Conservation Genetics Resources (2026) 18:48
https://doi.org/10.1007/s12686-026-01452-7

METHODS AND RESOURCES

From genome to field: a SNP panel for non-invasive studies of leopard (*Panthera pardus*)

Faruk Mamugy[1,2] · Laura D. Bertola[3] · Helena Königsson[1] · Amber Mertens De Vry[1] · Nicolas Dussex[4] · Bastian Schiffthaler[5] · Johanna L. A. Paijmans[7,8] · Michael Hofreiter[6] · Ryan Forbes[1,9] · Graham I. H. Kerley[9] · Kris Everatt[9,10,11] · Matthew Becker[12] · Scott Creel[12,13] · Stéphanie Bourgeois[14] · Twakundine Simpamba[15] · Clive Chifunte[15] · Reuben Kabungo[12] · Hana Raza[16] · Nabaz Horeni[16] · Bahez Farooq[16] · Marine Drouilly[11,17] · Göran Spong[1,12]

Received: 6 May 2026 / Accepted: 4 November 2026
© The Author(s) 2026

Abstract

To maximize the conservation impact of genetic resources, they should be applicable across the full distribution of a species, economically and logistically feasible, and validated for the diversity of sample types collected in the field. This is particularly relevant for large carnivores, which are elusive and wide-ranging, and whose individual movements and population dynamics often transcend administrative borders. Moreover, obtaining high-quality samples from such species is frequently challenging, necessitating the use of non-invasive samples of low quality. Here, we present a species-specific SNP panel for leopards (*Panthera pardus*). We used whole-genome data from across the global range as well as RAD sequencing data from Africa to identify candidate markers for assay development. Markers were validated using samples across seven African countries, and the final selection was based on genetic variation and performance on non-invasive samples. The resulting 96-assay (markers) panel comprises 85 autosomal, three X-linked, two Y-linked, one sex-associated, and five mitochondrial markers. The selected marker set exhibits variation across the species' global range (with an African focus), providing high power for individual identification ($PID_{sib} = 8.75 \times 10^{-18}$) and larger scale geographic assignment supported by a growing baseline of 649 individuals from 937 samples. Genotypes generated by this panel are straightforward to quality control and interpret without computationally demanding bioinformatic analyses. This makes the assay a cost-effective, accessible, and robust resource for leopard monitoring, conservation, research, and forensic applications.

Keywords SNP genotyping · Non-invasive genetic · Genetic monitoring · Individual identification · Capture–mark–recapture · Wildlife forensics

Introduction

Genomes hold vast amounts of information. By extracting parts of this information, individuals, their kin, behaviors, population features and history, evolutionary patterns and more can be revealed. Since their inception, genetic applications have increasingly been deployed to study wild populations and address questions difficult to assess via alternative means (Nichols and Spong 2017; Creel et al. 2019; Hohenlohe et al. 2021). As methods improve and costs decrease, the allure of genetic data persists. However, establishing protocols for extracting reliable genetic data from wild populations may require substantial investment. Once established, the protocol must negotiate cost, logistics, and sample specifics to accommodate sample influxes that may be constrained by time or unpredictable in size and quality. This makes it challenging to optimize the use of resources while providing predictable timelines and outputs (Bertola et al. 2024).

Minute quantities of DNA, sometimes collected non-invasively from the environment, permit the extraction of complete genotypes from individuals without direct

Extended author information available on the last page of the article

Published online: 17 November 2026

observation (Thomsen and Willerslev 2015). This possibility can be particularly useful when studying elusive species living at low population densities or to minimize disturbance to the subjects. Since sample collection can be done with simple means, it greatly facilitates large-scale collection efforts by management officials or the public (e.g. Giangregorio et al. 2019; Spitzer et al. 2016; Norman and Spong 2015). While relatively easy to collect, such as in the form of fecal samples, non-invasive sample sets typically include samples at the detection threshold for molecular analysis. Such samples have low amounts of intact DNA, and this may lead to difficulties in extracting data of sufficient quality. As for all data, genetic data also hold errors that may affect downstream inferences if not appropriately controlled (Creel et al. 2003). By running samples multiple times, the prevalence and manifestation of errors can be mapped. In some cases, building consensus genotypes may be a way to generate genotypes of sufficient quality and resolution (Solberg et al. 2006; Waits et al. 2001), but this method requires many replicates to make it statistically sound, making it demanding in resources (note that more recent sequencing applications such as GT-seq (amplicon sequencing; Campbell et al. 2015) in essence build a consensus genotype based on fragment prevalence). An alternative method is to use amplification success as a proxy for quality and simply discard samples below an empirically determined threshold, but for small sample sets this may not be an option. Whichever method is chosen to build genotypes, statistical validation of the final genotypes and their error rates should be performed. If genotyping error is underestimated, estimates of standing genetic variation or population size may be inflated (Creel et al. 2003). It might also cause loss of power to detect population structure and affect the precision of provenance assignments.

Any genetic marker panel should also be established for the population of interest, ideally by including samples from across the range from the very beginning of the development process. However, due to financial, logistical, and administrative constraints, such broadscale and transboundary sampling efforts remain rare, especially for wide-spread species (Kopatz et al. 2021). Yet, if the development of genetic markers only holds for a subset of the population, other subsets later included may suffer from incorrect inferences due to ascertainment bias. At its most extreme, lack of variation may prohibit identification of individuals. For species with a large, sometimes fragmented, population range, a lack of common genotype variation between subpopulations may thus limit the utility of a resource originally developed for a part of the species range (e.g. Giangregorio et al. 2019). But even when variation is sufficient for individual identification, many population genetic methods use the distribution of genetic variation to infer a range of processes such as dispersal and population size. Since such methods are generally based on the fundamental principles of population genetics (Wright 1931), differences caused by marker ascertainment may lead to erroneous inferences, directly affecting estimates of connectivity, population size and history (Palsbøll et al. 2013).

The leopard (*Panthera pardus*) has the most extensive geographic distribution of all large cats, with subspecies ranging across a variety of habitats from the African continent and the Arabian Peninsula, to the Southeast Asian islands and as far north as Eastern Russia. Eight leopard subspecies are recognized under the revised felid taxonomy of the IUCN SSC Cat Specialist Group, with African populations assigned to the nominate subspecies *Panthera pardus pardus* (Kitchener et al. 2017). Like most carnivore species, ranges have shrunk and become increasingly fragmented due to habitat loss, persecution, and illegal trade (Ripple et al. 2014). Leopard populations thus currently occupy only a fraction of their historic range (Jacobson et al. 2016). While its secretive nature and adaptability have permitted the leopard to eke out a living also in areas heavily dominated by humans (Pečnerová et al. 2021), the species is classified as Vulnerable in the 2024 IUCN Red List assessment (Stein et al. 2024). Population numbers and trends nevertheless remain poorly quantified across much of the range. There is therefore an urgent need for better monitoring tools to support conservation, management and law enforcement activities (Ripple et al. 2014).

Here we present a 96-marker SNP panel developed for non-invasive monitoring of leopards using geographically broad discovery datasets and validation samples from Africa. The panel combines autosomal, Y- and X-linked markers, and mitochondrial markers. They were selected for variation and performance on non-invasive material, and the resulting panel supports individual identification and description of population genetic structure.

Materials & methods

Our SNP-panel development combined an existing genomic dataset with new data to broaden geographic representation and support population-genetic analyses. First, we used SNP variants from published whole-genome data for 26 individuals, including historical samples, from across the leopard's range (Paijmans et al. 2021). Second, we identified SNPs by RAD sequencing of 152 tissue samples collected by the Zambian Carnivore Programme. We intersected these datasets to identify shared markers represented in the geographically broad whole-genome dataset. Third, final marker selection and performance assessment used microfluidic SNP genotyping (Fluidigm Biomark or Standard BioTools

X9) with tissue ($N=580$), fecal extracts ($N=295$), seizures, swabs, bone, teeth ($N=62$). After removal of duplicated samples from the same individual, high-quality genotypes of 649 individual leopards from seven African countries and Iraq now form a growing baseline.

Sample collection

The data and samples included in this study come from a range of projects and initiatives. Research tissue samples were collected during ecological field studies by biopsy darting or during handling (e.g. to fit radio collars or remove wire snares). All animal handling, sample collection, and shipping were done under relevant permits renewed yearly or issued prior to sample exports (e.g. ethical approval, code of conduct, animal handling, CITES, Nagoya, and other government approvals). We have also included samples collected by law enforcement and organizations fighting wildlife crime. These were taken from specimens of skin, bones, and teeth known or suspected to be leopard. Scat samples were collected as per Forbes et al. (2026).

Whole genome sequencing

For full details on the whole genome sequencing see (Paijmans et al. 2021). From these samples, a total of just under 1.4 M SNPs were identified. In short, we used the vcf output from the whole genome sequencing to identify SNPs with global variation that overlapped with the SNPs found by the RAD sequencing below.

RAD sequencing

Samples used for RAD sequencing ($n=152$) were extracted manually at the Luangwa field site of the Zambian Carnivore Programme, using the DNeasy Blood & Tissue Kit (Qiagen) or in our laboratory at SLU (Sweden) by a Qiagen Symphony extraction robot using the Qiagen Tissue Extraction kit according to the manufacturer's instructions. DNA quantity and purity were assessed by agarose gel and NanoDrop (Thermo Fisher Scientific).

We digested 0.5 µg DNA with the restriction enzyme EcoRI (Thermo Fisher Scientific). Digestion quality was visualized by gel electrophoresis. We submitted samples to the National Genomics Infrastructure (NGI, SciLifeLab, Stockholm), where fragments of 400–800 bp were excised for paired-end library construction (2×150 bp) and RAD-sequencing on an Illumina NovaSeq 6000.

SNP detection and filtering

Sequence quality of the demultiplexed RAD-seq reads was assessed using FastQC (Andrews 2010). Reads were trimmed and adapters removed with Trimmomatic v0.39 (Bolger et al. 2014) in paired-end mode; only reads longer than 100 bp after trimming were retained. De novo SNP detection used Stacks v2.53 (Rochette et al. 2019), with the pipeline components run individually. To compare RAD SNPs with the whole-genome dataset, Stacks catalogue sequences were aligned to the domestic-cat reference genome Felis_catus_6.2 using minimap2 (Li 2018), and RAD SNPs were projected onto reference coordinates. Overlapping SNPs were retained when their positions and reference and alternate alleles matched. Stacks identified 6,931 SNPs. Custom R scripts retained biallelic SNPs called in at least 75% of samples, with one SNP per retained RAD locus and at least 30 bp of flanking sequence on each side. Autosomal SNPs were required to have MAF > 0.2. At least five samples had to represent each genotype (aa, ab and bb), and the SNP had to be present in the geographically broad whole-genome dataset. These filters yielded 148 candidates. Of these, 133 passed in silico assay design by AH diagnostics and were taken forward for laboratory assessment on a Fluidigm Biomark or Standard BioTools X9. The final selection favoured assays with distinct genotype clusters across sample types.

Sex-associated markers comprised SMCY_147, SMCY_448 and yPp_AF2_416 (Table 1, S2, and S3). Sex assignments used the joint pattern of the three markers. Profiles were assigned as female when both SMCY_147 and SMCY_448 had no calls and yPp_AF2_416 was coded YY, and as male when both SMCY markers amplified and yPp_AF2_416 was coded XX. Other patterns were classified as unknown. Here, XX, XY and YY denote marker allele codes, not chromosomes. Calls incompatible with the assumed ploidy were excluded from the affected locus-level statistics after sex assignment. Five mitochondrial assays were included to investigate matrilineal variation and assist species confirmation, particularly for fecal samples.

SNP validation

The selected autosomal and X-linked markers were initially assessed by genotyping 119 leopard samples, two lion samples and one cheetah sample on the Standard BioTools Biomark platform, following the manufacturer's chemistry instructions and using 40 cycles for non-invasive material. The mitochondrial and sex-associated markers were tested in 31 leopards, 59 lions and one cheetah, including nine leopards of known sex (five females and four males). Positive controls (sample replicates) and no-template controls

Table 1 Genetic diversity of the 96-marker SNP panel

Marker class	No. SNPs	Valid *n* per SNP	Mean MAF	Mean H_E	Mean H_O	Mean h_{SNP}	Mean F_{ST}
Autosomal	85	575–649	0.334	0.427	0.394	NA	0.055
X-linked	3	461–468	0.408	0.480	0.417	NA	0.139
Sex-associated (male calls)	3	365–367	0.005	NA	NA	0.009	0.008
Mitochondrial	5	615–644	0.194	NA	NA	0.280	0.248

MAF—minor-allele frequency; H_E—pooled expected heterozygosity; H_O—observed heterozygosity; h_{SNP}—pooled haploid gene diversity (2pq); F_{ST}—finite-sample-corrected Nei-type differentiation estimate (Nei 1987); NA—not applicable. Summaries use 649 representative profiles from the allelematch analysis with alleleMismatch = 19 and consensusMethod = 1: 111 assigned females, 367 assigned males and 171 profiles of unknown sex. Autosomal and mitochondrial summaries include all sex categories. The sex-associated class contains the two SMCY markers and the ZFX/ZFY-homologue marker yPp_AF2_416; its statistics use assigned males only. These calls and mitochondrial calls retain haploid coding, with incompatible XY calls excluded. X-linked MAF and H_E use two chromosome copies per female and one per male; unknown-sex profiles and male XY calls are excluded. X-linked H_O uses 110–111 called females per SNP, and X-linked F_{ST} uses females only. Class means use unrounded per-SNP values from Table S1. F_{ST} is averaged over estimable loci using six country-based sampling-origin groups, with west Africa pooled, rather than inferred genetic groups

were included. We assessed sample call rates and visually inspected genotype-cluster separation in allelic-discrimination plots. Assays were rejected if controls behaved unexpectedly or genotype clusters were poorly separated. For diploid loci, all genotype combinations (XX, XY and YY) were also required to be represented; this criterion was adjusted for haploid markers. Control performance and cluster-quality requirements applied to every assay.

The selected final SNPs ($N = 96$) were used for a final genotyping run of sequenced individuals and additional leopard samples across Africa ($N = 937$). Samples were collected from different countries, covering broader variation in the African leopard (Pečnerová et al. 2021), as shown in Fig. 1. Sample replicates and descriptive statistics, such as heterozygosity and amplification success, were used to quality control data. Prior to assignment analyses, we removed samples with low amplification rate (more than 12.5% missing data; Fig. 2). We also removed replicates of individuals (individuals sampled more than once) by running the package Allelematch (Galpern et al 2012), resulting in a final set of 649 individuals. To test the SNPs performance as markers for investigations of population structure, we conducted a Discriminant Analysis of Principal Components (DAPC) analysis using the Adegenet package in R 4.3.2. We analyzed structure without a priori spatial information to explore the number of genetic groups in the data. Individuals were then assigned to the identified genetic groups, with geographic origin then used to summarize correspondence with the inferred genetic groups (Fig. 5).

SNP performance

We assessed genotype repeatability by genotyping 13 individuals three times, giving 39 replicates in total, using both tissue and fecal extracts. Paired tissue and fecal samples from the same known individuals were unavailable. We also examined empirical pairwise allelic differences in a separate diagnostic dataset that retained repeated sample records (Fig. 2). Laboratory controls and records identified as lion or cheetah were excluded. Of 1,025 sample runs, 875 records passed a missing-data threshold filter of 12.5%. Missingness was calculated across the 93 non-Y-labelled markers. Pairwise differences were summed across jointly called loci among the 85 autosomal SNPs, scoring a heterozygote–homozygote comparison as one difference and opposite homozygotes as two. Missing calls were omitted without rescaling the sums, and pairs with no jointly called locus were excluded. Filtering reduced the low and intermediate differences associated with poor call availability and made the low-distance component distinct from the main distribution.

Statistical analyses

Genetic-structure analyses used the adegenet framework (Jombart 2008) in R 4.3.2. Marker-level summaries in Table 1 and S1 were recalculated using custom scripts in R 4.3.3 from the 649 representative profiles using the sex and ploidy rules defined above. X-linked H_O used females only. H_E and H_O were not reported for haploid-coded assays, whose pooled allele-frequency diversity is reported as h_{SNP}. Missing-data proportions counted no-calls among all 649 profiles; ploidy-incompatible calls were counted separately. Gene-diversity partitioning followed Nei (1973), with the finite-sample estimator specified below. Differentiation used six country-based sampling-origin groups: Mozambique, South Africa, Senegal, Tanzania, Zambia and pooled west Africa. Per-locus F_{ST} was calculated as $(H_T - H_S)/H_T$ using the finite-sample-corrected within-group and total gene diversities of Nei (1987), following the equations implemented in hierfstat (Goudet 2005). Groups were weighted equally, with the harmonic mean of called genotype counts used in the corrections. X-linked differentiation used females only; the sex-associated and mitochondrial

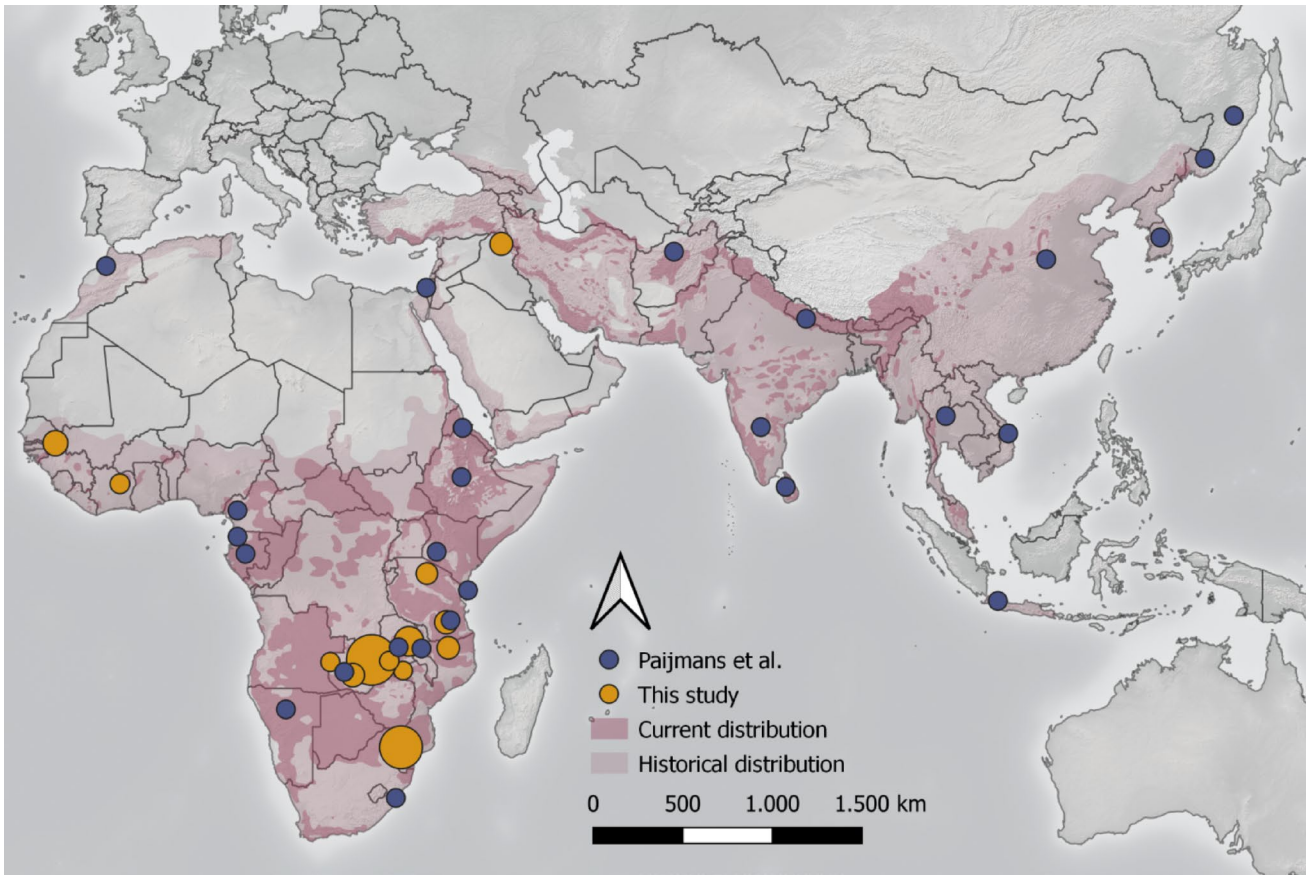


Fig. 1 Sample localities for leopard samples used in the design and testing of the SNP panel. Blue sample locations indicate whole-genome sequencing data (1 sample per locality), orange sample localities indicate newly integrated data (larger circles represent more samples). Leopard distribution derived from the IUCN Red List (Stein et al. 2024), with current distribution using both 'resident' categories ('extant' and 'possibly extant') and historic using 'extinct' and 'possibly extinct'

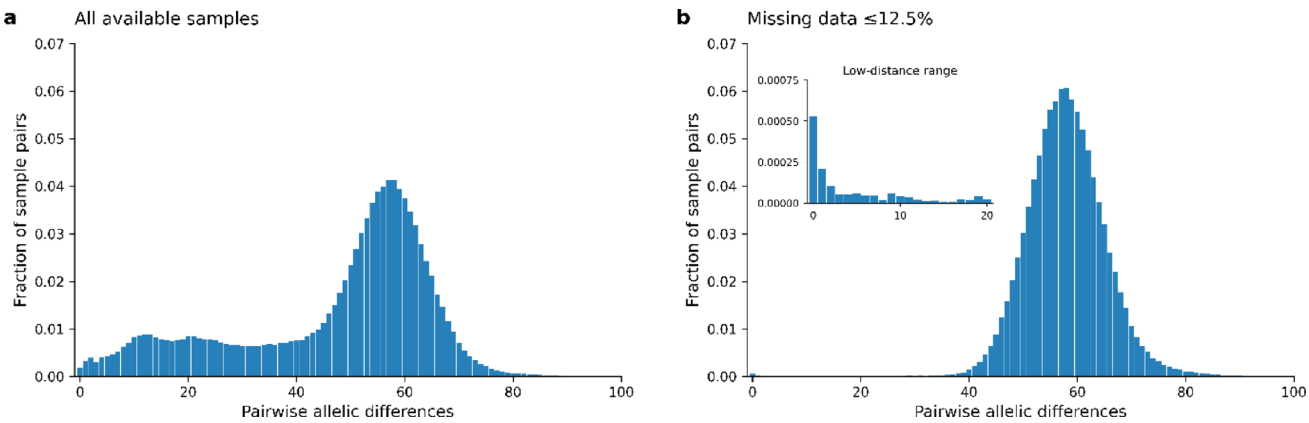


Fig. 2 Empirical distributions of pairwise allelic differences among leopard sample records. **a** All records with at least one called autosomal SNP. **b** Records with ≤12.5% missing data across 93 non-Y-labelled markers. Differences were summed across jointly called loci among 85 autosomal SNPs (heterozygote–homozygote = 1; opposite homozygotes = 2), without scaling for missing calls. Pairs without jointly called loci were omitted. Bars show the fraction of estimable sample pairs; the inset in b shows 0–20 differences. A mismatch threshold that is too stringent can split repeated samples into spurious individuals (ghosts), whereas an overly permissive threshold can merge distinct individuals (shadows; see Creel et al. 2003)

classes used the haploid form. Undefined estimates were reported as NA, negative finite estimates were retained, and class means were calculated from unrounded values over estimable loci.

Population structure was explored using discriminant analysis of principal components (DAPC; Jombart et al. 2010). The number of genetic clusters (K) was selected from the Bayesian information criterion in find.clusters, and a DAPC model was fitted using the inferred groups. Correspondence between these clusters and known geographic origin was summarized in a contingency table and visualized as a heatmap of individual counts. This describes the samples used to infer the clusters and is not an independent test of geographic assignment accuracy. Multivariate separation among clusters was illustrated with DAPC scatterplots.

Graphical content

Figures were generated in R using ggplot2, patchwork, plotrix, gridGraphics and grid. Figures were exported in PNG format with colour schemes chosen for accessibility.

Results

Individual genetic variation ranged from heterozygosities below 0.2 to nearly 0.7 (Fig. 3). Individuals with low heterozygosity were found in several populations (Figs. 3 and 4). A weak negative association between heterozygosity and missing data is seen in Fig. 3b, consistent with possible allelic dropout in some samples. However, low-heterozygosity profiles occurred across missing-data proportions from 0 to 0.12, indicating that poor call availability alone does not explain the pattern. Sibling probability of identity was 8.75×10^{-18}, indicating high allele-frequency-based discriminatory power; empirical differences among profiles provide a separate assessment of genotype quality (Fig. 2).

Panel-wide marker statistics are provided in Table S1, assay information for the full 96-marker panel in Table S2, and reference-sequence correspondences in Table S3. The sex-determination assays correctly assigned all nine samples of known sex (five females and four males). Among the 367 male-assigned profiles used for the marker summaries, SMCY_448 was polymorphic (MAF = 0.014), whereas valid SMCY_147 calls were monomorphic. The male-associated yPp_AF2_416 state was fixed by the assignment rule and therefore does not independently measure sex-assignment accuracy. X-linked heterozygosity in female-assigned profiles provides additional individual-level variation. Table 1 summarizes the marker classes, and Fig. 4 shows heterozygosity and missing data by country.

Iraqi samples were excluded from the structure analysis due to low sample number. Based on the SNP-genotyped samples with known provenance from Africa, four genetic clusters were identified (Fig. 5). These included one cluster of samples from west Africa (Genetic group 1 in Fig. 5). The second cluster held samples primarily from Zambia and Mozambique (Genetic group 2). Group 3 held most Tanzanian samples, together with samples from Mozambique and Zambia, most of these coming from the Nyerere National Park in Tanzania, the adjacent Niassa National Park in Mozambique and the nearby Luangwa National Park in Zambia. While partially separated by lake Malawi, dispersal corridors are found both to the north and to the south of this water body. Group 4 primarily held samples from the Greater Limpopo Transfrontier National Park in South Africa and Mozambique.

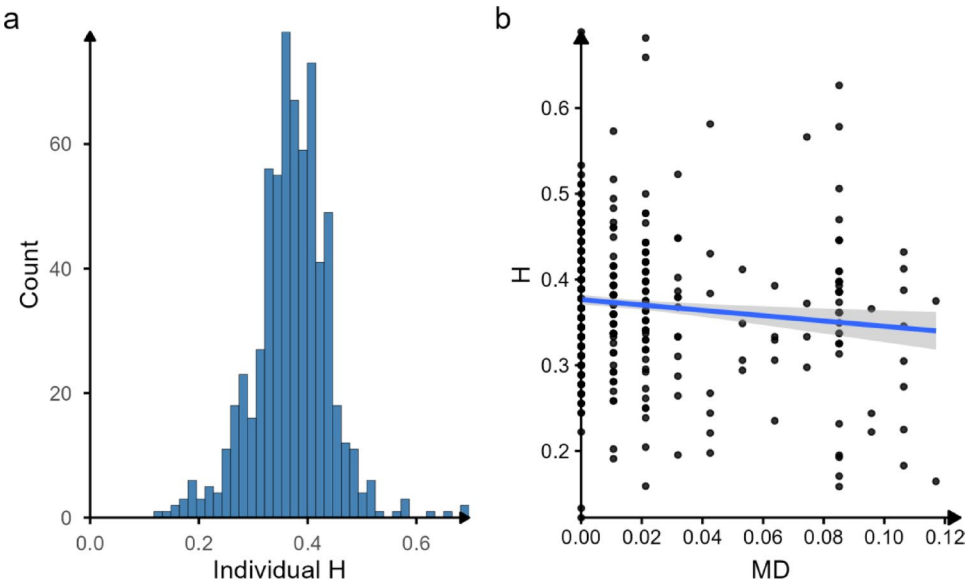


Fig. 3 **a** Distribution of individual heterozygosity. **b** Individual heterozygosity plotted against missing-data proportion. Profiles with > 12.5% missing data (call rate < 87.5%) were excluded to reduce the influence of poorly amplified samples

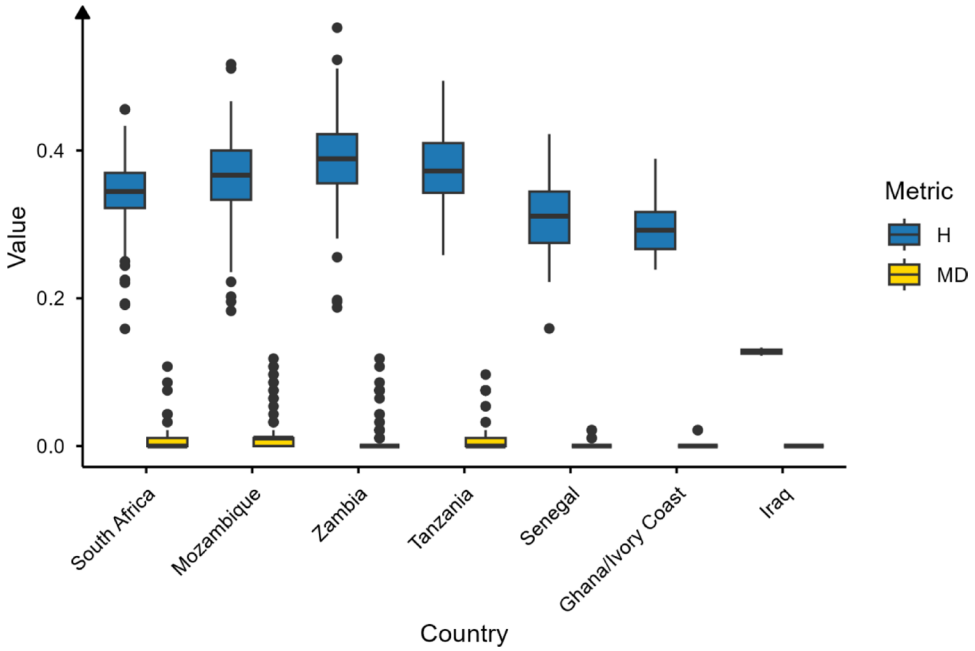


Fig. 4 Distribution of heterozygosity and missing data levels per country. Note that Ghana and Ivory Coast were combined as these samples largely originate from two parks in close proximity in the northern ranges of these countries but more precise coordinates are unavailable

Discussion

This study presents a validated 96-SNP panel for leopards, bridging the gap between genome-scale marker discovery and field-ready application for non-invasive monitoring of leopards across their global distribution. The panel shows high power for individual identification, reliable sex assignment, and forensic intelligence-level resolution of large-scale population structure, while remaining robust to the DNA degradation typical of fecal and other non-invasive samples. It also accurately assigns sex, and correctly identifies cheetah and lion through lack of heterozygosity and patterns of homozygosity in autosomal markers and mitochondrial haplotype. This is useful as field determination of scat donors has been shown to be unreliable (Forbes et al. 2026). A central result of our study is that robust genotypes can be obtained from challenging field material when marker choice is paired with an empirically defined amplification threshold that suppresses genotyping error to tolerable levels. In this way, the panel addresses one of the key constraints in wildlife genetics: how to retain analytical power while minimizing ghost and shadow effects, ultimately caused by genotyping error. The resulting assay therefore offers a practical compromise between marker number, cost, scalability, and data quality, without requiring computationally intensive downstream workflows. Taken together, our results show that a carefully curated SNP panel can provide a cost-effective and accessible platform for leopard monitoring, conservation, and forensic applications at broad spatial scales.

To save on costs, we used an existing genomic resource from a previously published study, to design a low-cost, accessible, and scalable resource. Reliable identification of individuals and sex, together with information on population structure, can support management, law enforcement and behavioral ecology. Potential applications include genetic capture–mark–recapture, connectivity and within-population diversity analyses. Kinship, parentage and assignment of unknown samples to a region of origin require validation appropriate to the population and question. While additional markers would increase the resolution for more demanding analyses such as pedigree reconstruction and more precise estimates of provenance, they would also raise development costs and could increase downstream genotyping costs, depending on platform capacity and configuration. Once a set of markers has been developed, the cost of genotyping is to a large extent dependent on scale. So, if samples cannot be run in suitable batch sizes, the cost may quickly become prohibitive. Alternatively, delaying analyses until a sufficient sample number has been accrued may delay results that may urgently be needed, which is particularly relevant in forensic cases. Here, the ability to run a smaller set of samples and markers is useful, for example by using single tubes in a qPCR or systems that allow flexible plate configurations. For specific use cases, it is possible to build upon the existing panel, by expanding it to include more markers. Thinking about this in a modular approach, in which a base panel allows cost-effective genotyping for a relatively small number of SNPs, and an expansion which further increases resolution if needed, allows us to continue to build on a joint baseline, while making the panel flexible and useful across applications.

Our quality-control results are consistent with our previous applications of SNP genotyping in wildlife studies

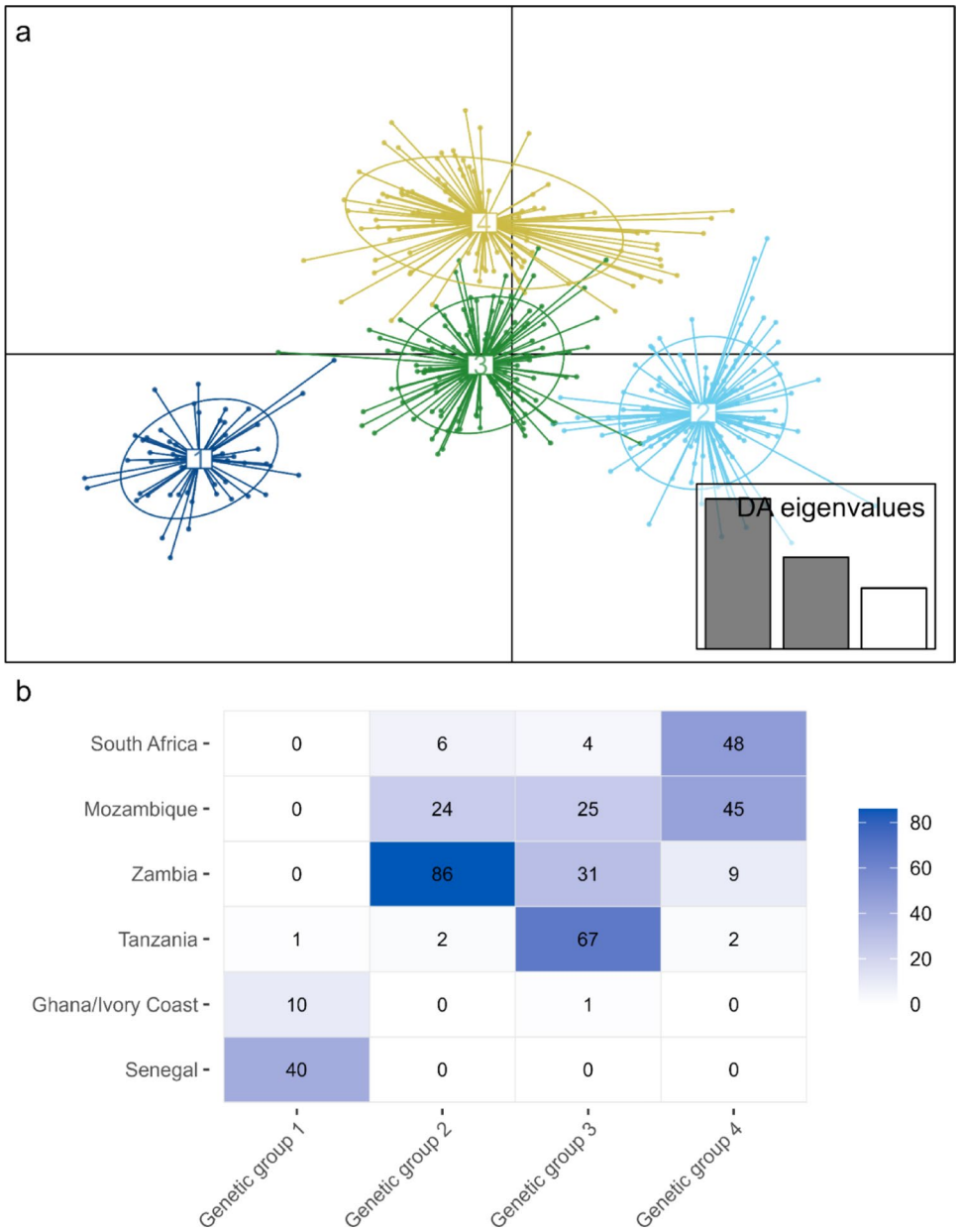


Fig. 5 Discriminant analysis of principal components (DAPC) for African samples of known provenance. **a** DAPC plot of the four inferred genetic groups. **b** Contingency table of sampling-origin groups and inferred genetic groups (1–4), containing 401 individuals. Cell values and colour intensity show counts. These are descriptive within-dataset associations, not independently validated geographic assignments

(Kleven et al. 2019; Norman and Spong 2015; Walton et al. 2021). In a direct comparison using noninvasive wolf, wildcat and brown bear samples, microfluidic SNP genotyping recovered more usable genotypes than microsatellite genotyping (von Thaden et al. 2017). For comparison, an eight-microsatellite leopard panel yielded a sibling probability of identity of 0.0005 (Mondol et al. 2009). Differences in populations and marker sets prevent a direct comparison of genotyping performance from these probabilities alone. During the optimization of this panel, we included scats collected along the South Africa/Mozambique border more than 8 years prior to DNA extraction and genotyping, with the dried scats stored in doubled air-tight plastic bags with silica gel in the inner bag. Additionally, temperatures in the storeroom in which the scats were stored regularly exceeded 40 °C during summer. Despite the suboptimal conditions for DNA preservation, high-quality genotypes were successfully acquired from 52% of the scats identified as leopards in the field. The high performance of our SNP panel with degraded DNA may enable researchers to proactively and opportunistically collect samples and store them, while funding is acquired, and still obtain high genotyping success rates. The empirical comparison of filtered and unfiltered profiles supports the use of call availability and genotype differences to assess individual-identification thresholds (Fig. 2). This is a critical requirement for most applications, but in particular when reliable individual identity is

of paramount importance (e.g. Capture-Mark-Recapture or forensics).

The probability of identity among siblings (PID_{sib}) is calculated by combining allele frequency–based genotype match probabilities across loci, taking their product under an assumption of independence among loci (Waits et al. 2001). A sufficiently informative marker set can therefore yield a low probability of an incidental genotype match. However, a low probability of identity alone does not guarantee accurate individual identification if genotyping errors (e.g., allelic dropout, false alleles) are present. In capture–mark–recapture (CMR) and pedigree reconstruction, even small error rates can bias individual counts or fail to establish relationships. Therefore, a low PID_{sib} must be accompanied by empirical assessment of genotype reliability.

The logistical challenges of working with wild populations, often of conservation concern, can be ameliorated by the protocols chosen. First, the discrete allele states of SNPs facilitate genotype comparison across laboratories and platforms when allele coding and assay performance are standardized. As such, and although whole-genome sequencing (WGS) is continually dropping in price, genotyping a preselected SNP panel is still a cost-effective tool for monitoring wildlife. Many questions of interest to wildlife managers do not require full genomes or a high number of SNP markers, but can be appropriately addressed by a panel of a few dozens of SNPs (Bertola et al. 2024). Here, more recent efforts of GT-seq may offer a viable alternative for larger sample sets (Campbell et al. 2015), although performance on low-quality material must be assessed for the assay and sample type concerned. Second, liquid based SNP platforms have been shown highly suitable for working with low quality samples (Norman and Spong 2015; von Thaden et al. 2017; Walton et al. 2021), such as non-invasively collected fecal samples, forensic traces, and historic samples. In research and management of wide ranging and elusive species, non-invasive samples are often the preferred or even the only available DNA source. Non-invasive sampling may also allow, in addition to minimizing disturbance to the animals, population-size estimation through genetic capture–mark–recapture (Forbes et al. 2025) or pedigree reconstruction (Spitzer et al. 2016). Third, the ability to reliably genotype individuals from non-invasively collected material may simplify permit logistics, depending on national regulations and sample type. This can significantly reduce the time from collection effort to result. International shipment of regulated specimens can nevertheless require coordinated permit and clearance procedures, which should be planned with the relevant authorities. Finally, a broadly representative SNP panel with openly documented assay sequences may help ameliorate the imbalance in international collaborations by reducing the dependence of colleagues in range states on sequencing and computational facilities held by external collaborators. Due to their modular character, SNP panels can be set up to address different relevant population genetics questions (Mueller et al. 2020), in manners compatible with locally available equipment and genotyping platforms (e.g. desktop sequencing or qPCR). Data storage and analysis do not require advanced computer clusters, and are more accessible for those without advanced bioinformatics training. Documented SNP panels that perform well across relevant sample types can therefore improve access to genetic data for managers and decision makers and support evidence-based conservation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-026-01452-7.

Acknowledgements This work has received financial support from NSF, VR (Swedish Research Council), Sida, USFWS, and Panthera. We also acknowledge sequencing support from the Swedish National Genomics Infrastructure (NGI) at the Science for Life Laboratory in Uppsala and Stockholm supported by the Swedish Research Council (Vetenskapsrådet) and the Knut and Alice Wallenberg Foundation. We are also grateful to a long list of management authorities and NGOs working across the range of states included in the current work. We are grateful for constructive and important feedback from two reviewers.

Author contributions All authors contributed to generating the sample set, to the discussion of the analysis set, and to the design of the SNP panel. Bastian Schiffthaler and Nicolas Dussex performed the bioinformatics of aligning the two sequencing efforts. The sample map was generated by Laura Bertola. Downstream analyses and the first draft of the manuscript was written by Faruk Mamugy and Göran Spong and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Swedish University of Agricultural Sciences. This work has received financial support from NSF, The Swedish Research Council (VR), Swedish International Development Cooperation Agency (Sida), U.S. Fish and Wildlife Service, and Panthera. We also acknowledge sequencing support from the Swedish National Genomics Infrastructure (NGI) at the Science for Life Laboratory in Uppsala and Stockholm supported by the Swedish Research Council (Vetenskapsrådet) and the Knut and Alice Wallenberg Foundation.

Data availability We are grateful to a long list of management authorities, law enforcement agencies, NGOs, and governments for their efforts in protecting our environment. As these samples come from a protected species and their ownership falls under a multitude of legislative frameworks, the raw data for exact location and associated genotypes are not provided here. Upon reasonable request, and contingent on approval and restrictions defined by relevant range states, raw data can be made available.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bertola LD, Brüniche-Olsen A, Kershaw F et al (2024) A pragmatic approach for integrating molecular tools into biodiversity conservation. Conserv Sci Pract 6(1):e13053. https://doi.org/10.1111/csp2.13053

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Campbell NR, Harmon SA, Narum SR (2015) Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol Ecol Resour 15(4):855–867. https://doi.org/10.1111/1755-0998.12357

Creel S, Spong G, Sands JL, Rotella J, Zeigle J, Joe L, Murphy KM, Smith D (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. Mol Ecol 12(7):2003–2009. https://doi.org/10.1046/j.1365-294X.2003.01868.x

Creel S, Spong G, Becker M, Simukonda C, Norman A, Schiffthaler B, Chifunte C (2019) Carnivores, competition and genetic connectivity in the Anthropocene. Sci Rep 9:16339. https://doi.org/10.1038/s41598-019-52904-0

Forbes RE, Kerley GIH, Everatt KT, Mamugy FPS, Spong G (2025) Estimating leopard population sizes in western Mozambique using SNP-based capture–mark–recapture models. J Mammal 106(3):632–638. https://doi.org/10.1093/jmammal/gyaf004

Forbes RE, Everatt KT, Andresen L, Spong G, Kerley GIH (2026) Who knows poo? Inaccuracy of scat identifications in the field necessitates genetic confirmation of scat donors. Mammalian Biology 106:577–588. https://doi.org/10.1007/s42991-025-00559-y

Galpern P, Manseau M, Hettinga P, Smith K, Wilson P (2012) Allelematch: An R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. Mol Ecol Resour 12(4):771–778. https://doi.org/10.1111/j.1755-0998.2012.03137.x

Giangregorio P, Norman AJ, Davoli F, Spong G (2019) Testing a new SNP-chip on the Alpine and Apennine brown bear (*Ursus arctos*) populations using non-invasive samples. Conserv Genet Resour 11(3):355–363. https://doi.org/10.1007/s12686-018-1017-0

Goudet J (2005) hierfstat, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5(1):184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Hohenlohe PA, Funk WC, Rajora OP (2021) Population genomics for wildlife conservation and management. Mol Ecol 30(1):62–82. https://doi.org/10.1111/mec.15720

Jacobson AP, Gerngross P, Lemeris JR Jr. et al (2016) Leopard (Panthera pardus) status, distribution, and the research efforts across its range. PeerJ 4:e1974. https://doi.org/10.7717/peerj.1974

Jombart T (2008) adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94

Kitchener AC, Breitenmoser-Würsten C, Eizirik E et al (2017) A revised taxonomy of the Felidae: The final report of the Cat Classification Task Force of the IUCN SSC Cat Specialist Group. Cat News. Special Issue 11:1–80

Kleven O, Ekblom R, Spong G et al (2019) Estimation of gene flow into the Scandinavian wolverine population. NINA Report 1617. Norwegian Institute for Nature Research, Trondheim. https://hdl.handle.net/11250/2592962

Kopatz A, Kleven O, Kojola I et al (2021) Restoration of transborder connectivity for Fennoscandian brown bears (*Ursus arctos*). Biol Conserv 253:108936. https://doi.org/10.1016/j.biocon.2020.108936

Li H (2018) Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Mondol S, Navya R, Athreya V, Sunagar K, Selvaraj VM, Ramakrishnan U (2009) A panel of microsatellites to individually identify leopards and its application to leopard monitoring in human dominated landscapes. BMC Genet 10:79. https://doi.org/10.1186/1471-2156-10-79

Mueller SA, Reiners TE, Steyer K, von Thaden A, Tiesmeyer A, Nowak C (2020) Revealing the origin of wildcat reappearance after presumed long-term absence. Eur J Wildl Res 66:94. https://doi.org/10.1007/s10344-020-01433-7

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70(12):3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nei M (1987) Molecular evolutionary genetics. Columbia University, New York

Nichols RV, Spong G (2017) An eDNA-based SNP assay for ungulate species and sex identification. Diversity 9(3):33. https://doi.org/10.3390/d9030033

Norman AJ, Spong G (2015) Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. Ecol Evol 5(15):3056–3065. https://doi.org/10.1002/ece3.1588

Paijmans JLA, Barlow A, Becker MS et al (2021) African and Asian leopards are highly differentiated at the genomic level. Curr Biol 31(9):1872–1882e5. https://doi.org/10.1016/j.cub.2021.03.084

Palsbøll PJ, Peery MZ, Olsen MT, Beissinger SR, Bérubé M (2013) Inferring recent historic abundance from current genetic diversity. Mol Ecol 22(1):22–40. https://doi.org/10.1111/mec.12094

Pečnerová P, Garcia-Erill G, Liu X et al (2021) High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. Curr Biol 31(9):1862–1871e5. https://doi.org/10.1016/j.cub.2021.01.064

Ripple WJ, Estes JA, Beschta RL et al (2014) Status and ecological effects of the world's largest carnivores. Science 343(6167):1241484. https://doi.org/10.1126/science.1241484

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol 28(21):4737–4754. https://doi.org/10.1111/mec.15253

Solberg KH, Bellemain E, Drageset O-M, Taberlet P, Swenson JE (2006) An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. Biol Conserv 128(2):158–168. https://doi.org/10.1016/j.biocon.2005.09.025

Spitzer R, Norman AJ, Schneider M, Spong G (2016) Estimating population size using single-nucleotide polymorphism-based

pedigree data. Ecol Evol 6(10):3174–3184. https://doi.org/10.1002/ece3.2076

Stein A, Gerngross P, Alhikmani H, Balme G, Bertola L, Drouilly M (2024) *Panthera pardus*. The IUCN red list of threatened species. https://www.iucnredlist.org/species/15954/254576956

Thomsen PF, Willerslev E (2015) Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183:4–18. https://doi.org/10.1016/j.biocon.2014.11.019

von Thaden A, Cocchiararo B, Jarausch A, Jüngling H, Karamanlidis AA, Tiesmeyer A, Nowak C, Muñoz-Fuentes V (2017) Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. Sci Rep 7:10768. https://doi.org/10.1038/s41598-017-10647-w

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. Mol Ecol 10(1):249–256. https://doi.org/10.1046/j.1365-294X.2001.01185.x

Walton Z, Hagenlund M, Østbye K, Samelius G, Odden M, Norman A, Willebrand T, Spong G (2021) Moving far, staying close: Red fox dispersal patterns revealed by SNP genotyping. Conserv Genet 22(2):249–257. https://doi.org/10.1007/s10592-021-01332-7

Wright S (1931) Evolution in Mendelian populations. Genetics 16(2):97–159. https://doi.org/10.1093/genetics/16.2.97

Statements & Declarations

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Faruk Mamugy[1,2] · Laura D. Bertola[3] · Helena Königsson[1] · Amber Mertens De Vry[1] · Nicolas Dussex[4] · Bastian Schiffthaler[5] · Johanna L. A. Paijmans[7,8] · Michael Hofreiter[6] · Ryan Forbes[1,9] · Graham I. H. Kerley[9] · Kris Everatt[9,10,11] · Matthew Becker[12] · Scott Creel[12,13] · Stéphanie Bourgeois[14] · Twakundine Simpamba[15] · Clive Chifunte[15] · Reuben Kabungo[12] · Hana Raza[16] · Nabaz Horeni[16] · Bahez Farooq[16] · Marine Drouilly[11,17] · Göran Spong[1,12]

✉ Göran Spong
goran.spong@slu.se

1 Molecular Ecology Group, SLU, 901 83 Umeå, Sweden

2 Eduardo Mondlane University, Maputo, Mozambique

3 National Centre for Biological Sciences, Bangalore, India

4 Unit for Population Analysis and Monitoring, Swedish Museum of Natural History, 106 91 Stockholm, Sweden

5 UPSC, SLU, 901 87 Umeå, Sweden

6 Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

7 School of Environmental and Natural Sciences, Bangor University, Bangor, England, UK

8 Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Cambridge, England, UK

9 Centre for African Conservation Ecology, Nelson Mandela University, Gqeberha, South Africa

10 Greater Limpopo Carnivore Programme, Limpopo National Park, Massingir, Mozambique

11 Panthera, New York, NY 10018, USA

12 Zambia Carnivore Programme, PO Box 80, Mfuwe, Eastern Province, Zambia

13 Department of Ecology, Montana State University, Bozeman, MT 59717, USA

14 Agence Nationale des Parcs Nationaux, Libreville, Gabon

15 Department of National Parks and Wildlife, Lusaka, Zambia

16 Leopards Beyond Borders, Sulaimani, Kurdistan Region, Iraq

17 Institute for Communities and Wildlife in Africa, University of Cape Town, Cape Town, South Africa