Characterization of 14 polymorphic microsatellite loci developed for an Afrotherian species endemic to southern Africa, *Elephantulus myurus* (Macroscelidea: Macroscelididae)

Sasha Hoffmann[1*], Heather Joy Webster[2*], Peter R. Teske[2], Heike Lutermann[1], Nigel C. Bennett[1] & Bettine Jansen van Vuuren[2#]

[1]Department of Zoology and Entomology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa,

[2]Molecular Zoology Laboratory, Department of Zoology, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa

*Both authors contributed equally to the work

[#]Corresponding author:

Bettine van Vuuren

Department of Zoology

University of Johannesburg

Auckland Park 2006

South Africa

bettinevv@uj.ac.za

Abstract

Fourteen microsatellite loci were developed for the eastern rock sengi, *Elephantulus myurus* Thomas & Schwann, 1906 by incorporating genetic diversity from across its range in South Africa. Sengis are small mammals belonging to the order Macroscelidea, which comprises 19 species, all of which are endemic to Africa. The loci were amplified in 66 individuals from six localities. An average of 10.5 alleles per locus were identified, with observed and expected heterozygosity values ranging from 0.081 – 0.909 and 0.484 – 0.885, respectively.

We also investigated cross-species amplification within the family and found variation in amplification success for five different species. The preliminary results from these amplification efforts could aid further studies into aspects of species diversity and biology. The markers described here represent the first set of variable nuclear markers for the genus *Elephantulus*, and, together with a set of 8 recently developed markers for *Rhynchocyon petersi,* Bocage 1880, the first markers for the Order Macroscelidea.

Keywords: SSR; Multiplex; 454-Sequencing; Macroscelidea; *Elephantulus*; Afrotheria

Introduction

Habitat fragmentation, harvesting by humans, adaptation to captivity and climate change are just a few of the factors which threaten conservation efforts. The danger of ineffective management strategies lies in not being able to effectively determine the extent to which populations are threatened or affected (Araki et al. 2007; Frankham 2010; Giger et al. 2008). The incorporation of genetic information vastly increases the power and accuracy with which important trends, such as inbreeding, dispersal and kin relationships, can be estimated (Allendorf et al. 2010; Hedrick 2001).

Habitat specialists or widely distributed species represent good models to study the impact of environmental change on population ecology and trends over time (Krauss et al. 2003). To this end, our study model is the eastern rock elephant shrew, or sengi, *Elephantulus myurus* Thomas & Schwann, 1906. This species, which is a member of the superorder Afrotheria (Order Macroscelidea), is endemic to Africa and occurs from eastern Mozambique southwards through parts of Zimbabwe, Botswana, Swaziland, Lesotho to South Africa (Skinner and Chimimba 2005). To date, the focus of the molecular work done on sengis has been on evolutionary relationships among taxa (both within the Afrotheria and within

Macroscelidea) (see e.g. Rathbun 2009; Smit et al. 2008, 2011) but limited information is available on intraspecific patterns (but see Smit et al. 2007, 2010). In addition, most studies have employed genetic markers that provide little information for intraspecific studies, including mitochondrial DNA and a few nuclear introns that were characterized by low variation.

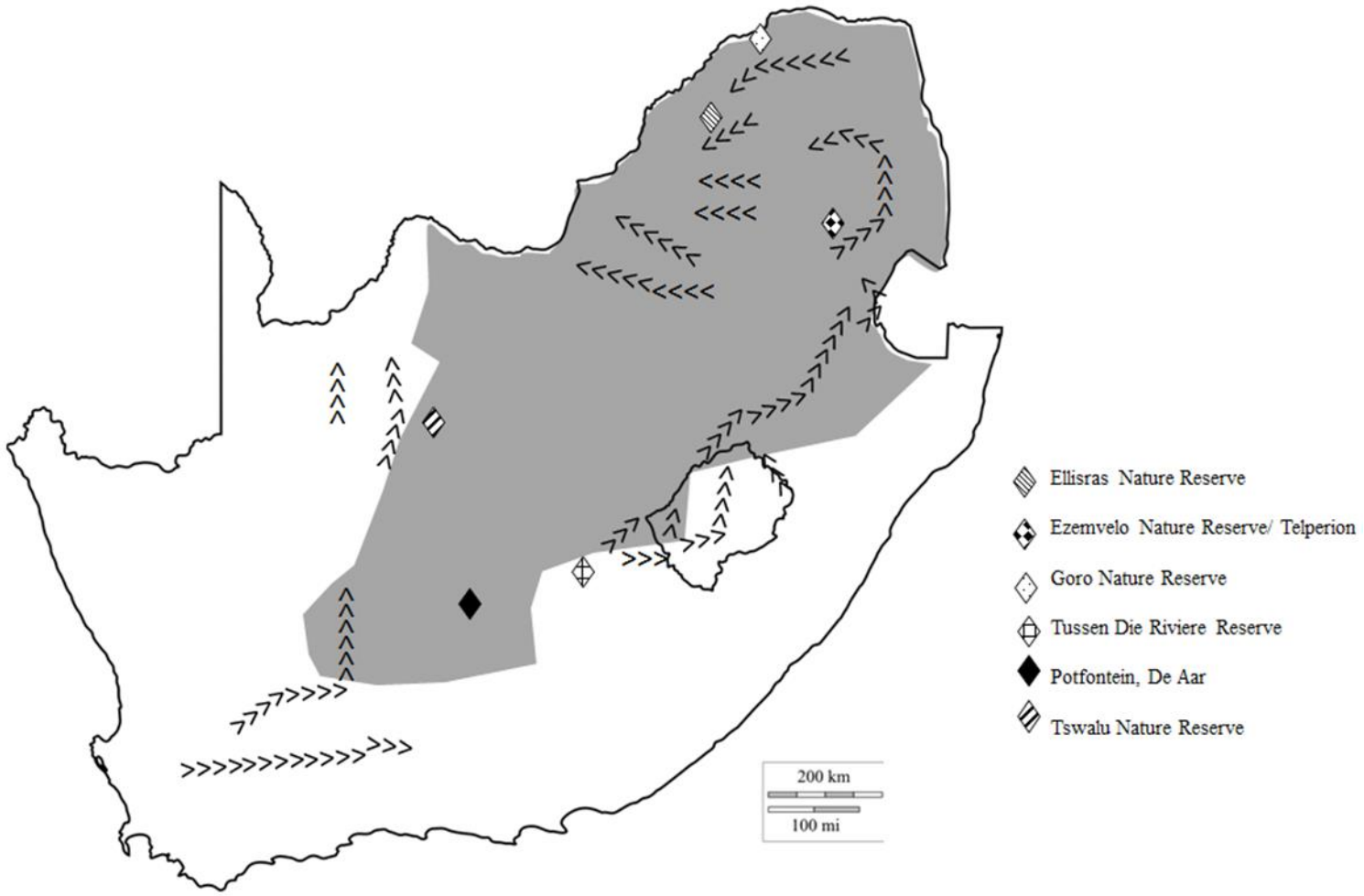
Rock sengis are considered highly territorial (Ribble and Perrin 2005) and their specific habitat preference often results in isolation of populations and possibly even individuals due to geographical barriers (see Smit et al. 2007). Such a pattern of isolation increases conservation priorities as one typically finds unique gene pools across their distributional range and it is important to also consider local adaptations (Kawecki and Ebert 2004). This is potentially true in our model species as specimens were, amongst other areas, sampled within the Vhembe biosphere reserve, a biodiversity hotspot experiencing increased threats due to local pressures from vulnerable people, a problem common to developing countries (Kasperson and Archer 2005; Pool-Stanvliet 2013; Steyn et al. 2010). Microsatellite markers make it possible to identify to the individual level which can help researchers compare genetic diversity at multiple scales, increasing the accuracy of observed trends (Allendorf et al. 2010; Bruford and Wayne 1993). The development of a suite of highly variable nuclear makers represents a significant way forward in studying intraspecific variation and familial relationships (Freeland et al. 2011). Here, we report on the development of a panel of 14 highly polymorphic microsatellite markers isolated from *E. myurus*.

Materials and methods

Tissue Acquisition and DNA protocol

Elephantulus myurus specimens were sampled between 2008 and 2016 from six different localities across the central to northern regions of South Africa (Figure 1). The number of

Fig. 1 A map showing the distribution of *Elephantulus myurus* across South Africa (in grey; major mountain ranges are shown by >>>>). The six sampling localities are indicated.



specimens per locality ranged from 2-8 (Lephalale -2; Potfontein – 2; Tswalu – 4; Tussen Die Riviere Reserve – 8) with the exception of Goro Game Reserve (n = 25) and Telperion/Ezemvelo Nature Reserve (n = 25). Total genomic DNA was extracted from tissue samples (ranging from ear clips to organ tissue samples; stored at -20°C either in absolute ethanol or at room temperature in DMSO salt-saturated solution) using the standard protocol suggested by the DNeasy Blood & Tissue Kit (Qiagen®, Whitehead Scientific (Pty) Ltd, Cape Town). DNA was stored at -20°C.

Library Development

DNA from 16 individuals, representing different sampling localities, were pooled for library construction done through Genoscreen© (Lille, France). One µg of DNA was used for the 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries (see Malausa et al. 2011 for detailed protocols). Eight probes (TG, TC, AAC, AAG, AGG, ACG, ACAT and ACTC) were used to produce an enriched DNA library of microsatellite motifs. The resulting sequences were manually analysed using BioEdit software (Hall 1999) to detect microsatellite motifs for PCR amplification and primer design. A total of 6943 sequences were examined of which 807 sequences containing microsatellite motifs; 40 loci were initially selected for optimization.

Microsatellite sequencing, validation and multiplex design

Primers were used in a working solution of 10 µM each. PCR reactions were performed in 10 µl reaction volume containing 5 µl QIAGEN Multiplex PCR Master Mix (2x), 1 µl of Q solution, 0.5 µl of forward and reverse primers, 1 µl of ddH_20 and 2 µl of DNA. PCR conditions consisted of a 15 minute initial denaturation at 95°C followed by 25 cycles of denaturation for 30 seconds at 94°C, annealing for 90 seconds at primer-specific annealing

temperatures (Online Resource 1) and strand extension for 50 seconds at 72°C. A final extension step for 30 minutes at 72°C completed the reactions.

Putative loci were initially screened for amplification on six DNA samples in simplex PCR reactions. Thirty-one primer pairs were validated and sent for a high sensitivity DNA assays and loci characterized by amplification failure, amplification of multiple bands, non-specific binding or noise, and low polymorphism was removed from subsequent tests. The remaining 21 loci were then pooled into three PCR multiplex sets based on fragment size, with the forward primers labelled with one of four commercial dyes (6-FAM, VIC, NED, PET; Applied Biosystems, Whitehead Scientific (Pty) Ltd, Cape Town). Following amplification in multiplexes, a further seven primers were discarded due to lack of polymorphism, non-specific binding, and excess stuttering. PCR products were subjected to a post PCR clean-up prior to electrophoresis using the Nucleofast ® 96 well PCR plate (Macherey-Nagel). Protocols were implemented as supplied by the manufacturer on a Tecan EVO® 150 robotic workstation. For analyses, 2 µl of cleaned PCR product was mixed with the appropriate internal size standard and Hi-Di™ Formamide prior to denaturing for 5 min at 95°C. Electrophoresis was performed on an ABI3130xl Genetic Analyzer using a 50 cm Capillary Array and POP-7™ polymer.

We also tested cross-amplification of our primers on five other sengi species; these were *E. brachyrhynchus* (A. Smith, 1836), *E. edwardii* (A. Smith, 1839), *E. intufi* (A. Smith, 1836), *E. rupestris* (A. Smith, 1831), and one representative from another genus of sengi, *Macroscelides proboscideus* (Shaw, 1800).

Data analyses

Alleles were analysed, scored and initial manual binning was inspected with the Geneious microsatellite plugin 1.4 (Kearse et al. 2012) and Tandem v1.09. Manual rounding of alleles

to the nearest number confirming the expected periodicity may cause underestimates of allelic richness and heterozygosity; such errors were minimized by comparing bins with Tandem scorring (Matschiner and Salzburger 2009). Manual binning was found to be very similar if not identical when compared to Tandem and manual binning was found to have, on average, lower rounding errors; a common measure for binning quality (Online Resource 2). All reads were scored twice (independently by two Southern Hemisphere and HW) to control for quality and selection of the trace peaks of the dataset. We used GENALEX v6.502 (Peakall and Smouse 2006; 2012) to estimate the number of alleles per locus (N_A), the observed (H_O) and expected (H_E) heterozygosities, the inbreeding coefficient (F_{IS}), and deviation from Hardy-Weinberg equilibrium (HWE). Summary statistics was also calculated for individuals from Goro and Telperion/Ezemvelo where sample sizes large (n = 25 individuals from each site); this was done to confirm overall results, to test the usuability of these markers on single populations, and to avoid possible biases introduced by spatial structure). For our two larger populations, linkage disequilibrium between loci was determined using Genepop v4.2 (Raymond and Rousset 1995). The presence of null alleles was investigated using Micro-Checker v 2.2.3 (van Oosterhout et al. 2004).

Results and discussion

From the 40 loci initially selected, 14 proved useful in documenting diversity within *Elephantulus myurus*. The final selection of these 14 markers was based on the analyses of 66 specimens. We detected 147 alleles across the 14 loci, with the average number of alleles per locus being 10.5 (Table 1). The effective average number of allele per locus was 5.2. Locus *EM30* had the highest number of alleles (18) while *EM25* had the lowest number of alleles (2) (Table 1). Observed and expected heterozygosity values ranged from 0.081 – 0.909 and from 0.484 – 0.885, respectively (Table 1).

Table 1 Characteristics of the 14 polymorphic microsatellite loci developed from *Elephantulus myurus.*

Locus	Primer Sequence (5' – 3')	Motif	Annealing T°	Dye	Size Range (bp)	Accession number	*N* = 46				
							N_A	H_O	H_E	F_{IS}	*p*HWE
Multiplex 1											
EM03	F: TGCCCACTAAGTCATCTTCCA R: GAGCTCCCTGGTGATGGTAA	$(TG)_{19}$	54°	VIC	219-245	In the process of submission to public genetics library	14	10.765	0.891	0.907	ns
EM05	F: TGGGCTTTCTATGAGTTGCAG R: CACAGAATCAGTGAGCTGGAAG	$(AAC)_{15}$	54°	NED	99-141		12	4.457	0.800	0.776	ns
EM27	F: GAGGGTCACCTATGGGTCAGT R: TCCAAAACGTTTCAAGGGAC	$(CAA)_{12}$	53°	VIC	104-134		9	2.366	0.523	0.577	***
EM 28	F: GGTGTTCAAAACCCAGCAGT R: TTTGGCACAGGACAGAACAG	$(TTG)_{13}$	52°	6-FAM	226-265		13	5.660	0.631	0.823	***
EM30	F: CCAGAGCGCCCATAATTAAA R: CGGGAGGGTAGGTGTGAGTA	$(AC)_{17}$	53°	PET	126-164		18	11.208	0.770	0.911	**
Multiplex 2											
EM04	F: CATGCTTGACTGGACAGCAC R: GTGCCGGGTCAATAAGAGAA	$(AAC)_{15}$	53°	6-FAM	126-177		8	3.439	0.530	0.709	***
EM23	F: CATCCTCCATGATGGGATTT R: GTCTCCTGCAACTCGAGGAC	$(TGT)_{13}$	53°	PET	103-130		8	4.022	0.682	0.751	**
EM29	F: CAATCAACTGTCTAAGGGCTTG R: TCTGTTGTTCTCAGGGTTGCT	$(GTT)_{13}$	53°	PET	185-209		9	3.424	0.538	0.708	ns
EM34	F: GGAGGGTCAGCAAGCAGTTA R: TGAGTCATCTATGAGTTGACAGCAC	$(TGT)_{18}$	56°	NED	125-158		12	7.030	0.785	0.858	***
EM36	F: GATATCAGCCCTGCACCCTA R: GGATACAAACCCACGTGTCA	$(GT)_8$	53°	VIC	153-173		11	4.456	0.652	0.776	**
Multiplex 3											
EM11	F: TCTGGGCAAATTTTCAGAAG R: CTGCCCTCTGTCCTGCTTAG	$(GA)_9$	52°	PET	150-154		3	1.679	0.182	0.404	***
EM25	F: AGCTCTCCCCAGAGCAATTT R: GGTTCAGCTATGAGTCGGTCA	$(TC)_9$	52°	VIC	118-120		2	1.939	0.081	0.484	***
EM26	F: TGAGGAATGAATACACAGGGG R: TGTGATTCCAAGGAGCACTG	$(AG)_7$	52°	NED	91-165		17	3.828	0.697	0.739	***
EM40	F: TGCAAACCCCTTATGACTCC R: ACCACAACCTTTGAGTTGGC	$(ACA)_{12}$	52°	6-FAM	137-170		11	8.729	0.909	0.885	ns

F denotes forward sequences and R denotes reverse sequences

N = number of individuals sampled, NA = number of alleles, HO = observed heterozygosity, HE = expected heterozygosity, FIS = fixation index and *p*HWE = *p* value for Hardy-Weinberg Equilibrium.

Significances are as follows, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ and ns denotes non-significance.

Genetic diversity in sengis is typically structured across their ranges because of the fragmented nature of their habitat (Smit et al. 2010). The 66 specimens included here originated from localities separated by as much as 1081 km, which may account for the high diversity observed in *E. myurus* compared to the notably lower levels of diversity reported for *Rhynchocyon petersi* Bocage 1880 (number of alleles per locus ranged from 2 to 6; Sabuni et al. 2015). Also, animals from a single locality was used in the Sabuni et al. study.

Ten loci showed significant deviation from Hardy Weinberg equilibrium (HWE) when all populations were pooled (Table 1). Equilibrium in populations might be affected by the presence of spatial structure. Indeed, when Goro and Telperion/Ezemvelo were analysed singly (sites with 25 individuals, Table 2), loci that did not conform to HWE decreased. Specifically, only five loci deviated significantly from HWE in specimens from Goro (*EM34* and *EM36* were the only loci with negative fixation index values (-0.118 and -0.052, respectively)). For Telperion/Ezemvelo, we found that five loci deviated significantly from HWE with only *EM28, EM30* and *EM26* being characterized by negative fixation index values (-0.040, -0.008 and -0.087, respectively). Micro-Checker detected possible null-alleles only at *EM36* for Goro and *EM25* for Telperion/Ezemvelo.

We assessed possible linkage disequilibrium in the two populations where we had high sample sizes (Goro and Telperion/Ezemvelo). In Goro *EM05-EM27, EM29-EM04*, *EM26-EM25* and *EM30-EM11* showed significant linkage ($p < 0.05$). In Telperion/Ezemvelo, we found significant linkages in *EM05-EM25, EM23-EM40*, and *EM34-EM40*. However, *EM11* and *EM25* had three or fewer alleles and hence this is potentially an artefact of excess homozygotes in the population. Similarly, the fact that the same pairs of loci were not consistently identified as possibly linked suggests an artefact of sample size rather than similar gene region contributions.

Table 2 Characteristics of the microsatellites in the two larger populations of *Elephantulus myurus* sampled during the study period.

	Locus	Size Range (bp)	*N* = 25				
			N_A	H_O	H_E	F_{IS}	*p*HWE
Goro	*EM03*	219-245	13.0	0.875	0.895	0.022	ns
	EM05	99-135	9.0	0.720	0.710	-0.015	ns
	EM27	107-125	6.0	0.480	0.620	0.226	***
	EM28	226-262	8.0	0.560	0.690	0.189	ns
	EM30	142-164	11.0	0.913	0.876	-0.042	ns
	EM04	126-156	4.0	0.520	0.560	0.071	ns
	EM23	103-118	5.0	0.680	0.625	-0.088	ns
	EM29	185-197	4.0	0.458	0.513	0.107	ns
	EM34	125-152	9.0	0.792	0.708	-0.118	***
	EM36	153-169	8.0	0.800	0.761	-0.052	*
	EM11	150-154	3.0	0.080	0.215	0.628	**
	EM25	118-120	2.0	0.000	0.499	1.000	***
	EM26	91-145	12.0	0.560	0.550	-0.019	ns
	EM40	137-164	9.0	0.880	0.846	-0.041	ns
Telperion/Ezemvelo	*EM03*	221-245	12.0	0.880	0.875	-0.005	ns
	EM05	102-141	10.0	0.960	0.826	-0.163	ns
	EM27	113-134	7.0	0.760	0.637	-0.193	ns
	EM28	232-262	10.0	0.880	0.846	-0.040	***
	EM30	126-162	12.0	0.792	0.786	-0.008	**
	EM04	147-177	8.0	0.800	0.794	-0.008	ns
	EM23	106-130	7.0	0.840	0.736	-0.141	ns
	EM29	185-203	6.0	0.680	0.773	0.120	ns
	EM34	128-158	11.0	0.960	0.815	-0.178	ns
	EM36	155-169	7.0	0.400	0.570	0.299	***
	EM11	150-152	2.0	0.320	0.499	0.359	ns
	EM25	118-120	2.0	0.174	0.340	0.489	*
	EM26	91-135	7.0	0.760	0.699	-0.087	***
	EM40	137-170	11.0	0.920	0.897	-0.026	ns

N = number of individuals sampled, NA = number of alleles, HO = observed heterozygosity, HE = expected heterozygosity, FIS = fixation index and *p*HWE = *p* value for Hardy-Weinberg Equilibrium.

Significances are as follows, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ and ns denotes non-significance.

None of the primers cross-amplified in the sister genus *Macroscelides*, and successful amplification varied across the different species of *Elephantulus*. The highest amplification success was in *E. rupestris* (66%) and the lowest in *E. edwardii* (4%, Table 3). The implication of successful cross-species amplification is that valuable information can now be generated at the population level for *Elephantulus* species, the majority of which (9 out of 11) are data deficient for population trends (IUCN Version 2015-4 2015). Highly variable nuclear markers, such as those described here, are invaluable in helping to shed light on population trends in this unique group of African small mammals.

In conclusion, we developed a suite of polymorphic microsatellite markers for *Elephantulus myurus.* These markers represent the first suite of microsatellite markers for *E. myurus* and indeed, the genus *Elephantulus*. These markers will be used to study population trends, spatial genetic structure and familial relationships linked to behaviour in our target species. In addition, the cross-species amplification, albeit with limited success, provides a first avenue to study comparative patterns of spatial diversity in *Elephantulus* species; information that will greatly add to our understanding of how landscapes shape the spatial genetic patterns.

Acknowledgements

Animals were collected under permit numbers 0089-RAP001-00001 and ZA/LP/HO/1633 issued by the Limpopo Provincial Government. Samples were also kindly donated by Duncan MacFadyen (E. Oppenheimer & Son) and Nico Avenant (National Museum, Bloemfontein). The study was conducted under ethics project approval from the University of Pretoria (EC037-14) and the University of Johannesburg (approved April 2015). The work was financially supported by the National Research Foundation of South Africa (BvV), the University of Johannesburg and the DST-NRF SARChi Chair of Behavioural Ecology and

Physiology (NCB). Electrophoresis of microsatellite markers was done at the Analytical Facility based at Stellenbosch University. Pierre-Alexandre Rastorgueff, Daniela M. Monsanto and Catharina W. Groenewald are thanked for valuable discussions and assistance. We also thank Dr. Tetsuo Gotoh and two anonymous reviewers for valuable comments during the review process.

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697–709.

Araki H, Cooper B, Blouin MS (2007) Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science 318:100–103.

Bruford MW, Wayne RK (1993) Microsatellites and their application to population genetic studies. Curr Opin Genet Dev 3:939–943.

Frankham R (2010) Where are we in conservation genetics and where do we need to go? Conserv Genet 11:661–663.

Freeland JR, Kirk H, Petersen SD (2011) Molecular Ecology, 2nd edn. John Wiley & Sons Ltd., West Sussex

Giger T, Excoffier L, Amstutz U, et al (2008) Population transcriptomics of life-history variation in the genus Salmo. Mol Ecol 17:3095–3108.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98.

Hedrick PW (2001) Conservation genetics: where are we now? Trends Ecol Evol 16:629–

636.

IUCN Version 2015-4 (2015) The IUCN Red List of Threatened Species. www.iucnredlist.org. Accessed 2015-05-13.

Kasperson RE, Archer ERM (2005) Vulnerable Peoples and Places. In: Norgaard R, Rapport D (eds) Ecosystems and human well-being: current state and trends: Findings of the Condition and Trends Working Group 1. p 143

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecol Lett 7:1225–1241.

Kearse M, Moir R, Wilson A, et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Krauss J, Steffan-Dewenter I, Tscharntke T (2003) How does landscape context contribute to effects of habitat fragmentation on diversity and population density of butterflies? J Biogeogr 30:889–900. doi: 10.1046/j.1365-2699.2003.00878.x

Malausa T, Gilles A, Meglécz E, et al (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. Mol Ecol Resour 11:638–644. doi: 10.1111/j.1755-0998.2011.02992.x

Matschiner M, Salzburger W (2009) TANDEM : integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25:1982–1983. doi: 10.1093/bioinformatics/btp303

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic

software for teaching and research. Mol Ecol Notes 6:288–295.

Pool-Stanvliet R (2013) A history of the UNESCO Man and the Biosphere Programme in South Africa. S Afr J Sci 109:6. doi: 10.1590/sajs.2013/a0035

Rathbun GB (2009) Why is there discordant diversity in sengi (Mammalia: Afrotheria: Macroscelidea) taxonomy and ecology? Afr J Ecol 47:1–13.

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

Ribble DO, Perrin MR (2005) Social organization of the eastern rock elephant-shrew (*Elephantulus myurus*): the evidence for mate guarding. Belgian J Zool 135:167–173.

Sabuni CA, Van Houtte N, Maganga SLS, et al (2015) Development of eight polymorphic microsatellite markers in the Black and Rufous sengi, *Rhynchocyon petersi*. Microsatellite Lett 7:193–195. doi: 10.1007/s12686-014-0330-5

Skinner JD, Chimimba CT (2005) The mammals of the southern African subregion. Cambridge University Press, Cambridge, UK

Smit HA, Jansen van Vuuren B, O'Brien PCM, et al (2011) Phylogenetic relationships of elephant-shrews (Afrotheria, Macroscelididae). J Zool 284:133–143.

Smit HA, Robinson TJ, Van Vuuren BJ (2007) Coalescence methods reveal the impact of vicariance on the spatial genetic structure of *Elephantulus edwardii* (Afrotheria, Macroscelidea). Mol Ecol 16:2680–2692.

Smit HA, Robinson TJ, Watson J, Jansen van Vuuren B (2008) A new species of elephant-shrew (Afrotheria: Macroscelidea: Elephantulus) from South Africa. J Mammal 89:1257–1269.

Smit HA, Watson J, Jansen van Vuuren B (2010) Relative importance of habitat connectivity in shaping the genetic profiles of two southern African elephant-shrews. J Biogeogr 37:857–864.

Steyn M, Nortje K, Funke N, et al (2010) What are the central issues of the livelihoods and biodiversity interface in our case areas? The case of the Greater Kruger area in South Africa. LiveDiverse Seventh Framework Programme. Project Co-ordinator.

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

Online Resource 1 - Table showing characteristics of the 40 putative loci selected from the 807 possible repeat sequence motifs as supplied by Genoscreen©

Label	Sequence Length	Forward (5'-3')	Reverse (5'-3')	Ta (°C)	Bp Repeat motif	Number of repeats
EM01	176	GATATTTGTAGGCCCAGCGA	GCTAATGCTTTCCTTCCACAA	51	**tg**	**10**
EM10	99	TGAGATGCTCCAGTCCATTG	ATAGCATGCTGGGAAAGGTG	52	**ct**	**5**
EM11	150	TCTGGGCAAATTTTCAGAAG	CTGCCCTCTGTCCTGCTTAG	52	**ga**	**10**
EM12	98	TGGCAAGTGAGAGTCAGGAA	TCCAATTTCTCCGGTACCAA	51	**ga**	**8**
EM13	212	ACAGTATGTAACCCGCCGTT	GAAGGAGCAGAAAAGGAGGC	53	**atcc**	**16**
EM14	138	GACCACTGTCCTAGTTTACTCCC	CCTGCCTATAATGAGCATTCAGT	56	**atcc**	**12**
EM15	251	TGCATGGATGGATATCTGGA	CCATTCACTCCCCACTGTTC	52	**gatg**	**11**
EM16	128	ATCAGCAATGCTATCCCTGG	CAGGGTCACTATGAGTCGCA	53	**ggt**	**5**
EM17	90	GGAGACCATCAAGAGAAATAGCA	AGTGCCTTTCTCCTTCACCA	53	**gag**	**7**
EM18	284	TCCAAGGCTCCTCATCTCTG	ATGGCCAGTTTCTGCTCAAC	53	**caaaa**	**17**
EM19	206	TTCTGACCTCTCCTCATGGC	AAGTGCAAACCACCTAACGG	53	**ctc**	**5**
EM02	91	TGCCTATTATTTAAGCCAGGGA	TTCTATAGCCACGTGATTGGT	51	**tg**	**18**
EM20	136	CATGATAAAGGGACGGAGGA	TCCCCGGTTATCATGAACTC	52	**gga**	**6**
EM21	269	CCTGAACCCTAGATGGAAACC	TTTAGGTCAGGGTTAGGGCA	53	**cctaac**	**6**
EM22	90	TGGCATAGGATAGGGCTTTG	AACATCTGGTCCAGGGCATA	52	**ag**	**8**
EM23	119	CATCCTCCATGATGGGATTT	GTCTCCTGCAACTCGAGGAC	53	**tgt**	**13**
EM24	111	CTCTTCCCCTTGGCATACAA	GGGGATCATGATAAAAGGGG	52	**cct**	**5**
EM25	119	AGCTCTCCCCAGAGCAATTT	GGTTCAGCTATGAGTCGGTCA	53	**tc**	**9**
EM26	94	TGAGGAATGAATACACAGGGG	TGTGATTCCAAGGAGCACTG	52	**ag**	**7**
EM27	135	GAGGGTCACCTATGGGTCAGT	TCCAAAACGTTTCAAGGGAC	53	**caa**	**12**
EM28	262	GGTGTTCAAAACCCAGCAGT	TTTGGCACAGGACAGAACAG	52	**ttg**	**13**
EM29	192	CAATCAACTGTCTAAGGGCTTG	TCTGTTGTTCTCAGGGTTGCT	53	**gtt**	**13**
EM03	232	TGCCCACTAAGTCATCTTCCA	GAGCTCCCTGGTGATGGTAA	54	**tg**	**19**
EM30	138	CCAGAGCGCCCATAATTAAA	CGGGAGGGTAGGTGTGAGTA	53	**ac**	**17**
EM31	140	GTGCTCGACTCAAAGGCAAT	TCTTTGAGTTGTGGTGCTGG	52	**aag**	**6**
EM32	240	CCAGTTCCCTGTCACTGGAC	AACCACGAAGCAACAAGGAC	54	**ac**	**21**
EM33	90	AATTTGTTGTCCAGGTCTCACA	ATTCAGTTCATGAGGGGTGG	51	**ac**	**12**
EM34	151	GGAGGGTCAGCAAGCAGTTA	TGAGTCATCTATGAGTTGACAGCAC	56	**tgt**	**18**
EM35	142	TTAATGGTGCTTTTGGCCTC	CAGTGATTATAACCAAAATAGCCAA	50	**gt**	**7**
EM36	162	GATATCAGCCCTGCACCCTA	GGATACAAACCCACGTGTCA	53	**gt**	**8**
EM37	90	TGAGTTAGAACTGACCTGGTGG	GAATAGTGAGCCACCTCGTGT	55	**tc**	**9**
EM38	106	GGAGGATGGAGTCATTAAAACA	GATCCTCTCACCGTCCTCCT	54	**agg**	**9**
EM39	148	TCCTTCTTCTCCTTCTGCTTCT	GCAATGGTGTCACTCAGGTG	53	**tcc**	**7**
EM04	176	CATGCTTGACTGGACAGCAC	GTGCCGGGTCAATAAGAGAA	53	**aac**	**15**
EM40	145	TGCAAACCCCTTATGACTCC	ACCACAACCTTTGAGTTGGC	52	**aca**	**12**
EM05	127	TGGGCTTTCTATGAGTTGCAG	CACAGAATCAGTGAGCTGGAAG	54	**aac**	**15**
EM06	102	AAATGGAATTGCAGCCTGAC	TGGGGAAGATTTTACTGGGA	50	**ctt**	**6**
EM07	111	CCAATCCAAAAGTCTGAGCC	CACAAGAACAACAAAAACAAATAACA	51	**ca**	**23**
EM08	264	TGACATACTTAAGACCACTGCTGC	GAGCTTGAGCAGAAGGAGGA	54	**ca**	**15**
EM09	128	CAGTTCGGGGAAGTTTGAGA	TTGAGCTCAGCTGGAACAGA	52	**ct**	**6**

Online Resource 2 - Binning protocol for the 14 microsatellite markers for *Elephantulus myurus*

When comparing Tandem binning to manual binning we found significant differences in number of alleles at *EM28, EM03, EM34* and *EM36.* In each instance manual rounding of alleles confirming to expected periodicity had lower rounding errors than Tandem (0.84 vs. 1.8; 0.57 vs. 1.05; 0.78 vs. 0.93 and 0.3 vs. 0.87 respectively). This is most likely due to these loci having large base pair ranges between different alleles sizes as can be seen for example in Locus 28 (Figure 1). Tandem, in the absences of user defined fix points, uses the shortest allele size as a fixed point whereas in manual binning we used the allele size with the highest frequency as our fixed point to minimise rounding error. Due to the large rounding errors observed in Tandem for these samples we opted to use our manual binning configuration as our standard binning protocol, also taking into account that for over 70% of the loci manual and Tandem binning correlated suggesting our manual binning configuration was optimal for these loci.

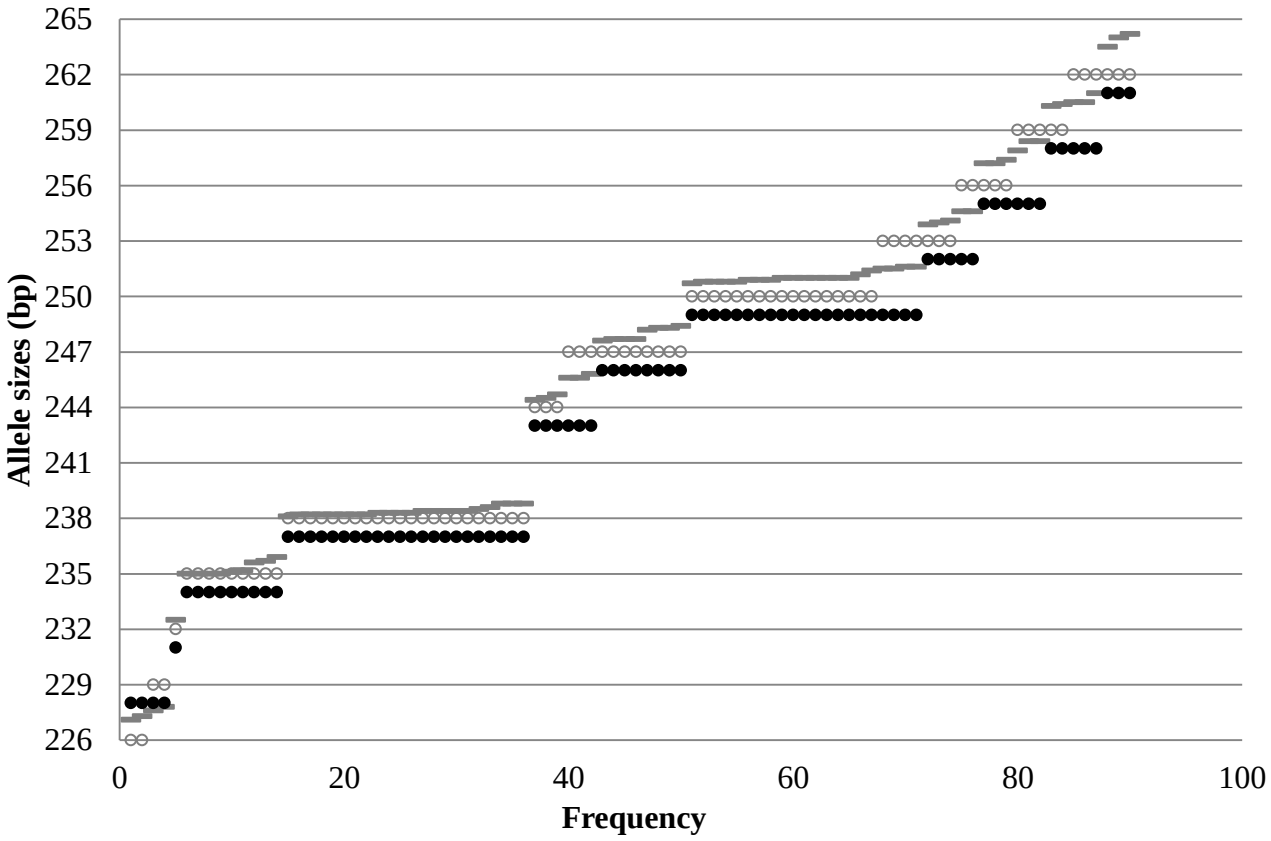


Figure 1 Rounded allele sizes comparing Tandem (filled circles) and manual binning (open circles) against raw allele sizes genotyped in Geneious (grey filled dashes).