SNP panel for non-invasive genotyping of leopard (*Panthera pardus*)

Authors: Faruk Mamugy[1,2], Laura D. Bertola[3], Amber Mertens De Vry[1], Nicolas Dussex[4], Bastian Shiffthaler[5], Joanna Paijmans[6], Michael Hofreiter[6], Ryan Forbes[1,7], Graham I.H. Kerley[7], Kris Everatt[7,8,9], Matthew Becker[10], Scott Creel[10, 11], Stéphanie Bourgeois[12], Marine Drouilly[8], Göran Spong[1,10]*

1) Molecular Ecology Group, VFM, 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, SE-106 91 Stockholm, Sweden
5) UPSC, Umeå University, 901 87 Umeå, Sweden
6) Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany
7) Centre for African Conservation Ecology, Nelson Mandela University, Gqeberha, South Africa
8) Greater Limpopo Carnivore Programme, Limpopo National Park, Mozambique
9) Panthera, New York, NY 10018, USA
10) Zambia Carnivore Programme, PO Box 80 Mfuwe, Eastern Province, Zambia
11) Department of Ecology, Montana State University, Bozeman MT 59717, USA
12) Agence Nationale des Parcs Nationaux, Libreville, Gabon

*Corresponding author: goran.spong@slu.se

Abstract

Genetic resources for species monitoring are ideally relevant for the species' full distribution range, feasible economically and logistically, and validated for the range of sample types collected from the field. This is particularly important for large carnivores that are elusive and wide-ranging, where individual and population processes often traverse administrative borders, and where obtaining high-quality samples can be challenging. Here we present a small species-specific SNP panel for leopards. We used whole genome data from across the global range and RAD sequence data from Zambian leopards to select markers for assay development. These were ascertained for 590 individual leopards from eight African countries and final selection was based on marker variation and performance on non-invasive samples. The final 96 marker panel holds 5 mitochondrial markers for species recognition, 3 Y-markers for determination of individual sex, 3 X-markers and 85 somatic markers, with an associated genetic baseline holding nearly 900 individuals. The selected autosomal markers hold variation across the global range with high power to identify individuals (PID=$2{,}45 \times 10^{-35}$) and in most cases their provenance with high assignment probability (>95%). Markers were also selected based on their performance on samples with low target DNA content, with distinct genotype separation in the output marker plots. The genotypes from this panel are thus straightforward to analyze and do not require computationally challenging bioinformatic resources, making this a low cost and accessible resource for leopard monitoring and research.

Introduction

Genomes hold vast amounts of information. By extracting parts of this information, individuals, their kin, reproductive patterns, 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 (REFSXX). As methods improve and costs decrease, the allure of genetic data persists. However, establishing protocols for extracting genetic data from wild populations may require a substantial development cost. Once established, the protocol must negotiate cost, logistics, and data quality and quantity to accommodate sample influx that may be constrained by time or unpredictable in size and quality.

Minute quantities of DNA, sometimes collected non-invasively from the environment, permit the extraction of complete genotypes from individuals without direct 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 such collection can be done without specialist knowledge or equipment, 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, Luikart, and Taberlet 2001), but this method requires many replicates to make it statistically sound, making it demanding in resources (note that sequencing applications such as GT-sequencing 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), false differences caused by marker ascertainment may lead to erroneous inferences, directly affecting estimates of connectivity, population size and history (Palsboll et al. 2013).

Leopards (*Panthera pardus*) have the most extensive geographic distribution of all large cats, with various 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 subspecies are recognized (Kitchener 2017, Stein et al. 2024) with all African populations considered to belong to the nominate form *P. pardus pardus.* The status of individual subspecies and populations vary dramatically: seven out of eight subspecies are threatened, with three being critically endangered, two endangered, and two vulnerable (Stein et al. 2024). 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. Its secretive nature and adaptability has permitted the leopard to eke out a living also in areas heavily dominated by humans (Pečnerová et al. 2021), IUCN still classifies its status as 'Vulnerable' due to apparently declining population trends. But actual population numbers, and hence rates or extents of decline, are virtually unknown. There is therefore an urgent need for better monitoring tools across the current range to support conservation, management and law enforcement activities (Ripple et al. 2014)).

Here we present a 96 marker SNP-panel optimized for non-invasive monitoring of leopards across their global range. The panel holds mt-, sex-, and nuclear markers where the vast majority show within population variation across the global range. Only markers that performed well on non-invasive material are included and the current selection provides high power to identify individuals, their sex, their kin, and their provenance.

Materials & Methods

Our SNP-panel development leverages an existing dataset and integrates new data to improve geographic representation, optimize the statistical power, and forms the start of a baseline for analyses of provenance. First, we used the SNP variant information from previously published whole genome data of 26 individuals (including historic samples) from across the global range of leopards (J. L. A. Paijmans et al. 2021)). Second, we identified SNPs by RAD sequencing of 152 tissue samples collected by the Zambian Carnivore Programme. These two data sets were then intersected to identify markers with variation across the global range. Third, a final marker selection step and assessment of marker performance was done on a fluid-based SNP-chip genotyping platform (Fluidigm Biomark or BioTools Juno X) using tissue and fecal extractions. Over 880 individual leopards from six African countries have now been genotyped with this panel.

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 radiocollars or remove wire snares). All animal handling, sample collection, and shipping were done under relevant permits renewed yearly or issued prior to sample exports. 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.* (2024).

Whole genome sequencing

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

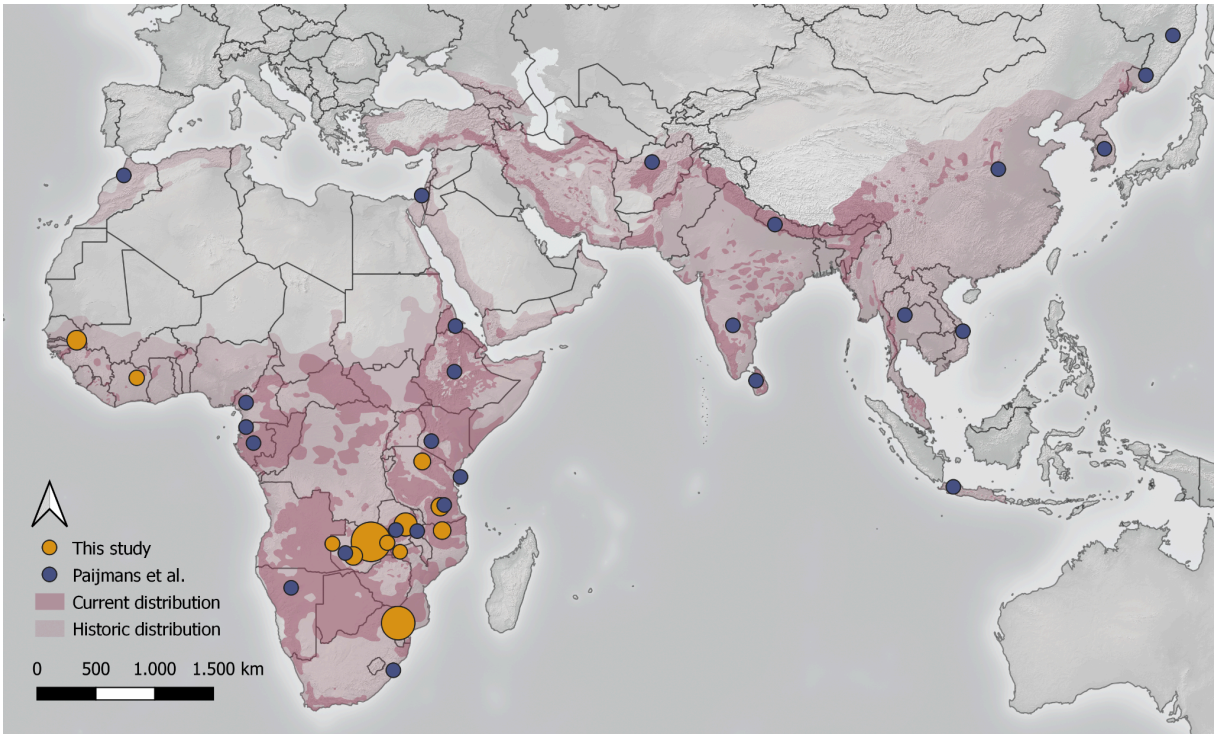


Figure 1. Sample localities for leopard samples used in the design and testing of the SNP panel. Blue sample locations indicate whole-genome sequencing data (1 sample per locality), orange sample localities indicate newly integrated data (larger circles represent a higher number of samples - see Suppl.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

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 0.5µg 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 400-800bp were excised for paired-end library construction (2x150bp) and RAD-sequencing on an Illumina NovaSeq 6000.

SNP detection and filtering

Sequence quality of the demultiplexed RADseq Illumina reads was assessed using FastQC (version, ref.). Raw reads were trimmed and adapters removed with Trimmomatic (0.39, ref.) in paired-end mode. Only reads with length >100 bp after trimming were retained for further processing. These remaining sequences were used as input for SNP detection using Stacks (2.53, ref.), running the individual components of the pipeline manually. A total of 6,931 SNPs were identified using the Stacks software. We filtered SNPs using customized R scripts to remove those of low quality. First, we selected only bi-allelic SNPs that were represented in at least 75% of the samples. Only stacks with one SNP with a minimum of 30bp up- and downstream (for assay development) were retained. with a minor allele frequency (MAF) > 0.2 and X-chromosome SNPs with MAF > 0.15. Finally, we selected for SNPs in which at least 5 samples were attributed to each genotype (i.e. aa, bb and ab) and that showed variation across the global range. After applying these filtering criteria, 148 SNPs were

selected for assay design. A total of 133 SNPs passed the *in silico* assay design, and were brought forward to *in vitro* assay development on a Fluidigm Biomark or Juno X.

Published Genbank sequences were used to identify Y-chromosome SNP markers that are informative of parental lineage (Table 1) and sex (Table 2). The selected sequences were aligned using BioEdit (v. 7.2, Tom Hall, Ibis Biosciences, Carlsbad, USA) and manually scanned to identify SNPs. The three sex-specific Y-chromosome SNPs simply detect the presence of the Y-chromosome (holding no variation). When these markers fail to amplify, samples are called female, and vice versa. Additionally, heterozygosity in the X-chromosome markers, only possible in females, is also used to verify sex assignments when possible. Mitochondrial SNPs were included to study matrilines and to distinguish from other sympatric carnivore species, such as lion *Panthera leo* and cheetah *Acinonyx jubatus,* particularly relevant for fecal species identification. Regions with known numts were excluded.

SNP validation

The autosomal and X-chromosome SNP markers were initially validated through genotyping a total of 119 leopard samples (2 lions and 1 cheetah) using the Fluidigm Biomark platform. The mitochondrial and Y-chromosome markers were validated by genotyping 31 leopards, (59 lions and 1 cheetah,) including 9 leopards of known sex (female n = 5, male n = 4). Both positive (sample replicates) and negative no template controls were included. We evaluated the call rate of the DNA samples and the performance of the assayed markers at successfully assigning a genotype to the samples. Every marker was visually assessed on the allele clustering of the DNA samples in the Biomark scatterplot. SNPs were invalidated if the controls did not work as expected, if the different genotypes clustered too near each other or if not all genotypes were present (i.e., XX, XY, YY). Only SNPs that passed these criteria were included on the final SNP-chip (with the exception of mitochondrial and Y-markers).

The selected final SNPs (N=96) were used for a final genotyping run of sequenced individuals and additional leopard samples across Africa. Samples were collected from different countries, covering broader variation in the African leopard (Pecnerove et al. 2021), as shown in Figure 1, and were extracted from a range of source material (blood, feces, hair, and tissue). Sample replicates and descriptive statistics, such as heterozygosity and amplification success, were used to quality control data. Prior to assignment analyses, we removed replicates of individuals (individuals sampled more than once) by running the package Allelematch. We also removed samples with low amplification rate (more than 5% missing data), resulting in a final set of 875 individuals prior to the assignment analysis. 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.

SNP performance

We assessed the reliability of genotype calls by running samples in replicates (N=39). We replicated both tissue and fecal extracts. Unfortunately we did not have access to both tissue and fecal samples from known individuals, but instead compared the distribution of individual allelic differences to the distribution of errors (Figure 2). This allowed us to empirically determine an amplification threshold (>98%) that suppresses the error rate to a point where the appearance of ghost or shadow genotypes is extremely unlikely ($p<0.001$).

Results

When samples hold enough DNA to result in high amplification rates, the individual genotypes provide reliable individual identification. For genotypes that passed the quality control filter, the mean missing data were <0.015.

The sex-determination markers correctly assigned all samples of known sex, resulting in an error rate of zero. Our Y markers hold no variation, but are included with the sole purpose of identifying sex. In contrast the X markers do hold variation, so in addition to providing additional individual and population resolution, the X markers can be used to further indicate sex- Males should appear homozygote for both Y and X, whereas females may be heterozygote for X (and lack any amplification for Y). For summary marker statistics, see Table 1.

Table 1. Summary marker statistics. H_O- Observed heterozygosity, H_T- Expected heterozygosity, F_{ST}- Standardized fixation index, MAF-Minor Allele frequency.

Marker type (N)	H_O	H_E	F_{STP}	MAF
mitochondrial	N/A	N/A	N/A	N/A
Y	N/A	N/A	N/A	0.25
X	0.22	0.16	-	0.09
Autosomal	0.58	0.56	0.02	0.17

Our quality control protocols provided empirical evidence that genotyping error can be controlled for successful individual identification (Figure 2), also from non-invasively collected samples. Even samples stored under suboptimal conditions were used to extract genotypes of sufficient quality. This is a critical requirement for most applications, but in particular when individual identity is of paramount importance (e.g. Capture-Mark-Recapture or forensics).

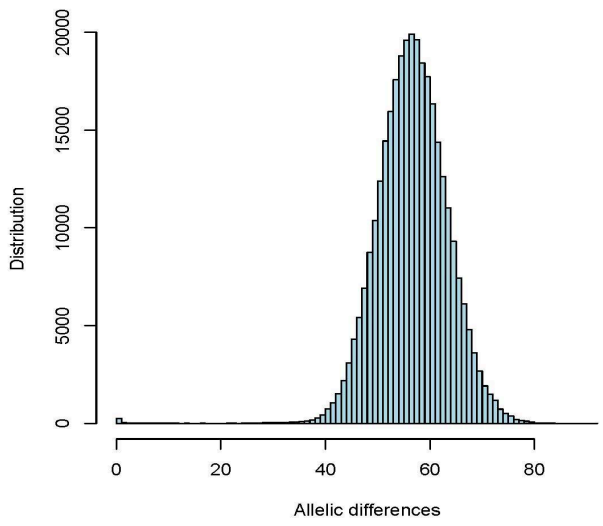


Figure 2. Pairwise allelic differences between all sample pairs. The large peak on the right, shows the distribution of genotype differences between unique individuals. The small peak on the left, shows discrepancies between sample pairs originating from the same individual, where the manifested genotyping error rate is 3% when excluding samples with an amplification rate <98% (for samples with an amplification rate as low as 90% the error rate is <0.06%). With lower amplification rates, sample quality may become prohibitive for reliable genotyping risking ghost and/or shadow effects in the binning of genotypes.

Based on the SNP genotyped samples, spanning a significant part of the African leopard distribution, three groups were identified, one holding west African samples, one holding mostly samples from South Africa and Mozambique (Greater Limpopo Area), and one large group holding samples from Tanzania, Zambia and northern Mozambique. When re-analysing the two latter groups (i.e., excluding west African samples), this group split into three genetic groups (Figure 3b and Table S2).

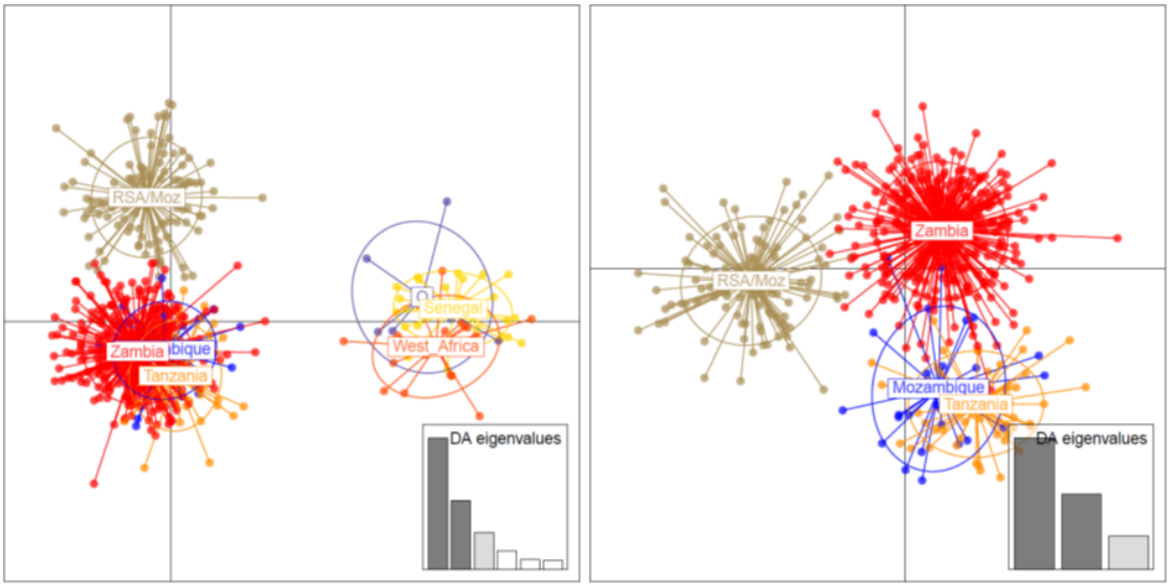


Figure 3. Discriminatory Analysis of Principal Components (DAPC). a) DAPC of all samples showing clear distinction between samples from west Africa. b) Subset of samples to only include East and southern Africa. Samples separate into groups with high alignment to geographical origin of samples (Table S2).

Discussion

Methods to extract genetic data from increasingly degraded sources are rapidly evolving and costs are declining. For applications across research, management and law enforcement these advances hold much promise. But moving from possibilities to practicalities when acquiring genetic data will invariably be a trade-off between **cost, genotype quality** and **logistics**. The optimal solution is also likely to change over time as methods improve and projects evolve. In this study we describe how we have resolved this trade-off to build a resource for research and conservation of leopards, applicable throughout the global range and optimized for low quality samples.

To save on **costs**, existing genomic resources, e.g. full genomes derived from previously published studies, can be repurposed to design cost-effective tools, which can meet the demand for cheap, accessible and scalable tools. Here we have presented a genotyping panel for leopards optimized for samples of low quality, partially dependent on previously published whole genome data and new RAD sequencing data. For most queries of management and law enforcement, including many questions in behavioral ecology, the ability to identify individuals, their sex, their kin, population structure and provenance reliably is what is needed. This allows for robust population estimates through capture-mark-recapture, estimates of connectivity, parentage analyses, assessments of within-population diversity, and assignment of unknown samples to a region of origin. While additional markers would likely further increase the resolution for more demanding analyses such as pedigree reconstruction and provenance, it would also raise costs of development and downstream genotyping. Once a set of markers has been developed, the cost of genotyping is to a large extent dependent on scale. 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. Here, the ability to run a smaller set of samples is useful, for example by using single tubes in a qPCR or systems that allow flexible plate configurations (e.g. Biosystems Juno).

Our **quality control** results align well with our previous work on other model systems (Creel et al. 2019; Kleven et al. 2019; Norman and Spong 2015; Walton et al. 2021)) and other studies showing that SNPs are better suited for working with degraded DNA than microsatellites (Morin et al. 2004). Additionally, liquid-based SNP arrays have shown to be better suited for low DNA content samples, than the more commonly used prespotted arrays (REFXX). During the optimization of this panel, we included scats collected along the South Africa/Mozambique border > 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 for example, funding is acquired, and still obtain high genotyping success rates.

The **logistical** challenges of working with wild populations, often of conservation concern, can be ameliorated by the protocols chosen. *First*, the inherent qualitative nature of SNPs allows for reliable and direct replication across laboratories and platforms. As such, and although whole-genome sequencing (WGS) is dropping in price rapidly, genotyping a preselected SNP panel is still a cost-effective and pragmatic tool for monitoring wildlife for the foreseeable future. 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. *Second*, liquid based SNP platforms are especially suitable for working with low quality samples (Carroll et al., 2018; A. von Thaden et al., 2017), 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 only available DNA source. Non-invasive sampling may also allow, in addition to minimizing disturbance to the animals, estimates of population size in a Capture-Mark-Recapture framework ((Spitzer et al. 2016). *Third*, the ability to reliably genotype individuals from non-invasively collected material also greatly facilitate transport logistics since these samples or DNA extracts are exempt from CITES. This significantly reduces the time from collection effort to result. In contrast, moving samples that fall under CITES are contingent on clearance procedures that sometimes can be slow or held up by administrative differences in how these co-dependent import- and export-permits are issued. Moreover, CITES permits are only valid for a specified number of samples, making it virtually impossible to apply for a permit before the collection effort has been completed. *Finally*, an open source SNP panel for the full distribution of the species, may ameliorate the imbalance in international collaborations by reducing the dependence of colleagues in range states on sequencing and computational facilities in the global north. Due to their modular character, SNP panels can be set up to address different relevant population genetics questions (Alina von Thaden 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.

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 and the Knut and Alice Wallenberg Foundation.

References

Creel, S., G. Spong, M. Becker, C. Simukonda, A. Norman, B. Schiffthaler, and C. Chifunte. 2019. "Carnivores, Competition and Genetic Connectivity in the Anthropocene." *Scientific Reports* 9 (November). https://doi.org/10.1038/s41598-019-52904-0.

Creel, S., G. Spong, J. L. Sands, J. Rotella, J. Zeigle, L. Joe, K. M. Murphy, and D. Smith. 2003. "Population Size Estimation in Yellowstone Wolves with Error-Prone Noninvasive Microsatellite Genotypes." *Molecular Ecology* 12 (7): 2003–9.

Giangregorio, P., A. J. Norman, F. Davoli, and G. Spong. 2019. "Testing a New SNP-Chip on the Alpine and Apennine Brown Bear (Ursus Arctos) Populations Using Non-Invasive Samples." *Conservation Genetics Resources* 11 (3): 355–63.

Kleven, O., R. Ekblom, G. Spong, G. Lansink, J. Aspi, S. Creel, I. Kojola, et al. 2019. "Estimation of Gene Flow into the Scandinavian Wolverine Population." 1504-3312. Vol. 1617. NINA Report. Norwegian Institute for Nature Research.

Kopatz, A., O. Kleven, I. Kojola, J. Aspi, A. J. Norman, G. Spong, N. Gyllenstrand, et al. 2021. "Restoration of Transborder Connectivity for Fennoscandian Brown Bears (Ursus Arctos)." *Biological Conservation* 253 (January). https://doi.org/10.1016/j.biocon.2020.108936.

Norman, A. J., and G. Spong. 2015. "Single Nucleotide Polymorphism-Based Dispersal Estimates Using Noninvasive Sampling." *Ecology and Evolution* 5 (15): 3056–65.

Paijmans, J., A. Barlow, M. Becker, J. Cahill, J. Fickel, D. Foerster, K. Gries, et al. 2021. "African and Asian Leopards Are Highly Distinct at the Genomic Level." *Current Biology: CB* 31:1–11.

Palsboll, P. J., M. Z. Peery, M. T. Olsen, S. R. Beissinger, and M. Berube. 2013. "Inferring Recent Historic Abundance from Current Genetic Diversity." *Molecular Ecology* 22 (1): 22–40.

Pečnerová, Patrícia, Genís Garcia-Erill, Xiaodong Liu, Casia Nursyifa, Ryan K. Waples, Cindy G. Santander, Liam Quinn, et al. 2021. "High Genetic Diversity and Low Differentiation Reflect the Ecological Versatility of the African Leopard." *Current Biology: CB* 31 (9): 1862–71.e5.

Ripple, W. J., J. A. Estes, R. L. Beschta, C. C. Wilmers, E. G. Ritchie, M. Hebblewhite, J. Berger, et al. 2014. "Status and Ecological Effects of the World's Largest Carnivores." *Science (New York, N.Y.)* 343 (6167): 151 – +.

Solberg, K. H., E. Bellemain, O. M. Drageset, P. Taberlet, and J. E. Swenson. 2006. "An Evaluation of Field and Non-Invasive Genetic Methods to Estimate Brown Bear (Ursus Arctos) Population Size." *Biological Conservation* 128 (2): 158–68.

Spitzer, R., A. J. Norman, M. Schneider, and G. Spong. 2016. "Estimating Population Size Using Single-Nucleotide Polymorphism-Based Pedigree Data." *Ecology and Evolution* 6 (10): 3174–84.

Thomsen, P. F., and E. Willerslev. 2015. "Environmental DNA - An Emerging Tool in Conservation for Monitoring Past and Present Biodiversity." *Biological Conservation* 183 (March):4–18.

Waits, L. P., G. Luikart, and P. Taberlet. 2001. "Estimating the Probability of Identity among Genotypes in Natural Populations: Cautions and Guidelines." *Molecular Ecology* 10 (1): 249–56.

Walton, Z., M. Hagenlund, K. Ostbye, G. Samelius, M. Odden, A. Norman, T. Willebrand, and G. Spong. 2021. "Moving Far, Staying Close: Red Fox Dispersal Patterns Revealed by SNP Genotyping." *Conservation Genetics* 22 (2): 249–57.

Wright, S. 1931. "Evolution in Mendelian Populations." *Genetics* 16:97–159.

Supplementary material

Table S1. Somatic marker statistics as calculated by the R package Adegenet. Ho: Observed heterozygosity, H_s: Expected heterozygosity within subpopulations, H_t: Total expected heterozygosity, D_{st}: Genetic diversity between subpopulations, D_{stp}: Proportion of genetic diversity between subpopulations, F_{stp}: Standardized fixation index, F_{is}: Inbreeding coefficient within subpopulations, D_{est}: Jost's measure of differentiation, MAF: Minor allele frequency.

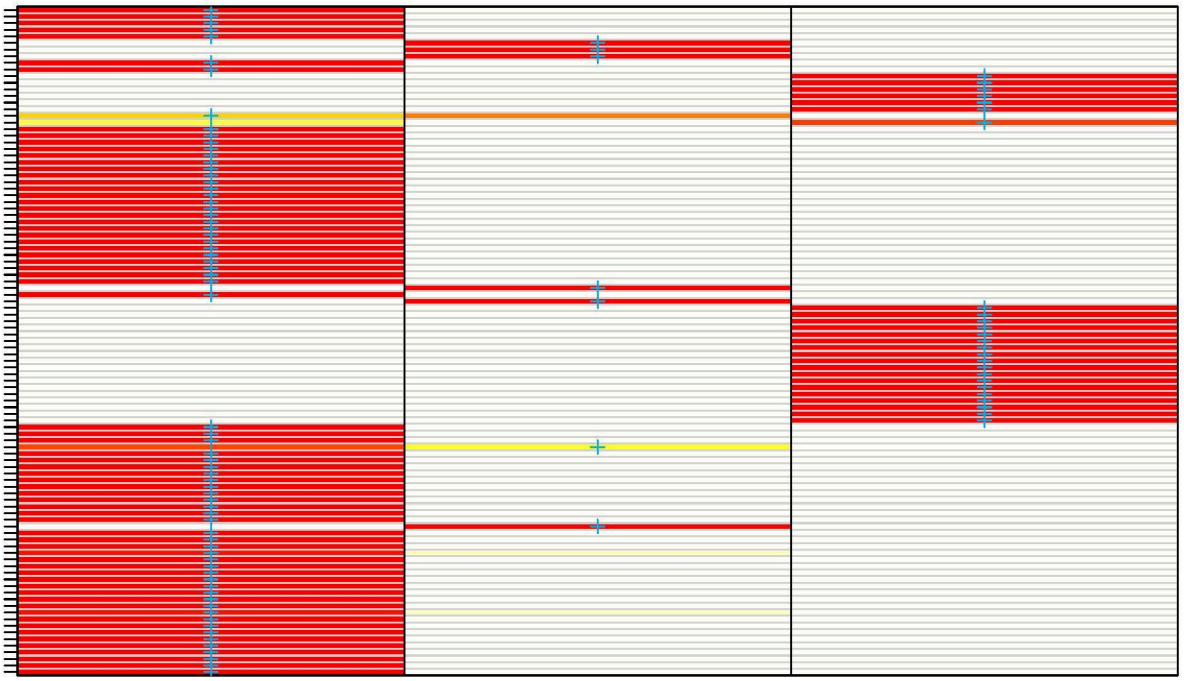
Marker	H_o	H_s	H_t	D_{st}	D_{stp}	F_{st}	F_{stp}	F_{is}	D_{est}	MAF
A1_ScFovFW_4_39801911	0.4020	0.5664	0.5764	0.0100	0.0116	0.0173	0.0201	-0.5717	0.0268	0.1708
B2_LB_scaffold230_250386	0.6603	0.5247	0.5269	0.0022	0.0025	0.0041	0.0048	-0.4012	0.0053	0.1553
B4scaff1534_2190118	0.6586	0.5018	0.5098	0.0081	0.0094	0.0158	0.0184	-0.3921	0.0189	0.1471
Pp_11592_67	0.6378	0.5968	0.5971	0.0004	0.0004	0.0006	0.0007	-0.5281	0.0010	0.1825
Pp_13073_67	0.5771	0.5746	0.5914	0.0168	0.0196	0.0283	0.0329	-0.6362	0.0460	0.1739
Pp_13684_59	0.4991	0.5813	0.5984	0.0171	0.0200	0.0286	0.0332	-0.5696	0.0477	0.1765
Pp_14371_58	0.6846	0.5356	0.5580	0.0225	0.0262	0.0403	0.0467	-0.5034	0.0564	0.1593
Pp_15206_41	0.7262	0.5040	0.5151	0.0112	0.0130	0.0217	0.0252	-0.4105	0.0262	0.1479
Pp_173448_40	0.6898	0.5587	0.5576	-0.0011	-0.0013	-0.0020	-0.0024	-0.4169	-0.0030	0.1678
Pp_175816_57	0.5147	0.5798	0.5900	0.0102	0.0119	0.0172	0.0200	-0.6014	0.0282	0.1759
Pp_1802_39	0.6984	0.5625	0.5637	0.0011	0.0013	0.0020	0.0024	-0.4400	0.0030	0.1693
Pp_1876_44	0.6655	0.5196	0.5240	0.0044	0.0052	0.0085	0.0099	-0.4012	0.0108	0.1534
Pp_19070_33	0.2444	0.3413	0.3413	-0.0001	-0.0001	-0.0002	-0.0002	-0.1756	-0.0001	0.0942
Pp_19553_43	0.5615	0.5911	0.5969	0.0057	0.0067	0.0096	0.0112	-0.5900	0.0163	0.1803
Pp_22460_56	0.5893	0.5904	0.5928	0.0024	0.0028	0.0041	0.0048	-0.5076	0.0069	0.1800
Pp_24442_37	0.6863	0.5752	0.5792	0.0040	0.0047	0.0069	0.0081	-0.4738	0.0110	0.1741
Pp_26016_36	0.6031	0.5855	0.5928	0.0073	0.0085	0.0123	0.0144	-0.5581	0.0206	0.1781
Pp_28960_47	0.5251	0.5810	0.5910	0.0100	0.0117	0.0170	0.0197	-0.6663	0.0279	0.1763
Pp_30116_42	0.6828	0.3415	0.3637	0.0223	0.0260	0.0612	0.0707	-0.2995	0.0394	0.0943
Pp_30396_46	0.5615	0.5713	0.5888	0.0175	0.0204	0.0298	0.0345	-0.5326	0.0477	0.1726

Pp_31480_43	0.7036	0.5396	0.5477	0.0081	0.0095	0.0149	0.0173	-0.4262	0.0206	0.1607
Pp_3169_46	0.7626	0.5408	0.5403	-0.0005	-0.0006	-0.0009	-0.0010	-0.4109	-0.0012	0.1612
Pp_34542_62	0.7678	0.5866	0.5896	0.0030	0.0034	0.0050	0.0058	-0.4995	0.0083	0.1785
Pp_35960_67	0.7972	0.5574	0.5585	0.0010	0.0012	0.0018	0.0021	-0.4634	0.0027	0.1674
Pp_36862_60	0.7036	0.5277	0.5333	0.0056	0.0066	0.0106	0.0123	-0.4001	0.0139	0.1564
Pp_36999_64	0.5182	0.4090	0.4578	0.0488	0.0570	0.1067	0.1223	-0.4333	0.0964	0.1156
Pp_384_47	0.6013	0.5142	0.5351	0.0209	0.0243	0.0390	0.0452	-0.4386	0.0501	0.1515
Pp_38839_43	0.6551	0.5596	0.5731	0.0134	0.0157	0.0235	0.0273	-0.5249	0.0356	0.1682
Pp_45419_53	0.6638	0.5699	0.5740	0.0042	0.0049	0.0073	0.0085	-0.4632	0.0114	0.1721
Pp_46191_47	0.6500	0.5767	0.5776	0.0009	0.0010	0.0016	0.0018	-0.4625	0.0025	0.1747
Pp_46576_61	0.6447	0.4983	0.5027	0.0045	0.0052	0.0089	0.0104	-0.3514	0.0104	0.1458
Pp_4658_45	0.3206	0.5751	0.5785	0.0033	0.0039	0.0058	0.0067	-0.6751	0.0092	0.1741
Pp_48057_37	0.6880	0.5869	0.5959	0.0089	0.0104	0.0150	0.0175	-0.5281	0.0252	0.1786
Pp_48568_50	0.6672	0.5881	0.5891	0.0011	0.0012	0.0018	0.0021	-0.4915	0.0030	0.1791
Pp_4921_66	0.6776	0.5401	0.5430	0.0029	0.0034	0.0053	0.0062	-0.4206	0.0074	0.1609
Pp_52173_36	0.5026	0.5747	0.5865	0.0118	0.0138	0.0201	0.0234	-0.6940	0.0324	0.1739
Pp_52521_39	0.6049	0.6077	0.6048	-0.0030	-0.0035	-0.0049	-0.0057	-0.5425	-0.0088	0.1868
Pp_53213_51	0.5737	0.5806	0.5837	0.0031	0.0036	0.0054	0.0062	-0.5859	0.0087	0.1762
Pp_5381_35	0.6326	0.5886	0.5942	0.0056	0.0066	0.0095	0.0111	-0.5639	0.0160	0.1793
Pp_54515_37	0.5494	0.5854	0.5961	0.0108	0.0126	0.0181	0.0210	-0.6171	0.0303	0.1781
Pp_54595_49	0.3154	0.5390	0.5627	0.0237	0.0277	0.0421	0.0488	-0.6260	0.0600	0.1605
Pp_55405_37	0.6395	0.5557	0.5750	0.0193	0.0225	0.0336	0.0389	-0.5377	0.0507	0.1667
Pp_55457_45	0.6066	0.5978	0.6006	0.0028	0.0032	0.0046	0.0054	-0.5330	0.0080	0.1829
Pp_55664_36	0.5927	0.6013	0.6014	0.0001	0.0002	0.0002	0.0003	-0.5257	0.0004	0.1843
Pp_55867_56	0.4020	0.6000	0.5967	-0.0033	-0.0038	-0.0055	-0.0064	-0.5978	-0.0095	0.1838
Pp_56318_48	0.6759	0.5737	0.5894	0.0157	0.0183	0.0266	0.0309	-0.5305	0.0429	0.1735

Pp_57098_41	0.6983	0.5566	0.5622	0.0055	0.0065	0.0098	0.0115	-0.4517	0.0146	0.1671
Pp_57306_61	0.4714	0.5820	0.5867	0.0047	0.0055	0.0081	0.0094	-0.6455	0.0132	0.1767
Pp_57395_51	0.7279	0.5643	0.5789	0.0146	0.0171	0.0253	0.0293	-0.5011	0.0392	0.1700
Pp_57509_41	0.5009	0.6068	0.6062	-0.0005	-0.0006	-0.0009	-0.0011	-0.5855	-0.0016	0.1865
Pp_57624_56	0.6083	0.5599	0.5601	0.0002	0.0002	0.0003	0.0004	-0.4283	0.0005	0.1683
Pp_58527_67	0.6950	0.5714	0.5702	-0.0012	-0.0014	-0.0022	-0.0025	-0.4413	-0.0034	0.1727
Pp_5858_32	0.6274	0.5982	0.6015	0.0033	0.0039	0.0055	0.0064	-0.5592	0.0096	0.1831
Pp_59375_34	0.3501	0.5762	0.5787	0.0025	0.0029	0.0043	0.0050	-0.7084	0.0068	0.1745
Pp_59917_38	0.5962	0.5420	0.5526	0.0106	0.0124	0.0192	0.0224	-0.4438	0.0271	0.1616
Pp_61843_33	0.6638	0.5384	0.5664	0.0279	0.0326	0.0493	0.0571	-0.5160	0.0706	0.1603
Pp_62972_54	0.7279	0.5207	0.5188	-0.0019	-0.0023	-0.0037	-0.0043	-0.3951	-0.0047	0.1538
Pp_64069_38	0.5927	0.5484	0.5577	0.0093	0.0108	0.0167	0.0194	-0.4574	0.0240	0.1640
Pp_65266_41	0.6932	0.5338	0.5355	0.0016	0.0019	0.0030	0.0035	-0.3882	0.0041	0.1586
Pp_66780_59	0.4437	0.5973	0.6050	0.0077	0.0090	0.0127	0.0148	-0.5698	0.0223	0.1827
Pp_67296_67	0.4939	0.5778	0.5821	0.0043	0.0051	0.0075	0.0087	-0.5237	0.0120	0.1751
Pp_67705_46	0.5996	0.4635	0.4715	0.0081	0.0094	0.0171	0.0199	-0.3563	0.0175	0.1338
Pp_67756_60	0.5892	0.5957	0.5962	0.0006	0.0007	0.0009	0.0011	-0.5339	0.0016	0.1821
Pp_68217_48	0.2045	0.5552	0.5537	-0.0015	-0.0017	-0.0026	-0.0031	-0.6108	-0.0038	0.1665
Pp_68521_44	0.5355	0.5754	0.5801	0.0048	0.0056	0.0082	0.0096	-0.6654	0.0131	0.1742
Pp_68948_55	0.6308	0.5248	0.5736	0.0487	0.0568	0.0849	0.0977	-0.6453	0.1196	0.1553
Pp_69731_50	0.3709	0.5954	0.6000	0.0046	0.0054	0.0077	0.0090	-0.6104	0.0134	0.1820
Pp_70083_60	0.5529	0.5854	0.5959	0.0105	0.0122	0.0176	0.0204	-0.5534	0.0294	0.1781
Pp_70360_32	0.5945	0.4973	0.5166	0.0194	0.0226	0.0375	0.0435	-0.4128	0.0449	0.1455
Pp_70430_56	0.6447	0.4645	0.4856	0.0211	0.0246	0.0435	0.0504	-0.3907	0.0460	0.1341
Pp_70707_59	0.6360	0.4728	0.4808	0.0079	0.0093	0.0165	0.0192	-0.3298	0.0176	0.1370
Pp_73146_38	0.6239	0.5367	0.5384	0.0017	0.0019	0.0031	0.0036	-0.4084	0.0042	0.1597

Pp_73453_33	0.5859	0.5585	0.5678	0.0093	0.0109	0.0164	0.0191	-0.4667	0.0246	0.1678
Pp_74789_47	0.5217	0.5714	0.5778	0.0064	0.0074	0.0110	0.0128	-0.5238	0.0173	0.1727
Pp_75372_48	0.5719	0.5660	0.5825	0.0165	0.0193	0.0284	0.0329	-0.6187	0.0444	0.1706
Pp_78818_46	0.6118	0.5840	0.5945	0.0105	0.0123	0.0177	0.0206	-0.6602	0.0295	0.1775
Pp_79566_67	0.6482	0.6016	0.6036	0.0020	0.0023	0.0033	0.0039	-0.5432	0.0058	0.1844
Pp_80090_42	0.4767	0.5783	0.5835	0.0052	0.0061	0.0089	0.0104	-0.5851	0.0144	0.1753
Pp_87650_38	0.6534	0.6020	0.6026	0.0006	0.0008	0.0011	0.0013	-0.5320	0.0019	0.1846
Pp_903281_51	0.7002	0.5686	0.5875	0.0189	0.0221	0.0322	0.0373	-0.6243	0.0511	0.1716
Pp_904633_55	0.3900	0.5893	0.5961	0.0068	0.0079	0.0114	0.0133	-0.5452	0.0193	0.1796
Pp_9315_64	0.5182	0.5678	0.5826	0.0148	0.0173	0.0254	0.0295	-0.6654	0.0399	0.1713
Pp_9376_58	0.5182	0.5649	0.5778	0.0129	0.0151	0.0224	0.0260	-0.7239	0.0346	0.1702
Pp_9565_61	0.6031	0.5586	0.5855	0.0269	0.0314	0.0460	0.0532	-0.6473	0.0711	0.1678
Pp_9729_66	0.3795	0.5891	0.5874	-0.0018	-0.0021	-0.0030	-0.0035	-0.5865	-0.0050	0.1795

Table S2. Individual assignment probabilities to the three genetic groups shown in figure 3b. Assignment probabilities >95% are colored in red.



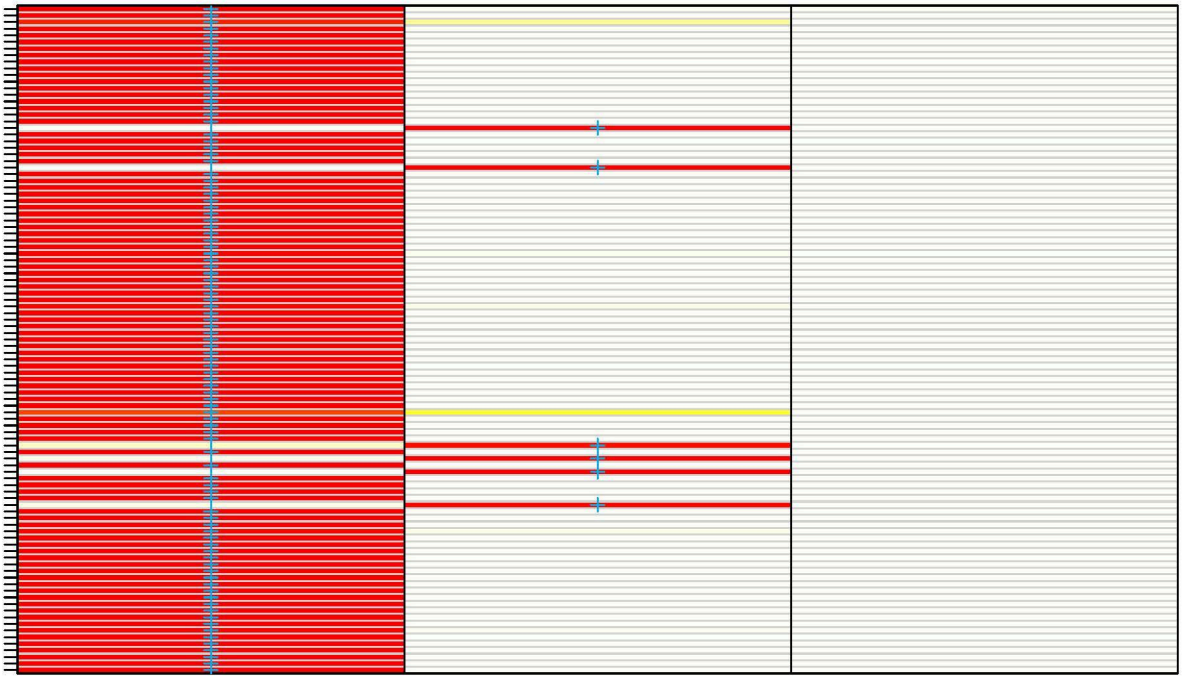


Table 3. Marker design.