007 IUCN red list of threatened species. 2007. <http://www.iucnredlist.org/>
- Janzen, DH. Tropical dry forests: The most endangered major tropical ecosystem. In: Wilson, EO.; Peter, FM., editors. *Biodiversity*. National Academy Press; Washington, DC: 1988.
- Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 2007;16:1099–1106. [PubMed: 17305863]
- Morin PA, Luikart G, Wayne RK. SNPs in ecology, evolution and conservation. *Trends Ecol Evol* 2004;19:208–216.
- Pounds S, Cheng C. Robust estimation of the false discovery rate. *Bioinformatics* 2006;22:1979–1987. [PubMed: 16777905]
- Raymond M, Rousset F. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 1995;86:248–249.
- Seddon JM, Parker HG, Ostrander EA, et al. SNPs in ecological and conservation studies: a testing the Scandinavian wolf population. *Mol Ecol* 2005;14:503–511. [PubMed: 15660941]

Vignal A, Milan D, SanCristobal M, et al. A review of SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 2002;34:275–305. [PubMed: 12081799]

41 SNPs successfully assayed* using SNPlex technology. SNP positions refer to position in consensus sequence submitted to dbGSS. Allele frequency and expected (H_E) and observed (H_O) heterozygosities were calculated in Genepop (Raymond and Rousset 1995); null allele frequency was estimated using Cervus (Kalinowski et al. 2007) using N chromosomes.

Table 1

Locus	SNP position	SNP id	N	Rare allele frequency	H_E	H_O	Null allele frequency	Accession Number
ThPISNP01	615	C/T	200	0.305	0.426	0.450	-0.030	ER936066
ThPISNP02	364	A/G	200	0.270	0.396	0.360	0.045	ER936067
ThPISNP03	259	G/C	200	0.375	0.471	0.550	-0.080	ER936068
ThPISNP06	400	G/C	202	0.045	0.086	0.089	-0.014	ER936070
ThPISNP07	221	C/T	200	0.040	0.077	0.080	-0.011	ER936071
ThPISNP08	210	G/C	198	0.187	0.305	0.333	-0.046	ER936072
ThPISNP10	134	C/T	200	0.315	0.434	0.410	0.026	ER936074
ThPISNP11	44	A/C	200	0.395	0.480	0.390	0.101	ER936075
ThPISNP12	134	C/T	196	0.281	0.406	0.418	-0.018	ER936076
ThPISNP13	293	C/T	188	0.287	0.412	0.426	-0.019	ER936077
ThPISNP15	298	A/G	200	0.340	0.451	0.520	-0.074	ER936079
ThPISNP16	169	A/G	200	0.430	0.493	0.460	0.032	ER936080
ThPISNP17	98	C/T	202	0.069	0.130	0.139	-0.027	ER936081
ThPISNP18	575	C/T	200	0.370	0.469	0.420	0.052	ER936082
ThPISNP19	109	A/G	194	0.041	0.079	0.082	-0.012	ER936083
ThPISNP21	308	A/C	198	0.455	0.498	0.424	0.078	ER936085
ThPISNP22	646	C/T	202	0.040	0.076	0.059	0.121	ER936086
ThPISNP24	118	C/T	198	0.045	0.087	0.091	-0.014	ER936088
ThPISNP25	349	C/T	156	0.372	0.470	0.564	-0.094	ER936089
ThPISNP26	128	A/G	202	0.144	0.247	0.228	0.038	ER936090
ThPISNP27	325	A/G	200	0.335	0.448	0.470	-0.027	ER936091
ThPISNP28	332	A/G	194	0.309	0.429	0.454	-0.030	ER936092
ThPISNP29	43	C/T	198	0.111	0.199	0.222	-0.053	ER936093
ThPISNP30	444	C/T	202	0.381	0.474	0.446	0.029	ER936094
ThPISNP31	143	A/T	192	0.453	0.498	0.490	0.006	ER936095
ThPISNP32	124	C/T	202	0.094	0.171	0.188	-0.043	ER936096
ThPISNP33	342	C/T	200	0.220	0.345	0.360	-0.024	ER936097
ThPISNP35	204	A/G	184	0.033	0.063	0.065	-0.008	ER936099
ThPISNP36	177	C/T	200	0.180	0.297	0.320	-0.040	ER936100
ThPISNP37	46	A/T	200	0.240	0.367	0.280	0.132	ER936101
ThPISNP38	133	A/G	198	0.303	0.425	0.444	-0.025	ER936102
ThPISNP39	493	A/G	202	0.356	0.461	0.475	-0.018	ER936103
ThPISNP41	429	A/G	190	0.226	0.352	0.368	-0.025	ER936104
ThPISNP43	379	A/G	202	0.376	0.472	0.475	-0.006	ER936106
ThPISNP44	92	A/T	200	0.060	0.113	0.120	-0.022	ER936107
ThPISNP45	237	C/T	202	0.109	0.195	0.178	0.043	ER936108
ThPISNP46	573	A/G	198	0.318	0.436	0.455	-0.023	ER936109
ThPISNP48	165	G/C	202	0.243	0.369	0.386	-0.025	ER936110
ThPISNP49	196	G/C	200	0.340	0.451	0.500	-0.054	ER936111
ThPISNP50	302	C/T	190	0.432	0.493	0.421	0.076	ER936112
ThPISNP51	273	A/G	200	0.235	0.361	0.290	0.107	ER936113

* Seven SNPs designed for this assay failed; they were [locus-position (Accession number)]: ThPISNP05-226 (ER936069); ThPISNP09-203 (ER936073); ThPISNP14-127 (ER936078); ThPISNP20-127 (ER936084); ThPISNP23-450 (ER936087); ThPISNP34-334 (ER936098); ThPISNP42-63 (ER936105).