Low levels of polymorphism at novel microsatellite loci developed for bathyergid mole-rats from South Africa

T.C. BRAY[a], N.C. BENNETT[b], and P. BLOOMER[a]

[a]Molecular Ecology and Evolution Programme, Department of Genetics, University of Pretoria, Pretoria, 0002, South Africa.

[b]Department of Zoology and Entomology, University of Pretoria, Pretoria, 0002, South Africa

Correspondence: T.C. Bray, Fax: +27 12 362 5327 E-mail: s29582832@tuks.co.za

Keywords: African mole-rats, Bathyergidae, *Bathyergus suillus*, cross-species, microsatellite

Summary

Eight microsatellite markers were developed for African mole-rats using the Fast Isolation by AFLP of Sequences Containing repeats (FIASCO) protocol and pyrosequencing. The markers were developed in *Bathyergus suillus* and applied to a selection of individuals from seven related taxa: *Bathyergus janetta*, *Cryptomys hottentotus hottentotus*, *C. h. pretoriae*, *Fukomys damarensis*, *F. darlingi*, *F. mechowii* and *Georychus capensis*. The markers displayed low to moderate variation with allele numbers ranging between one and six per species. We propose that larger repeat numbers at di-, tri- and tetranucleotide repeat loci generally yield higher levels of polymorphism.

The family Bathyergidae represents an endemic African radiation of subterranean rodents (Jarvis & Bennett, 1990). Mole-rat ecology is highly variable across species, varying with mode of social organisation and habitat preference (Bennett *et al.* 2000). There are still relatively few published species-specific molecular markers available for African mole-rats that may be used at a population level (Burland *et al.* 2001). As a consequence only a handful of population studies have been published to date (Reeve *et al.* 1990; Faulkes *et al.* 1997; Burland *et al.* 2002, 2004; Bishop *et al.* 2004; Patzenhauerova *et al.* 2010).

Bathyergus suillus is a solitary species and the largest of the southern African mole-rats, endemic to the Western Cape of South Africa (Skinner & Chimimba, 2005). Such solitary species have been largely neglected despite outnumbering their social cousins (Bennett *et al.* 2000).

DNA was extracted from a single *B. suillus* individual using the Qiagen tissue kit according to the manufacturer's instructions. Enrichment was performed using the Fast Isolation by AFLP of Sequences Containing repeats (FIASCO) protocol (Zane *et al.* 2002). Probes were chosen from among the most common repeats found in the rodent genome (Toth *et al.* 2007) and incorporated into two probe mixes: $AGG_{(6)}$, $ATA_{(6)}$, $CAA_{(6)}$ and $AAGG_{(6)}$, $AAAT_{(6)}$, $AGCA_{(6)}$. Mono- and di-nucleotide repeats were avoided due to increased scoring inaccuracies in these shorter repeat motifs as a result of point indels (i.e. Messier *et al.* 1996). Both enriched genomic DNA pools were combined and run on a Roche 454 GS-FLX platform using up to 5 µg of DNA. Sample preparation and analysis of run data were performed at Inqaba Biotechnical Industries (Pretoria, Gauteng, South Africa) using the manufacturer's protocol.

Pyrosequencing resulted in the generation of ~4400 and ~3700 reads of 40-500 base pairs with the tri- and tetra-nucleotide probe mixes respectively. The MSATCOMMANDER application (Faircloth 2008) was used to identify all di-to tetranucleotide repeats from the sequence data. In total 37 percent of the tri-nucleotide enrichment reads contained repeats of which approximately half corresponded to those probe motifs used (95% AGG). Of the tetra-nucleotide enrichment reads, 60 percent contained repeats (98% AAGG). Higher repeat numbers were chosen in order to isolate more polymorphic loci (Wierdl *et al.* 1997). 1998). Fifty loci were chosen for PCR amplification, optimisation, and polymorphism testing using primers designed in PRIMER3 (Rozen & Skaletsky. Of these loci, 31 failed to amplify consistently and 11 of the remaining loci were monomorphic in *B. suillus*. Those markers amplifying successfully in *B. suillus* with two or more alleles were subsequently applied to the seven other taxa: *B. janetta*, *Cryptomys hottentotus hottentotus*, *C. h. pretoriae*, *Fukomys damarensis*, *F. darlingi*, *F. mechowii* and *Georychus capensis*. The remaining eight polymorphic markers were applied in two multiplexes using the Qiagen Multiplex kit with standard conditions (45 cycles, 60˚C annealing temperature). Heterozygosity scores and linkage disequilibrium for all locus pairs were calculated in GENETIX4.05 (Belkhir *et al.* 2004). All loci were tested for conformation to Hardy Weinberg Equilibrium and for the presence of null alleles using GENEPOPv4.0.10 (Rousset 2008). GenBank accession numbers for the sequence of each locus in *B. suillus* are HQ186238-HQ186245 respectively.

Locus information and polymorphism in *B. suillus* is summarised in Table 1, differential amplification across taxa is detailed in Table 2. Only locus BS08 deviated from Hardy Weinberg Equilibrium in *B. suillus*, but after separating the two population samples this was no longer significant. Potentially significant frequencies of null alleles (i.e. ≥0.2 ; Dakin & Avise, 2004) were detected across three of the loci; BS03 (0.32 in *C. h. hottentotus*), BS04 (0.23 in *F. damarensis*), and BS05 (0.24 in *F. mechowi*). Low numbers of alleles in these small sample groups may have artificially raised the estimates for null alleles in this case. Two locus pairs were found to be at linkage disequilibrium according to Weir's R (1979) ($p>0.01$); BS01-BS03 and BS01-BS07. However, the best matches from a sequence comparison (Altschul *et al.* 1997) on the National Center for Biotechnology Information (NCBI) genomic database suggest that BS01 and BS07 are found on different *Mus musculus* chromosomes (10 and 13 respectively, Appendix 1). No significant matches were found for locus BS03. Correct fragment amplification and allele scoring was confirmed for up to three homozygotes in all loci except BS02 where the repeat region failed to sequence through completely (data not shown).

Poor amplification success and low numbers of alleles was also seen in *B. suillus* in a study by Burland *et al.* (2001), and may be characteristic of this species. Private alleles were found between the two *B. suillus* populations typed (data not shown) which provides evidence that these loci are of utility in the study of genetic variation within *B. suillus*. It is notable that one of the *B. suillus* populations used here is likely to have originated from few colonisation events predisposing it to low levels of genetic variation. Microsatellite loci with lower allele numbers have the associated advantage of low levels of homoplasy which has been shown to be a problem with more variable markers (Estoup *et al.* 2002).

We confirm that large numbers of repeat containing sequences can be generated using the pyrosequencing approach but we emphasise the need for careful consideration of repeat motifs chosen. Despite the choice to avoid di-nucleotide repeats here, two of the eight microsatellites described are, by necessity, of such a repeat class. This restates the potential for higher polymorphism of this shorter repeat class and recommend that it should be considered if low genetic variation is suspected.

Acknowledgements
The authors are grateful to the Department of Science and Technology and the National Research Foundation for funding from the SARCHi Chair for Mammal Behavioural Ecology and Physiology awarded to NCB. Hannah Thomas is thanked for providing sample tissues from animals. Funding was also provided by the University of Pretoria in provision of TCB's postdoctoral fellowship. This work is based upon research supported by the National Research Foundation.

References

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W. & Lipman D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Research* **25**, 3389-3402.

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (2004) GENETIX 4.05, Logiciel Sous Windows^{TM} Pour la Génétique Des Populations. Laboratoire Génome Populations, Interactions. CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Bennett N.C. & Faulkes C.G. (2000) *African Mole-rats ecology and eusociality*. Cambridge University Press, United Kingdom.

Bishop J.M., Jarvis J.U.M., Spinks A.C., Bennett N.C. & O'Ryan C (2004) Molecular insight into patterns of colony composition and paternity in the common mole-rat *Cryptomys hottentotus hottentotus. Molecular Ecology* **13**, 1217-1229.

Burland T.M., Bennett N.C., Jarvis J.U.M. & Faulkes C.G. (2004) Colony structure and parentage in wild colonies of cooperatively breeding Damaraland mole-rats suggest incest avoidance alone may not maintain reproductive skew. *Molecular Ecology* **13**, 2371-2379.

Burland T.M., Bennett NC., Jarvis J.U.M. & Faulkes C. (2002) Eusociality in African mole-rats: new insights from patterns of genetic relatedness in the Damaraland mole-rat (*Cryptomys damarensis*). *Proceedings of the Royal Society of London series B*, **269**, 1025-1030.

Burland T.M., Bishop J.M., Ooryana C. & Faulkes C.G. (2001) Microsatellite primers for the African mole-rat genus *Cryptomy*s and cross-species amplification within the family Bathyergidae. *Molecular Ecology Notes* **1**, 311-314.

Dakin E.E. & Avise J.C. (2004) Microsatellite null alleles in parentage analysis. *Heredity* **93**, 504-509.

Estoup A., Jarne P. & Cornuet J-M. (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591-1604.

Faircloth B.C. (2008) msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, **8**, 92-94.

Faulkes C.G., Abbott D.H., O'Brien G.H., Lau L., Roy M.R., Wayne R.K., & Bruford M.W. (1997) Micro- and macrogeographical genetical structure of colonies of naked mole-rats *Heterocephalus glaber*. *Molecular Ecology*, **6**, 615-628.

Jarvis J.U.M., & Bennett N.C. (1990) The evolutionary history, population biology and social structure of African mole-rats: Family Bathyergidae. In *Evolution of subterranean mammals at the organismal and molecular levels* (eds. E. Nevo and O.A.Reig), pp. 97-128, Wiley Liss, New York.

Messier,W., Li S.H. & Stewart C.B. (1996) The birth of microsatellites. *Nature*, **381**, 483.

Patzenhauerova H., Bryja J. & Sumbera R. (2010) Kinship structure and mating system in a solitary subterranean rodent, the silvery mole-rat. *Behavioural Ecology and Sociobiology*, **64**, 757-767.

Reeve H.K., Westneat D.F., Noon W.A., Sherman P.W. & Aquadro C.F. (1990) DNA "fingerprinting" reveals high levels of inbreeding in colonies of the eusocial naked mole-rat. *Proceedings of the National Academy of Science of the USA*, **87**, 2496-2500.

Rousset F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8** 103 - 106

Rozen S. & Skaletsky H.J. (1998) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Skinner J.D. & Chimimba C.T. (2005) 3rd edition: *Mammals of the southern African subregion*. Cambridge University Press.

Toth G., Gaspari Z. & Jurka J. (2007) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*, 967-981.

Weir B.S. (1979) Inferences about linkage disequilibrium. *Biometrics* **35**, 235-254.

Wierdl M., Dominska M. & Petes T.D. (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*.**146**(3), 769-79.

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

Table 1 Details of the eight microsatellite loci identified and typed in *Bathyergus suillus.*

Locus	Primer sequence (5'-3') (Forward primer sequence first)	Repeat motif	No. of alleles	Allele size range (bp)	H_O/ H_E
BS01	AAA CAG GAG GCC AGG AAT TT CAG GTT TCC AAG TTT TGT GGA	$(AAC)_4AAT(AAC)_2$	3	167-176	0.66/0.59
BS02	GAG TGC GTA GGG GTG AAG G GAG TCC TGA GTA ACT TAT TAG ATG ATG	$(AAAT)_6$	3	143-151	0.62/0.49
BS03	ACG AGT GCG TAT AGA GGA AAG TGT TGT CAC CGT AAG AGA GG	$(AGG)_5$	3	78-84	0.23/0.21
BS04	TGC CAG GAG GAA GAC AAG G ACT CGT TAG GCT GAG CTG	$(AGG)_6ACG(AGG)_2$	2	238-241	0.11/0.11
BS05	GGC CGG ATG CTT AAG GCT GG ATG TAC AGG TAC AGG TTC CC	$(TG)_4TT(TG)_5$	3	162-170	0.09/0.18
BS06	TTGAAGGTGGGGAGGGTAG TTGGGGTTATTGATTTTATTGG	$(AGG)_8$	5	164-176	0.45/0.47
BS07	TTCTCACCTGACTCCCAAGC GAGTATGCAAGGCCAGCAGT	$(CT)_9$	4	245-257	0.60/0.72
BS08	GGA AGG GTG GGA GAA AGG GAA ACT ATA CCC AAA CCA AAG C	$(AAGG)_6$	2	147-151	0.03/0.08

Table 2. Amplification success and levels of polymorphism seen across eight species of African mole-rat for the eight microsatellite markers identified in this study. For each locus the details included per species are as follows: allele number, allele size range, and proportion of individuals in which amplification was observed in parentheses. '-' denotes no peak, or multiple peaks.

Species \ Locus	BS01	BS02	BS03	BS04	BS05	BS06	BS07	BS08
Bathyergus suillus (34)	**3**	**3**	**3**	**2**	**3**	**5**	**4**	**2**
	167-176 (34)	143-151 (34)	78-84 (34)	238-241 (34)	162-170 (34)	164-176 (32)	245-257 (34)	147-151 (33)
B. janetta (3)	**3**	**1**	**1**	**2**	**3**	**2**	**2**	**-**
	167-179 (3)	151 (3)	84 (3)	238-241 (3)	156-168 (3)	164-167 (3)	245-249 (3)	
Cryptomys h. hottentotus (10)	**-**	**1**	**2**	**1**	**6**	**2**	**2**	**-**
		151 (4)	84-90 (10)	238 (7)	156-174 (10)	164-167 (9)	245-249 (10)	
Cryptomys h. pretoriae (10)	**-**	**2**	**1**	**2**	**1**	**1**	**1**	**-**
		151-159 (9)	84 (10)	229-238 (10)	154 (10)	164 (10)	245 (7)	
Fukomys damarensis (10)	**-**	**2**	**1**	**2**	**5**	**1**	**3**	**-**
		147-151 (8)	84 (8)	232-238 (9)	164-186 (9)	164-167 (7)	245-253 (7)	
Fukomys darlingi (10)	**-**	**2**	**2**	**1**	**5**	**2**	**2**	**-**
		147-151 (10)	81-84 (10)	232-238 (10)	170-182 (10)	167-170 (10)	245-249 (10)	
Fukomys mechowi (8)	**-**	**1**	**2**	**2**	**6**	**1**	**1**	**-**
		151 (7)	81-84 (8)	232-238 (7)	160-170 (8)	167 (7)	245 (3)	
Georychus capensis (10)	**-**	**3**	**2**	**1**	**4**	**2**	**3**	**-**
		143-151 (10)	81-84 (9)	229 (10)	166-178 (10)	164-167 (10)	245-253 (5)	

Appendices

Appendix 1 Basic Local Alignment Search Tool (BLAST) results for each locus when compared with the *Mus musculus* genome including identity, positional information and reference number. All results for the highest identity score are included for each locus.

Locus	Identity(%)	Chromosome	Position	Reference
BS01	96	10	26468027-26468057	NT_039492.7
BS01	96	10	11557068-11557098	NW_001030413.1
BS02	84	1	17076522- 17076569	NT_039185.7
BS03	-	-	-	
BS04	75	1	51025395-51025330	NT_078297.6
BS04	75	1	50530936- 50530871	NW_001030662.1
BS05	80	2	71910898-71910847	NT_039207.7
BS06	78	1	36056144-36056203	NW_001030662.1
BS07	72	13	13019829-13019723	NT_039590.7
BS07	72	13	1442086-1441980	NW_001030537.1
BS08	93	17	66345969- 66345939	NT_039649.7
BS08	93	17	23425282-23425252	NW_001030622.1