TECHNICAL NOTE (1003 words)

Identification and characterisation of ten microsatellite loci in the Noisy Scrub-bird *Atrichornis clamosus* using next-generation sequencing technology.

S. J. Cowen, R. J. N. Allcock, S. J. Comer, J. D. Wetherall and D. M. Groth

S. J. Cowen • J. D. Wetherall • D. M. Groth
Curtin Health Innovation Research Institute,
School of Biomedical Sciences, Curtin University,
GPO Box U1987,
Perth, WA 6845, Australia

R. J. N. Allcock
LotteryWest State Biomedical Facility - Genomics,
School of Pathology and Laboratory Medicine, University of Western Australia
Stirling Highway, Nedlands
Perth, WA 6000, Australia
and
Department of Clinical Immunology and Immunogenetics,
Royal Perth Hospital,
Perth, WA 6847, Australia

S. J. Comer
Department of Environment and Conservation,
120 Albany Highway,
Albany, WA 6330, Australia

Present Address:

S. J. Cowen
Department of Environment and Conservation,
120 Albany Highway,
Albany, WA 6330, Australia

Corresponding author: D.M. Groth email: D.Groth@curtin.edu.au

Keywords: next-generation sequencing • short-tandem repeats • microsatellites

Abstract

The Noisy Scrub-bird is an endangered species of songbird endemic to the south coast of Western Australia that has undergone a major and prolonged population bottleneck. Using shotgun 454 next-generation DNA sequencing we have identified and characterised ten polymorphic microsatellite loci in this species. Observed allelic diversity was relatively low (2-5 alleles per locus) and significant deviations from

Hardy-Weinberg Equilibrium observed, although the presence of null alleles was only postulated for two loci. The microsatellite loci characterised in this study will be useful in a future population genetics studies in this endangered species.

The Noisy Scrub-bird (NSB) (*Atrichornis clamosus* Gould, 1844) is a small semi-flightless passerine (Danks et al. 1996). The Action Plan for Australian Birds 2010 (Garnett et al. 2011) lists it as Endangered (B1ab(iii,v)+2ab(iii,v)) under IUCN Red List criteria (IUCN 2011). Discovered in 1842 in the Darling Range, south of Perth, and was found widespread throughout the south-west of WA. However, by the end of the 19th century, the NSB was designated "extinct" (Danks et al. 1996). In 1961, NSB was rediscovered at Mount Gardner (MG) in southwest WA. This severe contraction of range and population poses a genetic bottleneck risk. Hence, considerable importance in the ascertainment of levels of genetic variation and inbreeding in this species is important for future management strategies.

Microsatellites are short tandem repeating DNA which are frequently used as markers in population studies (Goldstein and Schlötterer 1999). However, they are relatively less abundant in the avian genome (Primmer et al. 1997). A previous study (McLoughlin, unpublished) used the screening of a small random shotgun fragment library method (Taylor et al, 1999) to identify four NSB di-nucleotide microsatellites and suggested that some genetic variability remained in the population. Enrichment techniques using oligonucleotide capture hybridisation can also identify microsatellite loci (Munyard et al. 2009). However, these methods are relatively time consuming and expensive (Saarinen and Austin 2010). The use of NGS for population genetics studies is increasing, particularly as it is able to generate considerable sequence information relatively economically (Csencsics et al. 2010).

DNA was extracted from blood samples using a DNeasy Kit (QIAGEN). DNA was also extracted from one feather sample (99M07) using the XytXtract – Animal (http://www.xytogen.com.au) (Castalanelli et al. 2010). One sample (male: NSB 01M04) was sequenced using a Roche 454 Genome Sequencer FLX at the Lotterywest State Biomedical Facility: Genomics unit (Royal Perth Hospital, Perth, Western Australia).

454 sequencing data was screened for tetra-nucleotide repeats (>4 repeats) using MSATcommander version 0.8.2 (Faircloth 2008). Primer design was performed using PRIMER3 version 0.4.0 (Rozen and Skaletsky 2000), screened for self-

complementarity, self-dimerization and hair-pin loops using Oligocalc (Kibbe 2007) and synthesised by Geneworks (Adelaide Australia).

PCR amplification was performed using 20 ng of genomic DNA, 0.05µl Bioline MyTaq DNA Polymerase (5U/µl) and 19.95µl Bioline MyTaq Reaction Buffer master mix containing 0.75-1µM each of forward and reverse primer using an Eppendorf thermocycler with parameters set at: denaturing 95°C, 5 minutes; 1 cycle of 95°C, 1 minute, 55°C,10 seconds and 72°C, 2 minutes; 15 cycles of 94°C, 20 seconds, 55°C, 10 seconds, 72°C, 1 minute; 23 cycles of 92°C, 20 seconds, 56°C,10 seconds, 72°C, 1 minute with a final extension at 72°C for 10 minutes. PCR was optimised by altering the annealing temperature, with the parameters varying from 54°C/55°C to 59°C/60°C for individual primer sets.

Fragment analysis was performed on a CEQ 8000 (Beckman Coulter, USA) in combination with the FRAG-4 program. Dye-labelled (Well-RED (Sigma-Aldrich)) forward (0.25µM) and non-labelled reverse primers (0.5µM) were used to amplify each locus using parameters previously described. The loading solution contained DNA Size Standard 400 (Beckman Coulter, USA) was added to the PCR products (39:1). Results were assessed for errors using the proprietary CEQ 8000 software (Beckman Coulter, USA). Deviation from Hardy-Weinberg equilibrium, linkage disequilibrium and observed heterozygosity (H_O) vs. expected heterozygosity (H_E) was determined using GENEPOP version 4.0.10 (Raymond and Rousset 1995). Loci were assessed for null alleles using Micro-checker 2.2.3 (van Oosterhout et al. 2004). Microsatellite sequences were screened for matches in the NCBI database using BLAST (Altschul et al. 1997) (http://blast.ncbi.nlm.nih.gov/).

The NGS yielded approximately 1,090,451 reads (average length of 278bp). Approximately 1450 tetra-nucleotide (>4 repeats), 2625 di-nucleotide (>7 repeats) and 989 tri-nucleotide (>7 repeats) microsatellites were identified (Table 1). The density in the NSB genome is similar to that estimated for avian genomes by Primmer et al. (1997). Thirty-seven, tetra-nucleotide repeat loci were selected for primer design, with 24 primer pairs evaluated by PCR and 10 pairs ultimately used. The 10 loci had 33 alleles with two loci being bi-allelic (shown in Table 2).

Genetic diversity was determined using 22 individuals from the MG subpopulation. Some loci had deviations from both HWE ($p < 0.0001$) and global tests for homozygosity ($p < 0.0001$). NSBSC1 and NSBSC13 were identified as having possible null alleles ($p < 0.005$). Locus NSBSC19 also contained possible null alleles but this was not consistent in multiple tests. When these loci are excluded from HWE exact tests, the population deviation from HWE is still significant ($p < 0.05$). Significant linkage disequilibrium was observed for NSBSC1 & NSBSC7 ($p < 0.005$). BLAST comparisons showed that NSBSC13 had similarity to lysyl oxidase 4 (LOX4) gene on chromosome 10 of the chicken (*Gallus gallus*) genome.

The 10 microsatellite loci identified and characterized using NGS technology in this study, show relatively low allelic diversity (2-5 alleles/locus) (Table 2), which is comparable with other bird species having undergone significant genetic bottlenecks (Tarr et al. 1998; Chan et al. 2011). An excess of homozygotes observed in this study may simply reflect the rapid population contraction and presumed extinction of the NSB in the late 1800s (Danks et al. 1996). Further analysis of these microsatellites in other NSB populations will test for the presence of null alleles at these loci. Heterozygote excess was observed for two loci (NSBSC11 & NSBSC22). However, differential allele frequencies in males and females in small randomly-mating populations can be due to binomial sampling error (Allendorf and Luikart 2007). This study has confirmed the utility of NGS as a tool for relatively quick and cost-effective marker generation.

Acknowledgments The authors would like to thank the Western Australian Department of Environment and Conservation's (DEC) South Coast Threatened Birds Recovery Team, the Lotterywest State Biomedical Facility Genomics for the use of the their 454 Life Sciences (Roche) FLX Sequencer and Birdlife Australia (formerly Birds Australia) for their support through the Stuart Leslie (SLBRA) and Professor Allan Keast (PAKRA) Bird Research Awards, which was invaluable to the completion of this study. We also thank the Guild of Mercers' Scholars Charitable Trust for their contribution towards field equipment.

References

Allendorf FW, Luikart G (2007) Conservation and the Genetics of Populations. Blackwell Publishing, Malden Massachusetts

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Castalanelli MA, Severtson DL, Brumley CJ, Szito A, Foottit RG, Grimm M, Munyard K, Groth DM (2010) A Rapid non-Destructive DNA extraction method for insects and other arthropods. J Asia Pacific Entomol 13:243-248

Chan C-H, Robertson HA, Saul EK, Nia LV, Phuong Luong V, Kong X, Zhao Y, Chambers GK (2011) Genetic variation in the kakerori (*Pomarea dimidiata*) an endangered endemic bird successfully recovering in the Cook Islands. Conserv Genet 12:441-447

Csencsics D, Brodbeck S, Holderegger R (2010) Cost-Effective, Species-Specific Microsatellite Development for the Endangered Dwarf Bulrush (*Typha minima*) Using Next-Generation Sequencing Technology. J Hered 101:789-793

Danks A, Burbidge AA, Burbidge AH, Smith GT (1996) Noisy Scrub-bird Recovery Plan. WA Department of Conservation and Land Management, Perth

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol Resour 8:92-94

Garnett ST, Szabo JK, Dutson G (2011) The action plan for Australian birds 2010. Birds Australia. CSIRO Publishing, Collingwood

Goldstein DB, Schlötterer C (eds) (1999) Microsatellites: evolution and applications. Oxford University Press, Oxford

IUCN, (2012) IUCN Red List of Threatened Species. Version 2012.1. http://www.iucnredlist.org. Accessed on 24 July 2012

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230-241

Kibbe WA (2007) OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35(suppl. 2):W43-W46

Munyard KA, Ledger JM, Lee CY, Babra C, Groth DM (2009) Characterization and multiplex genotyping of alpaca tetranucleotide microsatellite markers. Small Ruminant Res 85:153-156

Primmer CR, Raudsepp T, Chowdhary BP, Pape Moller A, Ellegren H (1997) Low Frequency of Microsatellites in the Avian Genome. Genome Res 7:471-482

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

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

Taylor, EL, Vercoe, P, Cockrem, J, Groth, D, Wetherall, JD, Martin, GB (1999) Isolation and characterization of microsatellite loci in the emu, *Dromaius novaehollandiae*, and cross-species amplification within Ratitae. Mol Ecol 8: 1963-1964

Tarr CL, Conant S, Fleischer RC (1998) Founder events and variation at microsatellite loci in an insular passerine bird, the Laysan Finch (*Telespiza cantans*). Mol Ecol 7:719-731

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

Walker JM, Rapley R (eds) (2009) Molecular Biology and Biotechnology, Royal Society of Chemistry, Cambridge, UK

Table 1 Search results from MSATcommander (version 0.8.2 (Faircloth 2008)) for di-, tri- and tetra-nucleotide microsatellite from ~1.1 million sequence reads yielded from NGS on Roche 454 Genome Sequencer FLX

Repeat Type	Minimum no. repeats	No. of sequence reads	Density in NSB genome *
Di-nucleotide	8	2625	~1 per 116Kbp
Tri-nucleotide	8	989	~1 per 309Kbp
Tetra-nucleotide	5	1450	~1 per 210Kbp

* density in genome estimated from (total sequence reads/microsatelite sequence reads) x average read length

Table 2 Primer sequences, repeat motifs and levels of diversity of 10 microsatellites characterized from the noisy scrub-bird (*Atrichornis clamosus*) from Two Peoples Bay Nature Reserve, Western Australia (n = 22)

Locus	GenBank Accession No.	Repeat Motif	Primer sequence (5' to 3')	No. alleles	Allele size (bp)	H_O	H_E	p
NSBSC1	KC355442	$(GACA)_9$	F - ACACAGCTTCCTTCCAAACAA	3	288-296	0.364	0.664	<0.001
			R - GTGTGCGAGCTTGGCTGAT					
NSBSC7	KC355443	$(GACA)_7$	F - CAGCAAGTCCACAAGGGTTT	3	178-186	0.545	0.664	0.026
			R - AAGGTGCTAGGGGCTGATGT					
NSBSC8	KC355444	$(AGGG)_7$	F - ACAGGAGCACCTCCAGCTC	4	194-214	0.500	0.709	0.131
			R - CTGCTCACCAGATGCTTCAA					
NSBSC11	KC355445	$(TAAA)_5$	F - GCTTTGCAAAACACCCATTT	2	226-230	0.545	0.405	0.143
			R - TTCCTTCCTGTGGGTACCTG					
NSBSC13	KC355446	$(AGGG)_6$	F - TCCCGGGAATTTAATCCTGT	4	144-160	0.182	0.441	0.002
			R - GTCTGTCCATCCCATCCATC					
NSBSC14	KC355447	$(AGAT)_6$	F - CCAGACAGGGACACCAACTC	5	148-172	0.727	0.729	0.961
			R - ACGGGCAGTAGTGTGAGACC					
NSBSC17	KC355448	$(GGAT)_{10}$	F - CAAACTCAGTGCATGAGGAGA	3	154-162	0.227	0.282	0.427
			R - CCATAAATTGCTAAAATACACCATTG					
NSBSC19	KC355449	$(GGAT)_8$	F - TGCTCCTGACTCTTCCTTCC	2	170-186	0.182	0.445	0.009
			R - TGTCCGATTCCGTCCATC					
NSBSC21	KC355450	$(ATCC)_9$	F - TTCAGATAGACCCGGAGAGC	4	244-256	0.500	0.609	0.204
			R - GCACATCACTGCCTGAAATC					
NSBSC22	KC355451	$(ATCC)_{10}$	F - TGCAGTGGGTGTGTAAGCTC	3	212-224	0.318	0.282	1.000
			R - AACAAATCAGGAGCCTGTGG					

bp length in base-pairs, H_E expected heterozygosity, H_O observed heterozygosity, p probability value for departure from Hardy-Weinberg Equilibrium