Molecular Ecology Notes (2005) **5**, 16–20 doi: 10.1111/j.1471-8286.2004.00816.x

PRIMER NOTE

Polymorphic microsatellite loci in a plural breeder, the grey-capped social weaver (*Pseudonigrita arnaudi*), isolated with an improved enrichment protocol using fragment size-selection

SUSAN B. MCRAE,*† STEPHEN T. EMLEN,* DUSTIN R. RUBENSTEIN* and STEVEN M. BOGDANOWICZ‡

*Department of Neurobiology and Behaviour, Seeley Mudd Hall, Cornell University, Ithaca, NY 14853, USA; †Department of Biology, East Carolina University, Greenville, NC 27858–4353, USA; ‡Department of Ecology and Evolutionary Biology, Corson Hall, Cornell University, Ithaca, NY 14853, USA

Abstract

We isolated polymorphic microsatellite loci (mean number of alleles = 12) from the grey-capped social weaver (*Pseudonigrita arnaudi*) genome. We designed an experiment to test whether the addition of a fragment size-selection step at either of two different stages in an enrichment protocol increased the average insert size in transformants. Fragment size-selection produced significantly larger average insert sizes among our positive (microsatellite-containing) transformants. Size-selection just before ligation to the cloning vector was most effective. Eighteen independently-assorting polymorphic loci were characterized. Some amplified in other ploceid weavers and related species. Collectively, the markers provide a mean exclusion probability of 0.9999 for parentage analysis, and will be a powerful tool for reconstructing extended family membership in this highly social bird.

Keywords: enrichment, microsatellite, *Passer domesticus*, Ploceidae, *Pseudonigrita arnaudi*, size-selection

Received 2 July 2004; revision received 26 August 2004; accepted 26 August 2004

The grey-capped social weaver (class Aves, family Ploceidae, *Pseudonigrita arnaudi*), hereafter GCSW, lives in extended family groups throughout its range in East Africa. Family groups range in size from two to nine members. Groups are typically composed of a dominant pair and adult helpers, mostly mature sons of the dominants and their mates. These subordinate pairs commonly breed. As part of a comprehensive study of the social organization and reproductive skew in these family groups, we are examining the genetic relationships among group members in order to compare the behaviour of different categories of relatives, to determine dispersal patterns, and examine patterns of reproductive partitioning within social groups.

We isolated new microsatellite markers for GCSWs using an enrichment protocol incorporating magnetic capture of streptavidin beads with biotinylated probes bound to microsatellite-containing genomic fragments (Kijas *et al*. 1994; Fleischer & Loew 1995). Enrichment protocols are highly effective in increasing the proportion of microsatellite-containing inserts among positive transformants (Zane *et al*. 2002). However, they can introduce the problem of reducing average insert length, possibly because of added PCR steps that favour amplification of smaller templates when large and small templates are mixed together. In the first libraries we constructed, the mean insert size was small: 83% ($N = 95$) of the inserts we sequenced were less than 300 bp in length. This is problematic when trying to design primers for microsatellite analysis because there is often insufficient sequence flanking the repeat region from which to design primers. Therefore, in subsequent libraries, we added a size-selection step at two different stages in the procedure in an attempt to increase the average insert size in our transformants.

Blood was collected in Queen's lysis buffer (Seutin *et al*. 1991) at study sites near the Mpala Research Centre (Nanyuki, Kenya). Genomic DNA was isolated from blood samples

Correspondence: S. B. McRae, Department of Biology, East Carolina University, Greenville, NC 27858-4353, USA. Fax: +1 252 328 4178; Email: mcraes@mail.ecu.edu

from four individual GCSWs using a DNeasy Tissue Kit (Qiagen). The DNA was combined for restriction digestion. Four aliquots of combined sample were each digested with a different restriction enzyme: *Alu*I, *Hae*III, *Hin*cII or *Rsa*I, and simultaneously ligated to SNX linker sequences (Hamilton *et al*. 1999). The digests were combined for magnetic capture of microsatellite-enriched fragments using biotinylated probes bound to streptavidin beads (Kijas *et al*. 1994). Among 10 separate libraries, we enriched for $(CT)_n$ $(GT)_n$ $(GTG)_n$ $(CTG)_n$ $(CTC)_n$ $(CTT)_n$ $(ATC)_n$ and $(CTA)_n$.

Four of the 10 libraries were produced using a fragment size-selection (SS) step. For two of these libraries enriched for GT, CT, ATC and CTA, respectively, we size-selected the original restriction digests run on a 1% agarose gel by excising from the gel DNA fragments in the 300–1000 bp range (early SS treatment). In two identical libraries, we size-selected in the same way DNA fragments in the same size range after the enrichment procedure (late SS treatment). This was done following a vent (exo-) polymerase (New England Biolabs) PCR (polymerase chain reaction) step to render the single-stranded products of magnetic capture double-stranded, for ligation to the cloning vector (Hamilton *et al*. 1999). We purified the gel-bound fragments using a gel extraction kit (Qiagen). In both control and experimental libraries, fragments enriched for microsatellites were ligated to pUC19 vector, used to transform *Escherichia coli* competent cells (Life Technologies). Transformation efficiency was consistent among the libraries with two thirds or more of the transformants containing inserts.

We screened each of the libraries using oligos end-labelled with [γ^{33}P] ATP as in Bogdanowicz *et al*. (1997), but hybridizations and washes were carried out at 52 °C.

We produced minipreps (Sambrook *et al*. 1989) for 287 positive clones. We amplified inserts from all minipreps and ran these on a 1% agarose gel to determine their relative sizes. Size-selected libraries produced more than twice as many inserts larger than 300 base pairs in length than did control libraries, and this difference was significant (percent inserts from minipreps > 300 bp (N): early SS treatment = 40% (55), late SS treatment = 46% (41), control (No SS) = 20% (191); $\chi^2 = 16.26$, 2 d.f., $P = 0.0003$). Note that for statistical analyses, sequences derived from different libraries are pooled by treatment. We sequenced 104 of the largest inserts and found a significant effect of the size-selection treatments on mean insert length. Among sequenced inserts, mean insert size was significantly larger among the size-selected treatments than in controls (mean insert size (in base pairs) for early SS = 338 ($N = 30$), late SS = 380 ($N = 22$), control = 290 ($N = 52$); ANOVA $F_{2,101} = 5.87$, $P = 0.004$). The effect was strongest for libraries constructed using the late SS treatment where size-selection was conducted just prior to ligation (Fisher's least significant difference early SS vs. no SS, mean difference = 47.6, $P = 0.05$; late SS vs. no SS, mean difference = 89.7, $P = 0.001$).

Primers were designed for 64 microsatellite-containing sequences. We tested marker variability on an ABI PRISM 377 DNA sequencer, and scored fragment sizes using GENESCAN™ software (Applied Biosystems). Most of the genotype data were generated on an ABI PRISM 3100 capillary DNA sequencer. Genotype data were processed using GENEMAPPER™ software (Applied Biosystems). Population statistics were generated using CERVUS (Marshall *et al*. 1998) and the web version of GENEPOP (Raymond & Rousset 1995).

Eighteen polymorphic loci were characterized for genetic analyses of GCSW families (Table 2). We genotyped 801 individuals from three main study sites. PCR reactions were carried out in MJ Research dyad thermocyclers, in 0.2 mL microtitre plates (MJ Research or Marsh), sealed with MJ Research 'A' sealing film. PCR reactions (10 μL) used 0.5 Units *Taq* DNA polymerase (Invitrogen) and reaction buffer (20 mM Tris-HCl (pH 8.4), 50 mM KCl (provided by Invitrogen), and 0.2 mM dNTPs. Most of the loci were amplified in multiplex PCR reactions: GCSW8 with GCSW10 (primers in ratio of 5:1.6), GCSW13 with GCSW35 (2:1), GCSW18 with GCSW55 (1:1), GCSW41 with GCSW45 (5:4), GCSW15 with GCSW50 (5:4), GCSW51 with GCSW47 (1.4:1). GCSW31 and GCSW39 were optimized for multiplex PCR together, but GCSW39 amplified poorly in combination, so they were separated. The loci GCSW4, GCSW20, GCSW57 and GCSW58 were always amplified alone. For optimized $MgCl_2$ concentration and maximum annealing temperatures, T_a, see Table 1. A typical reaction profile was as follows: 95 °C for 2 min, then five cycles of 92 °C for 40 s; T_a for 1 min; 72 °C for 1 min, followed by 25 cycles of 90 °C for 30 s (T_a –2 °C) for 30 s; 72 °C for 1 min, and a final extension of 72 °C for 30 min.

GCSW45 was determined to be sex-linked based on the fact that all 433 females genotyped appeared homozygous at this locus. Sex in this monomorphic species was determined independently by amplification of a section of the CHD gene (Griffiths *et al*. 1998). The observed heterozygosity for 365 males in the sample was 0.80 ($H_E = 0.78$). Our observations are consistent with this locus being on the avian Z chromosome because, in birds, females are the heterogametic sex and have a single copy of the Z chromosome.

There were no serious problems with null alleles among our loci. GCSW41 had the highest frequency of null alleles: 0.06, as estimated by CERVUS. Null allele frequency estimates for other autosomal loci ranged from 0.001–0.03. However, when we included all samples genotyped to date ($N = 801$), GENEPOP indicated each of the loci except GCSW31 was found to deviate significantly from the Hardy–Weinberg equilibrium, and 109 out of 153 locus dyads (71%) showed significant statistical linkage. We suspected that this might be due to the population being composed of extended families with large numbers of first-order relatives. Thus, we subsampled the population to exclude most of

Table 1 Characterization of polymorphic microsatellite loci for GCSWs

Locus	Repeat motif	Primer sequence 5′–3′	T_a (°C)	[$MgCl_2$] (mM)	Product size	Allele size range	No. alleles	N	H_O	H_E	Ex
GCSW4	$(AC)_{15}(TC)_3T_2(TC)_6$	F: CATTGCCTGTGCACATAAAC R: TGGCACATAAGATGATGAAAAT	57	2.0	125	113–143	14	801	0.89	0.89	0.63
GCSW8	$(AC)_{18}$	F: ACCACAATCAGAATTAGTAGAT R: ACCCTGCTGGTCATAGATATTGTAT	61	2.0	296	270–296	12	783	0.81	0.83	0.50
GCSW10	$(CA)_{15}$	F: CCATGACACGGACTCACTTC R: CATTTTGTAAGGCCCAGCTC	61	2.0	201	181–203	11	799	0.78	0.74	0.36
GCSW13	$(ATT)_8(ATC)_6$	F: AAGGCTGTAGAAACACCTTACA R: TGTCTAACTTTAGTCAACATAATGC	60	3.5	175	157–185	9	801	0.70	0.71	0.29
GCSW15	$(TAG)_{10}$	F: GGAGCCCAAACAGGATGTAT R: TCAAGATATGTGAAGATGAGGTGA	60	2.0	155	147–188	9	795	0.54	0.53	0.15
GCSW18	$(TG)_{10}$	F: CAGCAGCCTCAGTGATCCTA R: ATCCAGTGAGTGCTGGTGAG	60	1.4	202	192–212	7	765	0.51	0.54	0.16
GCSW20	$(AC)_{13}$	F: ACCCTAATTTGGGTCTCAAT R: GTTCAAAGATAGTTTAACATA	57	5.0	144	126–158	15	795	0.66	0.66	0.28
GCSW31	$(TG)_{14}$	F: GCACAGACACACTGCTACTGG R: CTTTGTAGCATGAGGACAGCA	60	2.0	301	298–332	15	799	0.70	0.67	0.28
GCSW35	$(CA)_{15}$	F: AAATGATTGCCACTCCATGA R: AACCTGTTCCTGGTCACACC	60	3.5	237	225–252	10	800	0.70	0.73	0.34
GCSW39	$(GT)_{12}ATGTTT(GT)_2GC(GT)_8$	F: TCCCCTGGACTTCTCAAATG R: AGGAATTGAACCGATAGAATGTG	60	2.0	219	202–245	23	793	0.86	0.86	0.58
GCSW41	$(GT)_{15}$	F: AGGGACACACACAGCATTACA R: GGTTTTGTCGTTGGGATTTG	60	1.4	163	150–174	11	776	0.68	0.78	0.41
GCSW45*	$(TG)_{18}$	F: TGCTTTGGTATGCAGTGTCC R: AGGAACTACAGTTTTGCATACGA	60	1.4	221	207–225	7	796	0.36	0.77	
GCSW47	$(CAT)_{15}$	F: GGCTTCTCTGGTTGCATGTC R: ACAGTAATCCCCAGCCATCA	60	2.0	233	210–237	9	799	0.68	0.67	0.24
GCSW50	$(CAA)_8A_7$	F: ACCAACCATTAGCTCTGACATTT R: TGTTTTCCATGATGGGATGA	60	2.0	223	198–219	7	788	0.61	0.62	0.21
GCSW51	$(AC)_{13}$	F: GGGATTTGTTCTGCTTGAATG R: AGACTTGGATGGAACTGCAAA	60	2.0	165	158–184	14	799	0.81	0.83	0.49
GCSW55	$(TG)_{14}AG(TG)_8$	F: AGTGATTGCCCCTGTGTGA R: GGGAATATGGTGGTTCTGGA	60	1.4	277	249–303	10	752	0.76	0.75	0.35
GCSW57	$(AC)_{23}$	F: TTGCTGTTCTTCCACACTGC R: TCAGATGGGTGAATTTCTTGG	60	1.0	123	99–149	16	786	0.85	0.89	0.63
GCSW58	$(AC)_{16}C_4(AC)_2C_2(AC)_5$	F: CTGTAGGAGGAAGAACAA R: TACTCCCCATATTTACACA	50	1.0	110	98–147	15	792	0.84	0.84	0.51

*Sex-linked locus (avian Z chromosome).

Recommended maximum annealing temperatures (T_a) and $MgCl_2$ concentrations are listed along with the product size for the repeat motif shown. The allele size range and number of alleles are given for sample sizes of N individuals. Observed (H_O) and expected (H_E) heterozygosity are listed for each locus. The probability of excluding a parent (Ex) is listed for all autosomal loci. Cloned sequences have been deposited with GenBank under accession numbers AY622335–AY622352.

the known offspring. We included all adults that bred in 2001 in all three main colonies (N = 102), and found that only GCSW41 and GCSW45 (the Z-linked locus) showed significant deviation from Hardy–Weinberg. Furthermore, after Bonferroni correction, there was no indication of significant linkage disequilibrium among loci (GENEPOP).

Cross-species amplification of the GCSW markers was successful in five other species, including three other ploceid weavers (Table 2). Each of the loci amplified product in at least two of the species, and many showed polymorphism. As expected, amplification of polymorphic products was most successful in closely-related weaverbirds. At least one locus (GCSW13) appears to be more variable in white-browed sparrow weavers *Plocepasser mahali* than in GCSWs. Sparrow weavers are distributed over a greater range, and are more abundant than GCSWs (Fry & Keith 2004). This may explain the greater variation in this neutral locus (Hartl & Clark 1989; Dallimer 1999).

Before developing the GCSW markers listed in Table 1, we tested whether eight microsatellite markers previously

Table 2 Cross-species amplification of GCSW microsatellite markers

Marker	White-browed sparrow weaver *Plocepasser mahali* *N* = 13	Sociable weaver *Philetarius socius* *N* = 6	Masked weaver *Ploceus velatus* *N* = 10	House sparrow *Passer domesticus* *N* = 3	Cut-throat finch *Amadina fasciata* *N* = 2
GCSW4	10	3	4	2	2
GCSW8	7	4	—	2	—
GCSW10	6	M	2	2	—
GCSW13	10	5	9	2	—
GCSW15	3	4	—	—	M?
GCSW18	2	—	2	—	2
GCSW20	—	3	P	4	—
GCSW31	6	5	P	2	—
GCSW35	9	5	10	M?	—
GCSW39	P	2	P	2	—
GCSW41	2	—	M	—	M
GCSW45	5	M	2	2	—
GCSW47	3	2	7	—	4
GCSW50	3	—	4	—	2
GCSW51	3	M	M	?	M
GCSW55	3	—	M	?	M?
GCSW57	8	8	4	3	2
GCSW58	—	—	5	M	—

The number of alleles amplified among a sample of *N* individuals is indicated. M denotes a monomorphic locus, and P indicates polymorphism where allele number is unclear due to amplification problems, and ? indicates the loci were not tested.

cloned from the white-browed sparrow weaver (McRae & Amos 1999) would amplify in GCSWs. We tested 10 individual GCSWs from four study sites. One Z-linked locus (WBSW10) did not amplify at all in GCSWs. Three autosomal loci appeared to be monomorphic (WBSW2, WBSW8, and WBSW9). The other four showed polymorphism (WBSW1, WBSW4, WBSW7 and WBSW11). However, only one (WBSW4 with four alleles among the GCSW samples) amplified consistently. The other three polymorphic loci showed above average homozygosity levels and a lack of amplification for some individuals, suggesting null alleles. In general, we find that null alleles are a common problem with cross-species amplification. This is of particular concern because we require precise relatedness estimates for our study of extended families. The presence of a significant proportion of null alleles could introduce error in both relatedness estimates and parentage assignment.

In conclusion, the set of 18 loci described here are sufficiently variable (mean allele frequency = 12), and reliable in their amplification (mean proportion of individuals typed across loci = 0.99) to allow the determination of relatedness among GCSWs. For parentage assignment, we estimate this set of loci has a mean exclusionary power (for first parent) of 0.9999. Hence, we are confident that we will be able to discriminate between parental candidates even when they are first-order relatives, a scenario that occurs frequently with extended family units.

Our enriched libraries constructed with a size-selection step produced a significantly higher proportion of microsatellite sequences over 300 bp. We recommend size-selection of fragments just prior to ligation to the cloning vector to yield enriched libraries with larger average insert lengths. We argue that this will improve yield of microsatellite markers by greatly enhancing the probability of having sufficient flanking sequence with which to design primers. Having longer insert sequences further allows greater flexibility to develop loci with fragments of compatible lengths for multiplexing, and to design primers with compatible PCR conditions for combining in multiplex PCR reactions.

Acknowledgements

The GCSW markers were cloned and tested in the Evolutionary Genetics Core Facility of the Department of Ecology and Evolutionary Biology at Cornell University. We are grateful to Irby Lovette for inviting us to genotype in his Evolutionary Biology Laboratory at the Cornell Laboratory of Ornithology. We thank Bernard Amakobe, Bernard Chege, Titus Imboma, Sylvester Karimi, and Wanyoike Wamiti for collecting the GCSW blood, and also Dana Hawley and Mark Anderson for contributing samples. The GCSW project is a collaborative effort of the Cornell University, the National Museums of Kenya (Nairobi) and the Mpala Research Centre. Funding was provided for by National Geographic Society grants #6947–01 and 7249–02, and National Science Foundation grant #IBN-0133795 to STE.

References

Bogdanowicz SM, Mastro VC, Prasher DC, Harrison RG (1997) Microsatellite DNA variation among Asian and North American gypsy moths (Lepidoptera: Lymantriidae). *Annals of the Entomological Society of America*, **90**, 768–775.

Dallimer M (1999) Cross-species amplification success of avian microsatellites in the red-billed quelea, *Quelea quelea*. *Molecular Ecology*, **8**, 685–702.

Fleischer RC, Loew S (1995) Construction and screening of microsatellite-enriched loci from a social lizard, *Egernia stokesii*, using a modified enrichment procedure. *Journal of Heredity*, **90**, 301–304.

Fry CH, Keith S (2004) *The Birds of Africa*, Vol. VII. Princeton University Press, Princeton, NJ.

Griffiths R, Double MC, Orr K, Dawson RJG (1998) A DNA test to sex most birds. *Molecular Ecology*, **7**, 1071–1075.

Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic libraries enriched for microsatellites. *Biotechniques*, **27**, 500–507.

Hartl DL, Clark AG (1989) *Principles of Population Genetics*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *Biotechniques*, **16**, 656.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

McRae SB, Amos W (1999) Characterization of hypervariable microsatellites in the cooperatively breeding white-browed sparrow weaver *Plocepasser mahali*. *Molecular Ecology*, **8**, 895–906.

Raymond M, Rousset F (1995) GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring. Harbor Laboratory Press, New York.

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissues samples for DNA analysis. *Canadian Journal of Zoology*, **69**, 82–90.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**, 1–16.