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

FRASER J. COMBE[1], EVELYN TAYLOR-COX[1], GRAEME FOX[1], TOMMY SANDRI[1,3], NICK DAVIS[2], MARTIN J. JONES[1], BRADLEY CAIN[1], DAVID MALLON[1] and W.EDWIN HARRIS[1]*

[1] *Division of Biology and Conservation Ecology, Manchester Metropolitan University, Manchester, M1 5GD, UK*
[2] *North of England Zoological Society, Chester Zoo, CH2 1EU Chester, UK*
[3] *Tropical Biodiversity Section, MUSE–Museo delle Scienze, Corso del Lavoro e della Scienza 3, 38122 Trento, Italy*
*For correspondence. E-mail: E.harris@mmu.ac.uk.

Received 28 March 2017; revised 19 June 2017; accepted 29 June 2017; published online 4 May 2018

Abstract. High-throughput sequencing tools promise to revolutionize many aspects of genetic 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 and 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 nonmodel 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*), which has a clear and timely conservation imperative but lacks any described neutral markers.

Keywords. microsatellites; high-throughput sequencing; Galaxy server; mountain bongo; *Tragelaphus eurycerus isaaci*.

Introduction

The *Tragelaphus eurycerus isaaci* (hereafter bongo) is a critically endangered antelope inhabiting montane forest habitat in central Kenya (IUCN 2017). It 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 four (currently) unconnected remnant populations (IUCN 2017). 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 nonmodel 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 from Chester Zoo, UK ($n = 5$), Givskud Zoo, Denmark ($n = 5$), Marwell Zoo, UK ($n = 3$), Wolburn Safari Park, UK ($n = 4$), Howletts Zoo, UK ($n = 7$), and Knowsley Safari Park, UK ($n = 2$). 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 (*T. spekii*; $n = 1$ from Parco Natura Viva, Bussolengo, Italy) were collected 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, The Netherlands). DNA from hair samples was extracted using Quick-DNA Universal kit (Zymo Research, Irvine, USA), following manufacturer guidelines, with the addition of 20 μL of 1 M dithiothreitol during lysis. DNA from faecal samples was extracted using Qiamp Stool kit following manufacturer guidelines (Qiagen). 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 50 ng 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 ver. 0.11.4 to generate quality information for the sequencing data (Ward *et al.* 2016). Trimmomatic ver. 0.32 was used to trim low quality bases from reads and remove low quality reads (Bolger *et al.* 2014). Pal_finder ver. 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 (2–6 mer). Finally, PANDAseq was used to confirm that both forward and reverse primer sequences occur in the same region 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), with cycling conditions as follows: 95°C for 5 min; 33 cycles of 95°C for 30 s, 68°C for 45 s, 72°C for 30 s, with a final extension cycle of 72°C for 30 min. Some primers required a touchdown 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 the 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, Foster City, 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 ver. 3.7 (Applied Biosystems). 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 ver. 2.0b23 (Meirmans and 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). Microsatellite DNA sequences underlying these analyses have been deposited to GenBank under the accessions KY700832–KY700849.

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 $n = 18$ 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, TEU-22, TEU-25 and TEU-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 $n = 8$ of 19 (42%)

Table 1. Characterization of 18 polymorphic microsatellite loci for the mountain bongo, *T. e. isaaci*.

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

N, number of individuals; A, number of alleles observed; H_o, observed heterozyqogsity; H_e, expected heterozygosity; P_{HWE}: P value from testing HWE; T_a, optomized 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, *T. spekii* (Y or N).

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, *T. 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 eight identified loci amplified of the 18, 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 and Mastromonaco 2015). However, there is evidence suggesting that zoo captive breeding programmes 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 recognized (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.

References

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

Bolger A. M., Lohse M. and 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, Gland, Switzerland.

Bosley L. F. 2012 International studbook for eastern/mountain bongo (*Tragelaphus eurycerus isaaci*). WAZA, Gland, Switzerland.

Castoe T. A., Poole A. W., Koning A. P. J., de Jones K. L., Tomback D. F., Oyler-McCance S. J. *et al.* 2012 Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS One* **7**, e30953.

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

Estes L. D., Okin G. S., Mwangi A. G. and 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 Environ.* **112**, 2033–2050.

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. *Anim. Conserv.* **14**, 521–532.

Faria P. J., Kavembe G. D., Jung'a J. O., Kimwele C. N., Estes L. D., Reillo P. R. *et al.* 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. *et al.* 2016 A Galaxy-based bioinformatics pipeline for optimised, streamlined microsatellite development from Illumina next-generation sequencing data. *Conserv. Genet. Resour.* **8**, 481–486.

Henkel J. R., Jones K. L., Hereford S. G., Savoie M. L., Leibo S. P. and 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 SSC Antelope Specialist Group 2017 *Tragelaphus eurycerus* ssp. *isaaci*. The IUCN Red List of threatened species 2017:e.T22057A50197212. https://doi.org/10.2305/IUCN.UK.2017_2. RLTS. T22057A50197212.en. Downloaded on 9 Apr 2018.

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

Koressaar T. and 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. and Neufeld J. D. 2012 PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31.

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

Ouborg N. J., Pertoldi C., Loeschcke V., Bijlsma R. and Hedrick P. W. 2010 Conservation genetics in transition to conservation genomics. *Trends Genet.* **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. *Mol. Ecol. Resour.* **8**, 103–106.

Schulte-Hostedde A. I. and 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. *et al.* 2015 Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **30**, 78–87.

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

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

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

Corresponding editor: Umesh C. Lavania