Forensic Science International: Animals and Environments 6 (2024) 100093

Contents lists available at ScienceDirect

Forensic Science International: Animals and Environments

journal homepage: www.sciencedirect.com/journal/forensic-science-international-animals-and-environments

Original research

Validation of a 60K SNP chip for caribou (*Rangifer tarandus*) for use in wildlife forensics, conservation, and management

Trottier-Lavoie Mallorie [a,b], Prunier Julien [c], Poisson William [a,b], Carrier Alexandra [a,b], Gilbert Isabelle [a,b], Mastromonaco Gabriela [f], Albert Vicky [d], Cecilia Hernandez [d], Bourret Vincent [d], Taillon Joëlle [d], Droit Arnaud [c], Côté Steeve D. [e], Robert Claude [a,b,*]

[a] Département des sciences animales, Faculté des Sciences de l'Agriculture et de l'Alimentation, Université Laval, Québec, QC, Canada
[b] Centre de recherche en Reproduction, Développement et Santé Intergénérationnelle (CRDSI), Québec, QC, Canada
[c] Plateforme de Bio-Informatique, Département de médecine moléculaire, Faculté de médecine, Université Laval, Québec, QC, Canada
[d] Ministère de l'Environnement, de la Lutte contre les Changements Climatiques, de la Faune et des Parcs, Québec, QC, Canada
[e] Caribou Ungava, Département de biologie and Centre d'études nordiques, Faculté des Sciences et de Génie, Université Laval, Québec, QC, Canada
[f] Toronto Zoo, Toronto, ON, Canada

ARTICLE INFO

Keywords:
SNP chip
Genotyping
Rangifer tarandus
DNA extraction
DNA quality
Population assignment
Forensic

ABSTRACT

Large-scale genotyping platforms are currently being developed for several wild species. By querying thousands of polymorphic loci, genomics can be a useful ecological tool for describing and monitoring populations. Genomics is becoming increasingly useful as a forensic tool because of its ability to identify population of origin for purposes of enforcing anti-poaching laws. Our aim was to test the new SNP chip for caribou/reindeer (*Rangifer tarandus*) (Illumina iSelect caribou 60 K) under recommended and non-optimal sample conditions. Impact on signal detection (call rate) and error rate were assessed using reference samples. The SNP chip was shown to be robust, highly sensitive, reliable, and accurate at more than 10-fold below the recommended DNA input. Biological source of DNA had minor impact, even with fecal pellets given sufficient amount of host DNA. Hybridization of non-*Rangifer* samples as well as samples bearing DNA from two *Rangifer* samples both showed a drop in call rate and shifted levels of heterozygosity. Based on a population-targeted subset of SNPs included in the chip design, reassignment of 981 samples to a functional group (here to a caribou ecotype) was highly accurate (99.59 %) and the relative probability of reassignment error was estimated using the logarithm of odds score. Overall, the SNP chip is suitable for analysis of caribou/reindeer genomes even with suboptimal sampling and hence useful for population management and forensics.

1. Introduction

Genetic signatures are routinely used to estimate population connectivity, genetic diversity, effective population sizes and kinship, and to support forensic identification of animal provenance. DNA polymorphisms, mitochondrial DNA markers or microsatellite markers in the nuclear genome, can be used to determine the origin of single samples [34]. More recently, SNP panels have become tools of choice for this purpose, as technological platforms now allow querying of hundreds of thousands of loci for more comprehensive coverage of the genome [19,25,28,29]. Starting from a list of known single base polymorphic regions, probes flanking the polymorphic sites are designed, synthesized and printed on a solid support to allow genotyping following a single base extension reaction. Such a genotyping chip has recently been developed for *Rangifer tarandus* [6].

The term 'caribou' refers to the species *Rangifer tarandus* in North America, while 'reindeer' is used in Europe and Asia [10,33]. Woodland caribou is the only subspecies found in the province of Québec and is divided in three ecotypes: migratory, boreal and mountain [16]. Worldwide, many caribou and reindeer populations are declining, and in some cases rapidly [10,30]. Protection of this species is urgent as, in North America, several populations from different ecotypes are of concern declining in size and targeted for conservation actions [17]. Molecular genetic tools can support and improve population monitoring and conservation. Developed in collaboration with the wildlife department of the province of Québec, the genotyping chip targets a panel of

* Corresponding author at: Département des sciences animales, Faculté des Sciences de l'Agriculture et de l'Alimentation, Université Laval, Québec, QC, Canada.
E-mail address: claude.robert@fsaa.ulaval.ca (R. Claude).

https://doi.org/10.1016/j.fsiae.2024.100093
Received 14 June 2023; Received in revised form 5 August 2024; Accepted 12 September 2024
Available online 29 September 2024

63,336 SNPs including 49,725 SNPs distributed across the entire genome and known to be polymorphic in both caribou and reindeer. In addition, the SNP panel contains 7349 SNPs selected specifically to delineate the three ecotypes found in the province of Québec, as well as 1410 SNPs associated with behaviour [6,7].

Field sampling in remote locations or forensic situations can be challenging because storage and environmental conditions affect the quantity and quality of the recoverable DNA. This is especially true given that wildlife research is shifting from invasive to non-invasive sampling and monitoring methods [29,35]. Earlier genotyping platforms required DNA of consistent quality [20,21]. However, current genotyping platforms rely on hybridization of small, targeted DNA segments, which are recoverable from low quality DNA samples [29]. We therefore expected the *Rangifer* SNP genotyping chip to be robust and to perform well over a wide range of sample quality.

Here, we present the testing and validation of the *Rangifer* platform under field conditions. We examined the sensitivity, repeatability, robustness in terms of DNA quality and tissue types, sample mixing, and specificity. The choice of validation parameters was based on the standards and guidelines of the Society for Wildlife Forensic Science [18] in addition to the Scientific Working Group on DNA Analysis Methods (SWGDAM) [27]. We also tested the reliability of the procedure of assignment to the original ecotype subset in accordance with these validation parameters.

2. Materials and methods

2.1. Sample collection

Caribou were sampled from live captures (99 %) and opportunistically from wildlife officers' activities. Live wildlife sampling was performed in compliance with the Canadian Council on Animal Care guidelines and received approval from the Université Laval animal protection committee. Caribou samples (ear punches, hair follicles, blood swabs, tendons, and feces) were provided by the *Ministère de l'Environnement, de la Lutte contre les Changements Climatiques, de la Faune et des Parcs du Québec* hereinafter called the MELCCFP, who also provided muscle samples from white-tailed deer (*Odoicoleus virginianus*), moose (*Alces americanus*), and sheep (*Ovis aries*). Cell samples from American elk (*Cervus canadensis*), yak (*Bos grunniens*), European bison (*Bison bonasus*), wood bison (*Bison bison*), bighorn sheep (*Ovis canadensis*), and barbary sheep (*Ammotragus lervia*) were provided in cryopreserved pellet form (7×10^5 cells) by the Toronto Zoo (Ontario, Canada). Bovine and swine muscle samples were collected in local slaughterhouses (Québec, Canada). Human samples such as prostate and ovary DNA were purchased from OriGene Technologies, Inc. (Rockville, Maryland, USA). Samples of lichens *Cladonia rangiferina* and *Evernia furfuracea* were collected in the Lac-Saint-Jean area (Québec, Canada). All animal and plant samples were transportedand stored at −20°C. Cell pellets and human genomic DNA were received on dry ice and stored at −80°C.

2.2. Genomic DNA isolation

Genomic DNA was isolated from caribou muscle and tendon using MagAttract HMV DNA kits (Qiagen, Mississauga, ON, Canada) with 40 μL of proteinase K before the overnight incubation step. For ear punches, hair (~300 follicles/sample), blood swabs, and cell pellets, DNeasy Blood & Tissue kits (Qiagen) were used in accordance with, respectively, the *purification of total DNA from animal tissues*, *purification of DNA from nails, hair, or feathers*, *isolation of total DNA from surface and buccal swabs*, and *purification of total DNA from animal blood or cells* protocols. Lichen samples were homogenized in a bead disruptor homogenizer (VWR International, Mississauga, ON, Canada) with 800 μL of AP1 buffer (Qiagen). Qiagen DNeasy plant Mini kits were then used in accordance with the manufacturer's instructions. Genomic DNA was extracted from fecal samples using DNeasy Blood & Tissue kits (Qiagen) in accordance with an MELCCFP in-house protocol (adapted from [3]). Briefly, two frozen fecal pellets were thawed in phosphate buffered saline (Sigma-Aldrich, Saint-Louis, Missouri, USA) for 4–6 minutes. The mucus coat was removed carefully from the thawed fecal matter using a cotton swab, which was placed (tip only) in a 1.5 mL microcentrifuge tube with 360 μL of ATL buffer and 40 μL of proteinase K. After thorough vortex mixing, the tube was placed in a rotator oven (Robbins Scientific model 400) for 2 h at 56°C and 12 rpm. RNAse (4 μL, 100 mg/mL) was then mixed in and held for 2 min at room temperature, and 400 μL of AL buffer (Qiagen) was mixed in, followed by 10 min in the rotator oven at 56°C and 12 rpm. Ethyl alcohol anhydrous (400 μL) was then mixed in, the swab tip was removed from the tube with sterile forceps, placed on a DNeasy mini spin column and centrifuged for 1 min at 4200 x g. The swab was then removed and discarded, the buffer/EtOH mixture was placed in the column and centrifuged for 1 min at 12,100 x g. The eluted extracts were stored at −80°C.

2.3. Genomic DNA integrity, concentration measurement and genotyping

Total gDNA integrity and concentration were evaluated using the 4200TapeStation and genomic DNA kit (Agilent Technologies, Palo Alto, CA, USA). The DNA integrity number (DIN) is on a 1–10 scale representing DNA sample quality in terms of degree of fragmentation, 1 represents highly degraded samples and 10 indicates no fragmentation and hence the best possible quality.

All samples were genotyped using the caribou/reindeer 60 K SNP BeadChip from Illumina [6] developed using sequences from nearly a thousand samples aligned with a caribou reference genome [24]. Sample hybridizations and genotype calling were performed at the *Centre d'expertise et de services de Génome Québec* (Montreal, Québec, Canada).

2.4. Validation parameters

2.4.1. Sensitivity

Sensitivity was evaluated using three caribou designated: *RATA-F* (muscle), *CA-35* (tendon), and *0006801* (muscle). For *RATA-F*, gDNA was isolated from 6 samples (25 mg each). For *CA-35* and *0006801*, 12 samples were used. All were analyzed for integrity and concentration and then pooled, giving 39,75 ng of gDNA with an average DIN of 8.7 for *RATA-F*, 54,22 ng (DIN 8.8) for *CA-35*, and 100,00 ng (DIN 6.9) for *0006801*. The three pools were each concentrated to 350 ng/μL using a Speedvac, diluted with Tris-EDTA buffer to 200 ng/μL and then 2-fold serially down to 0.195 ng/μL. For genotyping, duplicates at 6.25 ng/μL to 350 ng/μL and triplicates at 0.195 ng/μ to L 3.125 ng/μL were plated. All pools were genotyped on different plates and on different runs to account for batch effect. A total of 85 samples were genotyped (20 μL each). Genotyping error percentage was evaluated by comparing each DNA quantity to its reference of 2000 ng.

2.4.2. Repeatability

Repeatability was evaluated using sensitivity duplicates and triplicates test samples, from 0.195 ng/μL to 200 ng/μL. Genotyping error percentage was evaluated by comparing each animal respective triplicate and duplicate genotypes for each DNA quantity.

2.4.3. Robustness (DNA quality and tissue type)

The effect of DNA quality on the robustness of the results was evaluated using 497 samples (ear punches or hair follicles) at concentrations ranging from 50 ng/μL to 100 ng/μL and DIN ranging from 1.0 to 9.7. Samples were distributed in 20 μL aliquots on 8 different plates and genotyped on different runs.

The impact of tissue type on the robustness of the DNA analysis was evaluated using ear punches (n = 6), hair follicles (n = 6), muscle (n = 6), tendon (n = 3), feces (n = 5), and blood swabs (n = 4). These samples represented various DNA integrity numbers and concentrations

Table 1
DNA integrity number (DIN) and DNA concentration in samples used to test the robustness of the caribou SNP chip.

Sample type	N	DIN	DNA Concentration (ng/μL)
Ear punch	864	2.7–9.0	100
Hair follicle	111	1.0–9.7	100
Muscle biopsy	6	6.0–8.8	100
Tendon	3	8.4–9.1	100
Fecal pellet	6	5.8–7.7	15–54
Blood swab	4	3.2–5.9	5–100

(Table 1). Samples were distributed in 20 μL aliquots on a single plate and genotyped at the same time.

2.4.4. Effect of sample mixing/contamination

Mixed samples contained DNA from two animals paired in ratios of 87.5/12.5, 75/25 and 50/50. The pairings considered mixing caribous from different ecotypes: migratory + migratory, migratory + boreal (twice), and boreal + boreal (4 animals in total). Each animal was genotyped independently to serve as a reference genotype. All samples were adjusted to 20 μL at 100 ng/μL, distributed on the same plate, and genotyped in the same run.

2.4.5. Specificity

The specificity of the analysis was evaluated by comparing 13 different species (white-tailed deer (*Odocoileus virginianus*); moose (*Alces americanus*), american elk (*Cervus canadensis*); yak (*Bos grunniens*), cattle (*Bos taurus*), wood bison (*Bison bison*), European bison (*Bison bonasus*), sheep (*Ovis aries*), bighorn sheep (*Ovis canadensis*), Barbary sheep (*Ammotragus lervia*), human (*Homo sapiens*), pig (*Sus scrofa domesticus*) and two potential environmental contaminants, caribou/reindeer lichen (*Cladonia portentosa*) and (*Evernia furfuracea*). Three samples from each species (two from yak, barbary sheep, and sheep) were genotyped on the caribou SNP chip. Samples (20 μL at 100 ng/μL) were distributed on two different plates and genotyped at different moments. Three caribou samples with 2000 ng of DNA and three with 4 ng of DNA (3 other animals) were used as high- and low-quality positive controls, respectively.

2.5. Ecotype/population assignment

Samples used to validate ecotype and population assignments were collected over several years in Québec. Telemetric data from radio collared animals was used to provide positive reference data for testing accuracy of population assignment. DNA was extracted from a total of 981 samples (864 ear punches, 111 hair follicles samples or 6 fecal pellets samples) and genotyped on the SNP chip.

2.6. Statistics

For each SNP, positive signals are based on genotype clusters defined by all signals obtained from every sample hybridized so far. For the project, data from more that 2000 samples were used to define signal position within clusters and to identify outliers for every SNP. Call rate and percent genotyping error were chosen as the basic quality control metrics. The call rate is the number of SNPs unambiguously genotyped (generating a positive signal within the genotype clusters) out of the number of SNPs targeted by the chip. The percent genotyping error is the proportion of called genotypes that are not matched (allele dropout or false alleles) with the reference sample positive signal in compliance with Illumina DNA quantity and quality criteria. The JASP statistics implementation software (https://jasp-stats.org/ v0.17) was used unless otherwise specified.

2.6.1. Sensitivity

The relationship between DNA quantity and call rate or genotyping error is expected to fit a sigmoid distribution as we can expect very poor results when DNA quantity is very low. However, the data distribution (see result section) clearly showed that we did not reach the point of very low call rate (or very high genotyping error) even using minimal DNA quantities that could be sampled in the field. Thus, we best modeled this relationship by an exponential regression using the *lm*() function from the *stats* R (v4.2.3) package. Because the sensitivity data were not normally distributed according to the Shapiro-Wilk test and were repeated measurements, and each group contained more than two samples, a Friedman ANOVA test for repeated measures followed by a Bonferroni post hoc test were performed.

2.6.2. Repeatability

The genotyping error was analyzed at 4 ng to 4000 ng of DNA. Although these data were normally distributed according to the Shapiro-Wilk test, repeated measurements on more than two samples per group led to a Friedman ANOVA test for repeated measures followed by a Bonferroni post hoc test.

2.6.3. Robustness

As SNP-chip are well-known for robustness against poor quality DNA (e.g. highly fragmented), we were expecting to find no relationship between call rate and DNA quality. This relationship was tested performing regressions with linear and generalized linear models using the *OLS* and *glm* functions from the Python (3.10.12) *statsmodel* library and estimating confidence intervals for the models slopes. Then, impact of tissue type for source of DNA (which was not normally distributed, based on the Shapiro-Wilk test) was assessed using a Kruskal-Wallis ANOVA followed by a Bonferroni post hoc test.

2.6.4. Mixed/contaminated samples

To test the impact of mixing two genomes within one sample, we investigated the ratio of the proportion of homozygosity to the call rate for the pure sample (100 %) and the same sample mixed at 87.5 %, 75 % and 50 % with DNA from another caribou. SNP homozygosity is a proportion of the number of homozygous genotypes, relative to the total number of SNPs genotyped (expressed as a percentage). A multivariate normal distribution analysis was performed using the *mvrnorm* function the *MASS* R package and the 95 % confidence intervals around the centroid were determined for each mixture using the *stat_ellipse* function from the *ggplot2* R package [32]. The relationship between mixing proportions and genotyping error was tested by linear regression using the R stat *lm*() function.

2.6.5. Specificity

According to a Shapiro-Wilk test, species data were not normally distributed, and because they were not repeated measurements and were based on more than two samples per group, a Kruskal-Wallis ANOVA followed by a Bonferroni post hoc test was performed to assess the impact of the species on call rate and overall heterozygosity.

2.6.6. Assignation statistics

Because the assignment of a field sample to a population may be used to enforce anti-poaching laws, it must be highly reliable, and the error rate must be known. Genetic signatures of 981 samples from the province of Québec were assigned using the R *assignPOP* package [8]. The boreal, migratory or mountain ecotype for each sample was ascertained from telemetry monitoring carried out by the MELCCFP. The assignment of a caribou to an ecotype was validated for each animal by telemetric monitoring conducted over more than one year. Telemetric tracking allowed us to validate the assignment of a caribou to an ecotype and, almost each time, to a population. The genetic pool of each ecotype was delineated first using a subset of 5195 SNPs out of the 7349 specifically selected for ecotype discrimination on the SNP chip [6]. The 2154 SNPs discriminating the isolated herd of Gaspésie were not included as they represent a very unique genetic background [34]. The reference cluster

was defined with 968 samples to which subsequent samples were compared. Then all samples, including the 13 not used for the reference clustering, were assigned blindly to each ecotype using no other knowledge than the genomic data. Of the three models (Bayesian, RandomForest or Support-Vector-Machine) available in the *assignPOP* package for new sample assignment to an ecotype, the model that gave the lowest error rate was retained. The logarithm of odds (LOD) score for each assignment was calculated as follows:

$$LOD = -\log10((1-P_i)/(1-P_{j\text{-}k}))$$

where P_i is the highest probability associated with the predicted ecotype and $P_{j\text{-}k}$ is the sum of the probabilities associated with the other ecotypes. This LOD score was intended to show the confidence in the prediction and to control for the low confidence results possibly leading to erroneous assignment. Because assignment concerns would be associated with low LOD scores, the full range was first assessed and then all LOD scores equal to $+\infty$ (when $P_i = 1$) were reported as the maximum score of 20, a limit arbitrarily set to better appreciate lower range values on the graph.

To test the extent of precision (granularity) of assignment based on the genomic data obtained using this newly developed SNP chip, the procedure was reproduced using population information instead of ecotypes, to assign individual animals to a recognized population based on the telemetry (see [17]).

3. Results

I think a small introduction would be nice.

3.1. Sensitivity

The manufacturer's standard protocol requires at least 2000 ng of input genomic DNA to perform both quality control and the hybridization. The lower and upper DNA input limits were tested. Although DNA quantity as small as 4 ng were tested, call rates did not drop below 85 % (Fig. 1). The call rate increased and the genotyping error decreased as the quantity of DNA increased over the range from 4 to 4000 ng (Fig. 1 and Fig. 2), although the changes were not significant above 62 ng according to the Bonferroni post hoc test (about 60,000 cells given there is about 6 pg of DNA per genome). The effect was most notable up to 31 ng. Although the recommended input is around 2000 ng, these results show that the chip is very sensitive even with a DNA input 30-fold less than this amount.

3.2. Repeatability

To assess the extent of variance when identical samples are genotyped and to determine the DNA input below which genotyping will not be accurate or reliable, the same samples as for the sensitivity test were used. At the lowest DNA input tested (4 ng), the error rate (wrong genotypes) was about 4 % (Fig. 2). Over the range of 125 ng to 4000 ng of input DNA, technical replicates generated near identical results (Fig. 2), indicating good repeatability of the assay, even with 16-fold less DNA than recommended. Below 62 ng, the variance increased significantly. Based on the sensitivity and repeatability tests, this SNP chip will generate the same data when used with 125 ng to 2000 ng of input DNA.

3.3. Robustness versus sample quality

The impact of DNA fragmentation on the genotyping results shows that the call rate dropped by only 5 % for highly degraded samples. The relationship between call rate and DIN was better modeled with a linear regression (BIC=-2391) than a generalized linear model (BIC=-2155) with a significant slope of 0.0032 (95 % CI [0.002–0.004]; (Fig. 3). Although this slope was significant because of the large sample size, call rates were stable for DIN values ranging from 6 to 10 according to a partial linear regression (slope = 0.0011, 95 % CI [0.000–0.002]. Variance increased at DIN values below 6, but the call rate remained high (> 89 %). A few outliers were observed possibly related to issues that may have arisen during library preparation prior to the hybridization of the samples on the chip. Because these outliers represent only 0.8 % of the samples, we conclude that DNA fragmentation does not influence unduly the call rate obtained using this chip.

3.4. Robustness versus tissue type

Because wildlife studies may involve collecting a large variety of tissue types, we tested the ability of the caribou/reindeer SNP

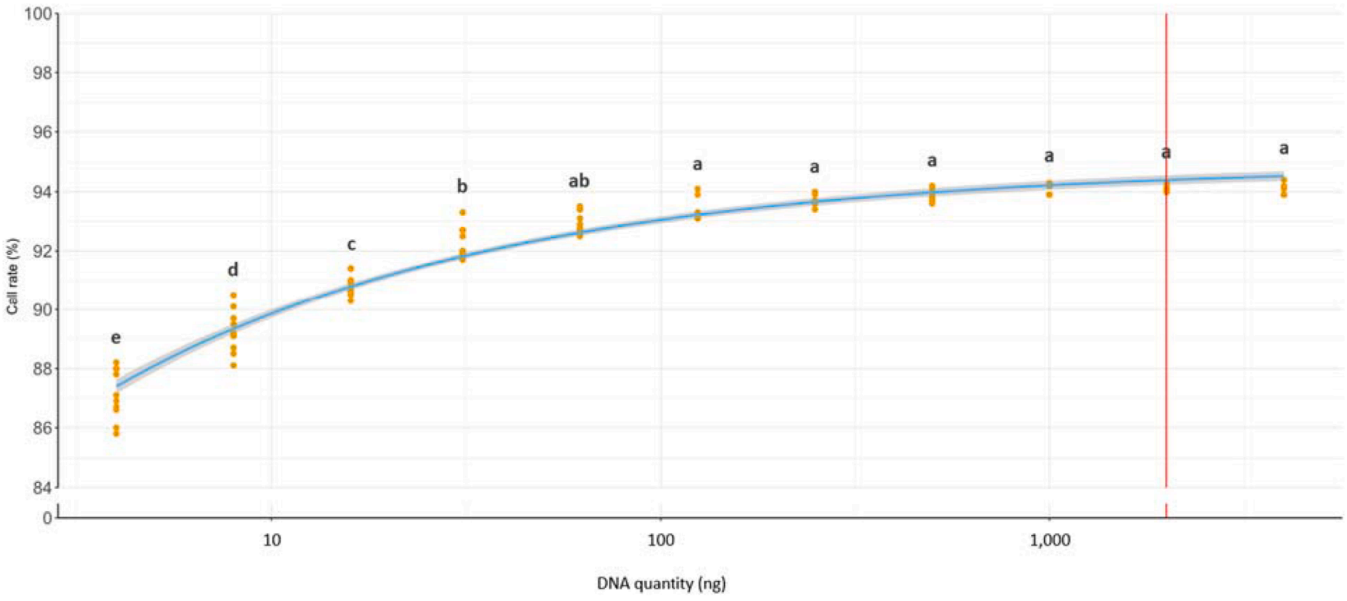

Fig. 1. Caribou SNP chip sensitivity (n = 85) based on ANOVA repeated measures analysis with Bonferroni post hoc test and exponential regression of call rate (A, $y = -13.61e^{-\log(x)} + 94.87$) versus DNA input from serial dilution samples from three animals. Red line indicates manufacturer's recommended input. Means denoted by a different letter indicate significant differences between treatments ($p < 0.05$).

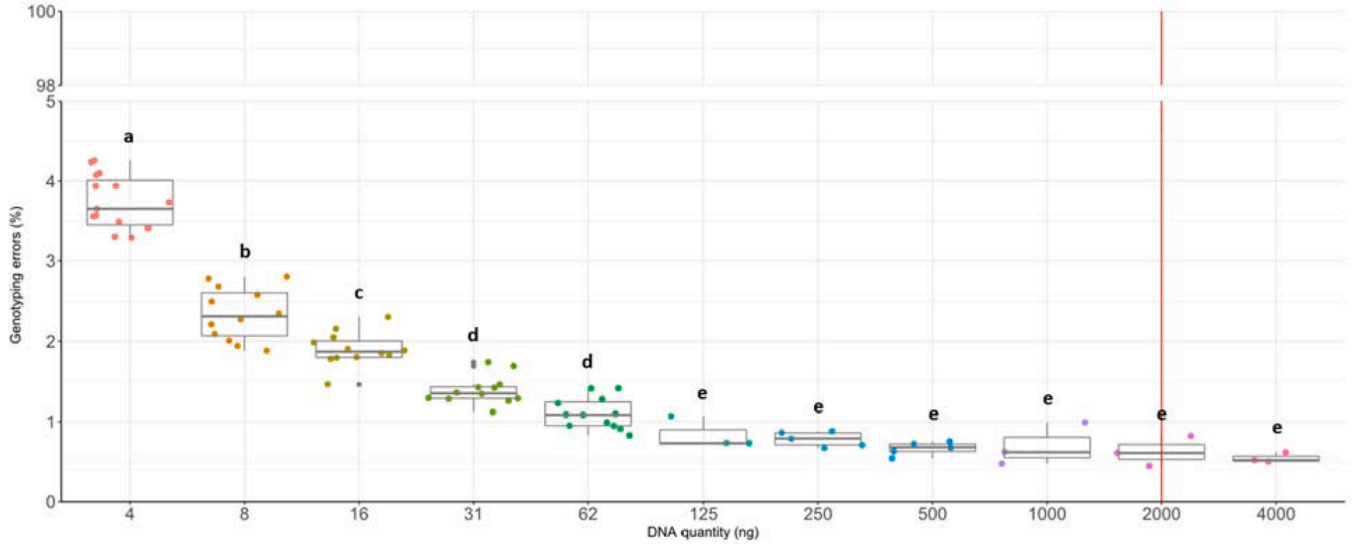

Fig. 2. Caribou SNP chip repeatability (n = 85) based on ANOVA repeated measures analysis with Bonferroni post hoc test of genotyping error versus DNA concentration in samples from three caribous. Red line indicates manufacturer's recommended input Means denoted by a different letter indicate significant differences between groups ($p < 0.05$).

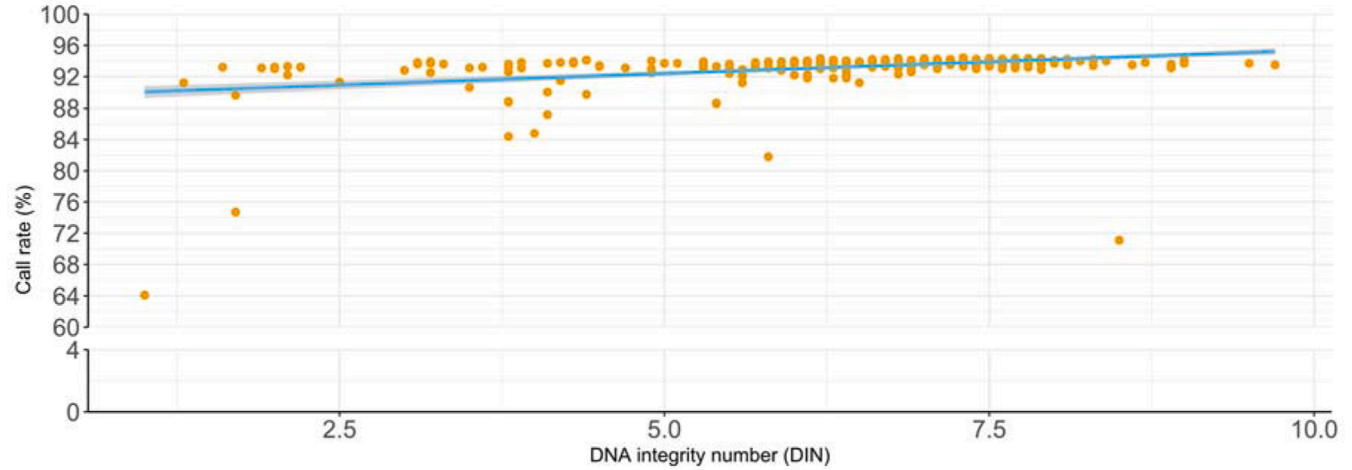

Fig. 3. Caribou SNP chip robustness when facing fragmented DNA samples (n = 497 samples). based on linear regression (slope = 0.0032 (CI: 0.002–0.004) and intercept = 0.9147)) of call rate versus sample DNA integrity number.

genotyping chip to function properly with blood, muscle, ear punch and tendon samples, hair follicles, and feces. The call rate averaged 94 % with most of these materials except for blood swab where two blood swabs had a call rate of 91 %, while the call rate of the two others was 13 %. Most tissue types thus had little or no impact on the results, whereas blood swab DNA may be an issue in some cases (Fig. 4).

3.5. Mixed/contaminated samples

In this test, the term "contamination" refers to the presence of the DNA of two animals of the same species and does not refer to environmental contamination or the presence of cross-species contamination. The effect of genome mixture ratio on the SNP chip call rate and detection of homozygosity was therefore tested. Overall, on the genotypes that were called, most were representative of the individual that was the most represented in the mix. Presence of DNA from more than one animal results in lower (< 75 %) and more variable call rates (Fig. 5). This loss of information is technical due to abnormal fluorescence ratios between both alleles triggering the analysis pipeline to indicate a "no call" result. As the proportion of DNA contamination increased, homozygosity decreased. Conversely, wrong genotypes, relative to the individual selected as reference, increased as the proportion of contaminating DNA increased (Fig. 6). Because of cluster overlapping, the SNP chip does not capture the mixing ratio, but clearly allows recognition of a mixed sample.

3.6. Specificity

The genotyping platform specificity to Rangifer was tested by comparing the call rates and heterozygous SNP percentages obtained for other *Cervidae* and some non-cervids. The call rate range differed significantly from caribou for all species except white-tailed deer, moose, and the American elk (Fig. 7A). However, these other cervids could be distinguished from caribou based on percent heterozygosity (below 4 % compared to 25 % for caribou no matter if the sample was optimal or suboptimal for input DNA) (Fig. 7B). The low heterozygosity observed in non-Rangifer species is likely due to the genomic position being monomorphic, or non-polymorphic, in those other species. The

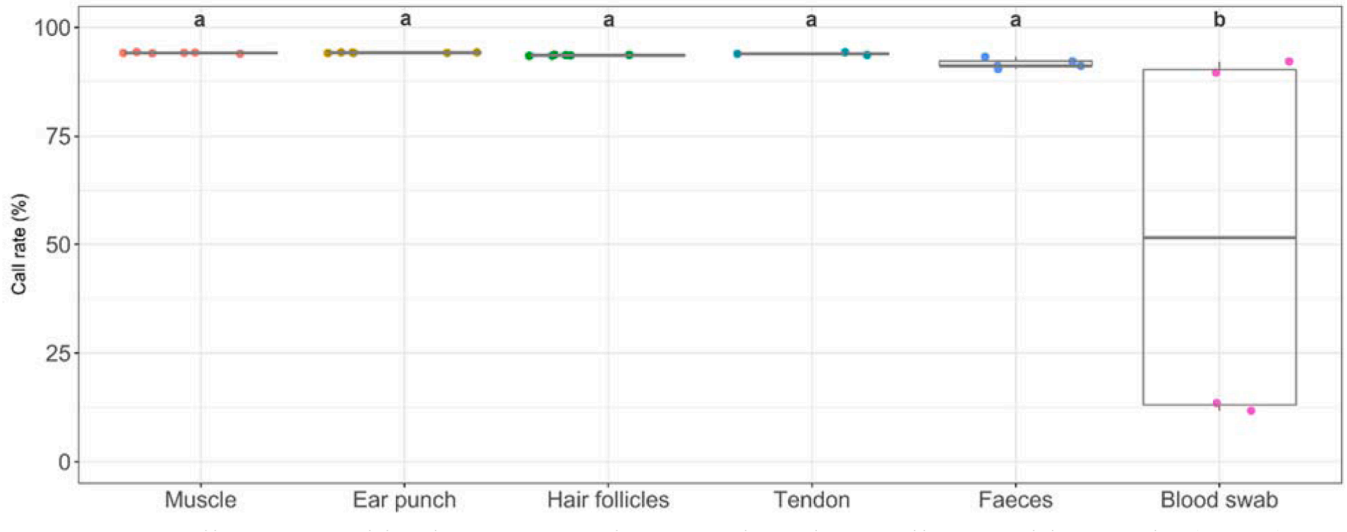

Fig. 4. Caribou SNP chip tissue type robustness based on call rate with muscle (n = 6), ear punches (n = 6), hair follicles (n = 6), tendon (n = 3), faeces (n = 5), and blood swabs (n = 4). Means denoted by a different letter indicate significant differences between treatments ($p < 0.05$).

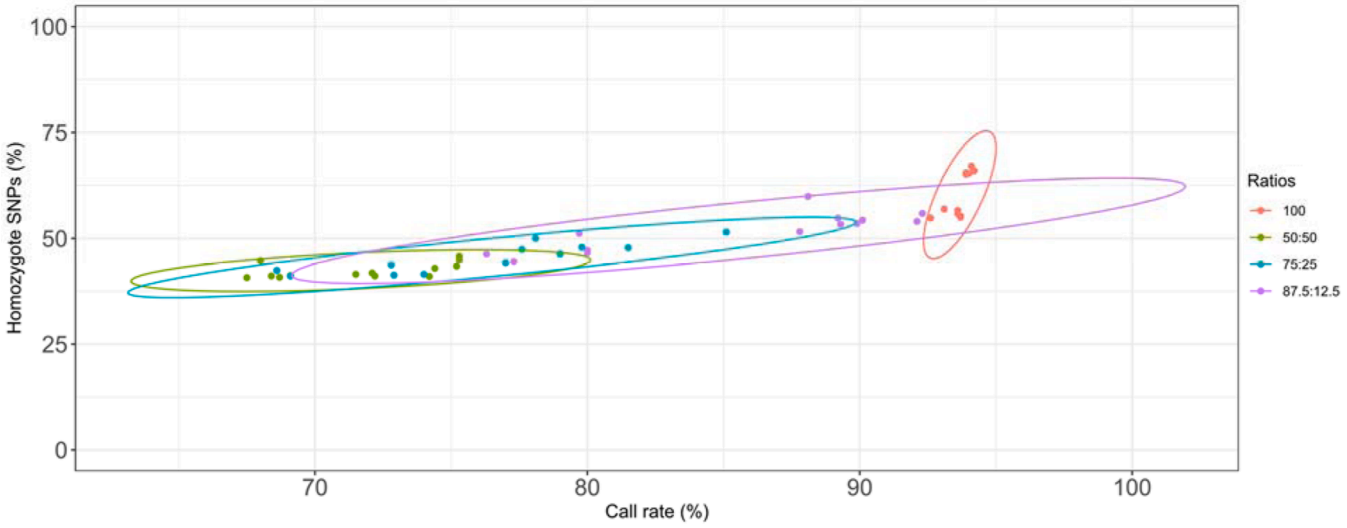

Fig. 5. Robustness of SNP genotyping of contaminated caribou samples based on multivariate normal distribution analysis of homozygosity and call rate as dependent response variables and mixing ratio (two animals-) as the explanatory variable. Ellipses indicate 95 % confidence intervals (n = 13 of each ratio).

caribou SNP chip therefore may be considered highly specific for *R. tarandus*.

3.7. Ecotype/population assignment

Among the three population-assignment models tested, the Bayesian model was the most accurate, based on the assignment error rate both in the unknown (n = 13) and the reference samples (n = 968). The overall correct assignment to the presumed ecotype based on telemetry data for each sample was 99.59 %. All 13 unknown samples were assigned correctly. Consistent with these results, the geographic distribution of the ecotypes showed an almost perfect split (Fig. 8A). Of the 981 animals, only four presented discordances between the genomic data and the telemetry data. Among these, three presented LOD scores below 5 and one presented a higher LOD score of 12.58 (green dot) (Fig. 8 A, B). All four caribous were of the boreal ecotype according to telemetry data but were inferred to be migratory based on genomic analysis.

To be able to visualize the lowest LOD values corresponding to potential assignment errors, the range maximum was set at 20. In total, 734 samples (74.8 %) reached LOD scores reaching the set maximum of 20 and 234 (23.9 %) samples presented scores between 5 and 15, while 13 samples (1.3 %) scored below 5 (Fig. 8B). Genetically less diverse populations, namely the mountain and boreal ecotypes, generated higher LOD scores than the migratory ecotype presented. This is indicative of more distinctive genetic signatures due to a more isolated and sedentary behaviour reducing allele sharing across populations.

To set minimal thresholds of acceptance, the impact of the validation parameters on assignment accuracy was explored. Using 644 samples with confirmed telemetry data and DIN information as well as the serial dilution test samples (Fig. 1), the assignment LOD scores or its accuracy did not relate to the state of DNA fragmentation (Fig. 9A). Similarly, DNA input quantity (all 981 samples of Fig. 8 and the serial dilution test samples) had no significant impact on the assignment accuracy and LOD scores (Fig. 9B). On the other hand, the presence of two different genomes within the same sample (this type of contamination is herein referred to as “mixed” samples) was found to have an impact on the assignment accuracy and LOD scores (Fig. 9C) (n = 1023 samples; all 981 pure samples and 42 mixed samples in Fig. 5). Some samples

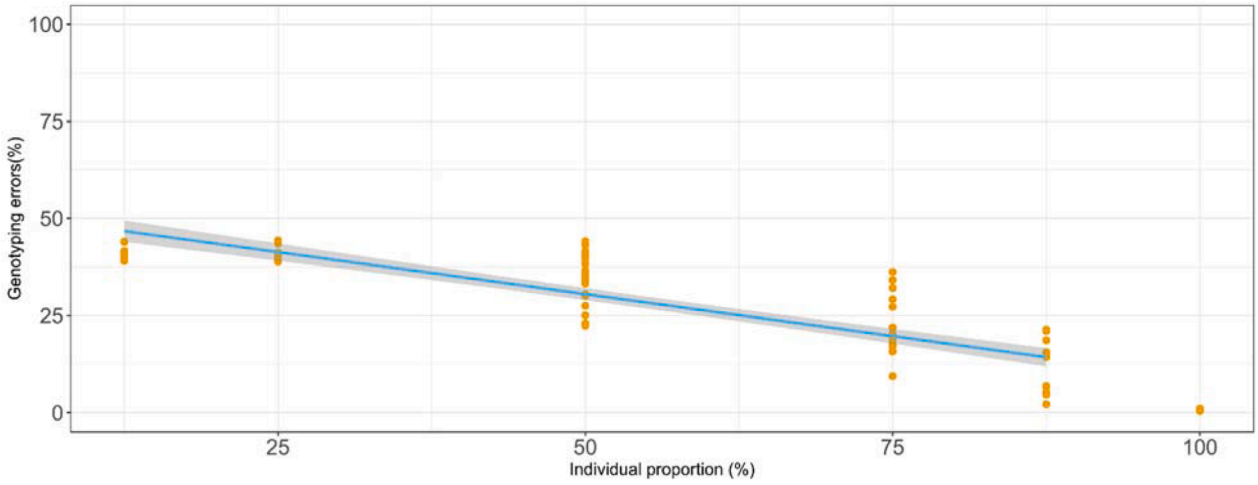

Fig. 6. SNP genotyping of mixed caribou samples (n = 65). Linear regression of genotyping error versus proportion of mixing of two caribous (y = −0.47x + 53.45).

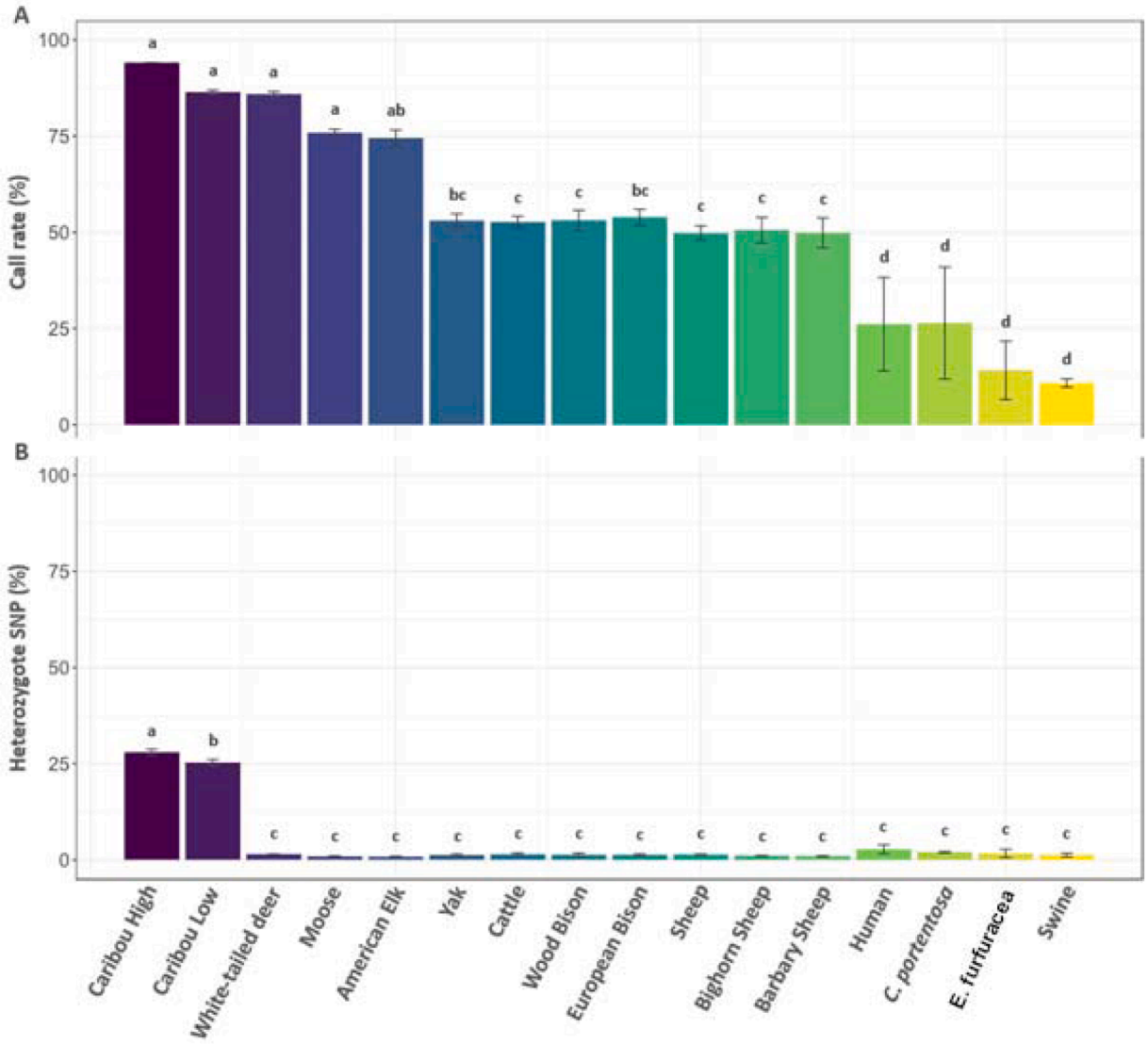

Fig. 7. Caribou SNP chip specificity (n = 2 or 3 per species) based on ANOVA analysis with Bonferroni post hoc test for call rate (A) and heterozygosity (B) of 13 genotyped species. 'High' and 'Low' refer to the DNA content of the positive controls. Means denoted by a different letter indicate significant differences between treatments ($p < 0.05$).

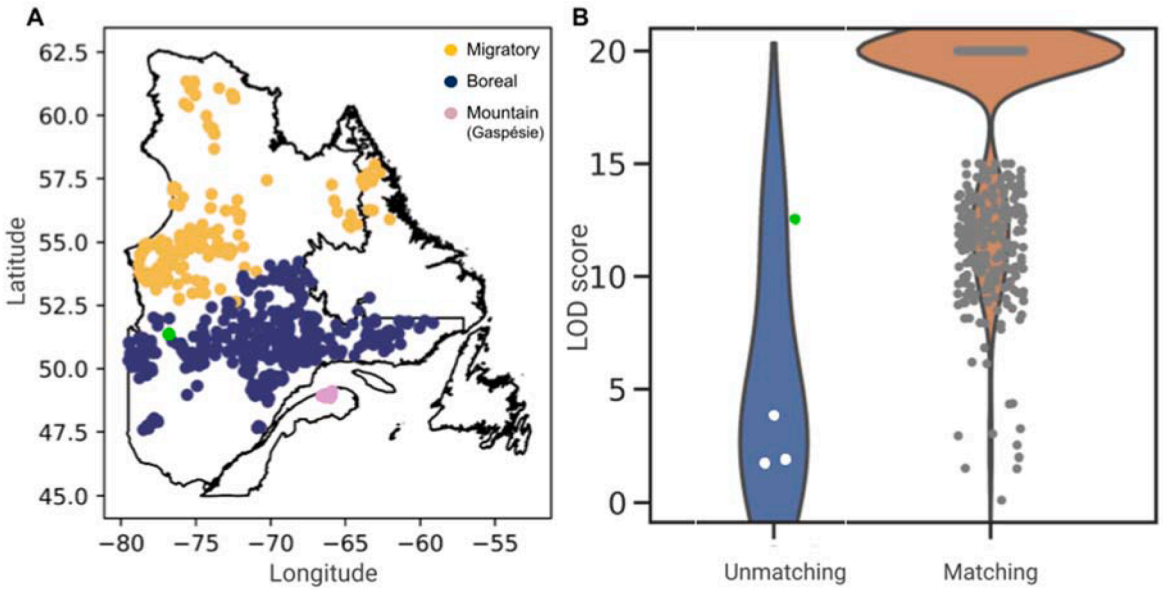

Fig. 8. Caribou ecotype assignment. 981 genotyped samples with confirmed telemetry data were tested for ecotype assignment. (A) Ecotype geographical distribution predicted by genotyping based on 5195 SNPs; (B) LOD scores associated with non-matching cases (n=4; predicted ecotype disagrees with telemetry data) and matching cases (n=977 genomic and telemetry data agree). Green dot represents geographical position of the disagreeing sample that generated the LOD score of 12.58. From telemetry data it is a boreal caribou and genomics assigned it as a migratory caribou.

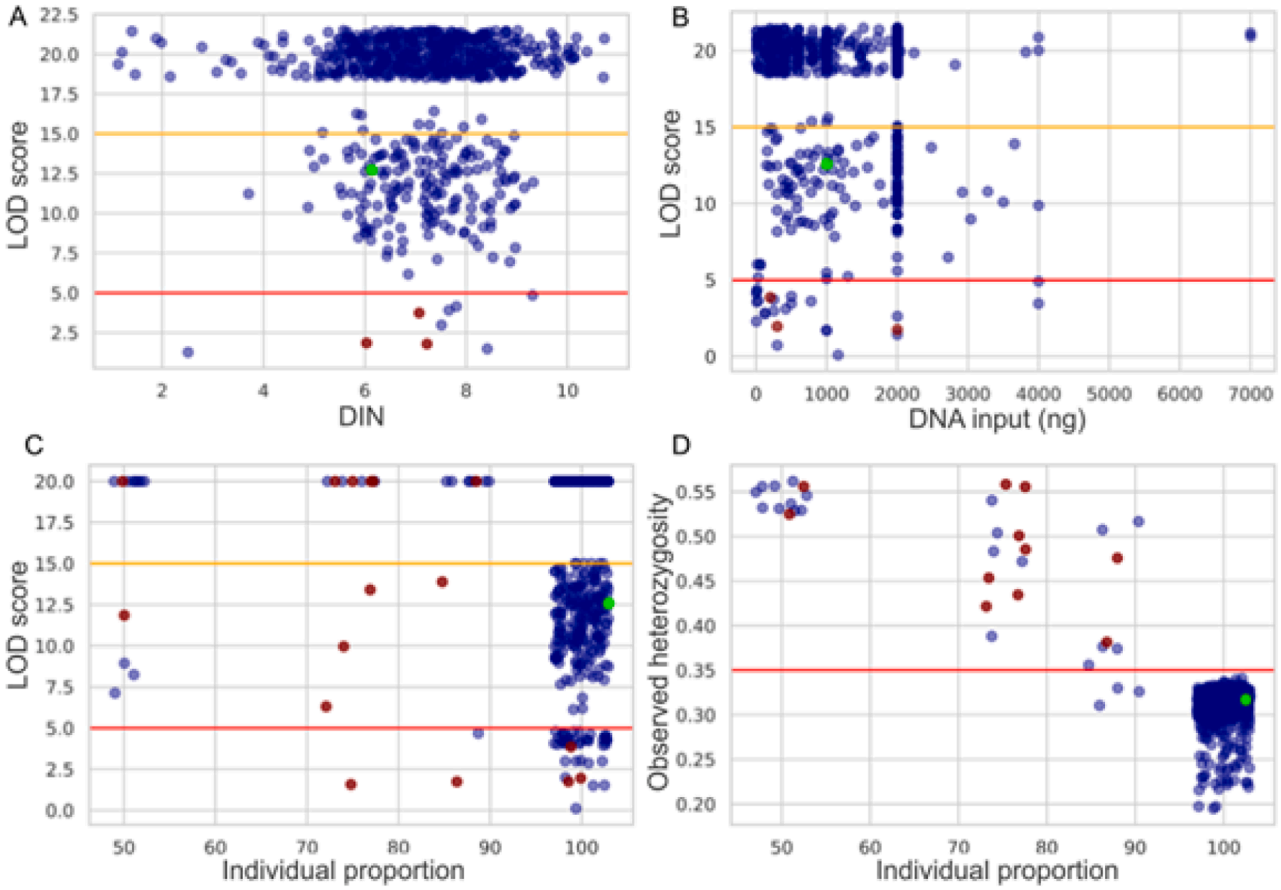

Fig. 9. Ecotype assignment robustness according to validation parameters. (A) Assignment LOD scores according to DNA quality (DIN) (n = 714 samples); (B) Assignment LOD scores according to DNA input (n = 1051 samples); (C, D) Heterozygosity and LOD scores according to individual proportion in mixed samples (n = 1023 samples). A value lower than 0.35 for heterozygosity can be used as a criterium to discriminate mixed from pure samples (C, D). Blue dots = accurate assignments; red dots = inaccurate assignments; green dot = inaccurate assignment generating LOD score of 12.58 in Fig. 8.

generating high LOD scores were found to be wrongly assigned (Fig. 9C). However, mixed samples display distinctive features allowing their identification (Figs. 5 and 9D). Mixed samples present a lower proportion of homozygote genotypes (Fig. 5) and conversely a higher proportion of heterozygosity (Fig. 9D). All wrongfully assigned mixed samples display a proportion of heterozygosity over 0.35 (Fig. 9D). Setting sample quality thresholds to a minimum LOD of 5 and $\leq$ 35 % of heterozygosity would eliminate all wrong ecotype assessments from mixed samples. Overall, only one sample meeting these minimal thresholds would remain discordant between the ecotype determined from telemetry data and from the genomic signature (the green dot sample with a 12.58 LOD in Figs. 8 and 9).

The capacity to use the genomic information to assign at the population level was tested using the population of origin confirmed by telemetry data rather than the ecotype to define models and clusters. Overall, a 93.58 % agreement was found between genomic data and telemetry data (Fig. 10). For nearly all misassignment cases (n = 67 (6.8 % of samples)), the animal was assigned to a neighboring population, suggesting that it was a recent migrant, whereas one misassignment (0.1 %) was to a more distant population and may be truly faulty in view of the very low LOD score of 0.59. As a result, the maps representing the sample population assignments based on genomic data and population information were highly similar (Fig. 10).

4. Discussion

Commercial SNP chips manufacturers provide conservative guidelines for optimal performance using optimal samples. However, field samples raise the technical challenge of analyzing DNA that may be scant or degraded [29] as well as relying on the analysis of trace samples [14]. The project's objective was to test the caribou SNP Chip [6] using experimental and field samples. The validation steps were chosen to determine the impact of suboptimal sample characteristics including DNA input and quality on call rate, error rate and population assignment precision. Overall, the genotyping platform proved to be sensitive, robust and reliable for samples that are far from the acceptable range recommended by SNP chip manufacturers. For the Illumina Bead chip, a minimal genomic DNA input of 2000 ng is recommended without integrity parameters (Infinium Assay Guide) whereas the other SNP chip manufacturer, Thermo-Fisher/Axiom, requires 200 ng of high quality (>10 kb fragment size) genomic DNA (Axiom Assay User Manual). We found that the caribou SNP chip can generate reliable data with much less DNA (16-fold less) than the recommended minimum. It has been observed previously that human genotypes can be confirmed with 10-fold less DNA than the chip manufacturer recommendation [22]. The proportion of genotyping errors was very low even when the DNA was diluted to 30-fold less than recommendations validating that the called genotypes remain reliable. The plateau nearing 94 % call rate is typical of this genotyping platform [6] and this small proportion of faulty probes was not removed from the analysis.

The performance of the caribou SNP chip was largely unaffected by DNA fragmentation, the call rate dropping by only 5 % for highly degraded samples. The lowest DIN value corresponded to an average fragment size of 250 bp whereas the highest value fragment size was over 50,000 bp [2]. It has been shown previously that DNA degraded to fragments of less than 75 bp nevertheless allowed reporting of 99–100 % of the SNPs [4]. As mentioned in other studies, SNP markers are more suitable than microsatellite markers for genotyping degraded samples since the targeted polymorphisms are single nucleotides, and shorter detection probes may be used [12,15,19,4]. This makes SNP chip platforms robust to process a wide range of DNA samples varying in concentration and/or quality.

Minute amounts of starting material are required and small biopsies of any tissues (few mm) and hair follicles (about 300) provide sufficient DNA. Similar amounts of DNA generated similar call rates, as reported previously [9]. Blood swabs were found to be more challenging. Results were most variable for blood swabs, which are collected on various surfaces by wildlife officers. While significant amounts of DNA can be retained on cotton swabs [1], contrarily to a whole blood sample where volume is known and cells can be isolated, the amount of blood present on swabs is unknown and may differ among samples. Moreover, DNA yield is unpredictable with blood swabs since the number of leukocytes per mL vary naturally depending on various conditions, including the health of the animal [23]. Genotyping failure is therefore more frequent

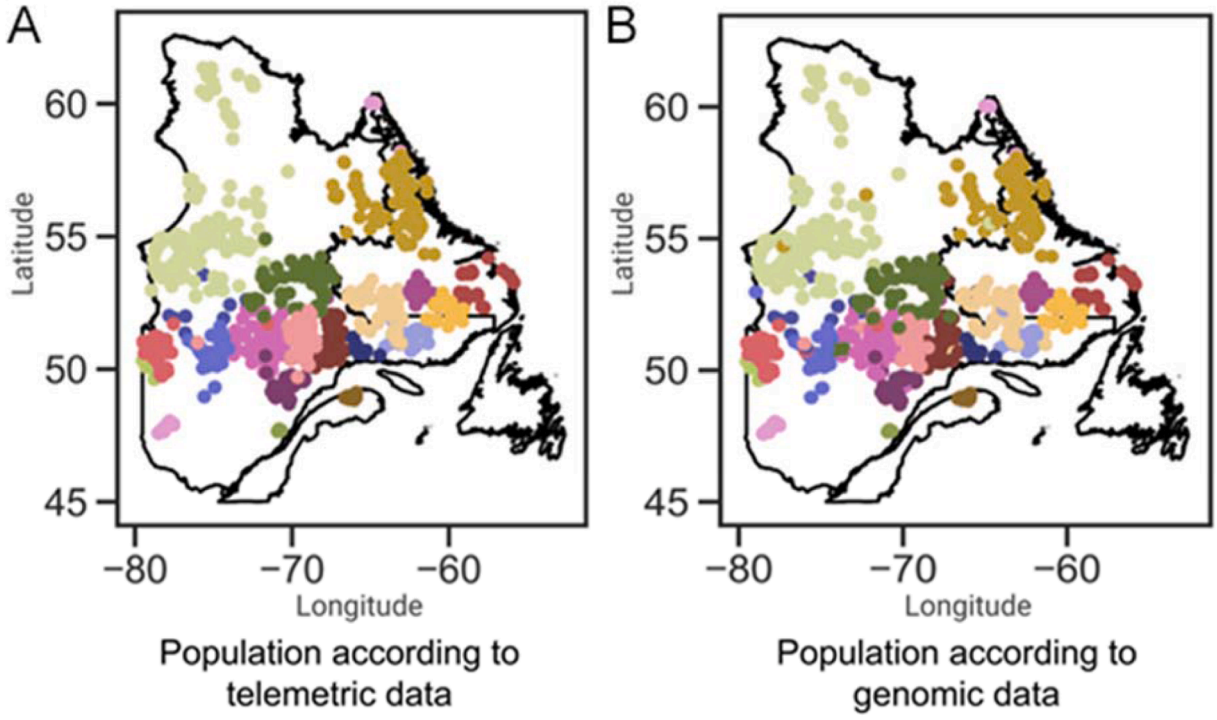

Fig. 10. Geographical distribution of caribou populations according to A) telemetry data and B) genomic data.

with blood swab samples.

One of the main challenges with fecal pellets is not knowing how much of the extracted DNA is from the host rather than plant or microbial origin. The DNA from two lichens generated minimal genotype calls. It is expected that in absence of host DNA, call rates will be very low (around 25 %) indicative of a failed sample. To maximize fecal samples success rate, collecting pellets and keeping them intact to avoid disturbing the mucus coat that will be rehydrated and carefully removed in the laboratory to extract the DNA is crucial. In this context, the amount of animal DNA present in the sample is the most important factor on genotyping results.

Since forensic samples are often collected without confirmation of the species of origin, the caribou SNP chip was therefore tested for the likelihood of generating misleading calls when facing non-*Rangifer* species. For distant species, the call rate was significantly lower. However, using the call rate alone, white-tailed deer, moose and elk could be mistaken for an extremely low quantity caribou sample (4 ng). The genetic proximity between these four species is well known [11]. Fortunately, the level of heterozygosity distinguishes caribou from all other species. The loci selected for the caribou SNP chip are mostly monomorphic in the three other cervid species, which are thus detected easily as excessive homozygosity or insufficient heterozygosity. This makes confusion unlikely even when the sample is of poor quality.

Overall, considering genotyping call and error rates, minimal input could be set to at least 62 ng of DNA without any strict consideration to DNA integrity. Based on the validation tests done herein, it is expected to provide quality data sufficient for most ecological and forensic study.

4.1. Ecotype assignment and acceptable thresholds

Analyzing the subpanel of 5195 SNPs selected for ecotype discrimination proved to be very efficient. The assignment accuracy was very high, and the robustness was linked to the LOD scores. The following intervals are therefore proposed where a LOD score below 5 indicates limited robustness or inconclusive results, 5–15 strongly supports the assignment, while a LOD $>$ 15 indicates near certainty of the assignment. By comparison, the use of a panel of 16 microsatellites [34] generated median LOD scores ranging from 2.5 to 6.2 (MELCCFP unpublished data). The use of 5195 biallelic loci provided much more discriminatory power. Comparing genomic-based and telemetry-based assignments to populations of origin further showed the power of the genotyping platform at a finer level of granularity. Almost all caribous were assigned correctly to the population assignment from telemetry [17], and misassignments were mostly to neighbouring populations, hence likely representing recent migrants or hybridization with such migrants.

Again, DNA integrity and input quantity at the tested amounts had no impact on assignment accuracy even at extremely low levels. Contrastingly, mixed samples (bearing the DNA of two animals of the same species) were often yielding inaccurate results. Forensic samples collected on hunting or investigation sites may contain DNA from more than one animal. This can of course confound genotyping results. The proportion of each animal has an impact where the genotype calls are biased towards the DNA present in the highest proportion. This could be problematic because erroneous genotyping can lead to wrongful assignment. However, mixed/contaminated samples show distinctive patterns and can be identified by using the extent of homozygosity in relation to the call rate. This provides a solution to the previously reported difficulty of identifying mixed samples when using bi-allelic SNPs [31]. In presence of two DNA donors a higher proportion of heterozygotes is expected due to the increased number of alleles present [13,4]. Given that all samples were of good quality in terms of DNA input and integrity, the drop in call rates when two donors were present was unexpected. It is caused by the genotyping platform's set parameters for genotype calling. It is based on ratios of a reference allele to an alternative allele [28]. In a single individual sample, the expected proportions are 100 % for homozygotes and 50 % for heterozygotes. However, in a mixed sample, lopsided proportions appear, leading the software to decline to call and thus increase the proportion of "no calls" [26]. For example, a sample containing 75 % heterozygote AB and 25 % homozygote AA will generate the irregular proportions 62.5 % A and 37.5 % B, which the software cannot interpret with confidence and therefore does not make a call.

When considering the following thresholds for sample acceptance and ensuing data interpretation: LOD score $\geq$ 5, heterozygosity value $\leq$ 35 % and call rate $\geq$ 90 %, only one sample remains discordant (12.58 LOD score; green dot in Figs. 8 and 9) between the ecotype defined by telemetry (boreal) and the one derived from its genomic signature (migratory). Because prosecution of poaching depends on the certainty of identifying the boreal ecotype, wrongfully assigning a migratory caribou to the boreal ecotype could be prejudicial. Conversely, as the case here, where a sample collected within the ecological range of the boreal ecotype was assigned to the migratory ecotype could most likely be due to past incursions by migratory caribous into the boreal ecotype populations [5]. With the current data, it is not possible to determine with certainty if this animal is a migratory caribou that joined a boreal population.

5. Conclusion

The 60 K caribou/reindeer SNP chip [6] was tested in accordance with the Society for Wildlife Forensic Science guidelines for DNA analysis methods [18]. Overall, this genomic tool proved to be highly robust and specific. Poor sampling conditions could affect the reliability of the analysis, but their main impact appears to be a decrease in call rate (loss of genotypic information) with minimal introduction of genotypic errors. Even mixed samples provide results that are sufficiently distinctive to be identified and discarded before ecotype assignment. Conservative sample and data quality thresholds are proposed. The use of thousands of SNPs provides unprecedented power for discriminating between closely related cervid specimens. The analysis defined clear probabilistic thresholds to categorize the robustness of ecotype assignment as inconclusive, strongly supportive, or near certain. These thresholds are intended to guide forensics and wildlife management experts in the use of the 60 K Rangifer genotyping platform.

Funding

This project was funded through a project grant awarded to C Robert, SD Côté, and A Droit in the Genomic Applications Partnership Program of Genome Canada (project title: "Use of Genomics to Manage and Protect Caribou Populations". We acknowledge co-financing by Génome Québec and the Ministère de l'Environnement, de la Lutte aux Changements Climatiques, de la Faune et des Parcs du Québec.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge Laurence Auger, Yohan Harton, David Landry, Rémi Couture, and Aïda Fall for their contribution to sample DNA extraction. We also thank the Centre d'Expertise et des Services Génome Québec at the Centre Hospitalier de Sainte-Justine, Montréal (QC, Canada), namely François Bacot, for their contribution to genotypic data production. The 60 K *Rangifer tarandus* SNP Chip is commercially available through the CES (https://cesgq.com/). We also thank ComputeCanada and CalculQuébec for access to large computing resources. Finally, we sincerely thank the reviewers for helpful comments and

suggestions that improved the manuscript.

Ethics approval and consent to participate

Nearly all samples were collected from live captures as part of caribou population monitoring by "Ministère de l'Environnement, de la Lutte aux Changements Climatiques, de la Faune et des Parcs" (MELCCFP), the ministry in charge of wildlife management in the province of Québec (Canada). Live capture and sample collection followed the Canadian Council on Animal Care guidelines, as well as the ARRIVE guidelines (https://arriveguidelines.org), and all procedures in the study were approved by the MELCCFP and the Animal Care Committees (CPA-FAUNE 18 − 04).

Consent for publication

Not applicable.

References

[1] M.S. Adamowicz, D.M. Stasulli, E.M. Sobestanovich, T.W. Bille, Evaluation of methods to improve the extraction and recovery of DNA from cotton swabs for forensic analysis, PloS One *9* (12) (2014) e116351, https://doi.org/10.1371/journal.pone.0116351.

[2] Agilent Technologies Inc, DNA Integrity Number (DIN) For the Assessment of Genomic DNA Samples in Real-Time Quantitative PCR (qPCR) Experiments, 2015. ⟨https://www.agilent.com/cs/library/applications/5991-6368EN.pdf⟩.

[3] M.C. Ball, R. Pither, M. Manseau, J. Clark, S.D. Petersen, S. Kingston, N. Morrill, P. Wilson, Characterization of target nuclear DNA from faeces reduces technical issues associated with the assumptions of low-quality and quantity template, Conserv. Genet. *8* (2007) 577–586, https://doi.org/10.1007/s10592-006-9193-y.

[4] N. Bose, K. Carlberg, G. Sensabaugh, H. Erlich, C. Calloway, Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples, Forensic Sci. Int. Genet. *34* (2018) 186–196, https://doi.org/10.1016/j.fsigen.2018.01.010.

[5] M. Boulet, S. Couturier, S.D. Côté, R.D. Otto, L. Bernatchez, Integrative use of spatial, genetic, and demographic analyses for investigating genetic connectivity between migratory, montane, and sedentary caribou herds, Mol. Ecol. *16* (20) (2007) 4223–4240, https://doi.org/10.1111/j.1365-294X.2007.03476.x.

[6] A. Carrier, J. Prunier, W. Poisson, M. Trottier-Lavoie, I. Gilbert, M. Cavedon, K. Pokharel, J. Kantanen, M. Musiani, S.D. Côté, V. Albert, J. Taillon, V. Bourret, A. Droit, C. Robert, Design and validation of a 63K genome-wide SNP-genotyping platform for caribou/reindeer (Rangifer tarandus), BMC Genom. *23* (1) (2022) 687, https://doi.org/10.1186/s12864-022-08899-6.

[7] M. Cavedon, B. vonHoldt, M. Hebblewhite, T. Hegel, E. Heppenheimer, D. Hervieux, S. Mariani, H. Schwantje, R. Steenweg, M. Watters, M. Musiani, Selection of both habitat and genes in specialized and endangered caribou, Conserv. Biol.: J. Soc. Conserv. Biol. (2022), https://doi.org/10.1111/cobi.13900.

[8] K.-Y. Chen, E.A. Marschall, M.G. Sovic, A.C. Fries, H.L. Gibbs, S.A. Ludsin, assignPOP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework, Methods Ecol. Evol. *9* (2) (2018) 439–446, https://doi.org/10.1111/2041-210X.12897.

[9] C.L. Dagnall, L.M. Morton, B.D. Hicks, S. Li, W. Zhou, E. Karlins, K. Teshome, S. Chowdhury, K.S. Lashley, J.N. Sampson, L.L. Robison, G.T. Armstrong, S. Bhatia, G.A. Radloff, S.M. Davies, M.A. Tucker, M. Yeager, S.J. Chanock, Successful use of whole genome amplified DNA from multiple source types for high-density Illumina SNP microarrays, BMC Genom. *19* (1) (2018) 182, https://doi.org/10.1186/s12864-018-4572-6.

[10] M. Festa-Bianchet, J.C. Ray, S. Boutin, S.D. Côté, A. Gunn, Conservation of caribou (Rangifer tarandus) in Canada: An uncertain future, Can. J. Zool. *89* (5) (2011) 419–434.

[11] F. Fontana, M. Rubini, Chromosomal evolution in cervidae, Biosystems *24* (2) (1990) 157–174, https://doi.org/10.1016/0303-2647(90)90008-O.

[12] R. Fries, G. Durstewitz, Digital DNA signatures for animal tagging, Nat. Biotechnol. *19* (6) (2001) 508, https://doi.org/10.1038/89213.

[13] A.C. Jäger, M.L. Alvarez, C.P. Davis, E. Guzmán, Y. Han, L. Way, P. Walichiewicz, D. Silva, N. Pham, G. Caves, J. Bruand, F. Schlesinger, S.J.K. Pond, J. Varlaro, K. M. Stephens, C.L. Holt, Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories, Forensic Sci. Int. Genet. *28* (2017) 52–70, https://doi.org/10.1016/j.fsigen.2017.01.011.

[14] D.-O. Jaquet-Chiffelle, E. Casey, A formalized model of the Trace, Forensic Sci. Int. *327* (2021) 110941, https://doi.org/10.1016/j.forsciint.2021.110941.

[15] R. Martínez-Arias, F. Calafell, E. Mateu, D. Comas, A. Andrés, J. Bertranpetit, Sequence variability of a human pseudogene, Genome Res. *11* (6) (2001) 1071–1085, https://doi.org/10.1101/gr.167701.

[16] MFFP. (2021a). *Espèces fauniques menacées ou vulnérables au Québec—Caribou des bois, écotype forestier*. ⟨https://www3.mffp.gouv.qc.ca/faune/especes/menacees/fiche.asp?noEsp=53⟩.

[17] MFFP. (2021b) *Revue de littérature sur les facteurs impliqués dans le déclin des populations de caribous forestiers au Québec et de caribous montagnards de la Gaspésie*. ⟨https://mffp.gouv.qc.ca/wp-content/uploads/RevueLitteratureCaribou.pdf⟩.

[18] M.K. Moore, B.W. Baker, T.L. Bauman, M.K. Burnham-Curtis, E.O. Espinoza, C. S. Ferrell, G.J. Frankham, K. Frazier, J.L. Giles, D. Hawk, J.J. Rovie-Ryan, R. N. Johnson, T. Knott, I.L. Kornfield, C. Lindquist, W.D. Lord, K.L. Morgan, R. C. O'Brien, R. Ogden, T.-L. Prigge, P. Schwenke, F.T. Sitam, P. Trail, E. Wictum, P. J. Wilson, B.C. Yates, L.M.I. Webster, The society for wildlife forensic science standards and guidelines, Forensic Sci. Int.: Anim. Environ. (2021) 100015, https://doi.org/10.1016/j.fsiae.2021.100015.

[19] P.A. Morin, G. Luikart, R.K. Wayne, the SNP workshop group, SNPs in ecology, evolution and conservation, Trends Ecol. Evol. *19* (4) (2004) 208–216, https://doi.org/10.1016/j.tree.2004.01.009.

[20] P.A. Morin, M. Mccarthy, Highly accurate SNP genotyping from historical and low-quality samples, Mol. Ecol. Notes *7* (6) (2007) 937–946, https://doi.org/10.1111/j.1471-8286.2007.01804.x.

[21] R. Ogden, Unlocking the potential of genomic technologies for wildlife forensics, Mol. Ecol. Resour. *11* (s1) (2011) 109–116, https://doi.org/10.1111/j.1755-0998.2010.02954.x.

[22] C.Y. Okitsu, D.J. Van Den Berg, M.R. Lieber, C.-L. Hsieh, Reproducibility and reliability of SNP analysis using human cellular DNA at or near nanogram levels, BMC Res. Notes *6* (2013) 515, https://doi.org/10.1186/1756-0500-6-515.

[23] C. Palmer, M. Diehn, A.A. Alizadeh, P.O. Brown, Cell-type specific gene expression profiles of leukocytes in human peripheral blood, BMC Genom. *7* (1) (2006) 115, https://doi.org/10.1186/1471-2164-7-115.

[24] J. Prunier, A. Carrier, I. Gilbert, W. Poisson, V. Albert, J. Taillon, V. Bourret, S. D. Côté, A. Droit, C. Robert, CNVs with adaptive potential in Rangifer tarandus: Genome architecture and new annotated assembly, Life Sci. Alliance *5* (3) (2022) e202101207, https://doi.org/10.26508/lsa.202101207.

[25] C. Schlötterer, The evolution of molecular markers—Just a matter of fashion? Nat. Rev. Genet. *5* (1) (2004) 63–69, https://doi.org/10.1038/nrg1249.

[26] F.J. Steemers, K.L. Gunderson, Whole genome genotyping technologies on the BeadArray™ platform. Biotechnol. J. *2* (1) (2007) 41–49, https://doi.org/10.1002/biot.200600213.

[27] SWGDAM, Scientific working group on DNA analysis methods: guidelines for the validation of probabilistic genotyping systems, Sci. Work. Group DNA Anal. Methods (2015). ⟨https://www.swgdam.org/_files/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf⟩.

[28] A. Vignal, D. Milan, M. SanCristobal, A. Eggen, A review on SNP and other types of molecular markers and their use in animal genetics, Genet. Sel. Evol. *34* (3) (2002) 275, https://doi.org/10.1186/1297-9686-34-3-275.

[29] A. von Thaden, B. Cocchiararo, A. Jarausch, H. Jüngling, A.A. Karamanlidis, A. Tiesmeyer, C. Nowak, V. Muñoz-Fuentes, Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays, Sci. Rep. *7* (1) (2017) 10768, https://doi.org/10.1038/s41598-017-10647-w.

[30] L.S. Vors, M.S. Boyce, Global declines of caribou and reindeer, Glob. Change Biol. *15* (2009) 2626–2633, https://doi.org/10.1111/j.1365-2486.2009.01974.x.

[31] A.A. Westen, A.S. Matai, J.F.J. Laros, H.C. Meiland, M. Jasper, W.J.F. de Leeuw, P. de Knijff, T. Sijen, Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples, Forensic Sci. Int. Genet. *3* (4) (2009) 233–241, https://doi.org/10.1016/j.fsigen.2009.02.003.

[32] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016. ⟨https://ggplot2.tidyverse.org⟩.

[33] G. Yannic, J. Ortego, L. Pellissier, N. Lecomte, L. Bernatchez, S.D. Côté, Linking genetic and ecological differentiation in an ungulate with a circumpolar distribution, Ecography *41* (6) (2018) 922–937.

[34] G. Yannic, L. Pellissier, J. Ortego, N. Lecomte, S. Couturier, C. Cuyler, C. Dussault, K. Hundertmark, R. Irvine, D. Jenkins, L. Kolpashikov, K. Mager, M. Musiani, K. Parker, K. Røed, T. Sipko, S. Þórisson, B. Weckworth, A. Guisan, S.D. Côté, Genetic diversity in caribou linked to past and future climate change, Nat. Clim. Change *4* (2014) 132–137, https://doi.org/10.1038/nclimate2074.

[35] M.A. Zemanova, Towards more compassionate wildlife research through the 3Rs principles: moving from invasive to non-invasive methods, Wildl. Biol. *2020* (1) (2020) 1–17.