The role of introgression and ecotypic parallelism in delineating intra-specific conservation units

Rebecca Taylor[1], Micheline Manseau[2], Rebekah Horn[1], Sonesinh Keobouasone[2], Brian Golding[3], and Paul Wilson[1]

[1]Trent University
[2]Environment and Climate Change Canada National Wildlife Research Centre
[3]McMaster University

May 5, 2020

Abstract

Parallel evolution can occur through novel mutations, standing genetic variation, or adaptive introgression. Uncovering parallelism and introgressed populations can complicate management of threatened species, particularly as admixed populations are not generally considered under conservation legislations. We examined high coverage whole-genome sequences of 30 caribou (*Rangifer tarandus*) from across North America and Greenland, representing divergent intra-specific lineages, to investigate parallelism and levels of introgression contributing to the formation of ecotypes. Caribou are split into four subspecies and 11 extant conservation units, known as Designatable Units (DUs), in Canada. Using genomes from all four subspecies and six DUs, we undertake demographic reconstruction and confirm two previously inferred instances of parallel evolution in the woodland subspecies and uncover an additional instance of parallelism of the eastern migratory ecotype. Detailed investigations reveal introgression in the woodland subspecies, with introgressed regions found spread throughout the genomes encompassing both neutral and functional sites. Our comprehensive investigations using whole genomes highlight the difficulties in unequivocally demonstrating parallelism through adaptive introgression in non-model species with complex demographic histories, with standing variation and introgression both potentially involved. Additionally, the impact of parallelism and introgression on the designation of conservation units has not been widely considered, and the caribou designations will need amending in light of our results. Uncovering and decoupling parallelism and differential patterns of introgression will become prevalent with the availability of comprehensive genomic data from non-model species, and we highlight the need to incorporate this into conservation unit designations.

1 | INTRODUCTION

Parallel evolution is a process where divergent populations living in similar environments independently evolve the same or similar traits (Lamichhaney et al., 2017; Oke, Rolshausen, Leblond, & Hendry, 2017). Traditionally, it was assumed that cases of parallelism occurred by either new mutations, or from selection on standing genetic variation (Macpherson & Nuismer, 2017), however, adaptive introgression can also lead to the selection of beneficial alleles (Fraser & Whiting, 2019; Hedrick, 2013; Lee & Coop, 2017). Adaptive introgression between divergent lineages can facilitate parallel evolution, even if traits are controlled by more than one locus. Adaptive introgression could therefore be difficult to distinguish from parallel evolution from standing genetic variation (Bassham, Catchen, Lescak, von Hippel, & Cresko, 2018; Fraser & Whiting, 2019; Hedrick, 2013; Lee & Coop, 2017), and the two may also happen in concert (Bassham et al., 2018; Fraser & Whiting, 2019; Lee & Coop, 2017).

Recent findings show high levels of introgression among taxa (Kumar et al., 2017; Mallet, 2005; Taylor & Larson, 2019), or highly selective introgression of important adaptive genomic regions (e.g. Poelstra et al.,

2014; Song et al., 2011; The Heliconius Genome Consortium, 2012). However, admixed or hybrid populations are not generally considered under current conservation legislations (Fitzpatrick, Ryan, Johnson, Corush, & Carter, 2015; vonHoldt, Brzeski, Wilcove, & Rutledge, 2018) and when discussed, the focus is typically on inter-species hybrids and not conservation units below the species level (Fitzpatrick et al., 2015). Given the current extinction crisis under climate change also resulting in range shifts and increased secondary contact (Garroway et al., 2010; Gómez, González-Megías, Lorite, Abdelaziz, & Perfectti, 2015), new management frameworks will be required to encompass more complex evolutionary histories (vonHoldt et al., 2018) and novel adaptive potential.

Here we investigate intra-specific parallelism and levels of introgression contributing to adaptive evolution in the formation of caribou (*Rangifer tarandus*) ecotypes across North America and Greenland representing divergent intra-specific lineages. In Canada, there are four caribou subspecies largely based on morphology (Banfield, 1967). Caribou in Canada are distributed in widely different ecozones, including the High Arctic, mountains, taiga, and boreal forests (Banfield, 1967; COSEWIC, 2011). They display evidence of local adaptation, with differences in morphology, diet, behaviour, and life history in different regions, leading to the classification of 12 Designatable Units (DUs; 11 extant and 1 extinct; COSEWIC, 2011; Figure 1 and Figure S1), often referred to as ecotypes, by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC, 2011). Importantly, all 11 extant ecotypes are now listed as at risk of extinction (COSEWIC, 2011-2017) and many have been declining rapidly due to human-mediated disturbances including climate change (Festa-Bianchet, Ray, Boutin, Côté, & Gunn, 2011; Vors & Boyce, 2009; Weckworth, Hebblewhite, Mariani, & Musiani, 2018). Additionally, caribou are of huge cultural, spiritual, and economic significance to many indigenous communities (Festa-Bianchet et al., 2011; Polfus et al., 2016). It is also a keystone species for the ecosystem, important for vegetation structure, nitrogen cycling, and predator populations (Festa-Bianchet et al., 2011).

Previous mitochondrial DNA studies indicate two major phylogenetic lineages of Caribou in North America (Cronin, MacNeil, & Patton, 2005; Flagstad & Røed, 2003; Klütsch, Manseau, & Wilson, 2012; Weckworth, Musiani, Devitt, Hebblewhite, & Mariani, 2012). The range of the boreal DU extends from the east coast of Canada to the northern regions of the Northwest Territories, and in the central and eastern part of the range, the boreal caribou sit within the North American phylogenetic lineage, or NAL (Klütsch et al., 2012; Polfus, Manseau, Klütsch, Simmons, & Wilson, 2017). However, the northern mountain DU and boreal caribou from the northern part of the Northwest Territories belong to the Beringian-Eurasian phylogenetic lineage, or BEL, even though all boreal and northern mountain caribou are currently considered within the woodland subspecies, indicating potential parallel evolution (Polfus et al., 2017). Additionally, the eastern migratory DU has two disjunct ranges, one in northern Manitoba and Ontario and the other in northern Quebec and Labrador (Figure S1). Eastern migratory caribou from the Ontario and Manitoba region were found to be an admixture of boreal caribou from the NAL lineage and barrenground caribou from the BEL lineage (Klütsch, Manseau, Trim, Polfus, & Wilson, 2016). However, it is unknown if the Quebec and Labrador eastern migratory ecotype share the same origin.

We examined high coverage whole-genome sequences of 30 caribou in the most comprehensive study to date covering six DUs and all four subspecies (Figure 1). We used genome-wide variation using population and phylogenomic approaches to investigate instances of parallel evolution. We then elucidated the extent of introgression across the genome among caribou lineages. Issues of parallelism and complex patterns of introgression will certainly become more prevalent and we discuss how the definition and delineation of conservation units could be informed by our results.

2 | MATERIALS AND METHODS

2.1 | Sample collection, extraction and sequencing

Tissue was collected from 28 caribou from across Canada and two caribou from Greenland between 1992 and 2015, representing all four subspecies and six Canadian Designatable Units (DUs; Figure 1 and Table S1). Samples were collected on road kills or from harvested animals by biologists or veterinarians with the

British Columbia, Manitoba, and Ontario provincial governments, the Canadian federal government, the Greenland government, the Sahtú Renewable Resources Board, The Royal Ontario Museum, the University of Manitoba, and an independent consultant (See Table S1). Tissues were stored in RNA later ICE (Thermo Fisher Scientific, MA, USA). Phenol chloroform extractions were performed on three of the samples (The Pas, Snow Lake, and Ignace) using 0.2g of tissue, and eluted in Tris-ethylenediaminetetraacetic acid (TE) buffer at 100µl. The other samples were extracted using a Qiagen DNAeasy tissue extraction kit following the manufacturer's instructions (Qiagen, Hilden, Germany). The samples were run on a Qubit fluorometer (Thermo Fisher Scientific, MA, USA) using the High Sensitivity Assay Kit and normalized to 20ng/µl at a final volume of 50µl. The DNA was shipped to The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children (Toronto, Ontario) for library preparation and sequencing. The samples were each run on one lane of an Illumina HiSeq X (Illumina, San Diego, CA, USA), for a total of 30 lanes of sequencing. All raw reads will be made available on the NCBI by the time of publication.

2.2 | Filtering raw reads and variant calling

We used Trimgalore 0.4.2 (available here: https://github.com/FelixKrueger/TrimGalore), a wrapper script for CutAdapt (Martin, 2011), to remove sequencing adaptors and to trim low quality ends from reads with a phred quality score below 30. Reads were aligned to the caribou reference genome (Taylor et al., 2019) using Bowtie2 2.3.0 (Langmead & Salzberg, 2012), and the SAM file converted to a BAM file using Samtools 1.5 (Li et al., 2009). We removed duplicate reads and added correct read group information to each BAM file using Picard 2.17.3 (Available: http://broadinstitute.github.io/picard/). We then sorted the BAM file using Samtools 1.5, and built an index using Picard. All BAM files were checked using FastQC 0.11.8 (Andrews, 2010).

In addition to the 30 caribou genomes we sequenced, we also used a reindeer genome from a domesticated animal from Inner Mongolia, sequenced by Li et al., (2017). The reads were downloaded from NCBI (SRR5763125-SRR5763133) and mapped back to the caribou reference genome using the same methods as above. After using FastQC, adaptor contamination was detected, as well as duplicate reads, and so we used ClipReads in GatK to remove Nextera adaptor contamination, and removed duplicates using Picard. We then re-checked the file using FastQC and found we had successfully removed the contamination. Some sequence duplication was still detected, however, and so results from the reindeer sequence may need to be treated with caution.

We ran each BAM file through BUSCO (Benchmarking Universal Single-Copy Orthologs 3.0.2; Waterhouse et al., 2018) to reconstruct 4,104 conserved mammalian genes to assess the completeness of each genome. As our reference genome reconstructed 3,820 (93.1%; Taylor et al., 2019) complete mammalian BUSCO genes, this represents an upper limit for our re-sequenced individuals. We used Haplotype Caller in GATK 3.8 (McKenna et al., 2010) to call variants and produce a variant call format (VCF) file for each caribou. Individual VCF files were combined using the Combine GVCFs function, and then we performed joint genotyping using Genotype GVCFs, both in GATK, to produce a VCF file with all caribou and the reindeer. For some PCA's (see below), we also made VCF files containing subsets of individuals.

We downloaded the raw reads for a Sitka deer (*Odocoileus hemionus sitkensis*) genome from the NCBI database (Bioproject PRJNA476345, run SRR7407804) sequenced as part of the CanSeq150 Initiative, to use as an outgroup. We aligned and filtered the reads in the same way as for the caribou genomes to produce an individual VCF file. We then used the Combine GVCFs function, and performed joint genotyping using Genotype GVCFs, both in GATK, to produce a VCF file with all caribou, the reindeer, and the Sitka deer, for analyses requiring an outgroup.

We did some additional filtering on the combined VCF files to ensure quality. We used VCFtools 0.1.14 (Danecek et al., 2011) to do two rounds of filtering. Firstly, we removed indels, and any site with a depth of less than 10 or more than 80 (roughly double the average depth across the genome), and removed any low-quality genotype calls, with a score below 20, which in VCFtools are changed to missing data. In the second round, we filtered to remove genotypes with more than 10% missing data. We did not filter to remove

any SNP with a minor allele frequency (MAF) of less than 0.05 as we have only one or two individuals from each location and this results in removing the private sites, instead relying on very high depth and stringent filtering to ensure a high-quality dataset. However, we did conduct the PCAs using an MAF filter and these looked identical to the data without the MAF filter (Figures S2-S5). The combined VCF file used for analyses with all individuals apart from the Sitka deer contained 34,573,476 SNPs, and the VCF including the Sitka deer contained 65,412,957 SNPs

2.3 | Population and phylogenomic structure

We calculated the mean depth of coverage for each BAM file using Samtools. After filtering, we measured the mean depth, the frequency of missing data, and the inbreeding co-efficient, F, for each individual in the final VCF file of 30 caribou plus the reindeer using VCFtools. We performed principle component analyses in R 3.4.4 (R Development Core Team, 2006) using the packages vcfR (Knaus & Grünwald, 2017) and Adegenet (Jombart, 2008). The PCA was done on the VCF file containing all caribou and the reindeer (but not the Sitka deer). We then ran subsets of individuals on different PCA's to gain higher resolution of different lineages (see Results).

We used VCFkit (available here: https://vcf-kit.readthedocs.io/en/latest/, using numpy 1.14 as the programme does not work with newer versions) to generate a fasta file using the 'phylo fasta' command. The programme concatenates SNPs for each sample, using the first genotype of each allele, and replacing missing values with an N. We ran this on the VCF file without the Sitka deer to create an unrooted tree as including the Sitka deer pushed all caribou too closely together to discern the branches. The resulting file was input into RAxML 8 (Stamatakis, 2014), and run using the GTRGAMMA model and 1,000 bootstrap replicates. We visualised the best tree in FigTree 1.4.2 (https://github.com/rambaut/figtree). We also aligned each of the conserved mammalian genes extracted from the genomes using BUSCO (above) to construct phylogenies, from which we made a consensus tree. We used the Sitka deer outgroup to root the tree. We used MUSCLE (Edgar, 2004) to align the sequences for each individual to create a combined fasta file for each gene. We then used RAxML as above to create a gene tree for each file, and then used ASTRAL-III (Zhang, Rabiee, Sayyari, & Mirarab, 2018) to create a consensus tree which was visualised in FigTree.

We used the populations module in Stacks 2.4.1 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) to convert our VCF files (both with and without the Sitka deer) into an input file for Treemix 1.13 (Pickrell & Pritchard, 2012). We ran Treemix from 0-9 migration events, with three iterations of each, grouping the SNPs in windows to account for possible linkage using a block size of 1,000 for two of the iterations and 5,000 for one of the iterations (because the OptM package, below, must have different likelihood scores between iterations). We plotted the resulting trees, and the residual plots, in RStudio 1.0.136 (RStudio Team, 2015). We then used the R package OptM (Fitak, In Review.) to calculate the second order rate of change in the log-likelihood of the different migration events (the ad hoc statistic delta M) to help infer how many migration events to visualize.

2.4 | Demographic reconstruction and admixture analyses

We made a consensus fastq file for each caribou and the reindeer from their bam files, using the Samtools and BCFtools 1.5. This was converted into an input file and run in Pairwise Sequentially Markovian Coalescent (PSMC) model in PSMC (Li & Durbin, 2011) to investigate past effective population size changes. These were plotted using the general mammal mutation rate of 1.0E-9 (Li & Durbin, 2011) and a generation time of 7 years (COSEWIC 2014-2017).

To calculate admixture statistics, we used the R package admixr (Petr, Vernot, & Kelso, 2019) to run ADMIXTOOLS (Patterson et al., 2012). We converted our VCF file containing the Sitka deer (to use as an outgroup) into EIGENSTRAT format using a C++ script (found here: https://github.com/bodkan/vcf2eigenstrat). As the package does not work when including more than 600 scaffolds, we filtered the dataset to include SNPs found only on the 600 largest scaffolds, which encompassed over 98% of the reference genome assembly (the scaffold L90 is 285; Taylor et al., 2019). We used the EIGENSTRAT files to run f3, f4, and f4-ratio statistics. See Reich, Thangaraj, Patterson, Price, and Singh (2009) and Patterson et al., (2012) for full explanations

of these tests, but briefly, the f3 statistic is a three-population test that can calculate whether population 'C' is a mixture of two other populations, 'A' and 'B'. A negative f3 statistic indicates that population 'C' is a mixture of 'A' and 'B'. The f4 statistic is an ABBA BABBA test and acts similarly to D statistics. It is a four-population test which requires a phylogenetic set up including two sister groups, a test group to see if introgression has occurred into one of the two sister groups, and an outgroup (which we always set as the Sitka deer). An f4 statistic which significantly differs from 0 indicates gene flow, whether it is positive or negative tells you into which of the sister populations. In the f4-ratio test, alpha is calculated, which is the proportion of the genome in population 'X' that originates from population 'B' as opposed to population 'A' (the proportion of population 'A' is calculated as 1 – alpha).

For these tests, we grouped the four barrenground genomes from Bluenose and Qamanirjuaq as they show no differentiation and testing them separately made no difference to the results. The four boreal caribou genomes from Ontario and Manitoba were run separately as these do show differentiation and grouping them did affect the outcome. We focussed on using these tests to investigate 1) the amount of barrenground introgression into eastern migratory caribou in Ontario/Manitoba and Quebec/Labrador (f3, f4 and f4-ratio tests) separately as they have non-overlapping ranges, 2) introgression between eastern migratory caribou in Ontario/Manitoba and Quebec/Labrador (f4 test), 3) introgression between boreal caribou of NAL origin and the mountain caribou (f4 and f4-ratio tests for significant populations), and 4) introgression between boreal caribou of NAL origin and the Northwest Territories boreal caribou of BEL origin (f4 and f4-ratio tests), since our goal was to investigate the potential role of adaptive introgression in leading to parallel evolution.

We first investigated introgression from barrenground into the Manitoba and Ontario boreal populations (f4 test), and due to its current geographic isolation and low levels of introgression from the barrenground lineage, we used Ignace as the representative NAL boreal population. Similarly, to investigate introgression from the NAL into the BEL, we used the Grant's caribou as the sister group as these showed the least amount of introgression from the NAL lineage (f4 test). In the tests to investigate BEL introgression into the boreal caribou of NAL origin, we used eastern migratory Quebec/Labrador caribou as the sister group which had the lowest introgression from the BEL (f3, f4 and f4-ratio tests). For the full set up our tests, see Supporting Information.

To investigate patterns of introgression across the genome we used the programme Dsuit (Malinsky, Matschiner, & Svardal, 2019). The Dinvestigate function can be used to calculate introgression in windows across the genome, and this was used to calculate the $f_D$and f_{DM} statistics (Malinsky et al., 2015) for sliding windows of 1,000 SNPs incremented by 250 SNPs across the genome. We used this programme to investigate introgression between NAL boreal caribou and the mountain caribou as well as between NAL boreal caribou and BEL boreal caribou to further investigate the process of parallel evolution. Again, we used Ignace as the representative NAL boreal population and as it is the most geographically isolated and has low levels of introgression from BEL. Similarly, we found Grant's caribou to show the lowest levels of introgression from the NAL and so we used them as the sister group into the BEL boreal and Columbia North caribou. The Sitka deer was still used as the outgroup in all tests.

To further investigate the potential role of adaptive introgression in the parallel evolution of the BEL boreal caribou (see Results) we investigated the gene composition of the most introgressed regions within the BEL boreal caribou identified as having originated from Ignace. We compared these to the most introgressed regions from Ignace into all mountain populations as adaptive introgression is unlikely to have played a role in the parallel evolution of these populations due to the uncovered patterns of introgression (see Results). To do this, we extracted the sequences for all regions across the genome with an f_{DM} score over 0.2 (as it is the most conservative statistic) using Bedtools 2.29 (Quinlan & Hall, 2010). To make sure the sister group used in the set-up of the test didn't bias the results, we only included regions that were flagged as highly introgressed from the NAL group when using both Grant's and Peary caribou as the sister group. We used the command line version of BLAST 2.6 (Altschul, Gish, Miller, Myers, & Lipman, 1990) to search for the genes present in these introgressed regions and genes with mRNA or predicted mRNA hits in at least two

species and with an E score of 0.

3 | RESULTS

3.1 | Genome quality assessments and demographic history

We sequenced 28 caribou genomes from across Canada and two caribou from Greenland to high coverage (35.57 – 43.03X; Table S1) representing all four subspecies and six Canadian Designatable Units (DUs; Figure 1 and Table S1), and used an additional reindeer genome from a domesticated animal in Inner Mongolia, sequenced by Li et al., (2019). Our caribou genomes showed high quality and recovery of BUSCO genes in the assembly, ranging from 92.7% to 93.1% of the more than 4,000 conserved genes surveyed (Table S1), and missing data per individual was low at 0.3%-1.0% with the exception of the previously published reindeer genome at 16.0% (Table S1).

A reconstruction of *Rangifer* demography over time using PSMC indicated a major population size expansion starting approximately 100-200 kya with peak population sizes around the glacial interstitial stage of a largely ice-free North America 120 kya (Figure 2a-b). This timing corresponds to a divergence of lineages largely concordant with the expansion and intra-specific diversification proposed by Banfield (1961). These differential population trajectories correspond to contemporary subspecies and ecotypes. The NAL and the Greenland caribou have much lower population sizes than BEL caribou, including the boreal caribou from the Northwest Territories, with Peary caribou being intermediate to these groups (Figure 2a-b, Figure S6 for all plotted together) consistent with an earlier divergence (Klütsch, Manseau, Anderson, Sinkins, & Wilson, 2017). Population sizes for all caribou lineages declined during the Wisconsin glaciation which lasted between 75-11 kya (Figure 2a-b), with the exception of the reindeer which has a unique demographic trajectory likely as a result of domestication. Contemporary inbreeding estimates varied greatly between different individuals. For the North American caribou, they ranged from -0.009 to 0.311. They were highest for Greenland at 0.654, and the Inner Mongolia reindeer also had an elevated co-efficient at 0.177, again reflecting the origin of the latter as originating from a domesticated population (Li et al., 2019; Table S1).

3.2 | Population and phylogenomic structure

Principal component analyses (PCAs) revealed four major clusters corresponding to the NAL and BEL as well as Peary and Greenland clusters (Figure 3a). For the former two groupings, North America and Beringia, the genome clustering did not conform to current subspecies or ecotype designations. Specifically, in the North American cluster, some caribou populations of the woodland subspecies grouped consistently, i.e. boreal caribou and eastern migratory caribou, but mountain caribou grouped with Beringian lineages. The Beringian cluster contained barrenground, Grant's, northern mountain, southern mountain, and Northwest Territory boreal caribou (Figure 3a) consistent with their PSMC demographic history. These lineages provide evidence of the parallel evolution of similar ecotypes from distinct lineages and histories.

Finer resolution PCA of the NAL caribou showed all four boreal caribou were separated, particularly Ignace which may be due to genetic drift as it has a high inbreeding co-efficient (Table S1; Figure 3b). Eastern migratory caribou from Ontario/Manitoba clustered closest to Manitoba boreal caribou although all were well separated (Figure 3b). Eastern migratory caribou from Quebec/Labrador were closest to Cochrane boreal and not eastern migratory caribou from Ontario/Manitoba, and so indicates similar ecotypes may have evolved in parallel (Figure 3b; Table S1). Fine scale analysis of the BEL caribou, aside from Peary and Western Greenland, showed the Inner Mongolian reindeer and northern mountain caribou from Itcha-Ilgachuz separating, which again may be due to drift and inbreeding (Table S1; Figure 3c). Southern mountain caribou from Columbia North are also relatively well separated. The rest all formed a relatively tight cluster, with the Northwest Territories boreal caribou and the Grant's caribou slightly separated (Figure 3c). We ran the 14 genomes that sat closely together in another PCA, and found the four barrenground caribou clustered together and the others to separate, especially the Northwest Territories boreal (Figure 3d).

Phylogenomic reconstruction using SNPs in RAxML (Figure 4) and conserved gene sequences from BUSCO (Figure 5), showed similar patterns. Both separated the NAL lineage from all others and within the NAL clade

eastern migratory caribou from Ontario/Manitoba and those from Quebec/Labrador were not reconstructed as sister groups, again indicating parallel evolution of the eastern migratory ecotype. In the SNP phylogeny, which has been rooted based on the BUSCO phylogeny (Figure 5) and the Treemix analysis with the Sitka deer (Figure S8), within the NAL clade eastern migratory caribou from Quebec are reconstructed as sister to boreal caribou from Ontario, whereas eastern migratory caribou from Ontario were placed as sister to boreal caribou from Manitoba which matches the geography of the sampling locations (Figure 1; Figure 4). Within the BEL clade, the boreal caribou from the Northwest Territories are reconstructed as basal to all others. The rest were split into two clades, one of these with Northern mountain caribou from Itcha-Ilgachuz and the southern mountain Columbia North caribou. The other clade was further split into two, with the northern mountain caribou from the Northwest Territories, Atlin, and Frog in one, and the other with the barrenground caribou from the Northwest Territories and Manitoba, Grant's caribou, the Inner Mongolia reindeer, and Peary and Western Greenland caribou forming a sister clade within the group (Figure 4).

The BUSCO phylogeny shows similar patterns to the SNP reconstruction although with shorter branch lengths between groups and lower support of nodes which is unsurprising given that is was reconstructed from conserved mammalian genes. Within the NAL clade, boreal caribou from Snow Lake are basal to all others, and then eastern migratory caribou from Quebec are sister to boreal caribou from Cochrane in one clade, with the eastern migratory caribou from Ontario with boreal caribou from Ignace and The Pas in another (Figure 5). As with the SNP phylogeny, the Northwest Territories boreal and all mountain caribou sat within the BEL clade as further evidence for parallel evolution of the woodland ecotype. In the BEL clade, boreal caribou from the Northwest Territories and the reindeer are basal. There are three major clades within this group, one containing barrenground caribou from the Northwest Territories and Manitoba and the Western Greenland and Peary caribou as a sister group within that clade, one containing the northern mountain caribou from Itcha-Ilgachuz and southern mountain caribou from Columbia North, and another containing the Grant's caribou and the rest of the northern mountain populations (Figure 5).

3.3 | Patterns of introgression

To assess the contribution of admixture and introgression among lineages in positioning caribou lineages, we applied Treemix and f3, f4, and f4-ratio statistics. The Treemix phylogeny with no migration events gave a similar topology to the RAxML tree (Figure 6a). When visualising seven migration events, which shows the least standard error and has the highest delta m score (Figure S7), we see migration from the ancestor of Peary and Western Greenland into both Northwest Territories and Manitoba barrenground, and a migration even from the ancestor of the NAL lineage into southern mountain caribou (Figure 6b). The other migration events all occur within the NAL group, including into Snow Lake from an ancestral group, from the ancestor of Cochrane and Ignace into Eastern migratory Ontario, from Cochrane into an ancestor of Snow Lake and The Pas and Eastern migratory Ontario, and from Eastern migratory Quebec into Cochrane. The tree shows large drift parameters for those individuals with high inbreeding co-efficients (Table S1; Figure 6b).

The f3 results gave significant signatures of the genomes of eastern migratory caribou in Ontario/Manitoba resulting from admixture between NAL boreal caribou from Igance (our reference NAL genomes, see Methods) and barrenground as well as from other genomes from the BEL. The f4-ratio statistic shows the Ontario/Manitoba eastern migratory caribou genomes to be of 7% barrenground origin. In contrast, there were no negative f3 scores for eastern migratory caribou in Quebec/Labrador, including from barrenground, with no proportion of their genome of barrenground origin. These results indicate that the eastern migratory caribou from the two disjunct ranges have different demographic histories (see Supporting Information for all statistics). Given the f3 results, we used the f4 statistic to test for introgression between the disjunct eastern migratory caribou populations and found evidence for introgression from Quebec into Ontario/Manitoba eastern migratory but not the other way around (Supporting Information).

The f4 results did not show introgression from Northwest Territories boreal, southern mountain Columbia North or any northern mountain population into NAL boreal caribou (full results for these tests in Supporting Information). The f4 results showed signatures of introgression from the NAL boreal caribou into southern mountain Columbia North, with the f4-ratio statistic indicating that 13.3% of their genomes shows NAL

boreal origin. However, we find no strong evidence of introgression from NAL boreal into any of the northern mountain caribou (Supporting Information). The f4 results do indicate strong signatures of introgression from the NAL boreal caribou into Northwest Territories boreal caribou, with the f4-ratio test suggesting that 16.2% of their genomes originates from the NAL boreal caribou, indicating the possibility of parallel evolution of the same ecotype by adaptive introgression (Supporting Information).

We also used average genome-wide f_D and f_{DM} statistics to estimate the proportion of the genome resulting from introgression, comparable to the f4-ratio scores, and we found the same trends although generally lower with the f_{DM} statistic, likely because it is a conservative estimate (full results in Supporting Information). This analysis also gives results in sliding windows across the genome, and for all comparisons, the regions of introgression appeared spread out throughout the genome encompassing both neutral and functional sites (Figures S9-S14).

To investigate the possibility of adaptive introgression in the parallel evolution of the Northwest Territories boreal caribou, we looked at the most highly introgressed regions from the NAL boreal caribou with an f_{DM} score of at least 0.2. We compared the results with all mountain populations as these are unlikely to have undergone adaptive introgression in the process of parallel evolution given that they have varying levels of introgression depending on distance from the boreal populations, with those closest to the Northwest Territories boreal caribou having negligible levels. We found 49 highly introgressed regions (f_{DM} 0.2 or above) originating from the NAL into Northwest Territories boreal caribou. Within these regions there was a total of 118 genes, with an average of 2.46 genes per introgressed region (Supporting Information for regions and gene lists). In the southern mountain Columbia North population, which is closest geographically to the boreal populations out of our sample locations and has very similar overall levels of introgression as Northwest Territories boreal caribou, we find 64 comparable regions, containing 244 genes and an average of 3.81 genes per region. The northern mountain populations all have fewer of these large, highly introgressed regions as expected from their overall very low levels of introgression from NAL, however the few genomic regions which have introgressed do also contain numerous gene sequences (Itcha-Ilgachuz 14 regions with 39 genes and an average of 2.79 genes per region; Frog has 6 regions with 18 genes and an average of three genes per region; Atlin has 8 regions with 26 genes and an average 3.25 genes per region; Redstone has 9 regions with 44 genes and an average 4.89 genes per region; see Supporting Information for regions and gene lists).

4 | DISCUSSION

Genome sequences of 30 caribou from across North America provided a comprehensive dataset in a non-model terrestrial mammal species-at-risk. We had unprecedented power to reconstruct phylogenomic and demographic history and measure levels of introgression between ecotypes, and to investigate the potential role this introgression has played in parallel evolution. Our results are concordant with previous mtDNA studies (Cronin et al., 2005; Flagstad & Røed 2003; Klütsch et al., 2012; Klütsch et al., 2016; Weckworth et al., 2012) which found two major mitochondrial DNA phylogenetic lineages, NAL and BEL, which likely correspond to divergence within refugia during glacial cycles (Flagstad & Røed 2003; Weckworth et al., 2012). We also found Peary caribou to be genetically distinct from the others in the BEL lineage (Figure 3a), supporting previous evidence of an additional High Arctic refugium (Klütsch et al., 2017). Demographic reconstruction over time showed differential population trajectories of the lineages starting approximately 100-120 kya, indicating divergence to have started well before the Last Glacial Maximum (Figure 2a-b).

4.1 | Parallel evolution and introgression in caribou ecotypes

Our results confirm previous evidence that northern mountain and boreal caribou from the Northwest Territories are within the BEL genomic lineage, even though they are both currently within the woodland subspecies, confirming that the woodland ecotype appears to have arisen in parallel for both (Polfus et al., 2017). Our central mountain caribou are also within the BEL genomic lineage, and this population has been found to be highly admixed based on the two mtDNA lineages (McDevitt et al., 2009). We found evidence for another, as yet undocumented, case of parallel evolution within the eastern migratory ecotype. The eastern migratory caribou from Ontario/Manitoba and those from Quebec/Labrador are not sister groups (Figures

3b and 4) and have different demographic and introgressive histories.

Recent studies are highlighting that introgression between lineages is far more common than previously realised (Coates, Byrne, & Moritz, 2018; Hamilton & Miller, 2015), and the same appears to be true for caribou with introgression likely playing a role in the evolution of the ecotypes. We find more introgression between caribou from the different lineages than anticipated, for example the barrenground caribou have substantial introgression from the NAL. Introgression is also seen from the NAL into the mountain caribou with an isolation by distance pattern, with negligible levels of introgression into more northerly northern mountain caribou. It thus seems unlikely that introgression drove the parallel evolution of the woodland phenotype of the mountain caribou. High levels of introgression coupled with the finding of many introgressed genes makes a compelling case for parallel evolution through adaptive introgression in the Northwest Territories boreal caribou. However, when we compare the gene compliment of the most highly introgressed regions in the Northwest Territories boreal caribou to those found in the mountain caribou, we again find introgressed regions spread across the genome including many genes, even though there are fewer regions overall. There are a few explanations for this pattern, including incomplete lineage sorting (ILS). ILS would be difficult to exclude, especially given that they are closely related intra-specific ecotypes (Lamichhaney et al., 2017). Whether these regions are a result of ILS or introgression the high gene compliment suggests that they could have persisted in the genome due to selection, even if they have not been involved in the parallel evolution of phenotype, due to filtration for maintenance of adaptive genome segments. Additionally, when studying cases of adaptive introgression in inter-species comparisons, areas of introgression are often restricted to single genomic regions (Schweizer et al., 2018), however in intra-specific taxa we may see larger introgressed regions persisting across the genome because the fitness costs may be lessened. Fully teasing these patterns apart in this case may be complicated because multiple processes are likely acting in concert, including ILS and standing variation being selected upon, coupled with differing levels of introgression as the lineages have come into secondary contact. Additionally, given the PSMC results it is possible that there have been multiple bouts of introgression during glacial cycles over the last ˜120,000 years as the lineages repeatedly came into secondary contact. Demonstrating adaptive introgression is complicated and requires the demonstration of the adaptive function of introgressed regions, meaning most cases have thus far have been for well understood traits or those controlled by a single locus (Taylor & Larson, 2019). In contrast, investigating parallel evolution of ecotypes, which will inevitably involve many functional regions, in a non-model species with divergent intra-specific linages and complex demographic histories is a difficult task.

4.2 | Conservation unit designations in the light of complex demographic histories

Given current rates of extirpation and extinction, it is imperative to have strong, scientifically supported management frameworks, particularly given tight resources for conservation (Jackiw, Mandil, & Hager, 2015). Recent work shows that admixture between lineages is common (Coates et al., 2018; vonHoldt et al., 2017), and that new sequencing technologies are allowing us to uncover the complex demographic histories of threatened taxa (Supple & Shapiro, 2018; vonHoldt et al., 2017). Both for caribou and more broadly, now is the time to decide what this means for management and conservation unit designations.

Recent discussion has highlighted that 'hybrid' level gene flow is not always negative, particularly in inbred populations or those needing to adapt to rapid change where admixture could be an important source of variation (Supple & Shapiro, 2018, vonHoldt et al., 2017). For example, we find the barrenground caribou to be very admixed and also to have the lowest individual inbreeding co-efficients, and similarly eastern migratory caribou from Ontario/Manitoba have lower inbreeding co-efficients than the non-admixed individuals from Quebec/Labrador (Table S1). Some argue that gene flow could even be facilitated to aid populations under threat from climate change (i.e. genetic rescue; Hamilton & Miller, 2015), which would be easiest between intra-specific populations (Hedrick & Fredrickson, 2010). Good conservation unit designations with an understanding of natural patterns of admixture is key to assess the potential to use such a strategy (Coates et al., 2018). Most discussions have focussed on policy for inter-species hybridisation (but see Coates et al., 2018; Supple & Shapiro, 2018), but a clear framework for conservation unit designation of admixed intra-specific lineages is needed.

Conservation unit designations depend on the goal of conservation, and whether the focus is on the preservation of phenotypes (or 'pure' genomes), or evolutionary and ecological processes to maintain resilience of an ecosystem (Fitzpatrick et al., 2015; vonHoldt et al.. 2017; Waples & Lindley, 2018). The latter is likely more useful when attempting to designate units for non-discreet entities, such as we see in caribou. With this in mind, some authors have suggested a flexible approach with each case considered on a context specific basis (Jakiw et al., 2015), whereas others promote the need for a structured and uniform framework to decide on management decisions (Coates et al., 2018). For caribou, it seems appropriate for a structured approach in the naming of subspecies. Coates et al., (2018) suggest that subspecies show local adaptation with or without gene flow. Coupling this idea with our phylogenomic and population genomic results and results from previous studies, Canadian caribou appear to fit into three subspecies; those in the NAL, those in the BEL, and Peary caribou which sit phylogenetically in the BEL but show strong population genomic differences and clear local adaptation of phenotype (Banfield, 1961; COSEWIC, 2011).

The most relevant application of our findings is in the delineation of conservation units in a species with complex and admixed evolutionary histories. We recommend that previously defined Designatable Units based on subspecies and subspecific ecotypes be reconsidered: specifically, because the boreal caribou from the Northwest Territories sit within a different lineage to the other caribou within the boreal DU and appear to have evolved in parallel, they could be split into separate DUs. Further fine scale work will be needed to refine the boundary of the BEL boreal vs the NAL boreal DU. Similarly, given the apparent parallel evolution of the eastern migratory ecotype and the different levels of admixture of Ontario/Manitoba vs Quebec/Labrador populations with the BEL lineage, should be divided into separate DUs. Consideration of whether this will help maximise the resilience of the ecosystem is needed, but this would match the evolutionary processes which have led to the evolution of the groups. Confusingly, Grant's caribou and barrenground caribou are currently separate subspecies but one DU. Barrenground caribou are very admixed which contrasts with the Grant's caribou we sampled and so perhaps they warrant listing as separate DUs. Further sampling is needed to resolve the mountain caribou, especially the central mountain population which has been shown to have mitochondrial DNA from both the BEL and NAL lineages (McDevitt et al., 2009). Additionally, genomic data from the southern mountain, and all other DUs not included in this study, is needed to further resolve the complex evolutionary histories and patterns of introgression more broadly. These divisions have significant implications for the status listing of each DU as threat status is assessed based on criteria such as abundance, and priority for management is given to DUs at greatest risk of extinction (COSEWIC, 2015). Given recent rapid declines in both range and population sizes, efficient conservation strategies are needed for caribou.

Our guidelines add to the current discussion about management of admixed populations and those with complex demographic histories (Coates et al., 2018; Fitzpatrick et al., 2015; Hamilton & Miller, 2015; Jackiw et al., 2015; Supple & Shapiro, 2018; vonHoldt et al., 2017). Namely, that subspecies designations are useful and could follow a structured framework (Coates et al., 2018), but that conservation units below the subspecies level likely require a case by case consideration especially given different regulations in different countries (Coates et al., 2018, vonHoldt et al., 2017). Many taxa are facing an increasing threat from climate change and habitat destruction (Hoffman et al., 2017; Ikeda et al., 2017) and genomic data and appropriate conservation unit designations will help with prioritisation given limited resources. Further, genomic data are essential for decisions of genetic rescue strategies. A key next step to achieve these goals, including for caribou, is to investigate adaptive genomic variation to incorporate with demographic history information (Funk, McKay, Hohenloe, & Allendorf, 2012; Funk, Forester, Converse, Darst, & Moreys, 2019).

ACKNOWLEDGEMENTS

Funding was provided through an NSERC Collaborative Research & Development (CRD) grant, NSERC grant RGPIN-2015-04477, Manitoba Hydro, Saskatchewan Power, and Weyerhaeuser Inc. We are thankful to Bridget Redquest and Austin Thompson for extracting DNA, and The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children (Toronto, Ontario) for library preparation and sequencing, to the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and

Compute Canada/ Calcul Canada gme-665-ab.Compute Canada (RRG gme-665-ab), and Amazon Cloud Computing for high-performance computing services. We are also highly recognizant of the contribution of those who provided the tissue samples including Dr. Helen Schwantje, Government of British Columbia; Kent Whaley, Dale Cross, Vicki Trim and Daryll Hedman, Government of Manitoba, Government of Ontario, Guy Savard, Environment and Climate Change Canada; Christine Cuyler, Government of Greenland; Dr. Judith Eger and Lim Burton, The Royal Ontario Museum, Dr. Jean Polfus, University of Manitoba and the Sahtú Renewable Resources Board; and Dr. Mary Gamberg, Gamberg Consulting.

AUTHOR CONTRIBUTION

R.S.T., R.L.H., M.M., and P.J.W. conceived and designed the study. R.S.T., R.L.H., and S.K. performed bioinformatics analyses. R.S.T. wrote the manuscript and R.L.H., M.M., G.B.G., and P.J.W provided feedback and edited the manuscript.

DATA AVAILABILITY STATEMENT

All raw sequencing data will be submitted to the NCBI sequence read archive by the time of publication.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biol* ogy, *215* , 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

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

Banfield, A.W.F. (1967). *A Revision of the Reindeer and Caribou, Genus Rangifer* . National Museum of Canada, Bulletin No. 177, Queen's Printer: Ottawa, ON, Canada.

Bassham. S., Catchen, J., Lescak, E., von Hippel, F. A., & Cresko, W. A. (2018). Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodelling in stickleback. *Genetics,209* **,** 921–939. https://doi.org/10.1534/genetics.117.300610

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics.*Molecular Ecology* , *22* , 3124-3140. https://doi.org/10.1111/mec.12354

Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution, 6* , 165. https://doi.org/10.3389/fevo.2018.00165

COSEWIC (2016). *COSEWIC assessment and status report on the Caribou Rangifer tarandus, Barrenground population, in Canada.* Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

COSEWIC (2017). *COSEWIC assessment and status report on the Caribou Dolphin and Union population (Rangifer tarandus), in Canada.*Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

COSEWIC (2017). *COSEWIC assessment and status report on the Caribou Rangifer tarandus, Eastern Migratory population and Torngat Mountains population, in Canada.* Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

COSEWIC (2014). *COSEWIC assessment and status report on the Caribou Rangifer tarandus, Newfoundland population, Atlantic-Gaspésie population and Boreal population, in Canada.* Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

COSEWIC (2014). *COSEWIC assessment and status report on the Caribou Rangifer tarandus, Northern Mountain population, Central Mountain population and Southern Mountain population in Canada.* Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

COSEWIC (2015). *COSEWIC assessment and status report on the Peary Caribou Rangifer tarandus pearyi in Canada.* Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

COSEWIC (2015). *COSEWIC Assessment Process, Categories and Guidelines* . Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

COSEWIC (2011). *Designatable Units for Caribou (Rangifer tarandus) in Canada* . Ottawa, ON: Committee on the Status of Endangered Wildlife in Canada.

Cronin, M. A., MacNeil, M. D., & Patton, J. C. (2005). Variation in mitochondrial DNA and microsatellite DNA in caribou (*Rangifer tarandus*) in North America. *Journal of Mammalogy, 86* , 495–505. https://doi.org/10.1644/1545-1542(2005)86[495:VIMDAM]2.0.CO;2

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics, 27,*2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32,*1792–1797. https://doi.org/10.1093/nar/gkh340

Festa-Bianchet, M., Ray, J. C., Boutin, S., Côté, S. D., & Gunn, A. (2011). Conservation of caribou (*Rangifer tarandus*) in Canada: An uncertain future. *Canadian Journal of Zoology* , *89* , 419-434. https://doi.org/10.1139/z11-025

Fitak, R. R. (submitted). optM: an R package to optimize the number of migration edges using threshold models. *Journal of Heredity.*

Fitzpatrick, B. M., Ryan, M. E., Johnson, J. R., Corush, J., & Carter, E. T. (2015). Hybridization and the species problem in conservation. *Current Zoology, 61,* 206–216. https://doi.org/10.1093/czoolo/61.1.206

Flagstad, Ø., & Røed, K. H. (2003). Refugial origins of reindeer (*rangifer tarandus* L.) inferred from mitochondrial DNA sequences. *Evolution, 57,* 658–670. https://doi.org/10.1111/j.0014-3820.2003.tb01557.x

Fraser, B. A., & Whiting, J. R. (2019). What can be learned by scanning the genome for molecular convergence in wild populations? *Annals of the New York Academy of Sciences* . https://doi.org/10.1111/nyas.14177.

Funk, W. C., Forester, B. R., Converse, S. J., Darst, C., & Moreys, S. (2019). Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. *Conservation Genetics* , *20,* 115-134. https://doi.org/10.1007/s10592-018-1096-1

Funk, W. C., McKay, J. K., Hohenloe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution, 27,* 489-496. https://doi.org/10.1016/j.tree.2012.05.012

Garroway, C. J., Bowman, J., Cascaden, T. J., Holloway, G. L., Mahan, C. G., Malcolm, J. R., ... Wilson, P. J. (2010). Climate change induced hybridization in flying squirrels. *Global Change Biology,16,* 113–121. https://doi.org/10.1111/j.1365-2486.2009.01948.x

Gómez, J. M., González-Megías, A., Lorite, J., Abdelaziz, M., & Perfectti, F. (2015). The silent extinction: climate change and the potential hybridization-mediated extinction of endemic. *Biodiversity and Conservation* , *24,* 1843–1857. https://doi.org/10.1007/s10531-015-0909-5

Hamilton, J. A., & Miller, J. M. (2015). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology* , *30,* 33–41. https://doi.org/10.1111/cobi.12574

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology, 22* , 4606–4618. https://doi.org/10.1111/mec.12415

Hedrick, P. W., & Fredrickson, R. (2010). Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics* , *11* , 615–626. https://doi.org/10.1007/s10592-009-9999-5

Hoffman, A. A., Sgrò, C. M., & Kristensen, T. N. (2017). Revisiting adaptive potential, population size, and conservation. *Trends in Ecology and Evolution, 32,* 506-517. https://doi.org/10.1016/j.tree.2017.03.012

Ikeda, D. H., Max, T. L., Allan, G. J., Lau, M. K., Shuster, S. M., & Whitham, T. G. (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biology, 23,*164-176. https://doi.org/10.1111/gcb.13470

Jackiw, R. N., Mandil, G., & Hager, H. A. (2015). A framework to guide the conservation of species hybrids based on ethical and ecological considerations. *Conservation Biology, 29,* 1040–1051. https://doi.org/10.1111/cobi.12526

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

Klütsch, C. F. C., Manseau, M., Anderson, M., Sinkins, P. & Wilson, P. J. (2017). Evolutionary reconstruction supports the presence of a Pleistocene Arctic refugium for a large mammal species. *Journal of Biogeography, 44* , 2729–2739. https://doi.org/10.1111/jbi.13090

Klütsch, C. F. C., Manseau, M., Trim, V., Polfus, J. L., & Wilson, P. J. (2016). The eastern migratory caribou: the role of genetic introgression in ecotype evolution. *Royal Soceity Open Science ,3,* 150469. https://doi.org/10.1098/rsos.150469

Klütsch, C. F. C., Manseau, M., & Wilson, P. J. (2012). Phylogeographical analysis of mtDNA data indicates postglacial expansion from multiple glacial refugia in woodland caribou (*Rangifer tarandus caribou*). *PLOS ONE, 7,* e52661. https://doi.org/10.1371/journal.pone.0052661

Knaus, B. J., & Grüwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources, 17,* 44–53. https://doi.org/10.1111/1755-0998.12549

Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L., Nilsson, M. A., & Janke, A. (2017). The evolutionary history of bears is characterized by gene flow across species. *Scientific Reports,7* , 46487. https://doi.org/10.1038/srep46487

Lamichhaney, S., Fuentes-Pardo, A. P., Rafati, N., Ryman, N., McCracken, G. R., Bourne, C., ... Andersson, L. (2017). Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *PNAS, 14* , E3452-E3461. https://doi.org/10.1073/pnas.1617728114

Langmead, B., & Salzberg. S. (2012). Fast gapped-read alignment with Bowtie2. *Nature Methods, 9,* 357-359. https://doi.org/10.1038/nmeth.1923

Lee, K. M., & Coop, G. (2017). Distinguishing among modes of convergent adaptation using population genomic data. *Genetics,207* **,** 1591–1619. https://doi.org/10.1534/genetics.117.300417

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature, 475* , 493-496. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Profect Data Processing Subgroup. (2009). The sequence alignment / map format and SAMtools. *Bioinformatics ,25* **,** 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Z., Lin, K., Ba, H., Chen, L., Yang, Y, Wang, K., ... Li, G. (2017). Draft genome of the reindeer (*Rangifer tarandus*).*GigaScience,* 6, 1–5. https://doi.org/10.1093/gigascience/gix102

Macpherson, A., & Nuismer, S. L. (2017). The probability of parallel genetic evolution from standing genetic variation. *Journal of Evolutionary Biology* , *30* **,** 326–337. https://doi.org/10.1111/jeb.13006

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., ... Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake.*Science, 350* , 1493-1498. https://doi.org/10.1126/science.aac9927

Malinsky, M., Matschiner, M., & Svardal, H. (2019). Dsuit – fast D-statistics and related admixture evidence from VCF files.*bioRxiv,* https://doi.org/10.1101/634477.

Mallet, J. (2005). Hybridization as an invasion of the genome.*Trends in Ecology and Evolution, 20* **,** 229-237. https://doi.org/10.1016/j.tree.2005.02.010

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal, 17* , 10-12. https://doi.org/10.14806/ej.17.1.200

McDevitt, A. D., Mariani, S., Hebblewhite, M., Decesare, N. J., Morgantini, L., Seip, D., ... Musiani, M. (2009). Survival in the Rockies of an endangered hybrid swarm from diverged caribou (*Rangifer tarandus*) lineages. *Molecular Ecology* ,*18* , 665–679. https://doi.org/10.1111/j.1365-294X.2008.04050.x

Mckenna, A., Hanna, M., Banks, E., Sivachenki, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* , *20* , 1297–1303. https://doi.org/10.1101/gr.107524.110

Oke, K. B., Rolshausen, G., Leblond, C., & Hendry, A. P. How Parallel Is Parallel Evolution? A Comparative Analysis in Fishes. *American Naturalist, 190* **,** 1-16. https://doi.org/10.1086/691989

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., ... Reich, D. (2012). Ancient admixture in human history.*Genetics, 192* **,** 1065–1093. https://doi.org/10.1534/genetics.112.145037

Petr, M., Vernot, B., & Kelso, J. (2019). admixr - R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics,35* **,** 3194–3195. https://doi.org/10.1093/bioinformatics/btz030

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics, 8,* e1002967. https://doi.org/10.1371/journal.pgen.1002967

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., ... Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science,344* , 1410-1414. https://doi.org/10.1126/science.1253226

Polfus, J. L., Manseau, M., Klütsch, C. F. C., Simmons, D., & Wilson, P. J. (2017). Ancient diversification in glacial refugia leads to intraspecific diversity in a Holarctic mammal. *Journal of Biogeography* , *44* , 386–396. https://doi.org/10.1111/jbi.12918

Polfus, J. L., Manseau, M., Simmons, D., Neyelle, M., Bayha, W., Andrew, F., ... Wilson, P. J. (2016). Łeghagots'enete (learning together) the importance of indigenous perspectives in the identification of biological variation. *Ecology and Society, 21* , 18. http://dx.doi.org/10.5751/ES-08284-210218

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,26* , 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2015). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature,461* **,** 489-494. https://doi.org/10.1038/nature08365

RStudio Team. (2015). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Available at http://www.rstudio.com/.

Schweizer, M., Warmuth, V., Alaei Kakhki, N., Aliabedian, M., Förschler, M., Shirihai, H., ... Burri, R. (2019). Parallel plumage colour evolution and introgressive hybridization in wheatears. *Journal of Evolutionary Biology* , *32* , 100–110. https://doi.org/10.1111/jeb.13401

Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F. R., Shih, C. H., ... Kohn, M. H. (2011). Report adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology* , *21* , 1296–1301. https://doi.org/10.1016/j.cub.2011.06.043

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,30* **,** 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology* , *19* , 131. https://doi.org/10.1186/s13059-018-1520-3

Taylor, R. S., Horn, R. L., Zhang, X., Golding, G. B., Manseau, M., & Wilson, P. J. (2019). The caribou (*Rangifer tarandus*) genome.*Genes, 10* , 540. https://doi.org/10.3390/genes10070540

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature.*Nature Ecology and Evolution, 3,* 170-177. https://doi.org/10.1038/s41559-018-0777-y

The Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species.*Nature, 487* , 94-98. https://doi.org/10.1038/nature11041

vonHoldt, B. M., Brzeski, K. E., Wilcove, D. S., & Rutledge, L. Y. (2018). Redefining the role of admixture and genomics in species conservation. *Conservation Letters* , *11* **,** 1–6. https://doi.org/10.1111/conl.12371

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

Waples, R. S., & Lindley, S. T. (2018). Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids.*Evolutionary Applications* , 11, 1518–1526. https://doi.org/10.1111/eva.12687

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ionnidis, P., Klioutchnikov, G., ... Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics.*Molecular Biology and Evolution* , *35* **,** 543–548. https://doi.org/10.1093/molbev/msx319

Weckworth, B. V., Hebblewhite, M., Mariani, S., & Musiani, M. (2018). Lines on a map: conservation units, meta-population dynamics, and recovery of woodland caribou in Canada. *Ecosphere, 9* , e02323. https://doi.org/10.1002/ecs2.2323

Weckworth, B. V, Musiani, M., Devitt, A., Hebblewhite, M., & Mariani, S. (2012). Reconstruction of caribou evolutionary history in Western North America and its implications for conservation. *Molecular Ecology* , 21, 3610–3624. 10.1111/j.1365-294X.2012.05621.x

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics, 19* **,** 153. https://doi.org/10.1186/s12859-018-2129-y

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Figures

Hosted file

`image1.emf` available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

FIGURE 1 Range of caribou in North America. Background colours show the ranges of the four subspecies (*R. t. caribou; R. t. groenlandicus; R. t. pearyi; R. t. granti*). Circles indicate sampling locations for this study and are coloured by Designatable Unit. We also included two genomes from Greenland shown by the black circles.

Hosted file

`image2.emf` available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

Hosted file

image3.emf available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

FIGURE 2 Reconstruction of effective population size of caribou. Results have been split into two plots given the differences in peak effective population sizes, with (a) showing the NAL lineage and Peary and Western Greenland caribou and (b) showing all other BEL caribou. The effective populations sizes remain the same until 100-200 kya where demographic histories start to differ, with peak population sizes 120 kya. NAL caribou have smaller peak effective population sizes than BEL caribou, with Peary and Greenland caribou intermediate.

Hosted file

image4.emf available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

Hosted file

image5.emf available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

FIGURE 3 Principal component analyses of caribou genetic variation. All plots show PC1 (x axis) and PC2 (y axis) shown by the eigenvalues plot in the corners. The plots show PCA of all 30 caribou and the Inner Mongolian reindeer (a), fine scale analysis of the NAL caribou (b), fine scale analysis of the BEL caribou, aside from Peary and Western Greenland (c), and fine scale analysis of the 14 individuals clustered together from Figure 3c (d).

Hosted file

image6.emf available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

Figure 4 Maximum likelihood phylogenomic reconstruction from SNP data in RAxML of 30 caribou and the Inner Mongolian reindeer. We show the unrooted phylogeny for clarity, with the root fixed where indicated in analyses using the Sitka deer as an outgroup (See Figure 5 and Figure S8). Nodes show bootstrap support values.

Hosted file

image7.emf available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

Figure 5 Consensus maximum likelihood phylogenomic reconstruction from ˜4,000 conserved mammalian gene sequences from BUSCO of 30 caribou and the Inner Mongolia reindeer rooted using a Sitka deer outgroup**.** Nodes show bootstrap support values.

Hosted file

image8.emf available at https://authorea.com/users/304235/articles/434610-the-role-of-introgression-and-ecotypic-parallelism-in-delineating-intra-specific-conservation-units

Figure 6 Unrooted maximum likelihood phylogeny reconstructed in Treemix with no migration events added (a) and an unrooted maximum likelihood phylogeny reconstructed in Treemix with 7 migration events added (b) as indicated from the OptM results (Figure S7).