RESEARCH NOTE

Rapid isolation and characterization of microsatellites in the critically endangered Mountain Bongo (*Tragelaphus eurycerus isaaci)*

Authors: Fraser J Combe[1], Evelyn Taylor-Cox[1], Graeme Fox[1], Tommy Sandri[1,3], Nick Davis[2], Martin J Jones[1], Bradly Cain[1], David Mallon[1], W Edwin Harris[1]*

1. Division of Biology and Conservation Ecology, Manchester Metropolitan University, Manchester, M1 5GD, United Kingdom

2. North of England Zoological Society, Chester Zoo, Chester, United Kingdom

3. Tropical Biodiversity Section, MUSE–Museo delle Scienze, Corso del Lavoro e della Scienza 3, 38122 Trento, Italy

**Corresponding author. E-mail: E.harris@mmu.ac.uk*

Keywords: microsatellites, high-throughput sequencing, Galaxy server, mountain bongo, *Tragelaphus eurycerus isaaci*

Introduction

High-throughput sequencing tools promise to revolutionize many aspects of genetics research, e.g. by allowing the identification of functional adaptive genetic variation. However, the expense and expertise required to apply these tools to basic conservation questions is a challenge for applications outside academia, resulting in a so-called "conservation genomics gap" (Shafer et al. 2015). The conservation genetics paradigm is that basic information about inbreeding and gene flow are often critical to inform conservation management of small populations (Ouborg et al.

2010). This information is often needed quickly and ideally should be accessible to workers without special expertise in genomics (DeSalle & Amato 2004). While the inferential power of high-throughput sequencing to interrogate the genome is profound, the cost for population analysis is higher (though decreasing) than for traditional neutral markers. Thus, the use of neutral markers is still relevant in conservation applications. However, this assumes that neutral markers have been discovered and characterized for a given species of conservation concern, which is often untrue for non-model organisms. Here, we use a fast, cost-efficient, high-throughput sequencing method (Illumina MiSeq) to rapidly identify and characterize microsatellites in the mountain bongo (*Tragelaphus eurycerus isaaci*, hereafter bongo), which has a clear and timely conservation imperative but lacks any described neutral markers.

The bongo is a critically endangered antelope inhabiting montane forest habitat in central Kenya (IUCN 2008). The bongo has declined precipitously over the past several decades, likely as a result of poaching, human encroachment, habitat loss, and disease (Estes et al. 2008, 2011, Kingdon, 2013). The wild population is estimated to be 75-140 individuals, divided between 4 (currently) unconnected remnant populations (IUCN 2008). While the wild bongo population has declined, the captive management of bongo in zoos has been a remarkable success. Founded from around 62 wild individuals caught over several decades since the 1960s, the captive population has grown to over 750 living individuals internationally, creating a source pool for reintroduction or population augmentation (Bosley, 2012). This situation has motivated the bongo repatriation project, where 18 bongo were returned to Kenya from captivity in the US to the Mount Kenya Wildlife Conservancy, intended as source animals for reintroduction (Reillo 2002). While it is thought that the remaining wild population of bongo is genetically depauperate (based on mitochondrial gene sequencing; Faria et al. 2011), there is no information available about the extant genetic variation in the captive bongo population or about fine-scale population genetic variation represented in the wild (e.g. gleaned from microsatellites). Here, we present the first microsatellites isolated and characterized for the bongo, show cross-amplification results for these markers in two closely related species, and discuss the applicability of our approach to other non-model species.

Material and methods

Sample collection and DNA extraction

Blood samples used for high-throughput sequencing were from n=2 bongo from Paignton Zoo (UK), collected during routine vet inspection. Microsatellites were screened in n=26 hair samples collected from captive individuals in EU zoos. Hair samples were collected using a sterile comb brushed firmly along the animal from neck to rump, which is then placed into a sterile sample bag. Faecal samples from waterbuck *(Kobus ellipsiprymnus*; n=2 from Knowsley Safari Park, UK), and sitatunga (*Tragelaphus spekii*; n=1 from Parco Natura Viva, Bussolengo, Italy) were collected in order to test for cross-amplification of microsatellites. Samples were stored at -20°C until use. Total genomic DNA was extracted from blood samples using DNEasy Blood and Tissue Kit, following manufacturer guidelines (Qiagen, Venlo, Netherlands). DNA from hair samples was extracted using *Quick*-DNA™ Universal Kit (Zymo Research, USA), following manufacturer guidelines, with the addition of 20µl of 1M dithiothreitol during lysis. DNA from faecal samples was extracted using Qiamp Stool Kit following manufacturer guidelines (Qiagen, Venlo, Netherlands). After extraction, DNA was stored at -20°C.

Microsatellite characterization and development

Bioinformatics and primer design

Putative microsatellite loci were identified using Illumina MiSeq and a Galaxy Server pipeline optimized for microsatellite development, in a modified workflow described by Griffiths et al. (2016). Our method started with 50ng genomic DNA extracted from blood. We used MiSeq

Illumina shotgun, paired-end 2*250 sequencing to generate sequence data (Nextera DNA Library Preparation Kit, Illumina, San Diego, USA). A total of 8,980,510 raw sequencing reads were produced from the MiSeq run. We used FastQC version 0.11.4 to generate quality information for the sequencing data (Ward et al. 2016). Trimmomatic version 0.32 was used to trim low quality bases from reads and remove low quality reads (Bolger et al. 2014). Pal_finder version 0.02.04 was used to identify potential amplifiable microsatellite loci (Castoe et al. 2012). Primer3 v.4.0.0 (Koressaar and Remm 2012) was used to design microsatellite primers from identified loci. Primer design was optimized for Qiagen Type-it Microsatellite PCR kit. The minimum number of microsatellite repeats searched for was n=8 for all default repeat types (2-6mer). Finally, PANDAseq was used to confirm that both forward and reverse primer sequences occur in the same region in order to increase PCR success rate (Masella et al. 2012).

PCR amplification protocols and data analysis

To prepare samples for PCR, we used the Type-it Microsatellite PCR Kit (Qiagen, Venlo, Netherlands), with cycling conditions as follows: 95°C for 5 mins; 33 cycles of 95°C for 30 secs, 68°C for 45 secs, 72°C for 30 secs, with a final extension cycle of 72°C for 30 mins. Some primers required a touch-down PCR protocol for successful amplification (see Table 1). Microsatellites were amplified with the universal Tail C (5'-CAGGACCAGGCTACCGTG-3') in the three-primer method for the binding of fluorescent markers, as described by Blacket et al. (2012). Amplification was confirmed in 1.5% agarose gel and fragment length analysis was carried out on a ABI 3730 DNA analyser (Applied Biosystems, California, USA) with Genescan™ 500 LIZ size standard (Thermo Fisher Scientific, Carlsbad, USA). Genotyping was conducted independently by two individuals to ensure consistency of calls.

Allele peaks were scored using GENEMAPPER version 3.7 (Applied Biosystems, California, USA). For each screened locus we calculated allele fragment size range, the number of alleles per locus (A), and observed (H_o) and expected (H_e) heterozygosity with GenoDive version

2.0b23 (Meirmans & Van Tienderen 2004). Micro-Checker 2.2.3 (Van Oosterhout *et al.*, 2004) was used to establish whether any observed heterozygote deficiencies were attributable to null alleles, scoring errors, or large allelic dropout. Deviation from Hardy-Weinberg equilibrium (HWE) and calculation of the inbreeding coefficient was performed for each locus using GENEPOP 4.1 online (Rousset, 2008).

Results

In total, the bioinformatics pipeline identified n=491 primer pairs representing putative microsatellite loci. The resulting loci were ranked according to quality and the first n=30 best primer pairs were selected for screening. We successfully amplified n=18 out of 30 microsatellite loci (60%) for the 28 bongo samples (Table 1). For the 18 successful loci, we observed a 96% amplification rate overall in bongo samples. Allelic richness ranged from n=2-6 with a mean of 3.8 (SE ± 0.3; Table 2). The mean heterozygosity observed was 0.42 (SE ± 0.05), with the highest value being 0.79. The average expected heterozygosity across the 19 loci was 0.47 ± 0.04 (SE), with the highest value being 0.68. No significant deviation from HWE was detected (all $P < 0.05$), except for TEU-13, -22, -25 and -28, however this may be caused by low sample sizes a factor associated with critically endangered species. Micro-Checker provided no support for this excess of homozygosity being due to null alleles. The average estimated inbreeding coefficient (r) observed was 0.09. In the bongo congener sitatunga, we found n=8 out of 19 (42%) primers were successfully amplified. In the waterbuck, we found no successful amplification (0%) in our screened primers.

Discussion

Here, we present the first polymorphic microsatellite loci for the critically endangered mountain bongo, *Tragelaphus eurycerus isaaci*, using a rapid next generation sequencing (NGS) method and a repeatable, open-access bioinformatics analysis pipeline. Using this method, we identified

491 candidate microsatellites, of which 30 loci were screened for variation, resulting in the characterization of 18 informative loci. Our microsatellite analysis suggests a low mean genetic diversity (H_o = 0.42) for our samples. Our study used samples sourced haphazardly from the closed zoo population, which could introduce bias in our estimates. However, the global zoo population of bongo was founded from individuals sourced a small geographical area of the Aberdares National Park, Kenya (Bosley 2011), and thus our results are consistent to expectation based on informative loci screened for a population having experienced a known bottleneck (Spencer et al. 2000).

Here we also report the first putative microsatellite loci identified for the bongo congener, the sitatunga. While only 8 out of the 18 identified loci amplified, they are potentially useful for future work on this species for which no other neutral genetic markers have been identified, however, further work is required to confirm this. We also screened 18 loci for amplification in the waterbuck and none amplified successfully. The waterbuck is sympatric to wild bongo, and discrimination between bongo and waterbuck sign has ~30% error rate in the field (estimated from field identified samples checked by mtDNA sequencing; Faria et al. 2011). The fact that these microsatellites do not amplify in waterbuck provides an alternative to mtDNA sequencing for identification of bongo faecal samples in the field.

Management of animal populations in zoos largely has the aim to avoid inbreeding depression, but also, increasingly, to maintain the evolutionary integrity and genetic variation of captive populations (Schulte-Hostedde & Mastromonaco 2015). However, there is evidence suggesting that zoo captive breeding programs may fall short of preventing genetic decay in the long term (Lacy 2013). While the importance of explicitly integrating genetic information with studbook information in a conservation context is increasingly recognised (Henkel et al. 2012), the application of genetics tools is still uncommon outside of academia, especially with regard to genomics tools (Shafer et al. 2015). This is particularly important for conservation applications where wild populations are failing, and where a captive population is a potential source for reintroduction or augmentation efforts, which will contribute to the long-term persistence of a species in the wild. Thus, our results have critical implications for the successful conservation management in this species, aiding the breeding efforts in captive populations, assessing gene flow

and genetic diversity in wild populations and selection of founders for reintroductions. While we offer our identification of informative microsatellite loci to efforts in bongo conservation, it is anticipated that our method combining NGS and open source bioinformatics tools for the rapid assessment and characterization of microsatellites is useful in bridging the conservation genomics gap for other species of conservation concern as well.

Acknowledgements

We kindly thank the following zoos which donated DNA samples for this study: Chester Zoo (UK), Marwell Zoo (UK), Woburn Safari Park (UK), Paignton Zoo (UK), Howletts Wild Animal Park (UK), Knowsley Safari Park (UK), Givskud Zoo (Denmark) and Antwerp Zoo (Belgium). Thanks also to the DNA Sequencing Facility and the Genomic Technologies Core Facility at the University of Manchester, for their expert advice and services. Funding for this study came from an incentivisation grant from the Manchester Metropolitan University.

Availability of data and material

Microsatellite DNA sequences underlying these analyses have been deposited to GENBANK under the accessions KY700832-KY700849

References

Blacket, M.J., Robin, C., Good, R.T., Lee, S.F., Miller, A.D., 2012. Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. Molecular Ecology Resources 12, 456–463.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bosley, L.F., 2011. International studbook for eastern/mountain bongo (Tragelaphus eurycerus isaaci). WAZA.

Castoe, T.A., Poole, A.W., Koning, A.P.J. de, Jones, K.L., Tomback, D.F., Oyler-McCance, S.J., Fike, J.A., Lance, S.L., Streicher, J.W., Smith, E.N., Pollock, D.D., 2012. Rapid Microsatellite Identification from Illumina Paired-End Genomic Sequencing in Two Birds and a Snake. PLOS ONE 7, e30953.

DeSalle, R., Amato, G., 2004. The expansion of conservation genetics. Nat Rev Genet 5, 702–712.

Estes, L.D., Mwangi, A.G., Reillo, P.R., Shugart, H.H., 2011. Predictive distribution modeling with enhanced remote sensing and multiple validation techniques to support mountain bongo antelope recovery. Animal Conservation 14, 521–532.

Estes, L.D., Okin, G.S., Mwangi, A.G., Shugart, H.H., 2008. Habitat selection by a rare forest antelope: A multi-scale approach combining field data and imagery from three sensors. Remote Sensing of Environment, Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue 112, 2033–2050.

Faria, P.J., Kavembe, G.D., Jung'a, J.O., Kimwele, C.N., Estes, L.D., Reillo, P.R., Mwangi, A.G., Bruford, M.W., 2011. The use of non-invasive molecular techniques to confirm the presence of mountain bongo Tragelaphus eurycerus isaaci populations in Kenya and preliminary inference of their mitochondrial genetic variation. Conserv Genet 12, 745–751.

Griffiths, S.M., Fox, G., Briggs, P.J., Donaldson, I.J., Hood, S., Richardson, P., Leaver, G.W., Truelove, N.K., Preziosi, R.F., 2016. A Galaxy-based bioinformatics pipeline for optimised, streamlined microsatellite development from Illumina next-generation sequencing data. Conservation Genet Resour 8, 481–486.

Henkel, J.R., Jones, K.L., Hereford, S.G., Savoie, M.L., Leibo, S. p., Howard, J.J., 2012. Integrating microsatellite and pedigree analyses to facilitate the captive management of the endangered Mississippi sandhill crane (Grus canadensis pulla). Zoo Biol. 31, 322–335.

IUCN, 2008. Tragelaphus eurycerus ssp. isaaci: IUCN SSC Antelope Specialist Group: The IUCN Red List of Threatened Species 2008: e.T22057A9354511.

Kingdon, J., Happold, D., Butynski, T., Hoffmann, M., Happold, M., Kalina, J., 2013. Mammals of Africa. A&C Black.

Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23, 1289–1291.

Lacy, R.C., 2013. Achieving True Sustainability of Zoo Populations. Zoo Biol. 32, 19–26.

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., Neufeld, J.D., 2012. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13, 31.

Meirmans, P.G., Van Tienderen, P.H., 2004. genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4, 792–794.

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R. (Kuke), Hedrick, P.W., 2010. Conservation genetics in transition to conservation genomics. Trends in Genetics 26, 177–187.

Reillo, P.R., 2002. Repatriation of the mountain bongo to Kenya. Gnusletter 21, 11–15.

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

Schulte-Hostedde, A.I., Mastromonaco, G.F., 2015. Integrating evolution in the management of captive zoo populations. Evol Appl 8, 413–422.

Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K.D., Fior, S., Hajibabaei, M., Hill, J.A., Hoezel, A.R., Höglund, J., Jensen, E.L., Krause, J., Kristensen, T.N., Krützen, M., McKay, J.K., Norman, A.J., Ogden, R., Österling, E.M., Ouborg, N.J., Piccolo, J., Popović, D., Primmer, C.R., Reed, F.A., Roumet, M., Salmona, J., Schenekar, T., Schwartz, M.K., Segelbacher, G., Senn, H., Thaulow, J., Valtonen, M., Veale, A., Vergeer, P., Vijay, N., Vilà, C., Weissensteiner, M., Wennerström, L., Wheat, C.W., Zieliński, P., 2015. Genomics and the challenging translation into conservation practice. Trends in Ecology & Evolution 30, 78–87.

Spencer, C.C., Neigel, J.E., Leberg, P.L., 2000. Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Molecular Ecology 9, 1517–1528.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4, 535–538.

Ward, J., Cole, C., Febrer, M., Barton, G.J., 2016. AlmostSignificant: simplifying quality control of high-throughput sequencing data. Bioinformatics 32, 3850–3851.

Received 28 March 2017, final revised form 19 June 2017; accepted 29 June 2017

Unedited version published online 25 October 2017

Locus	Primer sequence (5'-3')	n	A	H_O	H_E	P_{HWE}	Repeat motif	Ta (°C)	Allele size range (bp)	S	GenBank Accession
TEU01	F- GCATCTATGTTCTTCACCAGTGATATTGGC	28	4	0.50	0.41	0.12	GAAGT	68^{TD}	196 - 204	Y	KY700832
	R- GTATGGGAGATCCAGGTTCAATCCC										
TEU02	F- TCTCAGGGTGCTTACTCCCTATCCC	28	2	0.36	0.34	0.65	GGAT	68	384 - 388	Y	KY700833
	R- CATTTAGGGTTAGAGAGGCTAGATGTTGG										
TEU03	F- TTTTGCATCTTTGGCACCTACC	28	2	0.32	0.27	0.45	ATGG	68	372 - 376	N	KY700834
	R- CTCCTGCATGCTGATACATGGG										
TEU04	F- GCACAGCATTCTGTAATCCTCCTACCC	28	2	0.18	0.22	0.4	ATCC	68	450 - 454	N	KY700835
	R-TCATGGATTTCCCCAGTATTTCATGG										
TEU05	F- ACTGCAGAGGGTCCAGGTTCAGTCC	26	6	0.62	0.66	0.16	AAAT	68^{TD}	251 - 267	Y	KY700836
	R- TTAAGGAGGGATCACCCCAATCAGG										
TEU11	F- TTTCCTGGCTTATAACCCGGTCTCC	26	6	0.65	0.63	0.59	CAA	68^{TD}	326 - 350	Y	KY700837
	R- CCATCATCATTCACACAGAAGGGATTC										
TEU13	F- GCATCTCTCAAGTTTGAGTTTCAAGAGTCC	27	3	0.11	0.52	0.00	AAAT	68	468 - 484	N	KY700838
	R- CCAGAGGAAGCTGGTCATCTGAAAG										
TEU14	F- ATAGCACTGGCTTTCTTGGGTCTCC	28	4	0.79	0.62	0.06	ACAT	68^{TD}	298 - 306	N	KY700839
	R- GATCTGAAACCTCCAAAATCATACAGG										
TEU16	F- TCACTGCCCTGAAACTCTTTTGTGC	23	3	0.26	0.24	1.00	CTT	68	339 - 342	Y	KY700840
	R- CCTTGCTCTTGGACTTGAAGCATGA										
TEU21	F- CCCATATGCTTTAGTTTGCCAACCC	27	6	0.56	0.67	0.05	GCT	68	350 - 360	Y	KY700841
	R- CCAAACTACCAGCCACATCTCCCCT										
TEU22	F- GGGCTCAGTATATATGGCATCATTGG	24	4	0.38	0.47	0.00	TGC	68	436 - 445	N	KY700842
	R- CCCAAGTGAAAGACATTATGTTGCTACCA										
TEU23	F- TGAATTTGTTACTGTCACTGCCACTGC	26	3	0.73	0.59	0.11	ATT	68	277 - 283	N	KY700843
	R- GGTGTGTGTGAGTTCACGATAGTGCC										
TEU24	F- TGTGTCTTCAAGGAAAACTCATGAAGGG	27	4	0.70	0.68	0.52	GGA	68	343 - 353	N	KY700844
	R- CCTAACTTCTGCAAGGCTGGCTCAG										
TEU25	F- CTTTAGGGCGTGGACTGTGTCAGC	26	4	0.27	0.47	0.00	ATT(	68	450 - 459	N	KY700845
	R- CCGATTTAACCCCGTTTGTGCC										
TEU26	F- TCACATTTTGGTCTTGTTACACACTCAGC	28	2	0.14	0.20	0.28	TGG	68	378 - 387	N	KY700846
	R- GCATGGAAGCCCAGACAGAAAGC										
TEU28	F- TTTAGAGATTTGCCACGGAGAAGGC	28	5	0.18	0.42	0.00	GCA	68^{TD}	343 - 353	Y	KY700847
	R- GCCAAAGGTCACACAGTTAGCAGGA										
TEU29	F- CCAAAAGCCAACATATCTCCAAACCC	28	4	0.43	0.59	0.05	TTG	68	414 - 423	N	KY700848
	R- CGAGACTTCACTGGCTCATCAATCTG										
TUE30	F- TCAAAGTGCACTCCAATCTGATAGC	26	4	0.77	0.65	0.18	TGC	68^{TD}	261 - 289	Y	KY700849
	R- CCTGTCCGACTCTGTAAAAGCCTGT										

Table 1: Characterization of 18 polymorphic microsatellite loci for the mountain bongo, *Tragelaphus eurycerus isaaci*.: n: number of individuals; A: Number of alleles observed; HO: Observed heterozyqogsity; HE: Expected heterozygosity; PHWE: P value from testing Hardy-Weinberg equilibrium (HWE); Ta:

Optomised annealing temperature (°C); TD: PCR conditions altered to touchdown protocol (decrease by 0.5°C in 11 cycles from 70-65°C); Allele size range (bp): Observed size range in characterised individuals; S: Successful amplification in sitatunga, Tragelaphus spekii (Y or N).