

MISS MARIA CAVEDON (Orcid ID : 0000-0001-7534-1561)

DR. BRIDGETT VONHOLDT (Orcid ID : 0000-0001-6908-1687)

DR. STEFANO MARIANI (Orcid ID : 0000-0002-5329-0553)

Article type : Original Article

Genomics, environment and balancing selection in behaviorally bimodal populations: the caribou case

Short running title: Genes and selection in bimodal populations

Maria Cavedon,¹ Chrysoula Goubili,^{1,2,3} Elizabeth Heppenheimer,⁴ Bridgett vonHoldt,⁴ Stefano Mariani,² Mark Hebblewhite,⁵ Troy Hegel,⁶ Dave Hervieux,⁷ Robert Serrouya,⁸ Robin Steenweg,⁷ Byron V. Weckworth,⁹ Marco Musiani,^{10*}

¹ Faculty of Environmental Design, University of Calgary, Calgary, AB T2N 1N4, Canada.

² School of Environment and Life Sciences, University of Salford, Salford M5 4WT, UK.

³ Hellenic Agricultural Organisation, Fisheries Research Institute, Nea Peramos, Kavala, Macedonia, 64007, Greece.

⁴ Department of Ecology & Evolutionary Biology, Princeton University, 106A Guyot Hall, Princeton, NJ 08544-2016, USA.

⁵ Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, College of Forestry and Conservation, University of Montana, MT 59812, USA.

⁶ Yukon Department of Environment, P.O. Box 2703, Whitehorse, YT Y1A 2C6, Canada.

⁷ Resource Management - Operations Division, Alberta Environment and Sustainable Resource Development, Grande Prairie, AB T8V 6J4, Canada.

⁸ Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada.

⁹ Panthera, New York, NY 10018, USA.

¹⁰ Dept. of Biological Sciences, Faculty of Science and Veterinary Medicine (Joint Appointment), University of Calgary, Calgary, AB T2N 1N4, Canada.

*Corresponding author. E-mail: mmusiani@ucalgary.ca

Abstract

Selection forces that favor different phenotypes in different environments can change frequencies of genes between populations along environmental clines. Clines are also compatible with balancing forces, such as negative frequency-dependent selection (NFDS), which maintains phenotypic polymorphisms within populations. For example, NFDS is hypothesized to maintain partial migration, a dimorphic behavioral trait prominent in species

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.15039

This article is protected by copyright. All rights reserved.

where only a fraction of the population seasonally migrates. Overall, NFDS is believed to be a common phenomenon in nature, yet, a scarcity of studies were published linking naturally occurring allelic variation with bimodal or multimodal phenotypes and balancing selection. We applied a Pool-seq approach and detected selection on alleles associated with environmental variables along a North-South gradient in western North American caribou, a species displaying partially migratory behavior. On 51 loci, we found a signature of balancing selection, which could be related to NFDS and ultimately the maintenance of the phenotypic polymorphisms known within these populations. Yet, remarkably, we detected directional selection on a locus when our sample was divided in two behaviorally distinctive groups regardless of geographic provenance (a subset of GPS-collared migratory or sedentary individuals), indicating that, within populations, phenotypically homogeneous groups were genetically distinctive. Loci under selection were linked to functional genes involved in oxidative stress response, body development and taste perception. Overall, results indicated genetic differentiation along an environmental gradient of caribou populations, which we found characterized by genes potentially undergoing balancing selection. We suggest that the underlining balancing force, NFDS plays a strong role within populations harboring multiple haplotypes and phenotypes, as it is the norm in animals, plants and humans too.

Keywords

Negative frequency-dependent selection, balancing selection, genomics, migration, environmental and climatic factors, caribou, *Rangifer tarandus*,

Introduction

Gene-by-environment selection involves a divergent force that favors contrasting phenotypes in different environments, and is expected to result in allele frequency changes between populations along an environmental or geographic cline (Savolainen *et al.* 2013). However, such allele frequency patterns can also be actively maintained by balancing selection forces, such as negative frequency-dependent selection (NFDS), which maintains allelic polymorphisms within populations (Endler 1973). NFDS is believed to be a common phenomenon (Ayala & Campbell 1974), yet, there is a deficiency in publication history linking naturally occurring allelic variation with bimodal or multimodal phenotypes and balancing selection. One of the few examples is in orchids and morning glory flowers, which demonstrated the maintenance of genetic polymorphisms that regulate different flower pigmentation (Subramaniam & Rausher 2000; Gigord *et al.* 2001). A few others have indicated a major role of NFDS in maintaining behavioral polymorphisms, from bimodal foraging behaviors in drosophila (Greene *et al.* 2016) to individual personal and social traits in humans (Nettle 2006; Penke *et al.* 2007). Finally, it has been suggested that fitness trade-offs and NFDS can together maintain life history variation; although how these processes interact across genomes and life-history variants remains an important venue for continued research in ecology and evolution (Christie *et al.* 2018). Therefore, there remains a gap in published accounts, particularly in non-model organisms.

Many species are polymorphic with regard to phenotypic traits, and the maintenance of phenotypic diversity will depend on the strength of selection acting on phenotypic variants (Zamudio *et al.* 2016). NFDS is hypothesized to be involved in the evolution of phenotypic variation such as that found for partial migration, a behavioral dimorphism occurring within a population where a fraction of the population seasonally migrates (i.e. selecting for resources available in seasonal ranges), while other individuals are sedentary throughout the year (i.e. selecting for resources available in yearly ranges) (Chapman *et al.* 2011; Lundberg 2013; Christie *et al.* 2018). Partial migration has been observed in different species, such as passerines (Lundberg 1988), fishes (Chapman *et al.* 2012) and ungulates (Hebblewhite & Merrill 2011), including the subject of this study, caribou (*Rangifer tarandus*) (McDevitt *et al.* 2009).

Partial migration in caribou has been proposed as a consequence of the species' phylogeography, including adaptation to historical and current environmental conditions. Historically, two caribou lineages evolved in separated glacial refugia located respectively north (i.e. the Beringian–Eurasian lineage [BEL]) and south (i.e. the North American lineage [NAL]) of the ice sheet during the last glacial maximum (McDevitt *et al.* 2009). The historical northern refugia was predominantly composed of tundra habitat where migratory caribou subsisted primarily on terrestrial lichens, with the southern portion of the species' range dominated by forested environments where sedentary southern caribou grazed on terrestrial lichens and some groups browsed on arboreal lichens too (Geist 1998; Yannic *et al.* 2013). After glaciation, the species has recolonized vast areas previously covered by ice. Currently, six groups are recognized in western North America: the Barren-ground

subspecies (*R. t. groenlandicus* subspecies); the Alaskan subspecies (*R. t. granti* subspecies); and, within the Woodland Caribou subspecies (*R. t. caribou*), four distinctive groups (or “ecotypes”, defined below), Boreal, Southern Mountain, Central Mountain and Northern Mountain (Geist 1998; Bergerud *et al.* 2007; COSEWIC 2011; Weckworth *et al.* 2012). These groups are described as Designatable Units for conservation by the Federal Canadian Government (DUs, a definition largely overlapping with the ecological concept of ecotypes). Gregor (1944) defined an “ecotype” as an intraspecific group of animals or plants arising as a response to a particular habitat. The concept was dealt with explicitly to determine caribou DUs in Canada: recognized DUs utilize and are considered by COSEWIC (2011) as adapted to habitats that have divergent environmental and climatic conditions along a North-South cline (Figs. 1, 2, Fig. S1, Supporting information). This cline likely reflects diversifying evolutionary processes as caribou expanded and adapted to new, post-glacial habitats (Weckworth *et al.* 2012). Likewise, as BEL and NAL populations merged along the Canadian Rockies, ancestral characteristics, such as migratory behavior (and access to seasonally highly productive alpine tundra by some individuals), may have played a synergistic role in successful adaptation of modern day mountain caribou. However, to our best knowledge, selection studies in relation to behaviour and environmental factors have never been explicitly conducted on this species testing for a) a cline variation of genetic polymorphisms following environmental and latitudinal gradients of populations; b) allelic differences between migratory and sedentary caribou and c) a signature of balancing selection, possibly related to NFDS and ultimately responsible for the maintenance of phenotypic polymorphisms. Further evidence on the occurrence of partial migration in caribou warrants addressing these hypotheses.

Partial migration in the Southern, Central, and Northern Mountain caribou groups of western Canada has been extensively documented (Bergerud *et al.* 2007; McDevitt *et al.* 2009). Moreover, in the study by McDevitt *et al.* (2009), propensity to migrate was higher in the North where tundra-like alpine habitats were more abundant than in the South, therefore indicating a clinal variation. To resolve historical and current climatic, environmental and biogeographic patterns, McDevitt *et al.* (2009) proposed a “hybrid” origin hypothesis, suggesting that partial migration originated from post-glacial reunification of the two lineages, with subsequent adaptation to the intermediate environment. In fact, western Canada caribou populations live under different environmental conditions and, within populations, harbor both forms of migratory and sedentary behaviors, and foraging for ground or tree lichens, and the genetic and selection mechanisms to maintain such phenotypic polymorphisms are unknown.

The distribution of these binary ancestral traits in caribou across western Canada suggests a mosaic of balancing selection forces. Further studies are needed to understand the mechanisms involved in maintenance of phenotypic polymorphisms in caribou. Additionally, presence of polymorphisms motivated the need to also identify genetic regions under selection that are associated with local environmental conditions and traits, such as migratory behavior (Stapley *et al.* 2010; Ripple *et al.* 2015; Tucker *et al.* 2018). Finally, there

has been a global decline of populations of large terrestrial migratory mammals, which constitutes a loss in biodiversity and also has other negative impacts, as migratory animals play pivotal roles in ecosystems (Harris *et al.* 2009; Bauer & Hoyer 2014). It is therefore important to understand the relationship between genetics and migration in a species at risk like caribou (COSEWIC 2011), for caribou conservation and also as a sample case of other biodiversity and ecosystem function losses.

Here, we apply a Pool-seq approach and detect signatures of selection on alleles and genes associated with environmental variables along a North-South gradient in western North American caribou populations. Specifically, we document balancing selection, which could be related to the action of NFDS and ultimately the maintenance of the phenotypic polymorphisms known within these partially migratory populations. On the other hand, we detect directional selection when our sample was divided in two behaviorally distinctive groups regardless of geographic provenance (a subset of GPS-collared migratory or sedentary individuals), indicating that, within populations, phenotypically homogeneous groups were genetically distinctive. Overall, we find differentiation between caribou populations across genes potentially undergoing balancing selection and along an environmental gradient, and differentiation of migratory and sedentary individuals within populations, and we suggest that NFDS plays a strong role in species harboring multiple haplotypes and phenotypes.

Materials and Methods

Sampling, DNA extraction and library preparation

We extracted genomic DNA from archived samples of blood, muscle and skin from 270 caribou. These samples were obtained from capture and hunting activities across western North America conducted since approximately 2000 or earlier. Briefly, genomic DNA was extracted using phenol-chloroform extraction protocol and was visualized on a 1% agarose gel to assess quality of each sample and quantified on a Qubit® 2.0 fluorometer (Thermo Fisher Scientific). Samples with high RNA concentration were RNase treated using Riboshredder™ RNase Blend (Epicentre). Samples were pooled prior to sequencing (Pool-seq- Schlötterer *et al.* 2014), with libraries then prepared using a restriction site assisted DNA digestion (RAD-seq – see below). We constructed 18 pools (P1, P2 etc.) with DNA from 15 individuals per pool. Of these, 14 pools represented caribou populations distributed in a

spatial heterogeneous environment belonging to recognized subspecies and DUs: the Alaskan subspecies (*R. t. granti*, $n_{\text{pools}}=2$), the Barren-ground subspecies (*R. t. groenlandicus*, $n_{\text{pools}}=1$), and each DU of the Woodland subspecies (*R. t. caribou*) (Boreal, $n_{\text{pools}}=5$; Northern Mountain, $n_{\text{pools}}=2$; Central Mountain, $n_{\text{pools}}=3$; Southern Mountain, $n_{\text{pools}}=1$) (Fig. 1). The last four pools were composed of a subset of GPS-collared migratory individuals ($n=30$) or sedentary individuals ($n=30$) from a single geographic location including multiple populations, the central Rocky Mountains (Fig. 1). To classify individual animals as sedentary or migratory for these last four pools (P2, P3, P7 and P8), we followed McDevitt *et al.* (2009), who first defined summer ranges (1 July–15 September; Dyer *et al.* 2001, 2002) and winter ranges (1 December–30 April). Then, a given caribou was considered as migratory if it showed nonoverlapping ranges, and as sedentary if seasonal ranges were overlapping. Sedentary and migratory specimens to form subgroups P2, P3, P7 and P8 were picked from the following partially migratory pools: P5 and P6. P5 and P6 represent multiple populations of the same caribou type (Central Mountain type): À La Pêche, Redrock-Prairie Creek, Narraway, Red Willow, Quintette, Kennedy, Moberly, Scott and Parsnip. Therefore, comparison of sedentary and migratory pools should not be confounded by differences in caribou types and frequency of caribou of different types in the pools. Finally, samples were not exclusive to pools P2, P3, P7, P8, and P5 and P6 as, for example, a given caribou of known behavior (for example migratory) would appear both in P3 or P7 (migratory pools) and P5 or P6 (pools that are geographically distinct).

Each pool was quantified using a fluorometer and standardized to 10 mg, with RAD library preparation completed following Etter *et al.* (2011) (Method S1, Supporting information). Briefly, each population sample (pool, P, defined above) was digested with 20 U SbfI (New England Biolabs®) and barcoded with a unique 5-bp sequence. All barcodes differed by at least two nucleotides to aid in the quality of in silico filtering. Adaptors designed for sequencing were ligated to the samples as in Method S1 (Supporting information). Two replicate libraries were prepared with approximately equal amounts of DNA from the 18 population pools (merging procedures to obtain *de novo* assembly are explained below). Finally, each library was sequenced for 100-bp paired-end reads in one lane of the Illumina HiSeq 2000 platform located at Edinburgh Genomics Ashworth Laboratories (Edinburgh, UK), The short reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under Accession number SRP128712 (<https://www.ncbi.nlm.nih.gov/sra/SRP128712>).

Filtering, de novo Assembly, alignment and SNP calling

We removed remnant adapter sequence with Cutadapt v. 1.13 (Martin 2011), and then demultiplexed and filtered for quality using process_radtags in Stacks v. 1.29 (Catchen *et al.* 2013) using default parameters. Following Mastretta-Yanes *et al.* (2015), we merged both libraries and ran the final *de novo assembly* using the following parameters: 3 as the minimum depth of coverage required to create a locus, and 4 and 3 as the number of

allowed mismatches between loci within individuals, and between loci between individuals, respectively. (for a description of Stacks parameters see Method S2, Supporting information). The *de novo* reference genome was indexed and reads were aligned to the reference using BWA v. 0.7.12 (Li & Durbin 2009). We used Picard tools v. 1.106 to sort coordinates of the alignment and to create 18 BAM files (one per pool). We then used Samtools v. 1.3 to filter reads by mapping quality, to remove PCR duplicates and to obtain 18 pileup files. For quality filtering, we used the parameter “-q 20” as suggested in Kofler *et al.* (2011b). Also following Kofler *et al.* (2011a, 2011b), we finally called SNPs within each pool with PoPoolation 1.2.2 using as parameters a minimum of two alleles and a minimum and maximum coverage of four and 100, respectively. We then used two subset of SNPs for the following analyses. Analyses on 14 populations were based on a subset of SNPs that were biallelic and were identified in all populations. Analyses on migratory vs. sedentary pools were based on a subset of SNPs that were also biallelic but identified in all four pools made of caribou individuals with known movement behavior (above).

Genome-based diversity and differentiation estimates

We used PoPoolation v. 1.2.2 to estimate patterns of heterozygosity, nucleotide diversity (Tajima’s π) and population mutation rate (Watterson’s θ , θ_W). We used the following parameters: a minimum of two alleles were required, a minimum and maximum coverage of four and 100, respectively, and a non-overlapping window size of 300 basepairs. The values were then averaged across all SNPs for each pool. We then used the program PoPoolation2 to estimate the fixation index values (F_{ST}) for each SNP among pools pairwise, and adhered to the following parameters indicated in Kofler *et al.* (2011b): “window size” and “minimum covered fraction” both of one (to obtain single SNPs). The F_{ST} analysis was conducted separately for two sets of pools: the 14 pools (Pools 1, 4-6, 9-18) representing caribou spatially distinct populations and the 4 pools formed by sedentary (Pools 2 and 8) or migratory individuals (Pools 3 and 7).

Signature of selection – Population Genomics approach

We used an empirical F_{ST} approach (Akey *et al.* 2010) and BayeScan v. 2.1 (Foll & Gaggiotti 2008) to identify potential SNPs under selection among the 14 pools representing caribou spatially distinct populations and the four pools formed by migratory or sedentary

individuals. When analyzing caribou populations using the empirical F_{ST} approach, we categorized SNPs as outliers, and thus under selection, when F_{ST} values were within the upper 0.5% tails in at least 10 of the 91 pairwise comparisons of the 14 geographically distinct pools, corresponding to >10% of comparisons (De Carvalho *et al.* 2010; Fischer *et al.* 2013; Gagnaire *et al.* 2013; Guo *et al.* 2015, 2016; Kim *et al.* 2016). There is ample debate in the literature with regard to outliers' false positives in general, and with regard to how the cutoff level selected influences the process (reviewed by Lotterhos & Whitlock 2014, 2015). We therefore picked cutoffs standardly used in the literature (De Carvalho *et al.* 2010; Fischer *et al.* 2013; Guo *et al.* 2015, 2016).

We also tested for hypothesized differences of migratory vs. sedentary caribou. When analyzing migratory vs. sedentary pools ($n=2$ and $n=2$, respectively), SNPs were also similarly considered outliers when F_{ST} values were within the upper 0.5% tails. However, due to limited number of pools compared, we had to pick a more conservative threshold of minimum pairwise comparisons than the 10% commonly used in the literature (explained above). In theory, four pools could provide a total of six pairwise comparisons, but in practice two comparisons (of a migratory pool vs. another pool also migratory and of a sedentary pool vs. another pool also sedentary) had to be excluded, as these would be testing caribou with identical behaviour –i.e. not our hypothesized differences. We therefore categorized SNPs as outliers if the conditions above were met with regard to the tails in at least 1 of the 4 pairwise comparisons that were possible (25%) –percentages of less than 25 were not feasible.

In BayeScan, we used a false discovery rate (FDR) threshold to identify loci under selection ($FDR < 0.05$) and three different prior odds (PO of 10, 1,000 and 10,000) to account for the eventual occurrence of false positives (Schweizer *et al.* 2016). We identified 1,761, 599 and 51 outlier SNPs using Prior Odd parameters of 10, 1,000 and 10,000 respectively. We obtained input files of BayeScan with PoPoolation2, which converts SNP allele count data in genepop throughout a simulation process. BayeScan has known limitations (Narum *et al.* 2011; Lotterhos & Whitlock 2014, 2015). First, the program assumes that samples are evolutionarily independent (the program is based on the multinomial Dirichlet model), violating this assumption may influence estimates of F_{ST} , alpha and selection type. In addition, even while focusing on a conservative statistical methodology to obtain a limited set of outlier SNPs (as we did), the program is known to provide false positives (especially in species like caribou that are subject to Isolation by Distance) (Gubili *et al.* 2016) that indicate balancing selection (Lotterhos & Whitlock 2014, 2015). Lotterhos and Whitlock (2014) recognize that for most types of landscape genomic data (like ours), significance of outliers should be viewed with caution, because genetic correlations among samples (for example, units not evolutionarily independent) may result in a violation of model assumptions. The authors also acknowledge that cases of nonindependence may be most extreme in populations that have expanded from a refugium but have not yet reached equilibrium. Our determination of outliers should therefore be viewed with caution. However, this study relies on multiple methodological approaches to detect outliers (empirical test, BayeScan

and BayPass) and concordance of information with approaches to detect SNPs with geographic cline variation (Hzar – see below). In addition, there is no indication that any of our caribou populations is expanding, therefore alleviating the nonindependence concern. Finally, we used a Wilcoxon test to compare H_E (Expected Heterozygosity) for SNPs that were detected as under balancing selection by BayeScan (see below) to H_E for the other SNPs on which we conducted analyses of differences between populations (H_E expected to be lower).

Signature of selection - Landscape genomics approach

We selected a set of 12 environmental layers to explore associations between genomic and environmental characteristics known to influence selection in terrestrial Mammals (Harrigan *et al.* 2014; Schweizer *et al.* 2016): snow cover during October and November (average of values for years 2004-2008), Normalized Difference Vegetation Index (NDVI) during July and August (average of values for years 2004-2008), elevation, percentage of forest loss (2000-2014) (Hansen *et al.* 2013), five variables representing temperature and three representing precipitation (Table S1, Supporting information). We averaged pixel values for each variable inside caribou population ranges (analysis performed in ArcGIS v.10.3.1). We also determined centroids of population ranges and we used them to identify geographic coordinates of caribou groups.

To test the association of SNPs to environmental variables, we used a Bayesian approach as implemented in BayPass v. 1.0 (Gautier 2015) under the standard covariate model. We first calculated the covariance matrix (Ω) of pools, which represents neutral correlation of allele frequencies resulting from their possibly unknown and complex shared history. Next, we run the association test, which accounts for the covariance matrix and thus prevents for the occurrence of false positive. The association test estimated the Bayes Factor (BF) for each SNP as a measure of the correlation between the marker and each environmental factor. We analyzed each SNP in the full dataset using BayPass with all 12 variables with the following parameters: 50,000 iterations, of which 30,000 were discarded as a burn-in, a thinning interval of 50, and the option scalecov to standardize variables. Jeffreys' rule (Jeffrey 1961) was used to quantify the association between SNPs and environmental variables: "strong evidence" when $10 \leq BF < 15$, "very strong evidence" when $15 \leq BF < 20$ and "decisive evidence" when $BF \geq 20$. We run BayPass with three different seeds and checked for convergence of the results. We considered only SNPs with an average BF above 10 across the three runs as potentially having environmental correlations. Numerous analytical methods to detect association between SNPs and environmental variables have been developed in past years, all with strengths and limitations (i.e. also including Bayesian statistics approaches, Rellstab *et al.* 2015). Forester *et al.* (2018) recently compared methods and promoted redundancy analysis (RDA). Nonetheless, the authors indicated important limitations of RDA too, in particular when used with structured populations,

which should be the case in our sample including multiple populations some of which are known to be isolated (Weckworth *et al.* 2013). Our method of choice (BayPass) is designed to accommodate a covariance matrix of allele frequency correlations, and thus it somehow accounts for population structure (Gautier 2015). Therefore, the program might be appropriate to use in our case, although in combination with other tests too (see below).

To further verify the effect of natural selection on environmentally correlated SNPs, as well as outlier SNPs identified with the empirical F_{ST} approach and BayeScan, we completed a cline analysis using the R package “Hzar v. 2.5” (Derryberry *et al.* 2014), which fits allele frequency of SNPs to equilibrium geographic cline. In Hzar, clines are modeled as sigmoidal curves with exponential decay tails and evaluate the frequency of traits (i.e. genotypes or phenotypes) along a one dimensional geographic transect. The inflection point of the curve, or cline center, describes the point on the geographic transect where the change in trait frequency is most rapid. Additionally, the inverse of the maximum slope, or cline width, describes the geographic area over which this rapid change occurs. In our study, pools were virtually arranged along a transect from the southernmost pool (Pool 18) to the northernmost (Pool 16) and the Euclidean distance between the southernmost and the remaining pools was calculated. The geographic cline variation of pools’ allele frequencies for each SNP (outlier and environmentally correlated) was tested with three different models as described in Derryberry *et al.* (2014): model I (elongated) estimates the center (c) and the width (w) of the cline using observed allele frequencies at the top (P_{max}) and bottom (P_{min}) of the cline; model II (sharp) estimates c , w , P_{min} and P_{max} ; and model III (elongated with tails) estimates c , w , P_{min} , P_{max} and both fitted tails of the cline (δL , τL , δR , τR). These models were compared to a null model of no clinal transition using the Akaike Information Criterion (AIC) with the best-fitting model having lowest AIC value.

Population structure, organization of groups and genetic differentiation between migratory vs. sedentary pools

We determined population structure of the 14 caribou pools on the basis of the pairwise F_{ST} matrices. To visualize the multi-locus patterns of population differentiation, we conducted a principle coordinate analysis (PCoA) from average F_{ST} values using the R package Labdsv 1.8 (<http://ecology.msu.montana.edu/labdsv/R/>). We completed a separate PCoA to evaluate genetic differentiation between sedentary (Pools 2 and 8) and migratory pools. PCoA for both caribou populations and migratory vs. sedentary pools were conducted on the basis of the pairwise F_{ST} matrices calculated on two sets of SNPs: all available SNPs and under selection ones. The SNPs under selection that we used included those identified with both empirical F_{ST} approach and BayPass as is standard (Guo *et al.* 2015, 2016), but not those identified with BayeScan. BayeScan detects SNPs under different types of selection

(balancing or directional) and therefore would have introduced a dishomogeneous pool of SNPs.

Additionally, we generated a neighbor-joining tree based on pairwise F_{ST} distance values from a simulated multi-locus dataset with 1,000 bootstrap replicates in Populations 1.2.33 (<http://bioinformatics.org/~tryphon/populations/>). We evaluated the effect of geography on population structure using SpaceMix v. 0.13, a novel Bayesian statistical framework for studying the spatial distribution of genetic variation and genetic admixture also suited for Pool-seq data. Within SpaceMix (Bradburd *et al.* 2016), a “geogenetic” map is produced, in which inferred distances between pools are proportional to their genetic differentiation and long distance relatedness is modelled as genetic admixture. SpaceMix was implemented with the model option “estimating geogenetic locations and admixture source locations” using the geographic coordinates of the pools as spatial priors. We performed ten independent short chains of 10^5 , then analyzed a long chain of 10^7 iterations sampled every 10^3 steps. Geogenetic maps were plotted showing 95% credible ellipses

Gene ontology analysis

In this study, we ended up using contigs (reads that have been assembled together, which might overlap to sequences of genes - Etter *et al.* 2011) containing outlier and environmentally correlated SNPs as queries in nucleotide searches with Blastx against the bovine (*Bos Taurus*) protein database. Initially, we had chosen the bovine database due to the phylogenetic proximity of *Bos Taurus* to caribou. We had also compared our contigs harboring SNPs to the “Non Redundant” database of NCBI (database with protein sequences of all available species). However, the number of contigs that matched did not change and the annotations were similar to the ones we described initially using the *Bos* database. Therefore, we decided to use only the bovine genome. A draft genome of the reindeer (i.e. a domesticated caribou conspecific from Eurasia) is available, and initial quality assessments suggest that the assembly has “good level of contiguity and correctness” (Li *et al.* 2017). As the authors suggest, this information will be useful in the future in analysis of the genetic basis of the unique characteristics of reindeer (and caribou).

For the Blast search, we specified an E-value $<1.0 \times 10^{-6}$ and, in case of multiple hits, we selected the best match for each contig. We obtained functional gene ontology (GO) annotations for contigs with significant Blastx hits using the Blast2GO v. 3.3.5 pipeline (Conesa *et al.* 2005), which conducts Blast similarity searches and maps GO terms to the homologous sequences detected. Only ontologies with E-value $<1.0 \times 10^{-6}$, annotation cutoff > 55 and a GO weight > 5 were considered for annotation. Finally, we plotted GO results with wego, an online tool to visualize annotation following the vocabularies and classifications provided by the GO Consortium (Ye *et al.* 2006).

Results

Genetic variation within and between caribou populations (all SNPs)

With the objective to evaluate genetic variation in caribou, we analyzed 14 pools representing populations distributed in a spatial heterogeneous environment belonging to recognized subspecies and DUs, and four pools composed of either sedentary or migratory individuals (2 pools for each behavior) regardless of geographic provenance. We obtained 79 million raw sequence reads for the first library (range: 2-9 million per pool), and 97 million for the second library (range: 1.5-15 million per pool) (Table S2). After filtering, we retained an average of 2 million reads pairs per pool (total number of 40 and 41 million reads for the first and second library, respectively) (Table S2, Supporting information). In total, 302,501 loci were detected after the *de novo* assembly, and those containing at least one SNP (n=95,585) were then used to create the final reference genome, which contained 80,809 contigs with an average length close to 400 bp (minimum length 200, maximum length 713). We aligned 162 million reads to the reference (average 9 million reads per pool). After removing PCR duplicates and reads with low mapping quality, we obtained an average of 19,000 reads per pool that were available for variant discovery using PoPoolation2. We

identified on average 14,000 SNP variants within each pool, with a subset of 6,041 biallelic SNPs in common and valid to conduct analyses across our 14 caribou population pools, and a subset of 2,926 biallelic SNPs in common and valid to conduct analyses of genomic differences between the pools of migratory individuals and sedentary individuals (2 and 2 pools, respectively).

Expected heterozygosity (H_E) for the 14,000 SNPs datasets (all SNPs) (Table S3, Supporting information) across all populations was 0.398 on average (intra-population range=0.397-0.399), with comparable average genome-wide nucleotide diversity estimates (Tajima's π =0.008, s.e.=0.000005, intra-population range=0.0082-0.0083; θ_W =0.008, s.e.=0.000005, intra-population range=0.0082-0.0083). The average level of differentiation between populations as reflected by pairwise F_{ST} values was 0.046 (s.e.=0.00008; range=0.032-0.068).

We found that a PCoA of F_{ST} values calculated from all SNPs revealed a clear separation of Barren-ground, Northern Mountain and to a lesser extent Alaskan groups, likely reflecting ecological differentiation (Fig. 3A). Separation of groups was also evident in a neighbor-joining tree (Fig. 3B) and in a geogenetic map (Fig. S2, Supporting information). The organization of subspecies and DUs and their relationships were visualized by a covariance matrix of population allele frequencies resulting from their possibly unknown and complex shared history: a hierarchical tree cluster (Fig. 3E) clearly indicated separation of Barren-ground, Alaskan and Northern Mountain populations; the other populations were on a separate branch. The latter was also divided into Boreal groups and, on another sub-branch, Central Mountain and Southern Mountain populations. The only exception to this clear organization was the Little Smoky population (P4, classified with the Boreal DU, but of disputed assignment) as this was together with Central Mountain and Southern Mountain.

Environmental factors and candidate SNPs under selection in caribou (differences between populations)

To test for signatures of selection on caribou alleles we used an empirical outlier detection approach and identified 28 SNPs that appeared at least once in the top 0.5% of pairwise F_{ST} values between populations, suggestive of selection. We also tested association of SNPs with 12 environmental variables (Fig. 2, Figs. S1 and S3 and Table S4, Supporting information). We therefore identified 54 SNPs using a Bayesian approach, with 18 of these

54 associated with multiple environmental variables and 36 SNPs a single environmental variable, while a small fraction only of these variants (n=4 SNPs) were also identified using the empirical F_{ST} approach (above; see Venn diagram in Fig. S4). In this analysis, the four variables that arose as associated with the greatest number of SNPs were all climatic, including: minimum temperature of coldest month (n=15; Fig. 2), temperature seasonality (n=12), precipitation of coldest quarter (n=11), and annual precipitation (n=10). Other environmental variables both climatic and related to vegetation were associated with two or more SNPs, including: snow cover and elevation (n=8 SNPs each), precipitation seasonality (n=5), NDVI (n=4), mean annual temperature and maximum temperature of warmest month (n=3 each), and percentage of forest loss and mean diurnal temperature range (n=2 each). Finally, 50 of these 54 environmentally correlated SNPs displayed a North-South cline in genetic variation (Fig. 2, Table S4, Supporting information). This cline was also noted for 25 of 28 outlier SNPs detected with empirical F_{ST} approach (above). Cline shape Model I (i.e. elongated cline) explained the allele frequency variation of 15 (environmentally correlated) and 8 (empirical F_{ST} outliers) SNPs, while Model II (i.e. sharp cline) explained the allele frequency variation of 35 (environmentally correlated) and 17 (empirical F_{ST} outliers) SNPs.

Using SNPs under selection only (n. 78, including SNPs identified with BayPass and with empirical F_{ST} approach, as opposed to all SNPs, above), we surveyed genetic differentiation among populations. The PCoA spatially differentiated three population pools: Pool 1 (Banff and Jasper), Pool 18 (Columbia South, Duncan, Kinbasket and Purcell) and Pool 13 (Redstone/North West Ter.; Fig. 3C). A neighbor-joining tree confirmed the classification of the Little Smoky population (P4) together with Central and Southern Mountain populations based on branching topology, whereas Alaskan populations (P15 and P16) were monophyletic, and other populations belonging to Barren-ground, Northern Mountain and Boreal ecotypes lacked monophyly (Fig. 3D).

To verify that the environmentally-correlated loci are indeed signals resulting from natural selection, we used a Bayesian model-based methodology implemented in the program BayeScan. We identified 1,761, 599 and 51 outlier SNPs using Prior Odds parameters of 10, 1,000 and 10,000 respectively, and focus our description on the 51 SNPs obtained with the most statistically conservative approach. Surprisingly, all 51 SNPs obtained with the most conservative parameterization had a negative alpha (average alpha=-2.009; range (-2.528) – (-1.639)), indicating a possible effect of balancing selection (Fig. 4A). We also compared H_E for these 51 SNPs to H_E for the other SNPs on which we conducted analyses of differences between populations (n. 5,990) and, as expected, found greater H_E for those under balancing selection (Wilcoxon's $W=228930$, $p\text{-value}= 4.987 \times 10^{-14}$). Only one of the 51 outlier SNPs identified was also an outlier locus in the F_{ST} empirical approach, and none was an outlier associated with environmental factors (above). Finally, analysis of these 51 SNPs showed a North-South cline variation of only 11 SNPs. Detected clines were best explained by Model II (7 SNPs) and Model I (4 SNPs) (Table S4, Supporting information).

Selection in migratory and sedentary caribou (differences within populations)

Caribou with divergent migratory behaviors are anticipated to show genetic divergence. We used a combination of clustering and model-based methods to quantify the degree of differentiation among migratory and sedentary individuals. Regardless of geographic provenance, subset of GPS-collared individuals within populations yielded higher genetic differentiation than that estimated between geographically disparate populations (average pairwise $F_{ST}=0.086$, $s.e.=0.004$ between migratory and sedentary pools; between-population range=0.000-0.482, median=0.045). The PCoA of all SNPs revealed a separation between sedentary and migratory pools (Fig. 5A). Using the empirical F_{ST} approach, we identified 90 outlier SNPs potentially under selection regarding the migratory behavioral

phenotype. On the other hand, only one outlier SNP was identified using Bayesian statistics in the program BayeScan (Fig. 4B). This SNP had a positive alpha (alpha value= 1.7293), indicative of directional selection acting within populations. This finding contrasted with the negative alpha values indicative of balancing selection, which were consistently detected for the 51 SNPs analyzed across geographically distinct populations of caribou (above). All 91 outliers, including those found with empirical F_{ST} approach and with BayeScan, were unique to migratory behaviors and not identified in comparisons between populations to detect loci correlated to local environmental variables. A PCoA of the 91 outlier SNPs confirmed separation between sedentary and migratory pools (although such analysis was limited to four groups and therefore partial), with coordinate 1 explaining 45% of the variation and polarizing the behavioral trait along this axis (Fig. 5B).

Gene Ontology of candidate SNPs under selection in caribou populations

As our sequence reads were assembled without the guidance of a reference genome, we blasted reads that contained outlier SNPs to the bovine protein database in order to explore the function of variants putatively under natural selection. We found that 11 of 77 contigs that harbored SNPs potentially under selection corresponded to known bovine proteins ($E\text{-value} < 1.0 \times 10^{-6}$), and contributed to the North-South cline variation, with eight contigs associated with environmental variables. GO functional annotation of these 11 contigs yielded 111 GO terms, which were classified into 34 functional groups belonging to three functional categories: cellular component, molecular function and biological process (Fig. 6A, Table S5, Supporting information). Some contigs belonged to more than one functional group (for example, a gene may perform both binding and catalytic functions), which sometimes resulted in a sum exceeding 100% in a category (e.g. molecular function). Among the contigs categorized as cellular components, 100% were classified as cell part (GO

term defining any constituent part of a cell, as described in GO Consortium) (Ye *et al.* 2006).

The majority of the contigs categorized as molecular functions were associated with binding (64%) and molecular transducer activity (36%). Most of the contigs categorized as biological process were involved in cellular process (73%), biological regulation (55%) and pigmentation (55%).

Ten contigs out of the 36 harboring SNPs potentially under balancing selection corresponded to known bovine proteins ($E\text{-value} < 1.0 \times 10^{-6}$). GO functional annotation of the 10 contigs yielded 83 GO terms classified into 32 functional groups (Fig. 6B, Table S5, Supporting information). Among the contigs categorized as cellular components, 90% were classified as cell part. The majority of the contigs categorized as molecular functions was associated with binding (90%), catalytic activity (30%) and molecular transducer activity (30%). Most of the contigs categorized as biological process were involved in cellular process (80%) and metabolic process (60%).

The 91 SNPs under selection identified between migratory vs. sedentary pools within populations were located in only 50 contigs, with 12 of these contigs harboring SNPs corresponding to known bovine proteins ($E\text{-value} < 1.0 \times 10^{-6}$). Six of these contigs overlapped with proteins annotated and detected in spatially distinct populations of caribou. Functional annotation of the contigs yielded 194 GO terms, which were classified into 36 functional groups (Fig. 6A, Table S5, Supporting information). Among the contigs categorized as cellular components, 100% were classified as cell part. The majority of the contigs categorized as molecular functions were associated with binding (83%) and transporter activity (59%), and most of the contigs categorized as biological process were involved in cellular process (92%) and biological regulation (83%).

Discussion

Environmentally correlated selection and cline variation between caribou populations

Many caribou populations are declining and at risk of local extinction, while other populations are already extirpated (Hebblewhite *et al.* 2010; Festa-Bianchet *et al.* 2011). It is therefore important to identify locally selected traits, to understand whether these too might be lost as populations disappear (Hoffmann & Willi 2008). Our results indicated candidate SNPs under selection in caribou populations, and furthermore demonstrated that selection of different SNPs occurred in the spatially distinct populations sampled from Alaska to the southern Canadian Rockies. This corroborates our base theory that populations were subject to different environmental conditions (we strategically sampled areas with markedly different environments; Figs. 1, 2, Fig. S1, Supporting information), within the uninterrupted distribution of the species and could hold gene to environment associations under selection.

The variables associated with the greatest number of SNPs were all climatic, including temperature indices, precipitation indices and elevation (which could also be correlated with climate). Other non-climatic variables were also significantly associated with SNPs (although fewer associations were detected) including Normalized Difference Vegetation Index (NDVI, an index of plant productivity) (Pettorelli *et al.* 2011) and forest loss (Hansen *et al.* 2013). Our findings therefore confirmed that differentiated populations of caribou are likely adapted to different climates (Schoville *et al.* 2012) and different plant communities, and are being affected by anthropogenic disturbance. The cline shaped Models II (i.e. sharp cline) and Model I (i.e. elongated cline) explained allele frequency variation of a vast majority of SNPs under selection (Fig. 2, Fig. S1, Supporting information), further supporting the notion of differentiation along the Rockies. Caribou, as a species, evolved in extremely cold harsh environments, and the evolution of specific traits to cope with such conditions have been proposed, such as the ability to see ultraviolet B light (Tyler *et al.* 2016). However, glacial-interglacial cycles, and the associated population retractions and expansions, also allowed caribou to live in milder environments (e.g. boreal forests) (Geist 1998; Hummel & Ray 2008). These historic extremes in environmental conditions along a North-South gradient also define the current condition across our study area, thus it is conceivable that selection at different alleles remains, providing for the persistence of local adaptations and allowing caribou populations to survive our current interglacial epoch.

Environmentally-correlated SNPs in caribou populations were associated to five genes annotated in the cow reference genome. Notably, we identified *GPC5* (glypican-5) and *PDE4DIP* (myomegalin isoform X18). *GPC5* is involved in body development and has been found associated to body weight in other species (Gu *et al.* 2011). Caribou body size is observed to increase North to South, perhaps following climatic gradients (Geist 1998), and this clinal variation is hypothesized to be, at least in part, genetically determined (Klein *et al.* 1987). Our findings of environmentally correlated SNPs that are linked to the *GPC5* gene could support this notion. *PDE4DIP* is involved in regulating bitter taste perception, where variation is recognized to contribute to the development of different food preferences

across species and populations (Li & Zhang 2013; Kim *et al.* 2013). Our caribou populations live in different habitats, across which the literature reports different food preferences and diets (Geist 1998). The variation of SNPs linked to *PDE4DIP* that we found might therefore have contributed to these preferences and to the known differences in foraging between caribou populations in the study area.

The role played by balancing selection within populations with multiple alleles and phenotype in general

Non-environmental based tests for selection indicated balancing selection across 51 SNPs that were largely undetected by other methods. The occurrence of balancing selection in our populations might indicate that multiple alleles were actively maintained at frequencies higher than expected from genetic drift alone (Charlesworth 2009). Our populations are indeed documented to have multiple haplotypes and phenotypes: for example, some individuals harbor mtDNA of BEL type and others of NAL type, some individuals migrate seasonally and others are resident throughout the year, some individuals feed on ground and others on arboreal lichens. Balancing selection might explain the maintenance of such different phenotypes in caribou and promote the occurrence of different migratory behaviours, also within the same population. Balancing selection forces, such as NFDS, have been postulated as major drivers in the evolution of partial migration (Lundberg 1988; Christie *et al.* 2018), but they have never been tested for ungulates. Predation by wolves in caribou's seasonal ranges may be a driver of partial migration. The major cause of caribou mortality is wolf predation, which can be significant also on calves, therefore influencing recruitment and population dynamics too (Hervieux *et al.* 2013). In caribou, migration is a known mechanism to displace predators during the summer season, also including calving (Bergerud *et al.* 2007). It is therefore possible that, if a just few caribou individuals migrate, then they (and their calves) incur lower predation rates. However, the success of predator displacement might be density dependent. Predation rates will be relatively higher on caribou as the number of migrants (and their proportion in the population) becomes greater, therefore attracting more predators to summering/calving areas (Bergerud 1992). As a result, survival and reproduction of migrants will decrease in a negative dependent fashion with frequency of migrants in the population augmenting.

The inheritance of migratory behaviour in partially migratory populations of other taxa has also been explained by a threshold model of quantitative genetics (Pulido *et al.* 1996) involving a threshold value in a continuous variable that is underlying the migration propensity. This model has also been linked to NFDS, although Cobben & van Noordwijk (2016) recently demonstrated with simulations that partial migration could be maintained, in theory, also without NFDS (Jesmer *et al.* 2018). Our results might represent the first genetic evidence for balancing selection and NFDS forces acting on bimodal partially migratory populations. NFDS is believed to be a common phenomenon in nature, yet, a

scarcity of studies were published linking naturally occurring allelic variation with bimodal or multimodal phenotypes and balancing selection (Subramaniam & Rausher 2000; Gigord *et al.* 2001; Nettle 2006; Penke *et al.* 2007). We suggest that NFDS plays a strong role within populations harboring multiple haplotypes and phenotypes, as it is the norm in animals, plants and humans too.

Our findings indicate a genetic basis of migratory behavior and support previously formulated suggestions about adaptation to heterogeneous environments. Interestingly, for these partial migrators no published report documents intra-lifespan behavioral switching, a trait observed in other ungulates (e.g. elk) (Eggeman *et al.* 2016) that alternate between sedentary and migratory behaviors. Links between genetics and adaptation (as opposed to behavioral plasticity) were hypothesized in caribou, as migratory behavior has been statistically associated with the BEL mitochondrial DNA haplotype and to habitats with mixed composition of both tundra-like alpine and forested areas (McDevitt *et al.* 2009; Weckworth *et al.* 2012; Yannic *et al.* 2013). Consistent with a gene to behavior and environment association concept, we detected directional selection on an outlier SNP when our sample was divided in two phenotypically distinctive groups regardless of geographic provenance: a subset of GPS-collared migratory or sedentary individuals (suggesting that one allele was selected in migratory individuals and the other allele in sedentary individuals). This result therefore indicated that the two groups were now phenotypically homogeneous for migratory habits (within a given group) and genetically distinctive: selection was likely acting differentially within populations, between individuals performing migration or not. This SNP had a positive alpha (indicating directional selection), a stark contrast to the negative alpha values (balancing selection) detected for the 51 SNPs analyzed among spatially distinct populations (above), with some of them possibly formed by both sedentary and migratory individuals.

Other results further supported differentiation of migratory and sedentary caribou, regardless of geographic provenance (i.e. a force acting within populations). The overall average pairwise F_{ST} estimate was nearly double between migratory and sedentary groups of individuals (irrespective of populations) as compared to among distinct populations (0.086 and 0.046, respectively). Potentially, such differences in F_{ST} could be promoted by spatial separation during rutting and assortative mating among similar behaviors (of migratory vs. sedentary caribou individuals), and our results do not reject this view. In our study area, the possibility was explored by using autosomal microsatellite markers by McDevitt *et al.* (2009), who concluded that “some differentiation was evidenced” that could be due to reproductive isolation of family groups, especially during the breeding season. A proper

investigation of the contribution of assortative mating to differentiation would require individual genomic information, which was not available in this Pool-seq study. However, all 91 outlier SNPs that were uniquely found in the pool of caribou with known movement behaviour were distinctive from the outliers found in discrete populations. Moreover, these outliers seemed to contribute to differentiation as the PCoA plot of the F_{ST} values revealed a separation between sedentary and migratory pools, particularly when the outliers SNPs were used (Fig. 6B), indicating that reproductive isolations and demography alone (as reflected by all-SNPs) were not responsible for the full differentiation of migratory and sedentary caribou –i.e. SNP-associated genes could be important.

Outlier SNPs in migratory vs. sedentary pools were associated to seven bovine genes (out of which two genes were also found associated to environmentally-correlated SNPs, above). Most notably, we identified *IPO5* (importin-5), which is involved in energy metabolism (Freitas & Cunha 2009), and *SLC8B1* (sodium potassium calcium exchanger mitochondrial isoform X5), which is involved in oxidative stress response (Khananashvili 2013) (both genes not environmentally-correlated). These genes might play a role in migratory mammals, like the genes that are being discovered in migratory birds (Franchini *et al.* 2017), which are known to require the storage and consumption of high quantities of energy as well as physiological responses to oxidative stress, which often accompany prolonged physical activity (Skrip *et al.* 2015). In conclusion, *IPO5* and *SLC8B1* were identified in the gene ontology of migratory vs. sedentary pools only and, due to their known functions, might indicate a genetic component that uniquely enables migratory behavior.

Application of genomic, phenotypical and environmental information in conservation planning

Caribou in the study area are listed as threatened under the Species at Risk Act of Canada, which has raised concerns regarding population isolation and loss of genetic diversity (COSEWIC 2011; Weckworth *et al.* 2012). However, the levels of expected heterozygosity (H_E) we found for caribou populations were in line with the literature and approaching levels considered as “high” in non-endangered populations of carnivores as well as other wild mammals, when genomic approaches were also used (see Robinson *et al.* 2016; Goossens *et al.* 2017). Levels for F_{ST} were also in line with the literature and approaching levels considered as “moderate” in other mammals (Schweizer *et al.* 2016;

Goossens *et al.* 2017). Previous studies on caribou using alternative genetic markers (e.g. microsatellites, mitochondrial DNA) found diversity levels also in line with the literature (McDevitt *et al.* 2009). Isolation and decline of caribou populations, especially more southerly ones, although dramatic, have been relatively recent (Vors & Boyce 2009). It is therefore possible that a detectable response on local populations' diversity, which theoretically requires many generations, has not yet arisen in our populations. Our finding does not negate concerns for the future. A study by Serrouya *et al.* (2012) indicates that population declines could result in biodiversity losses in the future, especially in small and isolated populations along the southern fringes of the species' distribution.

Our results on the complex, but strong linkages between genetic, ecological factors and behavioral traits, including migration have clear management implications. First of all, biodiversity monitoring should continue for the units so far delineated for this species at risk, with a particular focus on populations that have some divergent characteristics from their DU, for example the Little Smoky (Note S1, Supporting information). In the future, if translocation actions will be considered, ideal source populations will be those that are closer genetically (in particular with regard to selected traits), ecologically and behaviorally. Such conclusions are warranted inference from this study, but results in this direction are just emerging, as translocated caribou seem to keep migration patterns of their source populations and seem maladapted to local environments and predators (Leech *et al.* 2017). More broadly, our findings indicate that unique genetic traits under selection might exist in different caribou populations, and also within populations, and similar patterns might be present in other species characterized by differences among populations and their environments and by intraspecific trait variation.

Acknowledgments

We would like to thank Alaska Department of Fish and Game and WestGrid compute Canada for data sharing and storage respectively. Research was conducted under Alberta, BC, and Parks Canada, Universities of Calgary and Montana research and collection permits. We would like to thank Karmi Garbi for library preparation, Baocheng Guo and Owen S. Wangenstein for their suggestions on data analysis and Jocelyn Poissant for intellectual input. We are grateful to two anonymous reviewers for their helpful comments on the manuscript. Funding for this work was provided by the Alberta Upstream Petroleum Research Fund (AUPRF), PTAC Petroleum Technology Alliance Canada, Canadian Association of Petroleum Producers, Parks Canada, Shell, Governments of BC, Alberta and Yukon, Conoco-Phillips, Alberta Conservation Association, NSERC and Yukon Fish and Wildlife Enhancement Trust.

References

- Akey JM, Ruhe AL, Akey D *et al.* (2010) Tracking footprints of artificial selection in the dog genome. *Proceedings of the National Academy of Sciences*, **107**, 1160-1165.
- Ayala FJ, Campbell CA (1974) Frequency-dependent selection. *Annual Review of Ecology and Systematics*, **5**, 115-138.
- Bauer S, Hoyer J (2014) Migratory animals couple biodiversity and ecosystem functioning worldwide. *Science*, **344**, 1242552.
- Bergerud AT, Luttich SN, Camp L (2007) *Return of caribou to Ungava* (Vol. 50). McGill-Queen's Press-MQUP.
- Bergerud AT (1992) *Rareness as an Antipredator Strategy to Reduce Predation Risk for Moose and Caribou*. In: McCullough DR., Barrett RH. (eds) *Wildlife 2001: Populations*. Springer, Dordrecht.
- Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. *PLoS Genetics*, **12**, e1005703.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124-3140.
- Chapman BB, Brönmark C, Nilsson JÅ, Hansson LA (2011) The ecology and evolution of partial migration. *Oikos*, **120**, 1764-1775.
- Chapman BB, Skov C, Hulthén K *et al.* (2012) Partial migration in fishes: definitions, methodologies and taxonomic distribution. *Journal of Fish Biology*, **81**, 479-499.
- Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195-205.
- Christie MR, McNickle GG, French RA, Blouin MS (2018) Life history variation is maintained by fitness trade-offs and negative frequency-dependent selection. *Proceedings of the National Academy of Sciences*, **17**, 4441-4446.
- Cobben MMP, van Noordwijk AJ (2016) Stable partial migration under a genetic threshold model of migratory behaviour. *Ecography*, **39**, 1210-1215.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674-3676.
- COSEWIC (2011) *Designatable units for caribou (Rangifer tarandus) in Canada*. Committee on the Status of Endangered Wildlife in Canada. Ottawa. 88 pp.
- Dyer SJ, O'Neill JP, Wasel SM, Boutin S (2001) Avoidance of industrial development by woodland caribou. *The Journal of Wildlife Management*, **65**, 531-547.
- Dyer SJ, O'Neill JP, Wasel SM, Boutin S (2002) Quantifying barrier effects of roads and seismic lines on movements of female woodland caribou in northeastern Alberta. *Canadian Journal of Zoology*, **80**, 839-845.
- De Carvalho D, Ingvarsson PK, Joseph J, Suter L *et al.* (2010) Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology*, **19**, 1638-1650.
- Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, **14**, 652-663.
- Eggeman SL, Hebblewhite M, Bohm H, Whittington J, Merrill EH (2016) Behavioural flexibility in migratory behaviour in a long-lived large herbivore. *Journal of Animal Ecology*, **85**, 785-797.
- Endler JA (1973) Gene flow and population differentiation. *Science*, **179**, 243-250.
- Etter P, Bassham S, Hohenlohe P, Johnson E (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology*, **772**, 157-

- Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011) Local De Novo Assembly of RAD Paired-End Contigs Using Short Sequencing Reads. *PLoS ONE*, **6**: e18561.
- Festa-Bianchet M, Ray JC, Boutin S, Côté SD, Gunn A (2011) Conservation of caribou (*Rangifer tarandus*) in Canada: an uncertain future. *Canadian Journal of Zoology*, **89**, 419-434.
- Fischer MC, Rellstab C, Tedder A, Zoller, A. *et al.* (2013) Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular ecology*, **22**, 5594-5607
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977-993.
- Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, **27**, 2215-2233
- Franchini P, Irisarri I, Fudickar A *et al.* (2017) Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species. *Molecular Ecology*, **26**, 3204-3216 (2017).
- Freitas N, Cunha C (2009) Mechanisms and signals for the nuclear import of proteins. *Current Genomics*, **10**, 550-557.
- Gagnaire PA, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, **67**, 2483-2497.
- Gautier M (2015) Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, **201**, 1555-1579.
- Geist V (1998) *Deer of the World: Their Evolution, Behavior and Ecology*. Stackpole Press, Mechanisburg
- Gigord LDB, Macnair MR, Smithson A (2001) Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina* (L.) Soo. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 6253-6255.
- Goossens B, Sharma R, Othman N *et al.* (2016) Habitat fragmentation and genetic diversity in natural populations of the Bornean elephant: Implications for conservation. *Biological Conservation*, **196**, 80-92.
- Greene JS, Brown M, Dobosiewicz M *et al.* (2016) Balancing selection shapes density-dependent foraging behavior. *Nature*, **539**, 254.
- Gregor JW (1944) The ecotype. *Biological Reviews*, **19**, 20-30.
- Gubili C, Mariani S, Weckworth BV, Galpern P, McDevitt AD, Hebblewhite M, Nickel B, Musiani M (2016) Environmental and anthropogenic drivers of connectivity patterns: a basis for prioritizing conservation efforts for threatened populations. *Evolutionary Applications*, **2**, 199-211.
- Guo B, DeFaveri J, Sotelo G, Nair A, Merilä J (2015) Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC biology*, **13**, 19.
- Guo B, Li Z, Merilä J (2016) Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular ecology*, **25**, 2833-2852.
- Gu X, Feng C, Ma L *et al.* (2011) Genome-wide association study of body weight in chicken F2 resource population. *PLoS One*, **6**, e21872.

<https://doi.org/10.1371/journal.pone.0021872>

- Hansen MC, Potapov PV, Moore R *et al.* (2013) High-resolution global maps of 21st-century forest cover change. *Science*, **342**, 850-853.
- Harrigan RJ, Thomassen HA, Buermann W, Smith TB (2014) A continental risk assessment of West Nile virus under climate change. *Global Change Biology*, **20**, 2417–2425.
- Harris G, Thirgood S, Hopcraft JG, Crooms J, Berger J (2009) Global decline in aggregated migrations of large terrestrial mammals. *Endangered Species Research*, **7**, 55-76.
- Hebblewhite M, White C, Musiani M (2010) Revisiting extinction in national parks: mountain caribou in Banff. *Conservation Biology*, **24**, 341-344.
- Hebblewhite M, Merrill EH (2011) Demographic balancing of migrant and resident elk in a partially migratory population through forage–predation tradeoffs. *Oikos*, **120**, 1860-1870.
- Hervieux D, Hebblewhite M, DeCesare NJ *et al.* (2013) Widespread declines in woodland caribou (*Rangifer tarandus caribou*) continue in Alberta. *Canadian Journal of Zoology*, **91**, 872–882.
- Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, **9**, 421-432.
- Hummel M, Ray JC (2008) *Caribou and the North: a shared future*. Dundurn. Press Ltd, Toronto, Ont
- Khananshvil D (2013) The SLC8 gene family of sodium–calcium exchangers (NCX)—Structure, function, and regulation in health and disease. *Molecular Aspects of Medicine*, **34**, 220-235.
- Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, Rothschild MF (2016) Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, **116**, 255-264.
- Kim KS, Cha NH, Kim KW *et al.* (2013) Transcriptomic analysis of the bitter taste receptor-mediated glucagon-like peptide-1 stimulation effect of quinine. *Biochip Journal*, **7**, 386-392.
- Klein DR, Meldgaard M, Fancy SG (1987) Factors determining leg length in *Rangifer tarandus*. *Journal of Mammalogy*, **68**, 642-55.
- Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C (2011a) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PloS one*, **6**, e15925.
- Kofler R, Pandey RV, Schlötterer C (2011b) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, **27**, 3435-3436.
- Jeffreys H. (1961) *Theory of Probability*. Oxford University Press, London/New York/Oxford.
- Jesmer BR, Merkle JA, Goheen JR, Aikens EO *et al.* (2018) Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. *Science*, **361**:1023-1025.
- Leech H, Jelinski DE, DeGroot L, Kuzyk G (2017) The temporal niche and seasonal differences in predation risk to translocated and resident woodland caribou (*Rangifer tarandus caribou*). *Canadian Journal of Zoology*, **95**, 809-820.
- Li D, Zhang J (2013) Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Molecular Biology and Evolution*, **31**, 303–309.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows Wheeler Transform. *Bioinformatics*, **14**, 1754-1760.

- Li Z, Lin Z, Ba H, Chen L, Yang Y *et al.* (2017) Draft genome of the Reindeer (*Rangifer tarandus*). *GigaScience*, **6**, 1–5. 2017
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.
- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031–1046.
- Lundberg P (1988) The evolution of partial migration in birds. *Trends in Ecology & Evolution*, **3**, 172–175.
- Lundberg P (2013) On the evolutionary stability of partial migration. *Journal of Theoretical Biology*, **321**, 36–39.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*. doi: <http://dx.doi.org/10.14806/ej.17.1.200>
- Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.
- McDevitt AD, Mariani S, Hebblewhite M, DeCesare NJ *et al.* (2009) Survival in the Rockies of an endangered hybrid swarm from diverged caribou (*Rangifer tarandus*) lineages. *Molecular Ecology*, **18**, 665–679.
- Nettle D (2006) The evolution of personality variation in humans and other animals. *American Psychologist*, **61**, 622–631.
- Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular ecology resources*, **11**, 184–194.
- Penke L, Denissen JJ, Miller GF (2007) The evolutionary genetics of personality. *European Journal of Personality*, **21**, 549–587.
- Pettorelli N, Ryan S, Mueller T *et al.* (2011) The Normalized Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. *Climate Research*, **46**, 15–27.
- Pulido F, Berthold P, Van Noordwijk AJ (1996) Frequency of migrants and migratory activity are genetically correlated in a bird population: evolutionary implications. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 14642–14647.
- Reelstab C, Gugerli F, Eckert AJ, Hancock A M, Holderegger R. (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**, 4348–4370.
- Ripple WJ, Newsome TM, Wolf C, Dirzo R *et al.* (2015) Collapse of the world's largest herbivores. *Science Advances*, 2015 **1**, e1400103.
- Robinson JA, Ortega-Del Vecchyo D, Fan Z *et al.* (2016) Genomic flatlining in the endangered island fox. *Current Biology*, **26**, 1183–1189.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **4**, 807–820.
- Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S (2012) Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 23–43.
- Schweizer RM., Vonholdt BM, Harrigan R *et al.* (2016) Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology*, **25**, 380–402.
- Schlötterer, R. Tobler, R. Kofler, V. Nolte (2014) Sequencing pools of individuals—mining

genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.

- Serrouya R, Paetkau D, McLellan BN, Boutin S, Campbell M, Jenkins DA (2012) Population size and major valleys explain microsatellite variation better than taxonomic units for caribou in western Canada. *Molecular Ecology*, **21**, 2588–2601.
- Skríp MM, Bauchinger U, Goymann W *et al.* (2015) Migrating songbirds on stopover prepare for, and recover from, oxidative challenges posed by long-distance flight. *Ecology and Evolution*, **5**, 3198–3209.
- Stapley J, Reger J, Feulner PG *et al.* (2010) Adaptation Genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.
- Subramaniam B, Rausher MD (2000). Balancing selection on a floral polymorphism. *Evolution*, **54**, 691–695.
- Tucker MA, Böhning-Gaese K, Fagan WF, Fryxell JM *et al.* (2018) Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. *Science*, 2018 **359**, 466–469.
- Tyler NJC, Stokkan KA, Hogg CR, Nellemann C, Vistnes AI (2016) Cryptic impact: Visual detection of corona light and avoidance of power lines by reindeer. *Wildlife Society Bulletin*, **40**, 50–58.
- Vors LS, Boyce MS (2009) Global declines of caribou and reindeer. *Global Change Biology*, **15**, 2626–2633.
- Weckworth BV, Musiani M, McDevitt AD, Hebblewhite M, Mariani S (2012) Reconstruction of caribou evolutionary history in western North America and its implications for conservation. *Molecular Ecology*, **21**, 3610–3624.
- Weckworth BV, Musiani M, DeCesare NJ, McDevitt AD, Hebblewhite M, Mariani S (2013) Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proc. R. Soc. B*, **280**, 20131756.
- Yannic G, Pellissier L, Ortego J, Lecomte N *et al.* (2013) Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change*, **4**, 132–137.
- Ye J, Fang L, Zheng H *et al.* (2006) WEGO: a web tool for plotting GO annotations. *Nucleic acids research*, **34**, W293–W297.
- Zamudio KR, Bell RC, Mason NA (2016) Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences*, **113**, 8041–8048.

Data Accessibility

Raw sequences are deposited in the NCBI Sequence Read Archive (SRA) under Accession number SRP128712 (<https://www.ncbi.nlm.nih.gov/sra/SRP128712>). Sample locations, environmental data and SNP are also available on the Dryad Digital Repository at datadryad.org (<https://datadryad.org/review?doi=doi:10.5061/dryad.0d2q5v6>).

Author contributions

M.M., S.M., M.C. and C.G. conceptualized and designed the study. M.M., B.V.W., T.H., D.H., R.S. and M.H. collected the samples or data. M.C., E.H., C.G. and K.G. analyzed samples and data, with supervision by B.vH. and M.M.. M.C., M.M. and B.vH. wrote the manuscript with input from all authors.

Tables and Figures

Fig. 1

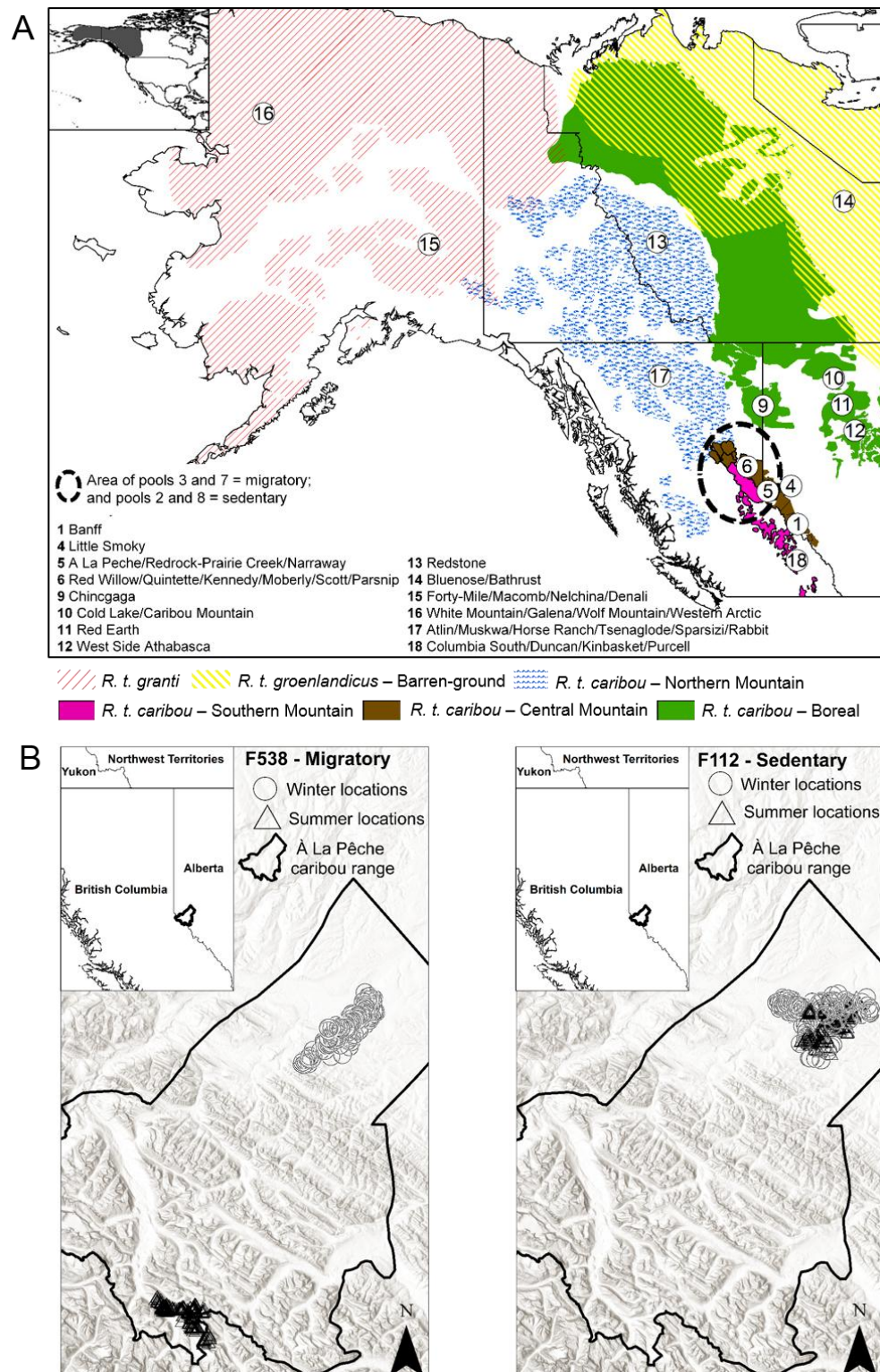


Fig. 1. Caribou populations sampled in western North America and movement patterns of individuals. (A) Map showing sampling sites of 18 caribou pools. Black numbered circles indicate sample locations. Colored polygons show a distribution of subspecies and ecotypes: red represents Alaska subspecies (*R. t. granti*); yellow represents Barren-ground subspecies (*R. t. groenlandicus*); pink, brown, blue and green represent Southern Mountain, Central

Mountain, Northern Mountain and Boreal ecotypes respectively within Woodland caribou (*R. t. caribou*). Dotted-line circle shows the source of four pools formed by either sedentary or migratory individuals from the same geographical area located in the Central Rocky Mountains. Sedentary and migratory specimens to form subgroups P2, P3, P7 and P8 were picked from the following partially migratory pools: P5 and P6. P5 and P6 represent multiple populations of the same caribou type (Central Mountain type): À La Pêche, Redrock-Prairie Creek, Narraway, Red Willow, Quintette, Kennedy, Moberly, Scott and Parsnip. (B) Map showing sample case of all GPS locations for a migratory and a sedentary caribou from the same population (À La Pêche). Triangles and circles represent summer and winter locations, respectively.

Fig. 2

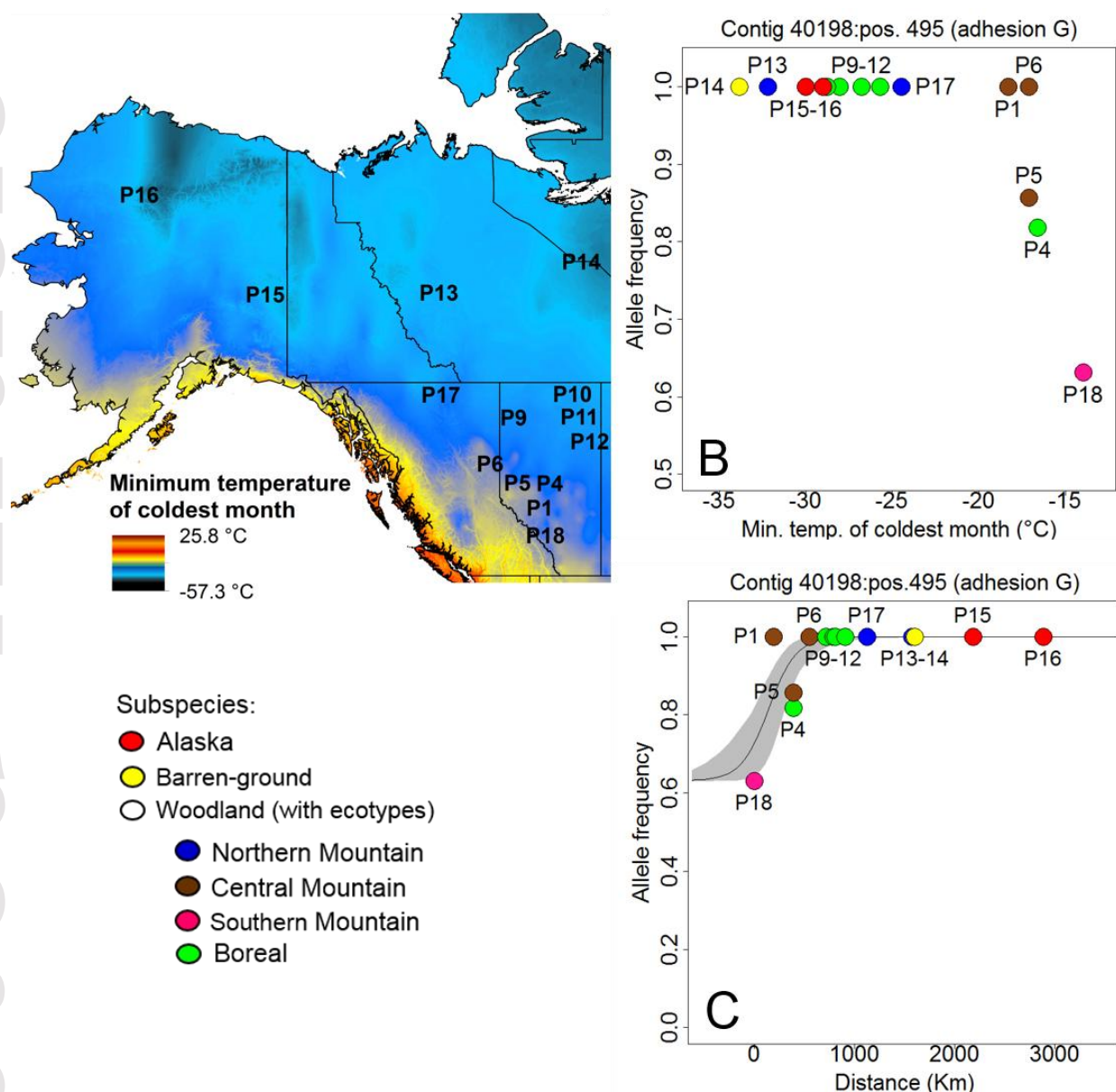


Fig. 2. Cline variation of a gene linked to a Single Nucleotide Polymorphism across caribou populations in western North America. (A) Map showing the variable associated with the greatest number of SNPs in our study: the minimum temperature of coldest month ($n=15$ SNPs). (B) Allele frequencies of the associated gene (contig belonging and position, as well gene association detected with Blast2GO) as these vary with minimum temperature of coldest month. (C) Geographic cline of the same gene shown in a cline shape plot (elongated cline). Abbreviated names are used as caribou populations' identifiers and colors represent recognized subspecies and ecotypes.

Fig. 3

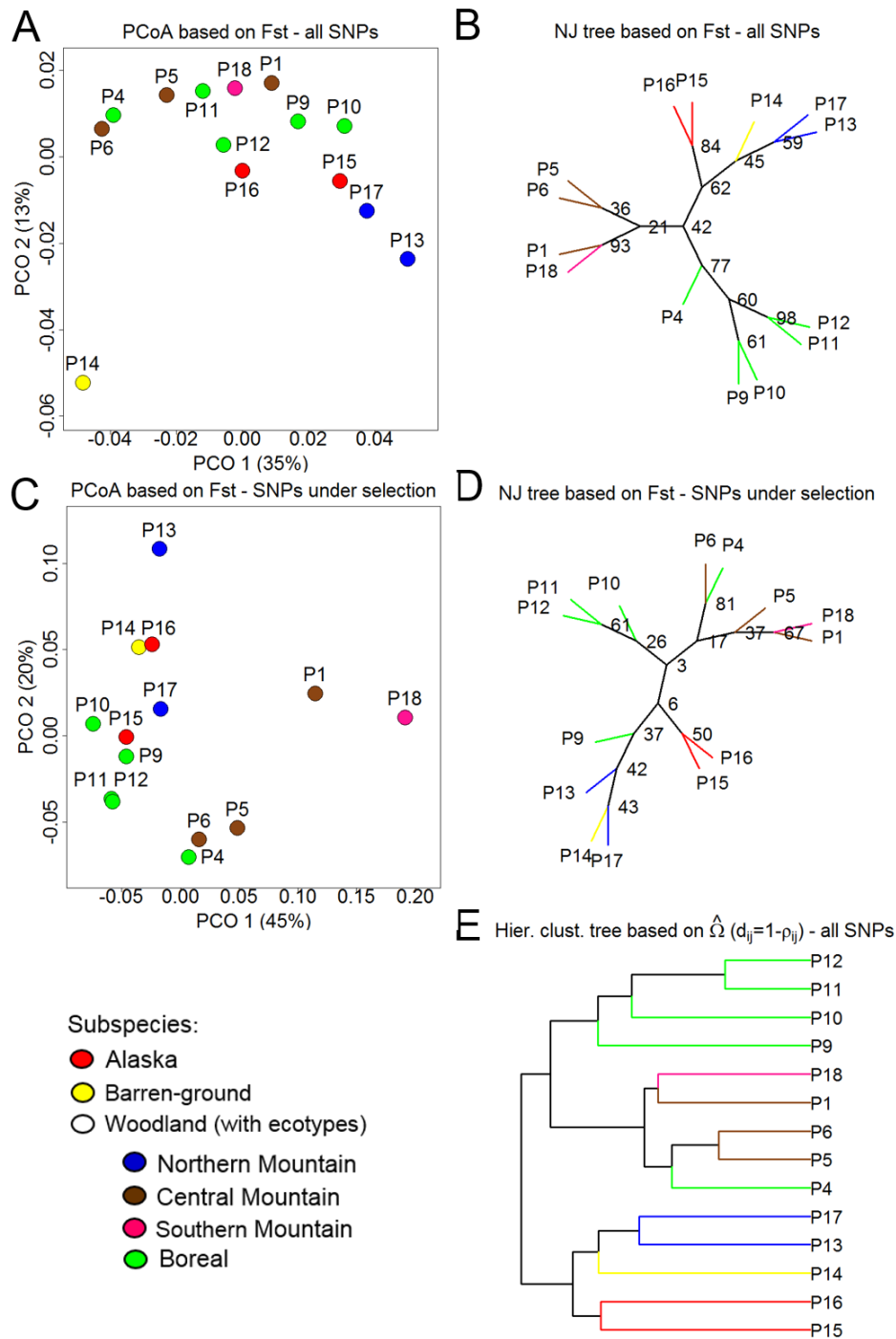


Fig. 3. Patterns of genetic differentiation of 14 caribou populations in western North America. Patterns estimated based upon all 6,041 biallelic SNPs (Panels A, B, E) and based upon 78 SNPs under selection (including outliers found with empirical F_{ST} approach and SNPs found associated to environmental variables; Panels C and D). Principal coordinates analysis (PCoA) plots of F_{ST} values are presented in panels A and C, and neighbour-joining trees in panels B and D, whereas a hierarchical tree cluster based on a covariance matrix (Ω) of population allele frequencies is in panel E. Abbreviated names are used as caribou populations' identifiers and colors represent recognized subspecies and ecotypes.

Fig. 4

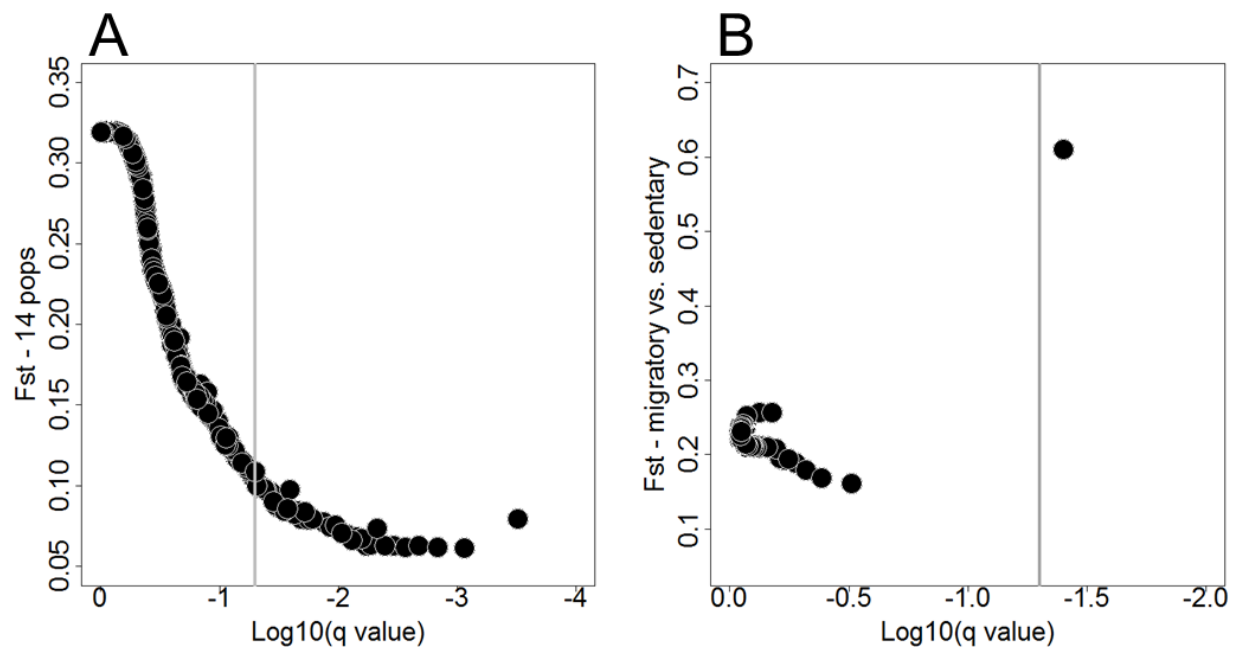


Fig. 4. Signatures of selection across 14 caribou populations in western North America and in migratory vs. sedentary individuals. Signatures of (A) balancing selection detected across 14 caribou populations, and (B) directional selection detected across pools of sedentary and migratory individuals using BayeScan. The horizontal axis indicates the log10 of the q value (the false discovery rate (FDR) analog to the P-value) and the vertical axis is the mean genetic differentiation (F_{ST}). Each point represents a single-nucleotide polymorphism (SNP) and significant outliers are visible on the right side of grey line.

Fig. 5

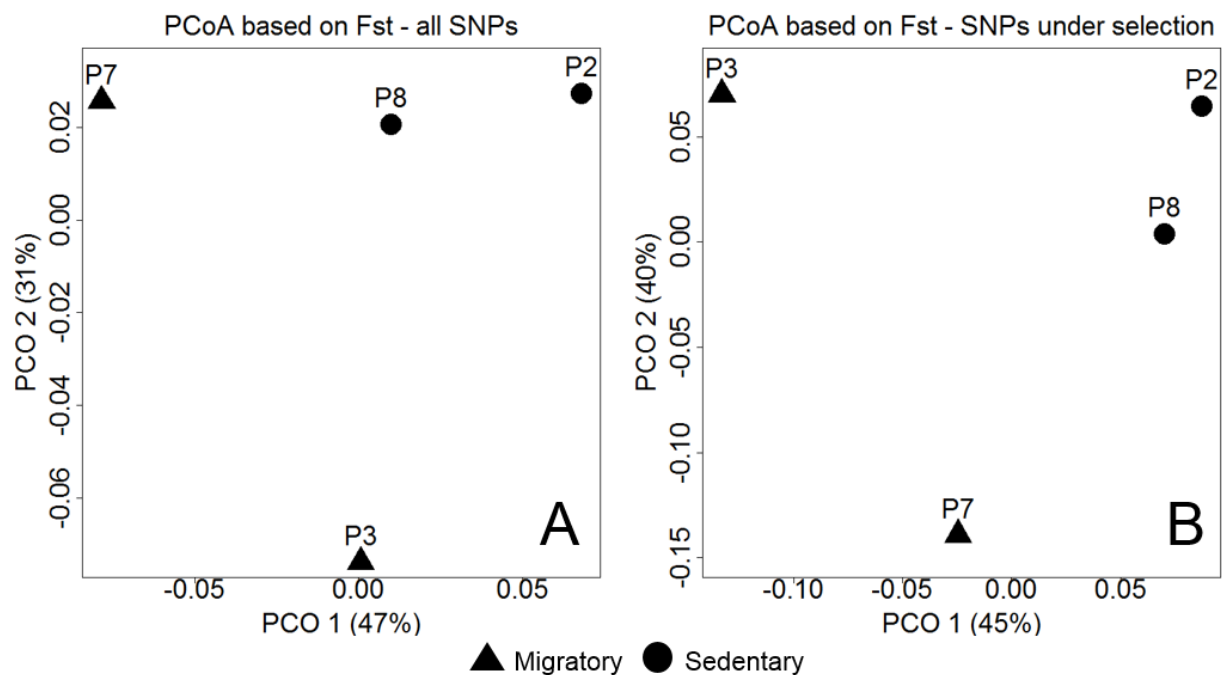


Fig. 5. Patterns of genetic differentiation between sedentary and migratory caribou individuals in the Central Rocky Mountains. (A) Principal coordinates analysis (PCoA) plot of F_{ST} values across migratory vs. sedentary pools based upon 2,926 biallelic SNPs. (B) PCoA plot based upon 91 outlier SNPs (including outliers found with empirical F_{ST} approach and BayeScan). Triangles represent Pool 3 and Pool 7, both formed by migratory individuals; circles represent Pool 2 and Pool 8, both formed by sedentary individuals.

Fig. 6

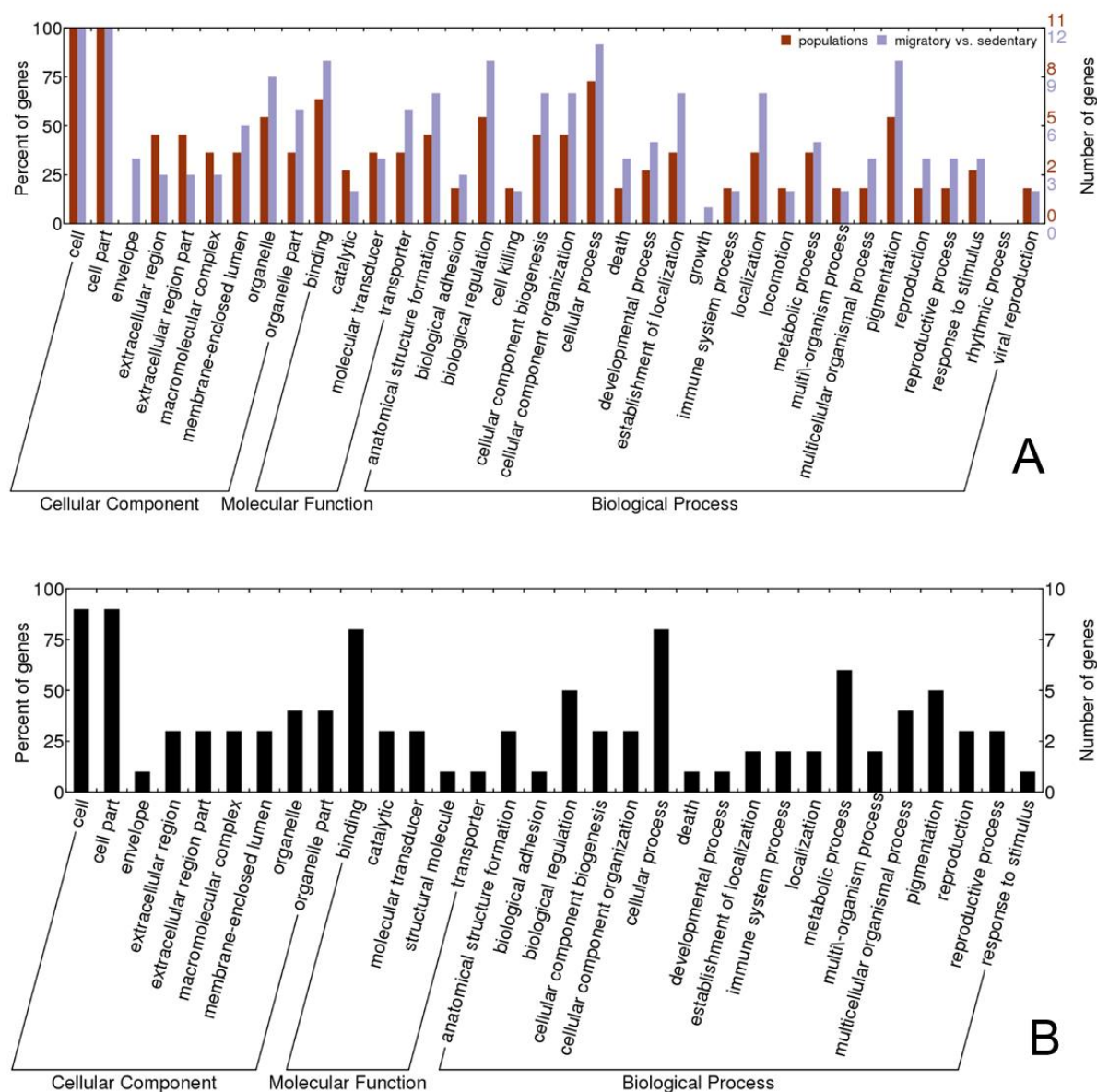


Fig. 6. Caribou gene ontology (GO) assignment plots based on bovine annotations. (A) Plot showing GO of genes harboring SNPs under selection (including outliers found with empirical F_{ST} approach and SNPs found associated to environmental variables) in caribou populations (blue color) and in migratory vs. sedentary pools (brown color). (B) Plot showing GO of genes harboring SNPs under balancing selection (outliers found with BayeScan) in caribou populations. Functional groups (along x-axis) were in three functional categories: cellular component, molecular function, and biological process. The y-axis represents number and percent of genes within a given functional category (e.g. molecular function) performing a specific function (e.g., binding, catalytic etc.). Some genes belong to more than one functional group (for example, a gene may perform both binding and catalytic functions), which may result in a sum exceeding 100% in a category (e.g. molecular function).