Supplemental Information

Robust forensic matching of confiscated horns to individual poached African rhinoceros

Cindy Harper, Anette Ludwig, Amy Clarke, Kagiso Makgopela, Andrey Yurchenko, Alan Guthrie, Pavel Dobrynin, Gaik Tamazian, Richard Emslie, Marile van Heerden, Markus Hofmeyr, Roderick Potter, Johannes Roets, Piet Beytell, Moses Otiende, Linus Kariuki, Raoul du Toit, Natasha Anderson, Joseph Okori, Alexey Antonik, Klaus-Peter Koepfli, Peter Thompson, and Stephen J. O'Brien

SUPPLEMENTAL METHODS

SUBJECT DETAILS

A total of 883 black rhinoceros (*Diceros bicornis*) and 3,085 white rhinoceros (*Ceratotherium simum*) samples were included in this study (Supplemental Figure S1A). Tissue specimens were submitted to the Veterinary Genetics Laboratory (VGL), University of Pretoria as part of the RhODIS® (Rhinoceros DNA Index System) database from various African rhinoceros range states. Recognized living subspecies of black rhinoceros are included: 1) 51 samples of the Eastern subspecies (*D.b. michaeli*); 2) 357 samples of the Southwestern subspecies (*D.b. bicornis*); and 3) 475 samples of the South-Central subspecies (*D.b. minor*). The white rhinoceros has two extant subspecies, the Southern white rhinoceros (*C.s. simum*) and the Northern white rhinoceros (*C.s. cottoni*) [S1]. Because only 3 animals remain of the Northern white subspecies they were not included in this study.

The majority of samples submitted to RhODIS® were collected in tamper-proof evidence collection bags which were produced and distributed by the VGL to environmental crime management authorities in South Africa and other countries. These sample kits contain multiple sample containers for tissue, horn, hair, blood (EDTA vacuum tubes) and toenail samples; in most cases multiple samples of each animal were collected. VGL samples were collected during forensic investigation of poaching scenes (~47%), during routine translocation, notching, dehorning for identification or hunting (~49%) or from rhinoceros horn stockpile identification operations (~4%) according to the RhODIS® guidelines. Samples received in the VGL were assigned individual barcode sample numbers. Sample quality varied from highly degraded, particularly in the case of samples from old carcasses, to highest quality blood samples from live animals.

METHOD DETAILS

DNA extraction and STR genotyping

Genomic DNA (gDNA) was extracted from blood and tissue samples using the Prepfiler®kit (ThermoFisher Scientific) according to the manufacturer's instructions [S2]. Hair samples were extracted using a modified alkaline extraction method [S3], where 1 to 3 hair roots were cut into a 1.5 ml tube and 100 µl of 0.2M NaOH was added and heated at 97°C for 15 minutes, following which 100µl of 0.2M Tris-HCl at pH 8.5 was added. PCR was performed in four multiplex reactions, using the 22 loci previously described [S2], with the addition of the Rh12 locus in multiplex 2 (Forward Primer, CTGGTGCATTCATCAGGGCT, Reverse Primer, AGAAGAGGTAGGAGAGGAAGTCA) (https://www.ncbi.nlm.nih.gov/nuccore/37496513) and the zinc finger (ZF) locus which was used to determine the gender of the animal from which the sample originated [S4]. STR analysis was performed using 4 multiplex reactions with between 5 and 8 markers included in each multiplex. Extracted DNA (1 µl diluted to approximately 30 ng/µl or undiluted at less than 30 ng/µl) was added to a PCR mastermix consisting of 5 µl of KAPA2G Fast Multiplex PCR Kit (Sigma-Aldrich) and 4 µl of primer mix in a 10 µl reaction volume. PCR was performed using a thermal cycler (GeneAmp® PCR System 9700, Life Technologies) with cycling conditions standardized as follows: 3 min at 95°C, 30 cycles of 95°C for 15 s, 60°C

for 30 s and 72°C for 30 s followed by an extension step at 72°C for 10 min. PCR product (0.5 µl) was loaded with 10 µl Hi-Di™ formamide and 0.25 µl GeneScan™ 500 LIZ® size standard (Life Technologies) and run on a 3130xl or 3500xl Genetic Analyzer (Applied Biosystems) and data transferred to a personal computer and analyzed using STRand software (University of California, Davis) [S5]. A set of bins for each locus within the four different panels were set up in STRand using fixed bin sizes to determine and standardize the allele calls between samples. Known control samples for both black and white rhinoceros were included with each sample set that was run to ensure the accuracy of allele calls between runs.

Chromosome assignment imputation

The rhinoceros chromosome position of each locus was imputed based upon identifying the primer and flanking sequence in the whole genome sequence of the Southern white rhinoceros (*C.s. simum*) (http://www.ncbi.nlm.nih.gov/genome/24631) and then using a reference assisted chromosome assembly of the white rhinoceros scaffolds aligned against the domestic horse (*Equus ferus caballus*) genome [S6] using Chromosomer [S7]. The chromosome assignment, albeit indirect involving two distantly related Perissodactyla species, allowed for an indication of likely chromosome linkage in detecting linkage disequilibria between STR loci that should be independent in a forensic analysis.

QUANTIFICATION AND STATISTICS

Measures of genetic diversity

Allele frequencies were obtained using Cervus (Version 3.03) [S8] and number of alleles (Na), number of effective alleles (Ne), observed heterozygosity (Ho) and expected heterozygosity (He) per locus (using the unbiased formula of Nei [S9]), the fixation index (F) and Hardy-Weinberg Equilibrium (HWE) using the Bonferroni correction were calculated in Cervus for all loci [S8]. Population structure was examined using three approaches: 1) an individual-based tree was constructed using NEIGHBOR of PHYLIP package [S10] based on allele-sharing, DPS (Proportion of shared alleles) [S11] distance-matrix generated in MSA 4.05 software [S12] with 1-DPS correction and visualized in FigTree software [S13], 2) Principal Component Analysis was performed in PAST 3 software [S14] using a variance-covariance matrix; before the analysis each allele for every loci was labelled as 0, 0.5, or 1 for allele absence, heterozygote or homozygote in a given individual and 3) we detected population partitions using the STRUCTURE algorithm which clusters individuals with minimal deviation from genetic and linkage equilibrium [S15]. For Supplemental Figure S1B to C, sex-linked markers and three loci with missing data were excluded leaving 18 loci (32A, DB44, 7B, 7C, BlRh1B, DB52, BR6, DB1, BlRh1C, 12F, BlRh37D, 32F, DB23, SR63, IR10, IR22, SR262, SR268). For STRUCTURE, K-values were evaluated for K=2 to K=8, with a burn-in of 50,000 iterations and 500,000 iterations at each value of K. Each K was run 10 times. The division of the black subspecies is supported by the Delta K value calculated in STRUCTURE Harvester using the Evanno method [S16]. Between population differentiation (F_{ST}) was determined using GenAlEx 6.5 [S17]. Differences between the pairwise F_{ST} were tested for significance using GENODIVE [S18] with 1000 permutations.

Forensic match application

Matching of specific DNA profiles provide evidentiary support that two samples are derived from the same individual if underlying data are available to permit an estimate of the rarity of the profile [S19]. Single locus match probability was calculated using the formulae of Balding and Nichols [S20]:

$$\Pr(A_u A_u | A_u A_u)_i = \frac{[2\theta_i + (1-\theta_i)P_u][3\theta_i + (1-\theta_i)P_u]}{(1+\theta_i)(1+2\theta_i)}$$

$$\Pr(A_u A_v | A_u A_v)_i = \frac{2[\theta_i + (1-\theta_i)P_u][\theta_i + (1-\theta_i)P_v]}{(1+\theta_i)(1+2\theta_i)}, u \neq v$$

The cumulative match probability across several STR loci was then obtained by the product rule. The between population allelic variation was quantified by F_{ST} and based on these data a Theta (θ) value of 0.1 was selected for use in calculation of match probability of African rhinoceros species in this study. The use of this sufficiently conservative θ also compensates departures from Hardy-Weinberg equilibrium at specific loci due to allelic variation in populations sampled [S21].

DATA AVAILABILITY

Genotypes are available at Mendeley Data Repository with the DOI: http://dx.doi.org/10.17632/d4tcjyxck6.1.

SUPPLEMENTAL RESULTS

A summary of the analyses of genetic population structure are depicted in Supplemental Figures S1B, C and D. These data support the recognition of one white rhinoceros subspecies (*Ceratotherium simum simum*), and three black rhinoceros subspecies, *D.b. bicornis*, *D.b. michaeli* and *D.b.minor*, with significant partitioning of the Zimbabwe versus KZN *D.b. minor* populations with animals derived from the Kruger National Park (KNP) being an admixture of these two. A STRUCTURE population analysis affirmed the genetic distinctiveness between the three subspecies of black rhinoceros and between the two *D.b. minor* groups (Supplemental Figure S1D). STRUCTURE (at K=4) resolves the Zimbabwe, the KZN and the admixed KNP population as a third distinct population. The relationship between the groups is presented in a neighbor joining tree (Supplemental Figure S1B). This tree shows that the black rhinoceros, *D.b .michaeli* subspecies, is basal to the more recently diverged *D.b. minor* and *D.b. bicornis* sister subspecies.

Supplemental Table S1 provides examples of the calculation of cumulative match probabilities for specific matching DNA profiles obtained from samples from two separate seizures of horns at airports and from the carcasses of the individual white and black rhinoceroses to which the horns were matched. Cumulative match probabilities for the white rhinoceros species were calculated excluding the monomorphic (BlRh37D, DB23 and IR22) and X-linked (IR12 and SR74) loci. Most STR loci within the three black rhinoceros subspecies diverged from HWE when the three subspecies were pooled due to the population subdivision. However, when the black rhinoceros subspecies and populations identified as panmictic were assessed separately, the polymorphic STR loci conformed to Hardy-Weinberg Equilibrium. Match probability calculations for black rhinoceros were calculated excluding the monomorphic (32F) and X-linked loci (IR12 and SR74).

SUPPLEMENTAL DISCUSSION

The overall genetic uniformity and panmixia of the white rhinoceros (*C.s. simum*) (Supplemental figure S1B and C) which comprise over 90% of the criminal cases received by the VGL, would allow forensic application of the product rule for this as a single species. The large dataset in this study confirmed the utility of monomorphic loci for species identification. The genotypic data, therefore, allows the assignment of an unknown sample as black or white rhinoceros. STRUCTURE analysis provides strong support for the classification of a sample into the three recognized black rhinoceros subspecies. DNA profile matches are made using all amplified loci and comparing the DNA profile to all genotype data on the RhODIS® database and confirmed manually and with electropherogram data. Match probabilities for specific white and black rhinoceros matches are done using the species specific allele frequencies and a conservative Theta (θ) of 0.1

following The Second National Research Council report on forensic DNA evidence recommendation 4.2 for estimating random match probabilities in human populations [S19]. As different loci are informative in white and black rhinoceros species and black rhinoceros subspecies or populations the 23 STR loci described here are the minimum set of markers that should be used for DNA forensic investigations for African rhinoceros.

Regulations for marking of rhinoceros and rhinoceros horns, under the National Environmental Management: Biodiversity Act (10/2004) were published in the South African Government Gazette in April 2012 [S22]. The regulations instruct that all rhinoceros should be sampled for DNA profiling when they are captured for identification, translocated or hunted and further that all stored rhinoceros horn is sampled. Tissue specimens must be sampled in specific kits and the DNA genotypes are to be added to the RhODIS® database. Reports are issued for forensic cases in which horns or horn products are recovered and linked to a specific carcass or where tools used in poaching incidents are recovered and associated blood traces linked to a carcass. The CITES (Convention on International Trade in Endangered Species of Fauna and Flora) Conference of Parties in Bangkok, Thailand, 3-14 March 2013, recommended (CoP16 Com II.24) that all CITES signatory countries should sample confiscated rhinoceros horn and submit this to an accredited forensic laboratory for DNA analysis. This imperative underpins the need to ensure that match probability estimations using a robust and uniform database are established in support of all international investigations. Robust, statistically significant genotype matches, prosecution, conviction and sentencing of wildlife traffickers in multiple cases validates the DNA matching approach and with sufficient public disclosure could discourage future crimes against rhinoceros species.

Author Contributions

Conceived and designed the experiments, CH, AG, PT, SJO; performed the experiments, AL, AC, KM; performed the data analysis, CH, AY, AG, PD, GT, AA, K-PK, SJO; field data and sample collection, RE, MvH, MH, RP, JR, PB, MO, LK, RdT, NA, JO; biobanking, CK, AL, AC, KM; wrote the manuscript, CH, AG, PT, SJO. All authors reviewed the manuscript.

ACKNOWLEDGEMENTS

The authors thank the many rhino management and law enforcement officers that facilitated, often under challenging circumstances, the collection of samples that enabled this study. George Duncan for review of an early draft of the manuscript. SANParks (South African National Parks), Ezemvelo KZN Wildlife, Kenya Wildlife Services, Lowveld Rhino Trust, Ministry of Environment and Tourism, Namibia, SANParks Honorary Rangers, WWF (World Wildlife Fund) African Rhino Program, Peace Parks Foundation, Castle Lager Boucher Legacy NPC and United Nations Environmental Programme - Global Environmental Facility (UNEP-GEF) through the South African Department of Environmental Affairs for support and funding of this study. SJO, AY, K-PK, and AA were supported, in part, by the Russian Science Foundation grant (project No. 17-14-01138) and by St. Petersburg State University (Genome Russia Grant no. 1.52.1647.2016).

SUPPLEMENTAL REFERENCES

S1. Emslie, R.H., and Knight, M.H. (2014). Update on African rhino status and trends: from IUCN SSC African Rhino Specialist Group (AfRSG). Report to CITES Standing Committee 65th Meeting.

S2. Harper, C.K., Vermeulen, G.J., Clarke, A.B., De Wet, J.I., and Guthrie, A.J. (2013). 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*, 428-433.

S3. Rudbeck, L., and Dissing, J. (1998). Rapid, simple alkaline extraction of human genomic DNA from whole blood, buccal epithelial cells, semen and forensic stains for PCR. Biotech *25*, 588-592.

S4. Peppin, L., McEwing, R., Ogden, R., Hermes, R., Harper, C., Guthrie, A., and Carvalho, G.R. (2010). Molecular sexing of African rhinoceros. Conserv Genet *11*, 1181-1184.

S5. Toonen, R.J., and Hughes, S. (2001). Increased throughput for fragment analysis on an ABI Prism® 377 automated sequencer using a membrane comb and STRand software. Biotech *31*, 1320-1324.
S6. Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., et al. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. Science *326*, 865-867.
S7. Tamazian, G., Dobrynin, P., Krasheninnikova, K., Komissarov, A., Koepfli, K.-P., and O'Brien, S.J. (2016). Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences. GigaScience *5*, 38.
S8. Kalinowski, S.T., Taper, M.L., and Marshall, T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol *16*, 1099-1106.
S9. Nei, M. (1987). Molecular evolutionary genetics, (New York: Columbia University Press).
S10. Felsenstein, J. (2005). PHYLIP (Phylogenetic Inference Package) version 3.6. (Department of Genome Sciences, University of Washington, Seattle).
S11. Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., and Cavalli-Sforza, L.L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. Nature *368*, 455-457.
S12. Dieringer, D., and Schlötterer, C. (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes *3*, 167-169.
S13. Rambaut, A. (2017). FigTree v1.4.3: Tree figure drawing tool.
S14. Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001). PAST: Palaeontological statistics software package for education and data analysis. Palaeontologia Electronica *4*, 9.
S15. Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics *164* 1567-1587.
S16. Earl, D.A., and von Holdt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour *4*, 359-361.
S17. Peakall, R., and Smouse, P.E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - An update. Bioinformatics *28*, 2537-2539.
S18. 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.
S19. Weir, B.S. (1996). The second National Research Council report on forensic DNA evidence. Am J Hum Genet *59 (3)*, 497–500.
S20. Balding, D.J., and Nichols, R.A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica *96*, 3-12.
S21. Buckleton, J., Curran, J., Goudet, J., Taylor, D., Thiery, A., and Weir, B.S. (2016). Population-specific FST values for forensic STR markers: A worldwide survey. Forensic Sci Int Genet *23*, 91-100.
S22. Anon., (2012). National Environmental Management: Biodiversity Act (10/2004): Norms and standards for the marking of rhinoceros and rhinoceros horn, and for the hunting of rhinoceros for trophy hunting purposes. In South African Government Gazette, Volume 562. pp. 3-7.

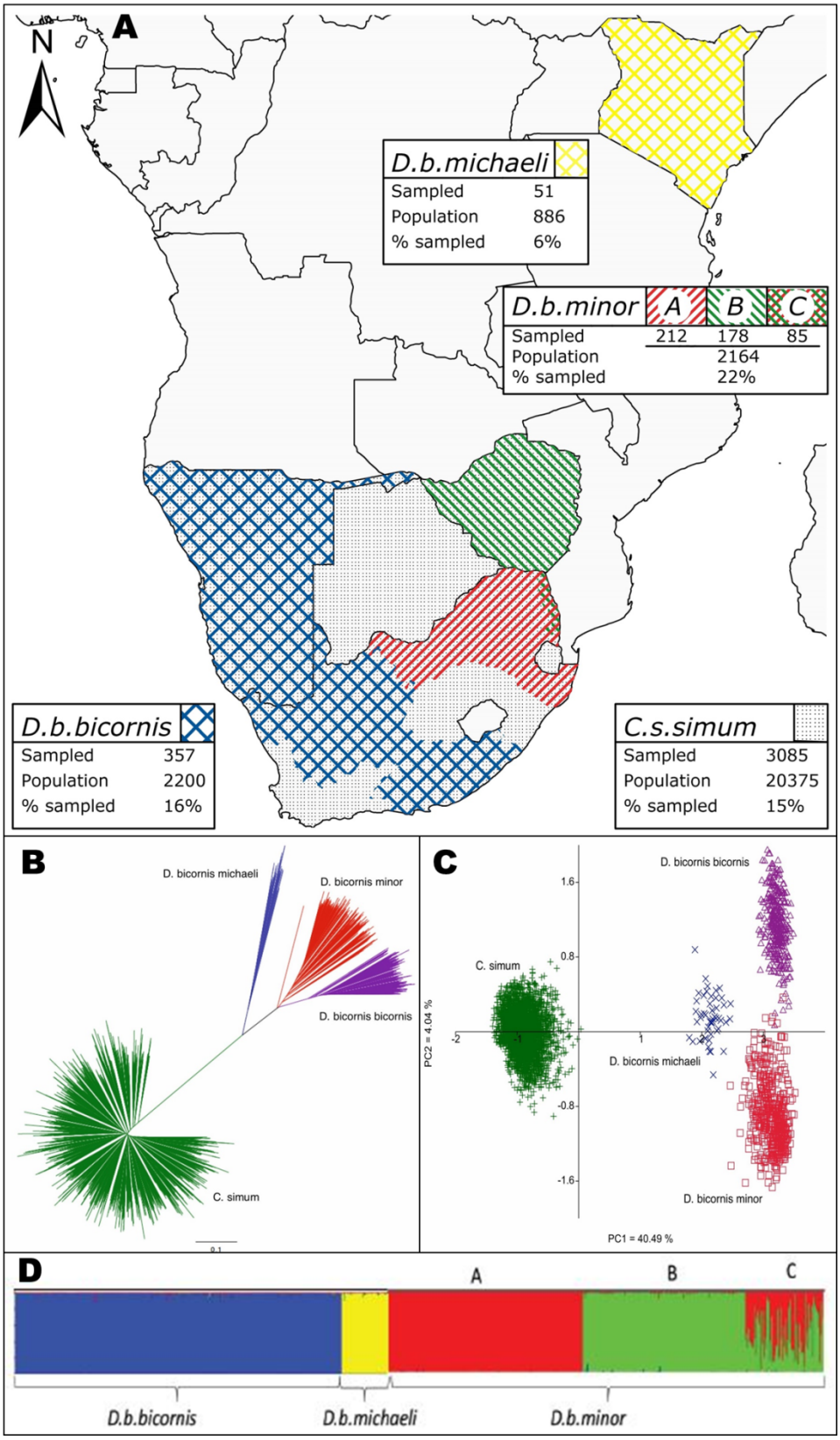

Figure S1: Map of distribution of rhinoceros species investigated with summary of population genetic analyses. A) Geographic origin, number and proportion of samples from the rhinoceros species and subspecies included in this study. B) Tree and C) Principle component analysis showing that the white rhinoceros (*C. simum simum)* comprises a single panmictic subspecies with the black rhinoceros subdivided into three subspecies *D.b. bicornis*, *D.b. michaeli* and *D.b. minor.* D) The STRUCTURE diagram supports the black rhinoceros subspecies subdivision with an additional partition of the *D.b. minor* subspecies originating in KwaZulu-Natal (KZN) (A), Zimbabwe (B), and a third group originating from the Kruger National Park that are an admixture of these two groups (C).

	Locus	White rhinoceros[a]						Black rhinoceros[b]					
		Al 1	Al 2	AF 1	AF 2	LMP	CMP	Al 1	Al 2	AF 1	AF 2	LMP	CMP
1	7B	229	229	0.26	0.26	0.18	1.80×10^{-1}	225	225	0.1	0.1	0.09	9.00×10^{-2}
2	7C	150	150	0.66	0.66	0.54	9.72×10^{-2}	158	168	0.38	0.12	0.14	1.26×10^{-2}
3	BlRh1B	240	240	0.79	0.79	0.7	6.80×10^{-2}	242	242	0.82	0.82	0.73	9.20×10^{-3}
4	BR6	132	150	0.64	0.21	0.29	1.97×10^{-2}	134	156	0.57	0.06	0.14	1.29×10^{-3}
5	DB1	160	162	0.86	0.14	0.3	5.92×10^{-3}	158	158	0.53	0.53	0.4	5.15×10^{-4}
6	BlRh1C	136	138	0.72	0.28	0.4	2.37×10^{-3}	126	128	0.47	0.17	0.2	1.03×10^{-4}
7	SR63	188	190	0.58	0.42	0.45	1.07×10^{-3}	194	194	0.4	0.4	0.28	2.88×10^{-5}
8	DB52	217	221	0.31	0.33	0.23	2.45×10^{-4}	215	219	0.14	0.44	0.17	4.90×10^{-6}
9	DB66	202	202	0.24	0.24	0.16	3.92×10^{-5}	198	208	0.09	0.2	0.07	3.43×10^{-7}
10	SR268	176	180	0.18	0.79	0.32	1.25×10^{-5}	196	196	0.27	0.27	0.18	6.18×10^{-8}
11	32A	196	196	0.5	0.5	0.36	4.52×10^{-6}	200	200	0.85	0.85	0.78	4.82×10^{-8}
12	DB44	218	218	0.76	0.76	0.66	2.98×10^{-6}	212	212	0.36	0.36	0.25	1.20×10^{-8}
13	12F	223	223	0.53	0.53	0.4	1.19×10^{-6}	229	235	0.2	0.07	0.07	8.43×10^{-10}
14	IR10	132	134	0.3	0.7	0.41	4.89×10^{-7}	136	136	0.73	0.73	0.63	5.31×10^{-10}
15	SR262	92	92	0.25	0.25	0.17	8.31×10^{-8}	98	100	0.45	0.11	0.15	7.97×10^{-11}
16	SR281	234	234	0.36	0.36	0.25	2.08×10^{-8}	238	238	0.8	0.8	0.71	5.66×10^{-11}
17	RH12	108	108	0.79	0.79	0.7	1.45×10^{-8}	124	124	0.35	0.35	0.24	1.36×10^{-11}
18	32F	175	175	0.13	0.13	0.1	1.45×10^{-9}	-	-	-	-	-	-
19	BlRh37D	-	-	-	-	-	-	201	203	0.6	0.39	0.44	5.98×10^{-12}
20	DB23	-	-	-	-	-	-	247	247	0.72	0.72	0.61	3.65×10^{-12}
21	IR22	-	-	-	-	-	-	217	217	0.19	0.19	0.13	4.74×10^{-13}

Table S1: Summary of cumulative match probability calculation for representative white and black rhinoceros cases where horns were seized at an airport and matched back to a poached carcass. Al1 – Allele 1, Al 2 – Allele 2, AF 1 – Species specific frequency of Allele 1, AF 2 – Species specific frequency of Allele 2, LMP – Locus specific match probability calculated using a conservative θ of 0.1 [S20], CMP – Cumulative match probability calculated using product rule. [a] and [b] refer to specific cases listed in Table 1.