Forensic Science International: Genetics 26 (2017) e25–e27

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Correspondence

Validation studies on dinucleotide STRs for forensic identification of black rhinoceros *Diceros bicornis*

Keywords:
Wildlife forensics
Rhinoceros
Diceros bicornis
STR
Validation

Dear Editor,

Rhinoceros populations are, for the second time in 50 years, undergoing a poaching crisis driven by the illegal trade in rhinoceros horn for use in traditional Asian medicine [1]. By 2013, on average three rhinoceros were illegally killed per day in South Africa [2]. Furthermore, horns have entered the illegal trade from other sources, including thefts from museums, private collections and government stockpiles [1,3]. The application of DNA techniques to assist law enforcement by tracing horn recovered in illegal trade is crucial [4], and yet the forensic validation of STR markers in endangered species is problematic for many reasons.

In choosing markers for an STR panel, tri- or tetranucleotides are preferred [5]. While dinucleotides are highly variable, they present challenges of increased stutter, altered heterozygote balance and non-standard, broken repeats, making them difficult to characterise. Yet, this higher level of variability observed in dinucleotide STRs persists in threatened species which exhibit low genetic variation. All five extant species of rhinoceros have either been through significant and recent bottlenecks or are close to extinction in the wild. Neutral genetic variation is therefore low, and the only published variable candidates for STR panels in this instance are dinucleotides, in spite of substantial effort to find alternative variable STRs (see [6]). Dinucleotide markers are therefore currently in use for rhinoceros wildlife forensic analysis in South Africa [4], and whilst the use of new technologies which enable improved intra-laboratory cooperation are desirable, these are not yet well established and validated within the forensic community. Here we present the results of some population genetic and laboratory based validation studies of STR markers within black rhinoceros, *Diceros bicornis*. These studies illustrate some of the pitfalls of working with these markers, but they also identify candidates that are demonstrably robust and could form part of a global STR panel.

In order for STRs to be used in a forensic panel, they must meet – or at least not overtly violate – population genetic assumptions required for a statistical assessment of the strength of a profile match. They should be unlinked, demonstrate Mendelian inheritance and broadly be in Hardy-Weinberg equilibrium. However, highly structured, small populations that may have been through a significant bottleneck will undoubtedly contravene some of these, and make others difficult to assess for compliance. It is important, therefore, to establish the characteristics of the population allele frequencies for STRs to be used in a forensic panel and discard problematic loci. Furthermore, obtaining samples from highly threatened species is particularly challenging. Therefore, bearing in mind the extremely structured nature of black rhinoceros, with several subspecies [7], this study focuses on one subspecies native to Kenya, *Diceros bicornis michaeli*.

Fifteen dinucleotide STR loci and a sexing marker (see Table S1), also included on a panel characterised by Harper et al.[4] were assessed against population genetic assumptions. The markers were combined into two multiplexes (see Table S2) to enable rapid genotyping of 52 (24 females, 32 males) wild Kenyan *D. bicornis michaeli* samples [8] (see Table S3 for allele frequency data). Between two and nine alleles were identified per locus, and polymorphism information content ranged between 0.173 and 0.841 (Table 1). Two loci were found to deviate from Hardy-Weinberg Equilibrium (BlRh37D $p = 0.0002$ and IR12 $p < 0.0001$; Micro-checker 2.2.3 [9]). Five individuals failed to amplify at BlRh37D despite repeated attempts, suggesting the presence of null alleles. IR12 was found to be in linkage disequilibrium with the sexing marker ZF1 ($p = 0.000001$, α threshold = 0.00041), and all male individuals ($n = 32$) were homozygous. Whilst the *C. simum* genome has not yet been assembled and therefore the chromosomal location of IR12 cannot be formally mapped, it does map to the X chromosome in the horse (*Equus caballus*) which is the closest relative with an assembled genome (Table S4). These multiple lines of evidence suggest that loci BlRh37D and IR12 fail to meet population genetic assumptions of random match probability (RMP) calculations and these markers should be excluded from an STR panel. DB1 and SR63 were also in linkage disequilibrium ($p = 0.0008$, α threshold = 0.00042), and were located on the same genomic scaffold using the preliminary assembly genome of *C. simum* (*GCA_000283155.1*, May 2012; Table S4). Laboratory studies were carried out for both DB1 and SR63 to discern which marker was easier to score, prior to excluding one from any final STR panel.

Laboratory validation studies were carried out on the remaining 13 markers in singleplex (see Table S2) using three blood samples from Port Lympne Wild Animal Park and 13 of the Kenyan samples included in the population study. Sanger sequencing of common alleles showed that fragment length reflected repeat number, with one exception (WR7C 33.2), and nomenclature was proposed accordingly (Table S5). WR7C allele 33.2 possessed an intermediate allele, likely caused by an indel outside of the repeat motif, but the sequence quality of the short flanking regions was too poor to identify this. Markers were highly repeatable across 13 individuals genotyped in triplicate, and typing was consistent both between

http://dx.doi.org/10.1016/j.fsigen.2016.10.016
1872-4973/

Table 1
Descriptive statistics of all loci, including number of alleles observed in *D. b. michael* (Na), expected heterozygosity (HE), observed heterozygosity (HO), polymorphism information content (PIC) and probability of deviation from HWE (significant deviations following Bonferroni correction indicated in bold).

Locus	Na	HE	HO	PIC	HWE (p)
BlRh1B	7	0.747	0.745	0.718	0.5400
BlRh1C	8	0.824	0.750	0.801	0.0418
BlRh37D	3	0.449	0.234	0.405	**0.0002**
BR6	4	0.659	0.673	0.594	0.1833
DB1	6	0.745	0.673	0.701	0.1633
DB23	4	0.564	0.462	0.494	0.0715
DB44	8	0.735	0.808	0.699	0.4952
DB52	9	0.812	0.788	0.787	0.3313
DB66	8	0.858	0.804	0.841	0.0798
IR12	8	0.716	0.269	0.693	**0.0000**
IR22	5	0.521	0.462	0.468	0.0219
SR63	5	0.596	0.538	0.540	0.0310
WR32A	7	0.828	0.750	0.805	0.3241
WR7B	2	0.189	0.173	0.171	0.4390
WR7C	8	0.797	0.723	0.772	0.1241

runs and analysts. All loci were robust to decreased T_A (−2 °C and −4 °C) and increased cycle number (up to +15), but increased T_A (+2 °C and +4 °C) reduced amplification success for some loci (Table S6). While some non-specific amplification was observed at high cycle number, this only occurred away from allelic bins and with a different peak shape from the marker, and therefore scoring results were not affected as the markers were assessed in singleplex. Increasing cycle number may, therefore, cause problems in a multiplex reaction. Sensitivity of the markers to template concentration (5 ng/μL down to 0.15 ng/μL) varied by locus and sample (Table S7), but mean minimum template concentration was below 0.9 ng/μL for all loci except WR7C (2.8 ng/μL). Nine mammalian (white and Indian rhinoceros, horse, badger, dog, fox, human, mink, mouse, otter, rabbit, red deer) and five non-mammalian species (chicken, eagle, shark tuna, turtle), representing closely related species and a broad range of other taxonomic groups which may be encountered in a wildlife forensic laboratory, were tested against the markers. Only WR7B and ZF1 (sexing locus) showed amplification in seven mammals (Table S8). Markers were chosen knowing that they would cross-amplify amongst rhinoceros species [4,6], but testing under our PCR conditions, even when T_A was relaxed to 58 °C, showed reduced amplification success in Indian rhinoceros (*Rhinoceros unicornis*, Table S8).

Precision and accuracy studies demonstrated that allele designation was unambiguous over triplicate genotypes of 13 individuals, despite all STR loci being dinucleotides with increased stutter and allelic drift. The greatest standard deviation of alleles observed in at least seven individuals was 0.164 bp and three times SD was 0.492 bp. Therefore, it would be appropriate to set allelic bins at 0.5 bp either side of the mean. Mean stutter ratio ranged from 15% (DB23) to 49% (BlRh1C) with a maximum observed ratio of 63% (IR22), and mean heterozygote balance was between 0.90 (DB44) and 0.71 (IR22) (Table 2). WR7C showed a large heterozygote imbalance of 0.36 on one occasion when alleles were separated by 26 bp (Table 2). Despite stutter and heterozygote balance being outside the limits typically used for tetranucleotides, allele calling remained consistent and unambiguous. We feel that the use of Liz600 would greatly reduce allelic drift; however, we would strongly advise against the use of markers with 1 bp variants. WR7C is therefore not recommended for inclusion in an STR panel. This locus also suffers from high sensitivity to increased temperature (Table S6), template concentration (Table S7), and the largest observed heterozygote imbalance (Table 2). The linked loci DB1 and SR63 both performed well throughout laboratory studies and ease of scoring was equivalent, thus either would be suitable for an STR panel.

Using the final 12 STRs (using DB1 and excluding SR63 in this instance) and the sexing locus, the probability of identity of unrelated individuals (P_{ID}) and of siblings ($P_{ID\text{-}SIBS}$) of the wild Kenyan samples were calculated in GenAlEx 6.5 [10,11] to be $P_{ID} = 2.6 \times 10^{-11}$ and $P_{ID\text{-}SIBS} = 7.2 \times 10^{-5}$. Cumulative likelihood random match probabilities (RMPs) [12] were calculated for the four wild samples profiled during sensitivity tests. The precise geographic origins of samples included in the population study were unknown, thus a conservative value of $\theta = 0.1$ was used to account for population structure in order to exceed estimates for this population by Muya et al. [8]. RMP values for the full profiles of these individuals ranged between 1.59×10^8 and 1.5×10^{10}.

At present, there is a lack of published markers for rhinoceros which meet the ideal standards for international validation. However, the validation studies we present demonstrate that dinucleotides can pass rigorous assessment. The studies demonstrated that four markers were not suitable for forensic application due to null alleles (BlRh37D), linkage (IR12 and SR63 or DB1) and characterisation issues (WR7C). We highlight the need to avoid loci with 1 bp variants wherever possible due to potential overlap of allelic bins and strongly encourage the development of an allelic ladder following selection of a final panel of STRs to facilitate inter-laboratory calibration. Nevertheless, the robustness of many of these markers to consistent scoring means they should not be discarded. They may not be ideal profiling markers, but in the absence of variable tetranucleotides, which because of their lower mutation rate than dinucleotides may be rare for species with reduced genetic variation, or an alternative profiling system, the validation studies presented here illustrate they can be robust and highly discriminatory for the purposes of individualisation. While

Table 2
Marker characterisation (fragment lengths and repeat motifs determined by allele sequencing) and results of precision and accuracy studies (mean standard deviation (SD) of all alleles at each locus, maximum SD observed in any allele, and mean heterozygote balance).

Marker	Fragment length range (bp)	Repeat motif	Reference allele	Mean SD	Max SD	Mean heterozygote balance
BlRh1B	236–250	$(GT)_9GCA(TG)_{3.1}$	14	0.185	0.316	0.813
BlRh1C	121–137	$(AC)_{15}(GCAC)_2(AC)_{3.1}$	22.1	0.061	0.071	0.732
BR6	141–147	$(CA)_{19}$	19	0.137	0.140	0.713
DB1	149–161	$(CA)_{14}$	14	0.111	0.154	0.828
DB23	241–253	$(CA)_{12}$	12	0.078	0.102	0.812
DB44	205–225	$(AC)_6AC(AC)_7G(CA)_4$	18.1	0.042	0.061	0.897
DB52	207–225	$(CA)_{17.1}$	17.1	0.038	0.056	0.774
DB66	182–208	$(CA)_{15}TA(CA)_8$	24	0.125	0.170	0.757
IR22	210–229	$(CA)_{13}$	13	0.033	0.044	0.705
SR63	193–203	$(AC)_{13}(TC)_3$	16	0.114	0.128	0.709
WR32A	197–209	$(AC)_6CCCCATACGCAA(AC)_{15.1}$	21	0.089	0.112	0.822
WR7B	223–225	$(TG)_{12}A(GT)_6$	18.1	0.144	0.148	0.790
WR7C	137–169	$(TC)_7CC(TC)_{17}(TA)_6$	31	0.120	0.255	0.727

this is unlikely to be the final set of markers used for global traceability of black rhinoceros horn recovered from illegal trade, it provides some insights into the problems faced when validating these tools in a critically endangered, subdivided species.

Acknowledgements

We would like to thank Mike Bruford (Cardiff University), Yoshan Moodley (University of Veterinary Medicine, Vienna), Antoinette Kotze (National Zoological Gardens of South Africa) and Port Lympne Wild Animal Park for helping us source samples for this project. Thanks also to Isa-Rita Russo (Cardiff University) for technical assistance and to Ross McEwing (TRACE) for advice planning a forensic validation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.10.016.

References

[1] T. Milliken J. Shaw, The South Africa—Viet Nam Rhino Horn Trade Nexus: A deadly combination of institutional lapses, corrupt wildlife industry professionals and Asian crime syndicates, TRAFFIC, Johannesburg, South Africa (2012).
[2] T. Milliken, Illegal Trade in Ivory and Rhino Horn: an Assessment Report to Improve Law Enforcement Under the Wildlife TRAPS Project, TRAFFIC, Cambridge, UK, https://www.usaid.gov/sites/default/files/documents/1865/W-TRAPS-Elephant-Rhino-report.pdf (accessed 08.07.16) (2014).
[3] R.H. Emslie, T. Milliken, B. Talukdar, African and Asian Rhinoceroses—Status, Conservation and Trade, CoP16 Doc. 54.2 (Rev. 1) Annex 2, CITES Secretariat, Geneva, Switzerland (2012).
[4] C.K. Harper, G.J. Vermeulen, A.B. Clarke, J.I. De Wet, A.J. Guthrie, Extraction of nuclear DNA from rhinoceros horn and characterization of DNA profiling systems for white (*Ceratotherium simum*) and black (*Diceros bicornis*) rhinoceros, Forensic Sci. Int. Genet. 7 (2013) 428–433, doi:http://dx.doi.org/10.1016/j.fsigen.2013.04.003.
[5] A. Linacre, L. Gusmão, W. Hecht, A.P. Hellmann, W.R. Mayr, W. Parson, M. Prinz, P. Schneider, N. Morling, ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations, Forensic Sci. Int. Genet. 5 (2011) 501–505, doi:http://dx.doi.org/10.1016/j.fsigen.2010.10.017.
[6] C. Scott, Microsatellite Variability in Four Contemporary Rhinocerous Species: Implications for Conservation, Queen's University, 2008.
[7] R. du Toit, African rhino systematics—the existing basis for subspecies classification of black and white rhinos, Pachyderm 9 (1987) 1–2.
[8] S.M. Muya, M.W. Bruford, A.W.T. Muigai, Z.B. Osiemo, E. Mwachiro, B. Okita-Ouma, B. Goossens, Substantial molecular variation and low genetic structure in Kenya's black rhinoceros: implications for conservation, Conserv. Genet. 12 (2011) 1575–1588, doi:http://dx.doi.org/10.1007/s10592-011-0256-3.
[9] C. Van Oosterhout, W.F. Hutchinson, D.P.M. Wills, P. Shipley, MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data, Mol. Ecol. Notes 4 (2004) 535–538, doi:http://dx.doi.org/10.1111/j.1471-8286.2004.00684.x.
[10] R. Peakall, P.E. Smouse, GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research, Mol. Ecol. Notes 6 (2006) 288–295, doi:http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x.
[11] R. Peakall, P.E. Smouse, GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update, Bioinformatics 28 (2012) 2537–2539, doi:http://dx.doi.org/10.1093/bioinformatics/bts460.
[12] B.S. Weir, Statistics for wildlife forensic DNA, in: J.E. Huffman, J.R. Wallace (Eds.), Wildl. Forensics Methods Appl., Wiley-Blackwell, Chichester, 2012, pp. 237–252, doi:http://dx.doi.org/10.1002/9781119953142.ch12.

K.L. Dicks
Biological Sciences, University of Chester, Chester, UK

L.M.I. Webster*
Science and Advice for Scottish Agriculture, Edinburgh, UK

I. McDowall
Biological Sciences, University of Chester, Chester, UK

S.M. Muya
Department of Zoology, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

J. Hopper
Aspinall Foundation, Port Lympne Wild Animal Park, Hythe, UK

P. O'Donoghue
Biological Sciences, University of Chester, Chester, UK

* Corresponding author.
E-mail address: Lucy.Webster@sasa.gsi.gov.uk (L. Webster).

Received 19 August 2016

Available online 24 October 2016