CONSERVATION GENOMICS OF THE LAKE SUPERIOR CARIBOU

A dissertation submitted to the Committee on Graduate Studies
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in the Faculty of Arts and Science

TRENT UNIVERSITY
Peterborough, Ontario, Canada

Environmental and Life Sciences Ph.D. Graduate Program
May 2024

ABSTRACT

CONSERVATION GENOMICS OF THE LAKE SUPERIOR CARIBOU

Kirsten Solmundson

Globally, wildlife populations are experiencing increasing rates of range loss, population decline, and extinction. Caribou *(Rangifer tarandus)* have experienced dramatic declines in both range and population size across Canada over the past century. Boreal caribou (*R. t. caribou*), one of twelve Designatable Units, have lost approximately half of their historic range in the last 150 years, particularly along the southern edge of their distribution. Despite this northward contraction, some populations have persisted at the trailing range edge, over 150 km south of the boreal continuous range (BCR) in Ontario, along the coast and near-shore islands of Lake Superior. Better understanding the population structure and evolutionary history of caribou in the Lake Superior range (LSR) could help to inform conservation and management actions, such as the delineation of conservation or management units or translocations between populations. In this thesis, I use whole genome sequences from boreal, eastern migratory and barren-ground caribou sampled in Manitoba, Ontario, and Quebec to investigate evolutionary history and population structure. I discovered that the LSR caribou form a distinct group but also some evidence of gene flow with the BCR. Notably, caribou from the LSR demonstrated relatively high levels of inbreeding (measured as Runs of Homozygosity; ROH) and genetic drift, which may contribute to the differentiation observed between caribou occupying the two ranges. Despite inbreeding, the LSR caribou retained Heterozygosity Rich Regions (HRR). I found genomic structure among caribou populations from the LSR and BCR but found these two ranges had similar demographic histories. My analyses indicate that the LSR caribou display distinct genomic characteristics but share ancestry with the BCR, with historical gene flow between these two ranges. Collectively, this dissertation characterizes the population structure

and evolutionary history of caribou from the southernmost range in Ontario, providing key insights for the conservation and management of these small and isolated populations.

KEYWORDS: conservation, genomics, whole genome sequencing, caribou, isolation

PREFACE

I have written my dissertation in manuscript format. Chapter 2 was published in *Ecology and Evolution*. An earlier draft of Chapter 3 was posted as a preprint; it may be published in it's current or a revised form. Chapter 4 has been prepared to submit to *Conservation Genetics*. Finally, I provide a general discussion and conclusions in Chapter 5.

The format of each chapter may vary slightly according to journal requirements. My research has been conducted in collaboration with others; I have identified my coauthors on the title page of each chapter and use the plural "we" throughout the main data chapters.

ACKNOWLEDGEMENTS

Firstly, I am grateful to my supervisors Dr. Jeff Bowman and Dr. Paul Wilson. Thank you for the opportunity to work on this important and interesting research project and for sharing your perspectives and advice. I also appreciate the contributions of my committee members Dr. Micheline Manseau and Dr. Brent Patterson, who have provided valuable insights throughout my research journey. I thank the members of the Bowman, Wilson, and Manseau lab groups who have provided support and encouragement over the years. I am especially grateful for Austin Thompson's assistance with the DNA extractions; I couldn't have done it without you (and the butterfly of luck!). Special thanks to fellow PhD Candidate Sasha Newar, who has been invaluable as a lab mate, friend, and roommate over the past 6 years. Thank-you also to the Trent professors who shared their knowledge and wisdom; I am particularly grateful for the insights shared by Dr. Joanna Freeland, Dr. Jim Schaefer, and Dr. Aaron Shafer, and their PhD students.

Thank you to my collaborators from Environment and Climate Change Canada, the Ontario Ministry of Natural Resources and Forestry, and the Canadian Museum of Nature for providing samples that contributed to this dissertation. This research was supported by funding from: the Ontario Species at Risk Stewardship Program, WCS Canada W. Garfield Weston Foundation, Ontario Graduate Scholarships, and the Genomic Applications Partnership Program of Genome Canada and Environment and Climate Change Canada.

Most importantly, none of my accomplishments would be possible without the unwavering love and support of my family and friends. I am so grateful to come from a family that values academics, science, and nature. To my mom and dad: thank you for encouraging me to be nature-loving, curious, and hard-working. To my sister: thank you for always being my best friend and biggest fan; you encourage me to be the best version of myself. To my hometown

friends: thank you for embracing my interests; I love sharing my passions with you, your families, and your classrooms. To my graduate school friends: thank you for all of the advice, inspiration, and outdoor adventures over the years; I have immensely enjoyed working on collaborative projects, helping with fieldwork, hiking, and camping with you all! Thank you to my partner Payton, who has always believed in me, my abilities, and my dreams. My cat Cub also deserves special recognition; he has been a constant companion through many coding marathons, virtual lectures, and writing sessions.

"*Last but not least, I want to thank me.*
I want to thank me for believing in me.
I want to thank me for doing all this hard work.
I want to thank me for having no days off.
I want to thank me for never quitting.
I want to thank me for always being a giver,
and trying to give more than I receive.
I want to thank me for trying to do more right than wrong."

- Calvin Cordozar Broadus Jr.
(*Snoop Dogg*)

TABLE OF CONTENTS

Abstract ii
Preface iv
Acknowledgements v
Table of Contents vii
List of Figures ix
List of Tables xi
Chapter 1: General Introduction 1
References 7
Chapter 2: Genomic population structure and inbreeding history of Lake Superior caribou 10
Abstract 11
Introduction 12
Methods 16
Results 23
Discussion 25
References 34
Tables 41
Figures 43
Declarations 48
Supplementary Materials 49
Chapter 3: Heterozygosity Rich Regions maintained in isolated caribou 58
Abstract 59
Introduction 60
Methods 61
Results 64
Discussion 66
References 70
Tables 74
Figures 74
Declarations 77
Supplementary Materials 78
Chapter 4: Whole genomes reveal the demographic history of Ontario's boreal caribou 87
Abstract 88
Introduction 89
Materials and Methods 92
Results 96
Discussion 98
References 102
Declarations 106
Tables 107
Figures 108
Supplementary Materials 110
Chapter 5: Conclusions 118
References 125
Supplementary Materials 128
Appendix 133

LIST OF FIGURES

Chapter 2

FIGURE 1 Caribou (*Rangifer tarandus*) surrounded by early morning mist on Michipicoten Island, Lake Superior, Ontario, Canada. Photo by Andy Silver (Ontario Ministry of Natural Resources and Forestry) 43

FIGURE 2 Sample sites of caribou (N = 20) in this study. Background colors show the ranges of three of the Canadian Designatable Units (DUs) included: barren- ground, eastern migratory, and boreal. Circles on map indicate sample locations and the colors indicate individual population assignment proportions under the best supported model of K = 2. The arrows on the map indicate migrations modeled by Treemix. The Lake Superior region is indicated by dashed borders; abbreviated labels indicate site names: Pukaskwa National Park (PNP), Neys Provincial Park (NPP), Pic Island (PI), The Slate Islands (SI), and Michipicoten Island (MI). X- axis label indicates individual IDs.....44

FIGURE 3 (a) Principal Component Analysis (PCA) visualizing genomic variation among caribou (N = 20). (b) Evolutionary tree (Treemix) without migration showing evolutionary relationships (k = 1000). Branch lengths indicate drift estimates. (c) Evolutionary tree (Treemix) migration model with the best support (k = 1000) depicting three migrations. Arrows indicate the direction of migration or gene flow; arrow colors indicate the strength of migration.....45

FIGURE 4 Inbreeding estimates in caribou based on Runs of Homozygosity (ROH) identified with PLINK (a, b) and ROHan (c, d) under the 250 kb size class. (a, c) FROH indicates the proportion of the genome classified as ROH. (b, d) Reflects the average length of ROH in kilobases, where shorter ROH indicate historical inbreeding and longer ROH indicate recent inbreeding. Error bars (c, d) represent the minimum and maximum estimates produced by the hidden Markov model in ROHan 46

FIGURE 5 Individual genetic diversity in caribou, calculated as genome- wide heterozygosity (Watterson's θ). Heterozygosity was calculated across the whole genome (including ROH; solid circles) and excluding regions in ROH (hollow triangles), using 250 kb windows and allowing a heterozygosity rate of 0.0005 within ROH. Error bars indicate the minimum and maximum estimates of the hidden Markov model.....47

Chapter 2 Supplementary

Figure S1. NGSAdmix plot for the next best supported K-values (3 and 6). Each bar represents an individual and colours represent population assignments..... 51

Figure S2. Cumulative variance for the 19 Principal Components of the PCA.....51

Figure S3. Additional Principal Component Axes: A) 1vs3 B) 2vs3 C) 1vs4 D) 2vs4.....52

Figure S4. Comparison of Treemix migration models. We performed 10 iterations at k-values of 500, 1000, and 2000 for each migration model (m=0-7). The likelihood and SD values (A) and comparison of Δm values (B) indicate the model with 3 migrations has the best support 53

Figure S5. Runs of homozygosity identified with PLINK under different sets of parameters. A-D) relaxed parameters, >250kb ROH. E-H) strict parameters, >250kb ROH. I-L) relaxed parameters, >1Mb ROH. M-P) strict parameters, >1Mb ROH.....54

Figure S6. Runs of homozygosity identified with ROHan under different sets of parameters. We used window sizes of 250kb (A-D) and 1Mb (E-H) under relaxed (A,B,E,F) and more strict (C,D,G,H) parameters. Error bars indicate minimum and maximum estimates produced by the Hidden Markov Model.....55

Figure S7. Size classes for Runs of Homozygosity (ROH) identified with ROHan under relaxed parameters with 250kb windows. Y-axis indicates individual caribou ID.....56

Chapter 3

Figure 1. Map of caribou sample (N=20) locations. Background layers reflect the ranges of the Designable Units (DU) of caribou: barren-ground, eastern migratory, and boreal. Within the boreal DU, we sampled caribou from the continuous range and the Lake Superior range.....74

Figure 2. Heterozygosity Rich Regions (HRR) identified using high-coverage whole genomes of 20 caribou. Points indicate heterozygosity, estimated as Waterson's θ with 1Mb windows across each genome. Grey error bars indicate the error bounds of the Markov Chain Monte Carlo. Note the Y-axis of Scaffold 7 (plot C) has a maximum of 0.04, whereas all other plots have a maximum Y of 0.03.....75

Figure 3. Summary of Gene Ontology analysis results for the A) molecular functions and B) biological processes associated with genes located in HRR.....76

Chapter 4

Figure 1. Sample locations and population assignment results from whole genomes of boreal caribou from Ontario, Canada (N=197). The background layers of the map indicate the current range distributions. The colours of map location points, PCA points, and the bar beneath the NGSAdmix plots reflect the sample groupings used for grouped genotyping: MICH = Michipicoten Island; PUK = Pukaskwa National Park; BCR ..... 1, BCR2, BCR3 = boreal continuous range of caribou in Ontario.....95

Figure 2. Historical effective population sizes of boreal caribou groups modelled using an Approximate Bayesian Computation approach with 500,000 simulations. The upper plot (A) displays deeper evolutionary history with the model starting 130,000 generations ago; the plot reflects the most recent 100,000 years. The lower (B) reflects more recent history, corresponding to the most recent 1000 generations before present. Dotted lines indicate the 5% and 95% quantiles and shaded regions indicate areas of high model uncertainty..... 96

LIST OF TABLES

Chapter 2

TABLE 1 Information for each caribou included in this study: sample individual reference number, subspecies classification, Canadian Designatable Unit, and approximate sampling location indicating herd or region sampled and the province. Mean depth refers to the average depth of coverage from filtered whole genome BAM files. Inbreeding was quantified as the proportion of the genome in Runs of Homozygosity (F_{ROH}) identified with PLINK and ROHan. Table values reflect ROH measured in kb, identified under the 250 kb size class and more relaxed set of parameters tested for each method.....41

Chapter 2 Supplementary

Table S1. Sample ID reference numbers and genomic reference numbers to access raw sequence data used in this study.....49

Table S2. Model probabilities based on the mean and SD of log likelihood values produced by NGSAdmix across an array of 10 runs. The model with the highest probability (K=2) is the best supported value of K 50

Table S3. Inbreeding estimates calculated using VCFtools, PLINK, and ROHan with a size class of 250 kb. We used the strictly filtered VCF file to calculate the inbreeding coefficient (F) in VCFtools and to identify ROH using PLINK with two sets of parameters (as specified in Methods). We also identified ROH in the individual BAM files using ROHan with two sets of parameters.....50

Chapter 3

Table 1. Heterozygosity Rich Regions identified in high-coverage whole genomes of caribou (N=20). Columns 1-3 indicate the locations of genetically diverse regions in BED format. Column 4 represents the number of predicted genes within each region, whereas column 5 indicates how many genes were identified, excluding duplicate entries 74

Chapter 3 Supplementary

Table S1. Metadata for all high coverage whole caribou genomes included in this study (N=20).....78

Table S2. Predicted genes within Heterozygosity Rich Regions 79

Table S3. Heterozygosity estimates (Watterson's θ) within Heterozygosity Rich Regions for each caribou (N-20). Columns 3-5 indicate the location of HRR windows in BED format. 82

Table S4. Molecular sub-functions associated with genes in Heterozygosity Rich Regions. 85

Table S5. Biological sub-processes associated with genes in Heterozygosity Rich Regions. 86

Chapter 4

Table 1. Summary of sample groupings of boreal caribou, determined through comparative analyses of low coverage whole genomes (N=197).....106

Chapter 4 Supplementary

Table S1. Metadata for genomic samples: identifier, tissue type, collection year, range area (BCR vs LSR), and average genome-wide sequencing depth 110

Table S2. Probability values for NGSAdmix models of K=2-8. Values were calculated from the log likelihood results from 10 arrays for each value of K.....117

Chapter 5 Supplementary

Table S1. Inbreeding estimates of caribou from Michipicoten Island. Samples were collected on the island prior to 2018 (Population group = Michipicoten Island) and during the translocation of caribou to the Slate and Caribou Islands.....130

Table S2. Pairwise relatedness (R_{AB}) estimates of caribou from Michipicoten Island. PCID indicates the individual genomic sample identifier.....132

CHAPTER 1

INTRODUCTION

Declining wildlife

Globally, wildlife are experiencing increasing rates of range loss, population decline, and extinction due to recent anthropogenic changes (Barnosky et al. 2011). An extensive review of 177 mammal species revealed that more than 40% of the species have experience severe population declines, and all have lost 30% or more of their historical geographic ranges (Ceballos et al. 2017). As a result of human activity, many species are persisting at low numbers or in isolated populations, and their survival now relies on human intervention, such as captive breeding programs and translocations to reestablish areas of range loss or genetically rescue dwindling populations (Lorimer 2015; Robinson et al. 2020).

Understanding the evolutionary history of a species is an important component in making informed management decisions, such as the delineation of species, subspecies, and populations into conservation and management units (COSEWIC 2020). The genetic composition of populations is influenced by past evolutionary events, including bottlenecks, admixture, inbreeding, and selective pressures, which results in variation among populations (Coates et al. 2018). In recognition of the importance of genetic variation, wildlife populations are typically divided into conservation management units that reflect their evolutionary lineages, which are sometimes referred to as Evolutionarily Significant Units or Distinct Population Segments (Coates et al. 2018; Funk et al. 2019). In Canada, wildlife populations that are considered to be both distinct and evolutionarily significant are categorized as Designatable Units (DUs) by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC 2020).

The caribou (*Rangifer tarandus*) is an iconic species with remarkable variation among populations, attributable to complex history of evolutionary events including local adaptation, bottlenecks, and introgression (Polfus et al. 2017; Taylor et al. 2022). The diversity of caribou found in Canada has resulted in the recognition of 12 Designatable Units (DUs) by COSEWIC (2011). Despite this diversity, all extant caribou in Canada have been recommended for listing as Species-at-Risk (Endangered, Threatened, or Special Concern; COSEWIC, 2014-2017) and the species is classified as Vulnerable by the IUCN throughout their circumpolar range (Gunn 2016).

Range loss

Over the past century, many wildlife species have been extirpated from significant portions of their historical ranges (Ceballos et al. 2017). In North American mammals, this range loss has largely occurred on the southern edge of the species range and is attributable to recent anthropogenic changes (Laliberte and Ripple 2004). As ranges shift and contract, remanent populations, which were previously contained within the core range, can become isolated (Schaefer 2003). These remanent populations are referred to as rear edge or trailing edge populations (Hampe and Petit 2005). In some species, range loss along the trailing edge can be offset by expansion along the leading edge or expanding edge; however, in species with a circumpolar distribution, such as the caribou, the species range cannot shift further northward (Hampe and Petit 2005; Festa-Bianchet et al. 2011).

Trailing edge

As ranges contract, conservation management typically assumes populations along the range periphery are less likely to persist than those in the range core (Wolf et al. 1996); however, most species persist in the periphery of their historical geographical ranges (Channell and Lomolino 2000). In fact, more than 70% of North American mammal species that have

experienced dramatic range loss have contracted towards the edge of their historic range rather than towards the range core (Laliberte and Ripple 2004). The boreal caribou (*R. t. caribou*), one of the twelve Designatable Units, has lost approximately half of its historic range in the last 150 years, particularly along the southern edge of its distribution. Despite this overall northward contraction, some populations have persisted at the southern trailing range edge, over 150 km south of the boreal continuous range (BCR) in Ontario, along the coast and near-shore islands of Lake Superior. The survival of caribou on Lake Superior islands, despite other nearby extirpations, is congruent with the findings of a multi-species review, which revealed that when a species' historical range includes both mainland and island sites, population persistence is expected to be the highest on islands(Channell and Lomolino 2000). Islands harbour greater proportions of threatened species than expected when compared to mainland habitats (Ricketts et al. 2005; Spatz et al. 2017); however, this disparity is often because islands are the last sites to be anthropogenically disturbed, allowing remnant populations to persist even when populations on the mainland have been extirpated (Lomolino and Channell 1998). This may be the case in the Lake Superior range (LSR), which is characterized by generally low levels of human disturbance provided by protected areas such as Pukaskwa National Park, and Neys and the Slate Islands Provincial Parks (Schaefer 2003).

In species at risk, small and isolated populations along the trailing edge can be disproportionately important for the species' survival and evolution as ranges shift and contract (Hampe and Petit 2005). Trailing edge populations are typically characterized by lowered within population genetic diversity, attributable to small population sizes and prolonged isolation; however, these populations also demonstrate disproportionately high levels of genetic differentiation in comparison to nearby populations (Hampe and Petit 2005). The conditions

created by trailing edge dynamics can encourage selection for local adaptation and reduced gene flow, producing remarkably distinct populations (Castric and Bernatchez 2003; Pérez-Tris et al. 2004; Hampe and Petit 2005). Despite the potential for local adaptation, rear edge populations face a high risk of local extinction (Hampe and Petit 2005).

Caribou in the LSR have faced multiple recent bottlenecks and translocation events, and are currently surviving at low abundance in isolation from the boreal continuous range of Ontario (Ontario Ministry of Natural Resources and Forestry 2018). Caribou historically inhabited the coast and islands of Lake Superior; however, the populations have struggled to persist for several decades (Schaefer 2003). By the 1980s, only one bull was confirmed to be present on Michipicoten Island when nine caribou were translocated from the Slate Islands to Michipicoten Island (Ontario Ministry of Natural Resources and Forestry 2018). This population grew to approximately 680 caribou by 2010; however, in the winter of 2014, an ice bridge formed, allowing wolves to colonize the island (Ontario Ministry of Natural Resources and Forestry 2018). Facing this new predation pressure, the Michipicoten Island caribou population rapidly declined until some of the few remaining caribou were translocated to the Slate and Caribou Islands in early 2018 (Ontario Ministry of Natural Resources and Forestry, 2018). The recent history of caribou in the LSR is well-documented, but their deeper evolutionary history remains unclear. Understanding the population structure and evolutionary history of caribou in the LSR is important for their conservation and management.

Conservation genomics

The field of genomics can provide new insights for wildlife conservation, allowing researchers to categorize individual or population-level variation and model recent and historic evolutionary events, such as inbreeding and introgression (Primmer 2009; Hohenlohe et al.

2021). Advances in technology now permit the sequencing of whole genomes from non-model wildlife species, enabling conservation genomic research to inform conservation status listings and recovery decisions (Funk et al. 2019). In this thesis, I use whole genomes from boreal, eastern migratory and barren-ground caribou sampled in Manitoba, Ontario, and Quebec to investigate evolutionary history and population structure.

Thesis objectives

In this dissertation, I use whole genome sequences to investigate the ecology and evolution of the Lake Superior caribou. My analyses use chromosome-level genomes generated using next-generation sequencing methods to investigate population history, adaptation, and inbreeding in comparison to caribou from several ranges and ecotypes. The overall objectives of my thesis are to:

1) Characterize the genomic population structure of LSR caribou in comparison to other populations

2) Investigate the evolutionary history and extent of inbreeding of caribou in the LSR

3) Identify putatively adaptive differences among caribou populations

4) Model the recent and ancient demographic history of caribou from the LSR and the boreal continuous range

5) Demonstrate the applicability of genomic data for wildlife conservation and management

Thesis structure

My dissertation is divided into 5 chapters; chapters 2, 3, 4 are the main data chapters. In Chapter 2, I use high coverage whole genomes from several caribou herds and ecotypes to investigate the population structure and history of the Lake Superior caribou. I identified the Lake Superior

caribou as a distinct genomic group that branched from the nearby northern continuous range. I detected some evidence of historical gene flow to the Lake Superior populations, but notably higher inbreeding in caribou from the Lake Superior range compared to other caribou herds and ecotypes. This chapter informed Chapters 3 and 4 where I subsequently investigate the adaptive differences and demographic histories among ranges. In Chapter 3, I investigate putatively adaptive regions. Using the high coverage data, I identified Heterozygosity Rich Regions, which can be indicative of balancing selection. I further investigate genes within these regions, as well as the associated molecular functions and biological pathways. In Chapter 4, I use low coverage whole genomes from boreal caribou in Ontario to investigate population structure among herds and their demographic history. This chapter reveals consistent genomic structure in caribou from Michipicoten Island and Pukaskwa National Park; however, we found similar demographic histories between the Lake Superior and boreal continuous ranges. Chapter 5 provides general conclusions on this dissertation research and the field of conservation genomics.

REFERENCES

Barnosky AD, Matzke N, Tomiya S, et al (2011) Has the Earth's sixth mass extinction already arrived? Nature 471:51–57. https://doi.org/10.1038/nature09678

Castric V, Bernatchez L (2003) The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchill). Genetics 163:983–996. https://doi.org/10.1093/genetics/163.3.983

Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc Natl Acad Sci 1–8. https://doi.org/10.1073/pnas.1704949114

Channell R, Lomolino M V. (2000) Dynamic biogeography and conservation of endangered species. Nature 403:84–86

Coates DJ, Byrne M, Moritz C (2018) Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. Front Ecol Evol 6:1–13. https://doi.org/10.3389/fevo.2018.00165

COSEWIC (2020) COSEWIC guidelines for recognizing designatable units. In: Gov. Canada. https://cosewic.ca/index.php/en-ca/reports/preparing-status-reports/guidelines-recognizing-designatable-units.html. Accessed 20 Feb 2023

COSEWIC (2011) Designatable units for caribou (*Rangifer tarandus*) in Canada. 88pp.

COSEWIC (2014a) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Newfoundland population, Atlantic-Gaspésie population and Boreal population, in Canada

COSEWIC (2014b) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Northern Mountain Population, Central Mountain Population, and Southern Mountain Population in Canada

COSEWIC (2015) COSEWIC assessment and status report on the Peary caribou *Rangifer tarandus pearyi*, in Canada

COSEWIC (2016) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Barren-Ground population, in Canada

COSEWIC (2017a) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Eastern Migratory population and Torngat Mountains population, in Canada

COSEWIC (2017b) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Dolphin and Union population, in Canada

Festa-Bianchet M, Ray JC, Boutin S, et al (2011) Conservation of caribou (*Rangifer tarandus*) in Canada: an uncertain future. Can J Zool 89:419–434. https://doi.org/10.1139/z11-025

Funk WC, Forester BR, Converse SJ, et al (2019) Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. Conserv Genet 20:115–134. https://doi.org/10.1007/s10592-018-1096-1

Gunn A (2016) *Rangifer tarandus*. The IUCN Red List of Threatened Species

Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: The rear edge matters. Ecol Lett 8:461–467. https://doi.org/10.1111/j.1461-0248.2005.00739.x

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

Laliberte AS, Ripple WJ (2004) Range contractions of North American carnivores and ungulates. Bioscience 54:123. https://doi.org/10.1641/0006-3568(2004)054[0123:RCONAC]2.0.CO;2

Lomolino M V., Channell R (1998) Range Collapse, Re-Introductions, and Biogeographic Guidelines for Conservation. Conserv Biol 12:481–484. https://doi.org/24.77.130.166

Lorimer J (2015) Wildlife in the Anthropocene: conservation after nature. U. of Minnestoa Press.

Ontario Ministry of Natural Resources and Forestry (2018) Seeking advice on the future of caribou in the Lake Superior coast range

Pérez-Tris J, Bensch S, Carbonell R, et al (2004) Historical diversification of migration patterns in a passerine bird. Evolution (N Y) 58:1819–1832. https://doi.org/10.1111/j.0014-3820.2004.tb00464.x

Polfus JL, Manseau M, Klütsch CFC, et al (2017) Ancient diversification in glacial refugia leads to intraspecific diversity in a Holarctic mammal. J Biogeogr 44:386–396. https://doi.org/10.1111/jbi.12918

Prentice MB, Bowman J, Khidas K, et al (2017) Selection and drift influence genetic differentiation of insular Canada lynx (*Lynx canadensis*) on Newfoundland and Cape Breton Island. Ecol Evol 7:3281–3294. https://doi.org/10.1002/ece3.2945

Primmer CR (2009) From conservation genetics to conservation genomics. Ann N Y Acad Sci 1162:357–368. https://doi.org/10.1111/j.1749-6632.2009.04444.x

Ricketts TH, Dinerstein E, Boucher T, et al (2005) Pinpointing and preventing imminent extinctions. Proc Natl Acad Sci U S A 102:18497–18501. https://doi.org/10.1073/pnas.0509060102

Robinson ZL, Bell DA, Dhendup T, et al (2020) Evaluating the outcomes of genetic rescue attempts. Conserv Biol 35:666–677. https://doi.org/10.1111/cobi.13596

Schaefer JA (2003) Long-term range recession and the persistence of caribou in the taiga. Conserv Biol 17:1435–1439. https://doi.org/10.1046/j.1523-1739.2003.02288.x

Spatz DR, Zilliacus KM, Holmes ND, et al (2017) Globally threatened vertebrates on islands with invasive species. Sci Adv 3:. https://doi.org/10.1126/sciadv.1603080

Supple MA, Shapiro B (2018) Conservation of biodiversity in the genomics era. Genome Biol 19:. https://doi.org/10.1186/s13059-018-1520-3

Taylor RS, Manseau M, Keobouasone S, et al (2022) Continent-wide phylogenomic framework reveals introgression as a driver of intra-specific diversity and enriched molecular pathways in caribou. bioRxiv. https://doi.org/10.1101/2022.12.19.521038

Wolf CM, Griffith B, Reed C, Temple SA (1996) Avian and Mammalian Translocations: Update and Reanalysis of 1987 Survey Data. Conserv Biol 10:1142–1154. https://doi.org/10.1046/j.1523-1739.1996.10041142.x

CHAPTER 2

Genomic population structure and inbreeding history of Lake Superior caribou

A VERSION OF THIS ARTICLE HAS BEEN PUBLISHED

Solmundson, K., Bowman, J., Manseau, M., Taylor, R. S., Keobouasone, S., & Wilson, P. J. (2023). Genomic population structure and inbreeding history of Lake Superior caribou. *Ecology and Evolution*, *13*(7), 1–14. https://doi.org/10.1002/ece3.10278

ABSTRACT

Caribou *(Rangifer tarandus)* have experienced dramatic declines in both range and population size across Canada over the past century. Boreal caribou (*R. t. caribou*), one of the twelve Designatable Units, have lost approximately half of their historic range in the last 150 years, particularly along the southern edge of their distribution. Despite this overall northward contraction, some populations have persisted at the trailing range edge, over 150 km south of the continuous boreal caribou range in Ontario, along the coast and near-shore islands of Lake Superior. The population history of caribou along Lake Superior remains unclear. It appears that these caribou likely represent a remnant distribution at the trailing edge of the receding population of boreal caribou, but they may also exhibit local adaptation to the coastal environment. Better understanding the population structure and history of caribou along Lake Superior is important for their conservation and management. Here, we use high-coverage whole genomes (N = 20) from boreal, eastern migratory and barren-ground caribou sampled in Manitoba, Ontario, and Quebec to investigate population structure and inbreeding histories. We discovered that the Lake Superior caribou form a distinct group but also some evidence of gene flow with the continuous boreal caribou range. Notably, caribou along Lake Superior demonstrated relatively high levels of inbreeding (measured as Runs of Homozygosity; ROH) and genetic drift, which may contribute to the differentiation observed between ranges. Despite inbreeding, caribou along Lake Superior retained high heterozygosity, particularly in genomic regions without ROH. These results suggest that they present distinct genomic characteristics but also some level of gene flow with the continuous range. Our study provides key insights into the genomics of the southernmost range of caribou in Ontario, beginning to unravel the evolutionary history of these small, isolated caribou populations.

KEYWORDS: Whole genomes, conservation genomics, population structure, inbreeding, caribou

1 | INTRODUCTION

The caribou (*Rangifer tarandus*), an iconic Canadian species, has experienced dramatic declines in both range and population size over the past century, raising conservation concerns (Laliberte and Ripple 2004; Festa-Bianchet *et al.* 2011). Caribou are ecologically diverse and central to the culture and livelihood of Indigenous peoples *(Festa-Bianchet et al. 2011; Polfus et al.* 2016). Caribou diversity is described by several subspecies and ecotypes, which differ in morphology and behaviour; for example, barren-ground caribou (*R. t. groenlandicus*) congregate in large, migratory groups on the tundra (COSEWIC 2016). Conversely, the woodland subspecies (*R. t. caribou*) has several ecotypes associated with different habitats, such as caribou found in the mountains across western Canada (COSEWIC 2014a), the eastern migratory caribou that migrate between the boreal forest and the tundra in eastern Canada (COSEWIC 2017a), and boreal caribou that are more sedentary and found throughout the boreal forest (COSEWIC 2014b). The diversity found in caribou has resulted in the recognition of 12 Designatable Units (DUs) by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC 2011). All extant caribou in Canada have been recommended for listing as Species-at-Risk (Endangered, Threatened, or Special Concern) by COSEWIC (COSEWIC, 2014-2017). The species is classified as Vulnerable by the IUCN and is estimated to have declined by 40% in the past 10-25 years across its circumpolar range (Gunn 2016).

The declining trends observed in caribou populations across Canada have raised conservation concerns, as small and isolated populations are more prone to inbreeding and may eventually fall into an “extinction vortex” and become extirpated (Gilpin and Soule 1986; Festa-Bianchet *et al.* 2011; Gagnon *et al.* 2019). The extent of inbreeding likely varies among populations however, especially in the context of historical population fluctuations and recent

declines. Additionally, recent phylogenomic analyses showed that the evolutionary lineages of caribou are not concordant with current DUs (Taylor *et al.* 2022), presenting further insights for conservation and management.

Declines in caribou ranges and population sizes have resulted in small and isolated populations, particularly within the southern mountain and boreal ecotypes (COSEWIC 2014b, a). A recent microsatellite study revealed genetic erosion, a decrease in connectivity, and an increase in inbreeding along the southern continuous range edge of boreal caribou in Ontario and Manitoba (Thompson *et al.* 2019). In Ontario, the southern continuous range edge of boreal caribou has been contracting northward for over a century, primarily due to anthropogenic habitat disturbance (Schaefer 2003). Boreal caribou rely on dense forest for sufficient forage and to avoid wolf predation when calving, and thus are limited by habitat loss and fragmentation in parts of their historic range (Festa-Bianchet *et al.* 2011). This range loss has resulted in isolated populations on the trailing range edge that have managed to persist along the coast and on nearshore islands of Lake Superior (Figure 1), over 150 km south of the continuous range edge (Figure 2; Ontario Ministry of Natural Resources, 2009; Schaefer, 2003). The recent history (1900s-present) of caribou along Lake Superior is well documented (e.g. Bergerud, 1985, 2001; Bergerud *et al.*, 2007, 2014; Carr *et al.*, 2012; Patterson *et al.*, 2014; Ontario Ministry of Natural Resources and Forestry, 2018); however their deeper evolutionary history remains unclear. Microsatellite analysis suggest some genetic structure within the region and low levels of gene flow between Lake Superior caribou from Pukaskwa National Park and caribou farther north in the continuous range (Drake *et al.* 2018).

Conservation management typically assumes populations along the range periphery are less likely to persist than those in the range core; however, an extensive multi-species review

revealed that most species persisted in the periphery of their historical geographical ranges (Channell and Lomolino 2000). Notably, when a species' historical range included both mainland and island sites, population persistence was highest on islands, even when island habitat patches were smaller than those on the mainland (Channell and Lomolino 2000). In general, islands harbour greater proportions of threatened species than expected when compared to mainland habitats (Ricketts *et al.* 2005; Spatz *et al*. 2017). However, this disparity is often because islands represent the last sites to be disturbed by anthropogenic factors, allowing remnant populations to persist even when populations on the mainland have been extirpated (Lomolino and Channell 1998). Another review demonstrated that as ranges contract, small and isolated populations along the rear edge can become disproportionately important for the species' survival and evolution (Hampe and Petit 2005). Small population size and prolonged isolation reduces within population genetic diversity; however, trailing edge populations also demonstrate disproportionately high levels of genetic differentiation when compared to nearby populations (Hampe & Petit, 2005). The conditions created by trailing edge dynamics can encourage selection for local adaptation and reduced gene flow, ultimately resulting in remarkably distinct populations (Castric and Bernatchez 2003; Pérez-Tris *et al.* 2004; Hampe and Petit 2005). Both drift and local adaptation can contribute to the unique characteristics of rear edge populations, and these effects can be difficult to disentangle (Prentice *et al.* 2017). Regardless, rear edge populations face a high risk of local extinction, especially when regional population dynamics such as immigration are impeded by isolation (Hampe and Petit, 2005).

In this study, we used high coverage whole genome sequences from 20 caribou to investigate population structure and inbreeding in small and isolated populations of boreal caribou from the Lake Superior range, boreal caribou from the continuous caribou range of

Ontario and Manitoba, eastern migratory caribou from Ontario and Quebec, and barren-ground caribou from northern Manitoba (Figure 2). We expected that population clustering among caribou would broadly reflect the ecotypes and sample locations. However, previous research suggested that eastern migratory caribou originated from introgression between barren-ground and boreal caribou (Klütsch *et al.* 2016), and a subsequent study indicated introgression has occurred among the barren-ground, eastern migratory, and boreal ecotypes (Taylor *et al*., 2020). Thus, this historic exchange of genetic material may be detected as migration or gene flow.

Caribou in the Lake Superior range persist in small and apparently isolated island populations (Drake *et al*., 2018; Schaefer, 2003), and we tested the hypothesis that these caribou exhibit characteristics of a trailing edge, including effects of genetic drift, low within-population diversity, and high differentiation from the continuous range. Therefore, we expected to observe high levels of inbreeding in the Lake Superior range and a high degree of differentiation with caribou from the continuous range. We expected to detect lower levels of inbreeding in boreal caribou from the continuous range of Ontario and Manitoba, as well as in the eastern migratory caribou; herds that have experienced recent declines but are not as small and isolated as the Lake Superior range (COSEWIC 2014b, 2017a). Further, we predicted barren-ground caribou from the Qamanirijuaq population ranging over northern Manitoba and Nunavut (Figure 2) would have the lowest inbreeding estimates, as they occur in large populations that have not experienced dramatic historical or recent declines (COSEWIC 2016). The lengths of genomic regions produced by inbreeding, called Runs of Homozygosity (ROH), indicate how recently inbreeding occurred, as continuous stretches of ROH are broken up during successive mating events (Ceballos *et al*. 2018). Thus, we predicted we would find the longest ROH in caribou from the Lake Superior range, reflecting recent inbreeding caused by anthropogenic range contraction

(Schaefer, 2003), but we may also find short ROH, representing historical inbreeding events, where long ROH have been broken up through mutation and recombination (Ceballos *et al.* 2018).

2| METHODS

2.1 | Caribou sampling

We sampled caribou from herds that differed in evolutionary history, demographic history, and extent of isolation. Broadly, caribou in North America can be divided into two lineages: the North American Lineage (NAL) which encompasses boreal and eastern migratory caribou (*R.t. caribou*), and the Berigan Lineage (BEL), represented in this study by barren-ground caribou (*R. t. groenlandicus*) (Klutsch *et al.* 2012; Polfus *et al*. 2017; Taylor *et al.* 2020). Boreal caribou samples (muscle, hide, hair, fecal pellet, shed antler; Table S1) were collected from the southern caribou range of Ontario by provincial biologists and sequenced for the study and can be retrieved from the National Centre for Biotechnology (NCBI) under the BioProject Accession no. PRJNA 984705. We also included previously sequenced whole genome raw reads (Taylor *et al*. 2020; BioProject Accession no. PRJNA 634908).

We included seven samples from the Lake Superior range in Ontario (Table 1): two samples from Michipicoten Island (LS39650, LS39651), two from the Slate Islands (LS21681, LS45994), one from the mainland area near Neys Provincial Park (LS39590), one from Pic Island of Neys Provincial Park (LS22426), and one from Pukaskwa National Park (LS39653). Over the past four decades, caribou herds along the coast and islands of Lake Superior have steadily declined and become increasingly isolated from the continuous caribou range of Ontario (Patterson *et al*., 2014, Shuter, Asselin, & Rodgers, 2016, Ontario Ministry of Natural Resources

and Forestry, 2018). There have been no caribou observed in the coastal Pukaskwa National Park in recent years, although some caribou have managed to persist on small islands. The island populations were founded by very few individuals, but in the absence of predation increased to high densities prior to recent declines. For instance, Michipicoten Island was founded by a single resident bull plus 8 caribou that were relocated from the Slate Islands in 1982-1989, and subsequently grew to an estimated population of 680 caribou by 2010 (Ontario Ministry of Natural Resources and Forestry 2018). However, the population quickly collapsed when predation pressure was introduced by wolves who immigrated to the island via an ice bridge (in 2014), prompting a relocation of some of the few remaining caribou to the Slate Islands in early 2018 (Ontario Ministry of Natural Resources and Forestry 2018). The Slate Islands once had the highest density caribou population in North America (Bergerud, Dalton, Butler, Camps, & Ferguson, 2007); however, over the past decade the population had also collapsed and was functionally extirpated at the time of relocation (i.e., there appeared to be only two resident bulls remaining). Our study includes two samples from Michipicoten Island, collected shortly after the population began to decline due to new predation pressure (2015, 2016). From the Slate Islands, we included one sample collected prior to the recent population declines (2009), and another collected shortly before caribou were relocated from Michipicoten Island (2017).

We also selected seven samples from the continuous boreal caribou range in Ontario (BO21401, BO22832, BO39654, BO45932, BO45933) and Manitoba (BO35324, BO35326). Within the eastern migratory ecotype, we included two samples from the George River herd (EM27689, EM27694) and two from the Pen Islands herd (EM20917, EM34590). The George River herd has experienced a dramatic population decline over recent decades from approximately 823 000 individuals in 1993 (Couturier *et al.* 1996), to approximately 8 900

individuals in 2016 (Gagnon *et al.* 2019); the samples included in this study were obtained in 2008 after the population had already begun to decline. The Pen Islands herd in northern Ontario was estimated to contain 16 638 individuals in 2011 (COSEWIC 2017a). Notably, the George River and Pen Islands herds are geographically isolated from each other (Figure 2) and recent research has revealed a divergent evolutionary history between these two populations *(Taylor et al.* 2020). We also included two barren-ground caribou samples from the Qamanirijuaq herd (BG21332, BG21350), a large population (estimated to contain 264 661 individuals in 2014) that has not experienced dramatic historical or recent declines (COSEWIC 2016).

2.2 | Genome sequencing, assembly, and quality control

DNA was extracted using the Qiagen DNeasy kit, following the manufacturer's protocols (Qiagen, Hilden, Germany),and quantified using a Qubit system (Thermo Fisher Scientific, MA, USA). The extracted DNA was then sent to The Centre for Applied Genomics (TCAG), at The Hospital for Sick Children (Toronto, ON). The samples were prepared with an Illumina library prep kit (Illumina, San Diego, CA, USA) with an insert size of 350bp and sequenced on the Illumina HiSeqX platform, yielding paired-end 150bp sequence reads. The raw sequence reads are available through the National Centre for Biotechnology (NCBI) BioProject Accession numbers PRJNA 634908 and PRJNA 984705 (Table S1).

We conducted all bioinformatic analyses using cloud computing resources from Compute Canada (RRG gme-665-ab) and Amazon Web Services (https://aws.amazon.com/). First, we removed sequencing adapters and low-quality bases (phred score <30) with Trimmomatic v0.38 (Bolger *et al.* 2014). We mapped the trimmed reads to a chromosome-level caribou reference

genome (Taylor *et al.* 2022) which has a N50 score of 64.42 Mb using Bowtie2 v2.3.0 (Langmead and Salzberg 2012).

We used Samtools v1.5 (Li *et al.* 2009) to convert the SAM files to BAM files and to sort the BAM files. We then removed duplicate reads and added read group information to each BAM file with Picard v2.17.3 (Broad Institute). We used Sambamba v0.8.0 (Streit *et al.* 2013) to retain only primary alignments and BamUtil v1.0.14 (https://github.com/statgen/bamUtil) to clip overlapping regions. We used Samtools to remove bases with a mapping quality (q) lower than 20 and index the BAM files. We checked the quality of each BAM file using FastQC v0.11.8 (Andrews 2010). Finally, we used Samtools to produce alignment statistics (flagstat) and to calculate the depth of coverage across each genome.

We used the GATK v4.0.2 (McKenna *et al.* 2010) Haplotype Caller to produce variant call format (GVCF) files for each caribou. We then used CombineGVCFs and GenotypeGVCFs in GATK to combine and genotype the GVCFs, producing grouped VCF files. We used VCFtools v0.1.14 (Danecek *et al.,* 2011) to select scaffolds and perform filtering. Although the reference genome used in this study does not have a sex chromosome characterized, several regions on Scaffold 36 had genes linked with sex chromosomes (Liu *et al.* 2019). Thus, we selected the 35 largest scaffolds (representing >99% of the genome) to focus our analyses on large autosomes and performed additional filtering: we removed sites with a depth <2 or >60, indels, non bi-allelic sites, low-quality genotype calls (GQ <20), and genotypes with more than 50% missing data (henceforth: filtered VCF). Finally, we produced a more strictly filtered version that contained no missing data (henceforth: strictly filtered VCF).

We attempted to retain as many informative sites as possible, as strict loci filtering can lead to irresolute conclusions and bioinformatics tools are becoming reliable when performing

under randomly distributed missing data (Huang and Lacey Knowles 2016; Hodel *et al.* 2017). Minor allele frequency (MAF) and Hardy-Weinberg equilibrium (HWE) estimates are conventionally calculated to identify putative sequencing artifacts (Chang 2020); however, excluding loci based on these parameters can lead to allelic dropout. Both parameters are highly dependent on sampling size and the population of origin, and can represent true evolutionary signals (Pearman *et al*. 2022). Outlier and linkage disequilibrium (LD) scans search for loci with allelic frequencies out of neutral, and therefore random, expectations. Thinning for LD is likely to exclude many diagnostic markers, decreasing the power of the analyses, including ROH identification (Meyermans *et al.* 2020a). Given divergent population histories and the small number of samples representing each population, we did not filter our data for MAF, HWE, or LD; however, we attempted to account for LD in our population history analyses as described below.

2.3 | Genomic population structure

We explored population structure using the two filtered VCF files. We used Atlas (Link *et al.* 2017) to convert the filtered VCF file to a Beagle file for NGSAdmix (Skotte *et al.* 2013). We then used NGSAdmix v32 to explore population groupings among individuals (K=2 - 9). We conducted 10 arrays at each K value and then used R to plot the outputs (Figure 2) and compare the log likelihood values across runs to select the best supported number of populations (K). Specifically, we used the Cluster Markov Packager Across K from Evanno (Evanno *et al.* 2005) via an R script provided by Bay *et al.* (2021) to select the best K value by dividing the mean log likelihood of each K by the standard deviation (Table S2).

We conducted Principal Component Analyses (PCA; Figure 3A) in R v4.0.2 (R Core Team 2023) using the strictly filtered VCF containing no missing data. We used Stacks v2.60 (Catchen 2013) to convert the filtered VCF file to input for Treemix v1.1. To account for possible linkage, we performed analyses with different sized groupings of SNPs (k=500, 1000, 2000). We created evolutionary trees (Figure 3) with and without migration events (m=0 - 7). We performed 10 arrays for each parameter and plotted the outputs in R. We then used the OptM package in R (Fitak 2021) to select the migration model with the most support.

2.4 | Genomic diversity & inbreeding

We calculated individual inbreeding coefficients (F) based on observed and expected heterozygosity using VCFtools (Danecek *et al*. 2011) with the strictly filtered VCF file (Table S3). We also quantified inbreeding as the amount of genome in Runs of Homozygosity (ROH) using PLINK v1.90b4.6 (Chang *et al.* 2015) and ROHan (Renaud *et al*. 2019). PLINK examines SNP data using a window-based observational approach to identify ROH segments, which are homozygous genomic regions where an individual has received the same copy of an allele from both parents due to inbreeding (Meyermans *et al*. 2020a). Conversely, ROHan combines a local Bayesian model and Hidden Markov Model (HMM) to identify ROH from individual mapped genomes (Renaud *et al.* 2019).

We assessed the robustness of our results by examining multiple parameters with two size categories and different rates of the number of heterozygous sites allowed under both ROH methods. For all PLINK analyses we used the strictly filtered VCF file and did not filter for MAF nor LD following recommendations from Meyermans *et al*. (2020). We selected parameters based on similar investigations of non-model chromosome-level genome assemblies

(e.g. Duntsch, Whibley, Brekke, Ewen, & Santure, 2021; Lavanchy & Goudet, 2023; Martin *et al.*, 2023; von Seth *et al.*, 2021). Specifically, we applied the following "strict" parameters: homozyg-window-snp 100, homozyg-window-het 1, homozyg-het 5, homozyg-gap 200, homozyg-density 50, homozyg-snp 100. We also conducted PLINK analyses with more "relaxed" parameters: homozyg-window-snp 50, homozyg-window-het 2, homozyg-het 10, homozyg-gap 1000, homozyg-density 100, homozyg-snp 50. Both sets of parameters included homozyg-window-threshold 0.05. Following the same approach as Martin *et al.* (2023), we applied these parameters under two ROH size categories by using homozyg-kb to identify ROH >250 kb and >1Mb in length. F_{ROH} was then calculated for each individual as the total length of ROH divided by the length of the 35 chromosomes examined (Figure S5).

For ROHan analysis, we used the unfiltered BAM files as the program takes base quality into account (Renaud *et al.* 2019). Under the relatively "strict" parameters, we allowed $5x10^{-5}$ heterozygous sites within ROH (--rohmu 5e-5), and the more "relaxed" parameters allowed $5x10^{-4}$ heterozygous sites within ROH. For all ROHan analyses, we specified a transition/transversion ratio of 2.09 based on a calculation from the strictly filtered VCF file (--tstv 2.09). Similar to our PLINK approach, we conducted analyses under both sets of parameters with 250kb (--size 250000) and 1Mb windows. The percent of genome in ROH reported by ROHan was converted to a proportion to represent F_{ROH} (Figure S6).

Finally, to measure genetic diversity, we calculated genome-wide heterozygosity (Watterson's θ), producing two estimates for each individual: heterozygosity across all genomic regions and excluding regions in ROH in ROHan (Renaud *et al.* 2019). For this analysis we used 250kb windows with a heterozygosity rate of $5x10^{-4}$ tolerated within ROH.

3 | RESULTS

3.1 | Genome quality control

All individual BAM files passed quality control with FastQC and Samtools flagstat. The final filtered BAM files had an average depth of 29x (Table 1). We created two versions of the grouped VCF to retain as many informative sites as possible: the filtered version contained 23 859 411 SNPs; and the strictly filtered VCF contained 9 338 805 SNPs.

3.2 | Population structure

We first explored genomic structure among populations of caribou ecotypes. We investigated population groupings with NGSAdmix and found the best supported model was K=2 (Figure 2), which had the highest log likelihood value and 100% convergence across runs (Table S2). We found barren-ground, eastern migratory, and boreal caribou from the continuous range were assigned to the first cluster, although the eastern migratory and boreal samples from Ontario also shared a small proportion of assignment to the second cluster (Figure 2). The Lake Superior caribou were mostly assigned to the second cluster, except for the individual from Pukaskwa National Park which was split between the two groups (53%). The next best supported model was K=3 (Figure S1), which also indicated the Lake Superior caribou cluster together, with Pukaskwa National Park and one other sample showing mixed assignment.

The Principal Component Analysis (PCA) revealed genomic groupings among samples (Figure 3A, Figure S3). We retained 19 Principal Components (PCs); PC1 and PC2 (Figure 3A) collectively explained 20% of the cumulative variance (Figure S2). We plotted comparisons of PCs 1-4 (Figure S3). Our results distinguished the barren-ground caribou from the other ecotypes present in the study, whereas eastern migratory caribou and boreal caribou from the

continuous range grouped together (Figure 3A, Figure S3). The Lake Superior caribou largely grouped together, with the exception of the sample from Pukaskwa National Park. The Treemix results were consistent throughout iterations across the regardless of the SNP grouping size (k-value), and revealed the Lake Superior caribou on a branch together, with Pukaskwa National Park representing an older branch based on its basal position, which has experienced considerable drift as indicated by the drift parameter (Figure 3). The best supported Treemix model across all k-values indicated 3 migrations (Figure S4); notably, all migrations originated from basal placements in the tree (rather than branch tips), which indicates the migration occurred historically or from a closely related unsampled population (Decker *et al.* 2014). We detected migration from Pukaskwa National Park to Michipicoten Island, from the nearby boreal continuous range (Nipigon) to Pic Island, and from barren-ground into eastern migratory caribou from the George River herd (Figure 3C).

3.3 | Genomic diversity & inbreeding

We estimated inbreeding as the inbreeding coefficient (F) and the proportion of the genome in ROH (F_{ROH}; Table S3). We identified ROH using two methods, under two sets of parameters, and at different size scales to identify shorter ROH associated with historical inbreeding, and longer ROH indicating recent inbreeding. Not surprisingly, we found fewer, but longer ROH when analyses are restricted to a larger size class with both methods (Figure S5, Figures S6). For instance, when PLINK is restricted to ROH >1Mb, we detect zero ROH in several individuals, even under relaxed parameters (Figure S5). Under the strict ROHan parameters, we detected little to no ROH in any individual (Table S3, Figure S6), underscoring the importance of examining results under multiple methods and parameters.

Across the methods and parameters we explored, barren-ground caribou consistently had the lowest inbreeding levels, and the highest inbreeding estimates were observed in caribou from the Lake Superior range (Table S3, Figure S5, Figure S6). These caribou had the highest inbreeding coefficents (F), the largest proportion of the genome in ROH (F_{ROH}), and the ROH were notably longer, indicating more recent inbreeding (Figure 4). We found an abundance of ROH in the Lake Superior caribou under both size classes (250kb and 1Mb; Figure S5, Figure S6), suggesting both recent and historical inbreeding has occurred. Notably, one individual from Pic Island had lower ROH estimates than the other Lake Superior caribou, reflecting values similar to the boreal caribou from the continuous range (Figure S5, Figure S6). This may be the result of low levels of gene flow with the continuous range (Figure 2), low levels of drift in comparison to other Lake Superior caribou (Figure 3), or historically lower inbreeding levels as the sample was collected prior to 2008 (Table 1). All methods corroborated that the inbreeding levels in eastern migratory and boreal caribou from the continous range are higher than those observed in barren-ground and lower than Lake Superior with little variation among individuals.

The lowest genomic diversity estimates, calculated as genome-wide heterozygosity, were observed in caribou from Lake Superior; however, some caribou from the Lake Superior range had relatively high heterozygosity estimates, with values similar to those observed in the continuous boreal range (Figure 5). Across all samples, most individuals showed no difference in heterozygosity inside and outside ROH, which is not surprising as many caribou had only a small amount of ROH identified. However, the Lake Superior caribou showed notably higher heterozygosity outside of ROH (Figure 5).

4 | DISCUSSION

We sampled caribou from the trailing edge of Ontario's caribou range, and as predicted, these caribou exhibited high levels of inbreeding relative to caribou in the continuous range farther north (e.g., Hampe and Petit 2005). The Lake Superior caribou also exhibited evidence of differentiation from caribou in the continuous range, and low within population genomic diversity. The Lake Superior range contains Ontario's southernmost caribou populations, which have become small and isolated from other caribou in Ontario through anthropogenic range contraction (Schaefer, 2003; Vors, Schaefer, Pond, Rodgers, & Patterson, 2007). We found an abundance of short (>250kb) and long (>1Mb) ROH, indicating both historical and recent inbreeding has occurred (Figure S7). All of the other caribou populations investigated had comparatively low levels of inbreeding (Figure 4) regardless of evolutionary origins (NAL or BEL lineage; Klutsch, Manseau, & Wilson, 2012; Polfus *et al*., 2017).

Broadly, the population groupings revealed by our analyses did not clearly reflect current management designations (DUs), concordant with other recent genomic studies (Taylor *et al*., 2022). Barren-ground caribou are distinct from the other populations sampled in this study (Figure 3), which was predicted as they are the only samples from the Beringian-Eurasian Lineage (BEL) included (COSEWIC 2016). However, within the North American Lineage (NAL; Polfus *et al.* 2017), there is little distinction between eastern migratory and boreal caribou based on variation across the whole genome (Figure 3A). Our results indicate that caribou from the Lake Superior range group together but not closely with other boreal caribou from the continuous range (Figure 2, Figure 3A). We observed some evidence that this differentiation is due to the isolation of Lake Superior caribou from the continuous range, but further research should explore the importance of local adaptation in these island caribou.

4.1 | Population structure

Our evolutionary tree revealed that caribou in the Lake Superior range form a consistent group that branches from the nearby continuous boreal range, with Pukaskwa National Park representing a basal branch (Figure 3). We detected evidence of gene flow from the continuous boreal range to the Lake Superior range, confirming a previous study that suggested remnant genetic connectivity between Lake Superior and the continuous range (Drake *et al*. 2018). We also detected weak historical gene flow from barren-ground caribou into eastern migratory caribou from George River, which is not surprising as previous research has indicated that the eastern migratory ecotype was formed by historical introgression between barren-ground and boreal caribou (Klütsch *et al.* 2016).

The PCA and Treemix analyses revealed barren-ground caribou are distinct from the other populations sampled in this study (Figure 3); however, we found little distinction between eastern migratory and boreal caribou. These results are consistent with other genomic research that revealed eastern migratory and boreal caribou from NAL cannot be divided into monophyletic lineages (Taylor *et al.* 2022). Further, when the samples are assigned to two population clusters (Figure 2), barren-ground, eastern migratory, and boreal caribou from the continuous range group together, which is likely due to historical introgression among these ecotypes (Taylor *et al*. 2020).

Our results revealed low differentiation within the Lake Superior range but high levels of differentiation from the continuous range, which is predicted for rear edge populations (Hampe and Petit 2005). In fact, we observed a greater distinction between the Lake Superior range and continuous range boreal caribou than we did between the boreal and eastern migratory ecotypes. We did not find evidence of high regional diversity among caribou within the Lake Superior

range, which can occur in isolated trailing edge populations due to high levels of genetic drift and a lack of gene flow among patches (Hampe and Petit 2005). Increased sampling of other patches along the trailing edge is likely required to investigate divergent drift patterns. However, our results suggest some connectivity exists, or recently existed, among Lake Superior islands and coastal regions. Previous research has also suggested connectivity exists within the Lake Superior range, as a caribou radio-collared on the Slate Islands traveled to Pukaskwa National Park, following the near-shore past other sites included within our study, near Neys Provincal Park and Pic Island (Bergerud, 1985; Bergerud *et al.*, 2007). The previous studies suggested that in the past these caribou have made long movements but always stayed near the Lake Superior shore (Bergerud, 1985; Bergerud *et al.*, 2007), which is supported by our findings demonstrating low differentiation within the Lake Superior range with high differentation from the continuous range.

The boreal and eastern migratory ecotypes both originate from the NAL and share extensive areas of habitat overlap, particularly in winter (COSEWIC, 2017b). However, these ecotypes are managed as distinct DUs based on differences in behaviour and life history strategies: eastern migratory caribou aggregate on the tundra during calving and are the only group of NAL caribou to migrate (COSEWIC, 2017b). Conversely, boreal caribou remain within the forest year-round and avoid conspecifics during calving, instead relying on dense woods to avoid predation (COSEWIC, 2014a).

Interestingly, the long-term persistence of caribou in the Lake Superior range is partially attributed to their calving strategy: instead of using the typical strategy of boreal caribou, who avoid wolf predation by using dense woodlots to space out from conspecifics when calving, the Lake Superior caribou use the shoreline and nearby islands to escape predation (Bergerud, 1985;

Bergerud, McLaren, Krysl, Wade, & Wyett, 2014). Another factor encouraging caribou persistence in this range is the presence of protected areas (e.g. Pukaskwa National Park, Neys PP) with low levels of anthropogenic disturbance (Schaefer, 2003). The islands and protected coastal areas may provide refugia from the negative impacts of human encroachment. As negative human impacts spread, areas along the range periphery and remote islands are less impacted by anthropogenic disturbances and thus, represent patches where persistence is more likely than it is in the core range, providing valuable opportunities for conservation (Channell and Lomolino 2000). Notably, the features demonstrated by the Lake Superior caribou, such as small population sizes, isolation from the core range, and associations with distinct habitat features can encourage local adaptation (Hampe and Petit 2005). Peripheral populations face more diverse environmental conditions than central populations and therefore may be more likely to be pre-adapted to anthropogenic disturbances or changing environmental conditions (Volis and Kark 1994; Lomolino and Channell 1998; Crandall *et al.* 2000; Cassel-Lundhagen *et al*. 2009).. However, adaptive processes may also be hindered by the high levels of inbreeding and drift experienced by the remaining caribou in the Lake Superior range.

4.2 | Inbreeding histories

We found inbreeding estimates produced across methods varied in magnitude but generally corroborated on inbreeding ranks among individuals. Our data met the minimum requirements of 10x depth for PLINK and 5x depth for ROHan to produce reliable ROH estimates (Renaud *et al*. 2019). As the field of conservation genomics rapidly expands, we urge researchers to ensure their data meets the minimum requirements for inbreeding analyses, as a high density of SNPs is required for accurate ROH identification and reduced genome coverages result in an

underestimation of F_{ROH} (Meyermans *et al.* 2020a; Lavanchy and Goudet 2023). Additionally, after assessing that the data meets the minimum depth and SNP density requirements, we suggest conducting analyses with multiple methods under different parameters to ensure the results are robust. In general, we observed a greater abundance but shorter ROH with PLINK than we did with ROHan (Figure 4). The reporting of shorter ROH could be due to differences in the input data or the underlying models, resulting in PLINK splitting ROH that are continuous under the ROHan model. For instance, PLINK used sliding window observations, whereas ROHan used a HMM approach; PLINK examined high quality variant sites across the genomes, whereas ROHan examined all mapped sites, resulting in more continuous data (Renaud *et al*. 2019). Additionally, the specific PLINK parameters used, such as the SNP density requirements, may bias the results towards shorter ROH.

Caribou from the Lake Superior range consistently had higher inbreeding estimates than the other populations sampled (Figure S5, Figure S6). These small coastal and island populations are relatively isolated and have experienced several bottlenecks (Bergerud *et al.*, 2007; Fletcher, 2017). The sample from Pukaskwa National Park demonstrated high levels of inbreeding, including notably long ROH indicating recent inbreeding. The caribou population in Pukaskwa National Park persisted at low levels for years and currently no caribou remain in the Park; however, one of the last caribou recorded with wildlife cameras in the park had malformed antlers, which was suggested to be evidence of inbreeding depression (Drake *et al.* 2018). The high F_{ROH} values observed in the Lake Superior caribou reflect their overall inbreeding levels, whereas the combination of short and long ROH likely reflects the historical and recent bottlenecks experienced by these populations.

Despite consistently elevated inbreeding levels, we found several caribou from the Lake Superior range maintained relatively high levels of genetic diversity. The lowest diversity levels correspond to caribou from the Slate Islands with high ROH estimates; however, our results indicated relatively high diversity has been maintained outside of ROH (Figure 5). This suggests genetic diversity may be maintained by natural selection in genomic regions where variation is important (Hedrick and Garcia-Dorado 2016; Selli *et al*. 2021).

Given the consistently high inbreeding levels observed in the Lake Superior range compared to the other populations, it is possible that divergent inbreeding histories are further driving the observed genomic differences between populations. However, one of the individuals with the highest inbreeding levels and the highest drift estimate, from Pukaskwa National Park, showed more similarities to the continuous range than the other Lake Superior caribou did; although these results may also suggest the sample from Pukaskwa National Park was more similar to the continuous range than it was to the other Lake Superior samples (Figure 2, Figure 3). If the distinctions between populations were largely driven by inbreeding or drift, we would expect the individual with the highest inbreeding and drift estimates to show the greatest distinction, whereas the Lake Superior samples with comparatively lower inbreeding and drift levels should demonstrate more similarities with the continuous range. Notably, the population structure patterns observed (Figure 2) may be indicative of three different evolutionary scenarios which can be difficult to disentangle (Falush *et al*. 2016; Garcia-Erill and Albrechtsen 2020). Specifically, the patterns demonstrated by the Lake Superior caribou may be due to multiple recent bottlenecks, such as a bottleneck when the Pukaskwa National Park population diverged followed by a subsequent bottleneck when the other Lake Superior caribou diverged. Alternatively, the population assignment observed in the sample from Pukaskwa National Park

may be reflecting recent admixture or ghost admixture from a historical lineage that has been lost or was not sampled (Falush *et al.* 2016; Garcia-Erill and Albrechtsen 2020). Our results may be affected by uneven sampling (Puechmaille 2016), especially as we are using a single sample from some locations. Future research should strive to sequence additional genomes to allow for more even sampling design; although, this may be challenging for some regions where caribou are now locally extinct (e.g., Pukaskwa National Park).

Significant efforts have been invested in the continued persistence of the Lake Superior caribou populations, including multiple relocations between islands (Bergerud *et al*., 2007; Ontario Ministry of Natural Resources and Forestry, 2018). Given the small number of caribou remaining and high degree of inbreeding, we recommend that future management decisions take inbreeding into consideration. Understanding individual inbreeding levels may be especially important in the context of relocations, and should be considered and monitored when re-establishing or supplementing populations (Scott *et al.* 2020). Thus, we are further investigating the level of inbreeding using a larger sample size from the different populations with a focus on caribou that have been recently relocated following rapid declines.

4.3 | Conclusions

We used high coverage whole genomes to delineate population structure and inbreeding histories in caribou from populations representing divergent evolutionary histories, differing in population size and extent of isolation. We found eastern migratory caribou and boreal caribou from the continuous range broadly cluster together under population genomic models. We found caribou from the Lake Superior range form a distinct group, however we also detected evidence of gene flow between Lake Superior and the continuous range of boreal caribou. Specifically, we

identified a nearby population in the continuous range with evidence of shared ancestry and historical gene flow to the Lake Superior range, which could be used to inform future management if restoring connectivity between the two ranges is a priority (Armstrong *et al.* 2010), and deemed appropriate given the potential for local adaptation.

We found the lowest levels of inbreeding in barren-ground caribou and relatively low inbreeding estimates in eastern migratory and boreal caribou from the continuous range. We observed consistently elevated inbreeding estimates in the Lake Superior populations, which have experienced historical bottlenecks, recent declines, and become increasingly isolated due to recent range contraction (Bergerud *et al*., 2007; Bergerud *et al.*, 2014; Schaefer, 2003). We observed an abundance of both long and short ROH in these isolated populations, confirming both historical and recent inbreeding has occurred. Given the results of our study, the high levels of inbreeding in the Lake Superior caribou may be further driving the observed distinctions between populations. To determine the significance of the observed population structure, future research should attempt to investigate local adaptation.

REFERENCES

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

Armstrong, T. E. R. et al. (2010) 'Caribou conservation and recovery in Ontario: development and implementation of the Caribou Conservation Plan', Rangifer, (20), pp. 145–147. Available at: http://www.unpri.org/signatories/signatories/.

Bay, R. et al. (2021) Marine Genomics, UC Davis Department of Evolution and Ecology. Available at: https://baylab.github.io/MarineGenomics/week-9--population-structure-using-ngsadmix.html#how-do-we-know-which-k-to-pick (Accessed: 31 May 2023).

Bergerud, A. T. (1985) 'Antipredator strategies of caribou: dispersion along shorelines', Canadian Journal of Zoology, 63(Dauphine 1981), pp. 1324–1329. doi: 10.1139/z84-229.

Bergerud, A. T. (2001) The Herbivores of the Slate Islands, 1974 to 2001. Terrace Bay ON.

Bergerud, A. T. et al. (2007) 'Woodland caribou persistence and extirpation in relic populations on Lake Superior', Rangifer, 27(17), pp. 57–78.

Bergerud, A. T. et al. (2014) 'Losing the Predator—Prey Space Race Leads to Extirpation of Woodland Caribou from Pukaskwa National Park', https://doi-org.proxy1.lib.trentu.ca/10.2980/21-(3-4)-3700. Centre d'études nordiques, Université Laval, 21(3–4), pp. 374–386. doi: 10.2980/21-(3-4)-3700.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', Bioinformatics. Oxford University Press, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Broad Institute (no date) 'Picard'. Available at: http://broadinstitute.github.io/picard.

Carr, N. L. et al. (2012) 'Comparative woodland caribou population surveys in Slate Islands Provincial', Rangifer, pp. 205–218.

Cassel-Lundhagen A, Tammaru T, Windig JJ, et al (2009) Are peripheral populations special? Congruent patterns in two butterfly species. Ecography (Cop) 32:591–600. https://doi.org/10.1111/j.1600-0587.2008.05685.x

Castric, V. and Bernatchez, L. (2003) 'The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchill)', Genetics, 163(3), pp. 983–996. doi: 10.1093/genetics/163.3.983.

Catchen, J. M. (2013) 'Stacks: an analysis tool set for population genomics', Molecular ecology, 22(11), pp. 3124–3140. doi: 10.1111/mec.12354.Stacks.

Ceballos, F. C. et al. (2018) 'Runs of homozygosity: Windows into population history and trait architecture', Nature Reviews Genetics. Nature Publishing Group, 19(4), pp. 220–234. doi: 10.1038/nrg.2017.109.

Chang, C. C. et al. (2015) 'Second-generation PLINK: Rising to the challenge of larger and richer datasets', GigaScience. BioMed Central Ltd., 4(1). doi: 10.1186/S13742-015-0047-8.

Chang, C. C. (2020) 'Data management and summary statistics with PLINK', Methods in Molecular Biology, 2090, pp. 49–65. doi: 10.1007/978-1-0716-0199-0_3.

Channell, R. and Lomolino, M. V. (2000) 'Dynamic biogeography and conservation of endangered species', Nature, 403(January), pp. 84–86.

COSEWIC (2011) 'Designatable units for caribou (*Rangifer tarandus*) in Canada', p. 88pp. Available at: http://www.cosewic.gc.ca/eng/sct12/COSEWIC_Caribou_DU_Report_23Dec2011.pdf.

COSEWIC (2014a) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Newfoundland population, Atlantic-Gaspésie population and Boreal population, in Canada. Available at: www.registrelep-sararegistry.gc.ca/default_e.cfm.

COSEWIC (2014b) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Northern Mountain Population, Central Mountain Population, and Southern Mountain Population in Canada. Available at: http://www.sararegistry.gc.ca/virtual_sara/files/cosewic/sr_Caribou_Northern_Central_Southern_2014_e.pdf.

COSEWIC (2015) COSEWIC assessment and status report on the Peary caribou *Rangifer tarandus pearyi*, in Canada.

COSEWIC (2016) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Barren-Ground population, in Canada. Available at: http://www.registrelep-sararegistry.gc.ca/default.asp?lang=en&n=24F7211B-1.

COSEWIC (2017a) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Dolphin and Union population, in Canada. Available at: https://wildlife-species.canada.ca/species-risk-registry/virtual_sara/files/cosewic/srCaribouDolphinUnion2017e1.pdf.

COSEWIC (2017b) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Eastern Migratory population and Torngat Mountains population, in Canada. Available at: http://www.registrelep-sararegistry.gc.ca/default.asp?lang=en&n=24F7211B-1.

Couturier, S. et al. (1996) 'Calving photocensus of the Rivière George Caribou Herd and comparison with an independent census', Rangifer, 16(4), p. 283. doi: 10.7557/2.16.4.1268.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. Trends Ecol Evol 15:290–295. https://doi.org/10.1016/S0169-5347(00)01876-0

Dallaire, X. and Mérot, C. (2022) 'A setback into a success: What can batch effects tell us about best practices in genomics?', Molecular Ecology Resources. Wiley. doi: 10.1111/1755-0998.13615.

Danecek, P. et al. (2011) 'The variant call format and VCFtools', Bioinformatics, 27(15), pp. 2156–2158. doi: 10.1093/bioinformatics/btr330.

Decker, J. E. et al. (2014) 'Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle', PLoS Genetics, 10(3). doi: 10.1371/journal.pgen.1004254.

Drake, C. C. et al. (2018) 'Does connectivity exist for remnant boreal caribou (*Rangifer tarandus caribou*) along the Lake Superior Coastal Range? Options for landscape restoration', Rangifer, 38(1), p. 13. doi: 10.7557/2.38.1.4124.

Duntsch, L. et al. (2021) 'Genomic data of different resolutions reveal consistent inbreeding estimates but contrasting homozygosity landscapes for the threatened Aotearoa New Zealand hihi', Molecular Ecology, (July), pp. 1–15. doi: 10.1111/mec.16068.

Evanno, G., Regnaut, S. and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study', Molecular Ecology, 14(8), pp. 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x.

Falush, D., Dorp, L. van and Lawson, D. (2016) 'A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots', bioRxiv, p. 066431. doi: 10.1101/066431.

Festa-Bianchet, M. et al. (2011) 'Conservation of caribou (*Rangifer tarandus*) in Canada: an uncertain future', Canadian Journal of Zoology, 89(5), pp. 419–434. doi: 10.1139/z11-025.

Fitak, R. R. (2021) 'OptM: Estimating the optimal number of migration edges on population trees using Treemix', Biology Methods and Protocols, 6(1), pp. 1–6. doi: 10.1093/biomethods/bpab017.

Fletcher, R. (2017) 'Wolf, caribou issue on Michipicoten Island raises red flags for many', Sault Star, 14 November. Available at: http://www.saultstar.com/2017/11/13/wolf-caribou-issue-on-michipicoten-island-raises-red-flags-for-many.

Gagnon, M. et al. (2019) 'No evidence of inbreeding depression in fast declining herds of migratory caribou', Journal of Evolutionary Biology, (August), pp. 1–14. doi: 10.1111/jeb.13533.

Garcia-Erill, G. and Albrechtsen, A. (2020) 'Evaluation of model fit of inferred admixture proportions', Molecular Ecology Resources, 20(4), pp. 936–949. doi: 10.1111/1755-0998.13171.

Gilpin, M. E. and Soule, M.E. (1986) 'Minimum viable populations: processes of extinction', in Soule, Me.E (ed.) Conservation Biology: The Science of Scarcity and Diversity. Sunderland, MA: Sinauer Associates, pp. 19–34.

Gunn, A. (2016) *Rangifer tarandus*. The IUCN Red List of Threatened Species. doi: https://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T29742A22167140.en.

Hampe, A. and Petit, R. J. (2005) 'Conserving biodiversity under climate change: The rear edge matters', Ecology Letters, 8(5), pp. 461–467. doi: 10.1111/j.1461-0248.2005.00739.x.

Hedrick, P. W. and Garcia-Dorado, A. (2016) 'Understanding inbreeding depression, purging, and genetic rescue', Trends in Ecology and Evolution. Elsevier Ltd, 31(12), pp. 940–952. doi: 10.1016/j.tree.2016.09.005.

Hodel, R. G. J. et al. (2017) 'Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: Comparing microsatellites and RAD-Seq and investigating loci filtering', Scientific Reports. Springer US, 7(1), pp. 1–14. doi: 10.1038/s41598-017-16810-7.

Huang, H. and Lacey Knowles, L. (2016) 'Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of rad sequences', Systematic Biology, 65(3), pp. 357–365. doi: 10.1093/sysbio/syu046.

Klütsch, C. F. C. et al. (2016) 'The eastern migratory caribou: the role of genetic introgression in ecotype evolution', Royal Society Open Science, 3, pp. 1–13. doi: 10.1098/rsos.150469.

Klutsch, C. F. C., Manseau, M. and 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(12), pp. 1–10. doi: 10.1371/journal.pone.0052661.

Laliberte, A. S. and Ripple, W. J. (2004) 'Range contractions of North American carnivores and ungulates', BioScience, 54(2), p. 123. doi: 10.1641/0006-3568(2004)054[0123:RCONAC]2.0.CO;2.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', Nature Methods, 9, pp. 357–359.

Lavanchy, E. and Goudet, J. (2023) 'Effect of reduced genomic representation on using runs of homozygosity for inbreeding characterization', Molecular Ecology, 00(December 2022), pp. 1–16. doi: 10.1111/1755-0998.13755.

Li, H. et al. (2009) 'The Sequence Alignment/Map format and SAMtools', Bioinformatics, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Link, V. et al. (2017) 'ATLAS: Analysis Tools for Low-depth and Ancient Samples', BioRxiv, 33(16), pp. 1–7.

Liu, R. et al. (2019) 'New insights into mammalian sex chromosome structure and evolution using high-quality sequences from bovine X and Y chromosomes', BMC Genomics. BMC Genomics, 20(1), pp. 1–11. doi: 10.1186/s12864-019-6364-z.

Lomolino, M. V. and Channell, R. (1998) ‘Range Collapse, Re-Introductions, and Biogeographic Guidelines for Conservation’, Conservation Biology, 12(2), pp. 481–484. doi: 24.77.130.166.

Lou, R. N. and Therkildsen, N. O. (2022) ‘Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation’, Molecular Ecology Resources. John Wiley & Sons, Ltd, 22(5), pp. 1678–1692. doi: 10.1111/1755-0998.13559.

Martin, C. et al. (2023) ‘Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging island populations of a bird’, Molecular Ecology, pp. 1–10. doi: 10.1111/mec.16865.

McKenna, A. et al. (2010) ‘The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data’, Genome Research, 20, pp. 297–303. doi: 10.1101/gr.107524.110.20.

Meyermans, R. et al. (2020) ‘How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species’, BMC Genomics. BioMed Central Ltd., 21(1), pp. 1–14. doi: 10.1186/s12864-020-6463-x.

Ontario Ministry of Natural Resources (2009) ‘Ontario’s woodland caribou conservation plan’, p. 28pp.

Ontario Ministry of Natural Resources and Forestry (2018) Seeking advice on the future of caribou in the Lake Superior coast range. Available at: https://ero.ontario.ca/notice/013-2587.

Patterson, L. D. et al. (2014) ‘Detecting a population decline of woodland caribou (*Rangifer tarandus caribou*) from non-standardized monitoring data in Pukaskwa national park, Ontario’, Wildlife Society Bulletin, 38(2), pp. 348–357. doi: 10.1002/wsb.402.

Pearman, W. S., Urban, L. and Alexander, A. (2022) ‘Commonly used Hardy–Weinberg equilibrium filtering schemes impact population structure inferences using RADseq data’, Molecular Ecology Resources, 22(7), pp. 2599–2613. doi: 10.1111/1755-0998.13646.

Pérez-Tris, J. et al. (2004) ‘Historical diversification of migration patterns in a passerine bird’, Evolution, 58(8), pp. 1819–1832. doi: 10.1111/j.0014-3820.2004.tb00464.x.

Polfus, J. L. et al. (2016) Łegha´ gots ’ enete¸ (learning together): the importance of indigenous perspectives in the identification of biological variation, Ecology and Society. doi: 10.5751/ES-08284-210218.

Polfus, J. L. et al. (2017) ‘Ancient diversification in glacial refugia leads to intraspecific diversity in a Holarctic mammal’, Journal of Biogeography, 44(2), pp. 386–396. doi: 10.1111/jbi.12918.

Prentice, M. B. et al. (2017) 'Selection and drift influence genetic differentiation of insular Canada lynx (*Lynx canadensis*) on Newfoundland and Cape Breton Island', Ecology and Evolution, 7(9), pp. 3281–3294. doi: 10.1002/ece3.2945.

Puechmaille, S. J. (2016) 'The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem', Molecular Ecology Resources, 16(3), pp. 608–627. doi: 10.1111/1755-0998.12512.

R Core Team, . (2020) 'R: A language and environment for statistical computing'. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.r-project.org/.

Renaud, G. et al. (2019) 'Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples', Genetics, 212(July), pp. 587–614.

Ricketts, T. H. et al. (2005) 'Pinpointing and preventing imminent extinctions', Proceedings of the National Academy of Sciences of the United States of America, 102(51), pp. 18497–18501. doi: 10.1073/pnas.0509060102.

Schaefer, J. A. (2003) 'Long-term range recession and the persistence of caribou in the taiga', Conservation Biology, 17(5), pp. 1435–1439. doi: 10.1046/j.1523-1739.2003.02288.x.

Scott, P. A. et al. (2020) 'Individual heterozygosity predicts translocation success in threatened desert tortoises', Science, 370(6520), pp. 1086–1089. doi: 10.1126/science.abb0421.

Selli, A. et al. (2021) 'Detection and Visualization of Heterozygosity-Rich Regions and Runs of Homozygosity in Worldwide Sheep Populations', 11, p. 2696. doi: 10.3390/ani11092696.

Volis S, Kark S (1994) Core and Peripheral Populations and Global Climate Change. Isr J Plant Sci 42:331–345. https://doi.org/10.1080/07929978.1994.10676584

von Seth, J. et al. (2021) 'Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations', Nature Communications. Springer US, 12(1), pp. 1–11. doi: 10.1038/s41467-021-22386-8.

Shuter, J., Asselin, N. C. and Rodgers, A. (2016) Results of the 2016 Lake Superior Coast Range (LSCR) caribou (*Rangifer tarandus caribou*) aerial survey. Thunder Bay, Ontario; Pukaskwa National Park, Heron Bay, Ontario.

Skotte, L., Korneliussen, T. S. and Albrechtsen, A. (2013) 'Estimating individual admixture proportions from next generation sequencing data', Genetics, 195(3), pp. 693–702. doi: 10.1534/genetics.113.154138.

Spatz, D. R. et al. (2017) 'Globally threatened vertebrates on islands with invasive species', Science Advances, 3(10). doi: 10.1126/sciadv.1603080.

Streit, K. et al. (2013) 'Sambamba: Runtime adaptive parallel execution', in ADAPT '13: Proceedings of the 3rd International Workshop on Adaptive Self-Tuning Computing

Systems. Berlin, Germany: Association for Computing Machinery, pp. 1–6. doi: 10.1145/2484904.2484911.

Taylor, R. S. et al. (2020) 'The role of introgression and ecotypic parallelism in delineating intra-specific conservation units', Molecular Ecology, (00), pp. 1–17. doi: 10.1111/mec.15522.

Taylor, R. S. et al. (2022) 'Continent-wide phylogenomic framework reveals introgression as a driver of intra-specific diversity and enriched molecular pathways in caribou', bioRxiv. doi: 10.1101/2022.12.19.521038.

Thompson, L. M. et al. (2019) 'Spatial differences in genetic diversity and northward migration suggest genetic erosion along the boreal caribou southern range limit and continued range retraction', Ecology and Evolution, (April), p. ece3.5269. doi: 10.1002/ece3.5269.

Vors, L. S. et al. (2007) 'Woodland Caribou Extirpation and Anthropogenic Landscape Disturbance in Ontario', Journal of Wildlife Management, 71(4), pp. 1249–1256. doi: 10.2193/2006-263.

Tables

Table 1. Information for each caribou included in this study: sample individual reference number, subspecies classification, Canadian Designatable Unit, and approximate sampling location indicating herd or region sampled and the province. Mean depth refers to the average depth of coverage from filtered whole genome BAM files. Inbreeding was quantified as the proportion of the genome in Runs of Homozygosity (F_{ROH}) identified with PLINK and ROHan. Table values reflect ROH measured in kb, identified under the 250 kb size class and more relaxed set of parameters tested for each method.

Sample ID	Subspecies	Designatable Unit	Sample Region[1]	Year	Mean Depth	F_{ROH} (PLINK)	Mean Length ROH (PLINK)	F_{ROH} (ROHan)	Mean Length ROH (ROHan)
BG21332	*R. t. groenlandicus*	Barren-ground	Brochet Junction area, MB	2008	38x	0.003	381	0.001	625
BG21350	*R. t. groenlandicus*	Barren-ground	Brochet Junction area, MB	2007	38x	0.002	372	0.002	1060
EM20917	*R. t. caribou*	Eastern migratory	Fort Severn, ON	NA	36x	0.023	376	0.009	688
EM34590	*R. t. caribou*	Eastern migratory	Pen Islands, ON	1992	37x	0.037	488	0.025	1320
EM27689	*R. t. caribou*	Eastern migratory	George River, NL	2008	38x	0.048	509	0.021	1930
EM27694	*R. t. caribou*	Eastern migratory	George River, NL	2008	39x	0.022	356	0.002	833
BO35324	*R. t. caribou*	Boreal	The Pas, MB	2008	35x	0.071	582	0.058	1170
BO35326	*R. t. caribou*	Boreal	Snow Lake, MB	2009	38x	0.041	561	0.013	1260
BO39654	*R. t. caribou*	Boreal	Cochrane, ON	2009	38x	0.048	531	0.014	1120
BO22832	*R. t. caribou*	Boreal	Hearst, ON	2009	19x	0.030	410	0.021	947
BO21401	*R. t. caribou*	Boreal	Red Lake, ON	2008	10x	0.028	389	0.017	799

BO45932	*R. t. caribou*	Boreal	Nipigon, ON	2011	13x	0.030	448	0.022	1010
BO45933	*R. t. caribou*	Boreal	Nipigon, ON	2012	18x	0.053	556	0.043	1470
LS39653	*R. t. caribou*	Boreal	Pukaskwa National Park, ON	1999	40x	0.420	1005	0.171	2320
LS22426	*R. t. caribou*	Boreal	Pic Island, ON	<2008	10x	0.052	522	0.040	1370
LS39590	*R. t. caribou*	Boreal	Neys area, ON	2011	35x	0.250	884	0.225	3160
LS39650	*R. t. caribou*	Boreal	Michipicoten Island, ON	2015	37x	0.215	800	0.161	2330
LS39651	*R. t. caribou*	Boreal	Michipicoten Island, ON	2016	40x	0.212	861	0.081	1790
LS21681	*R. t. caribou*	Boreal	Slate Islands, ON	2009	10x	0.188	711	0.179	3250
LS45994	*R. t. caribou*	Boreal	Slate Islands, ON	2017	17x	0.251	865	0.248	3810

[1]Abbreviations for Canadian provinces: MB=Manitoba, ON=Ontario, NL=Newfoundland and Labrador.

Figures

FIGURE 1 Caribou (*Rangifer tarandus*) surrounded by early morning mist on Michipicoten Island, Lake Superior, Ontario, Canada. Photo by Andy Silver (Ontario Ministry of Natural Resources and Forestry).

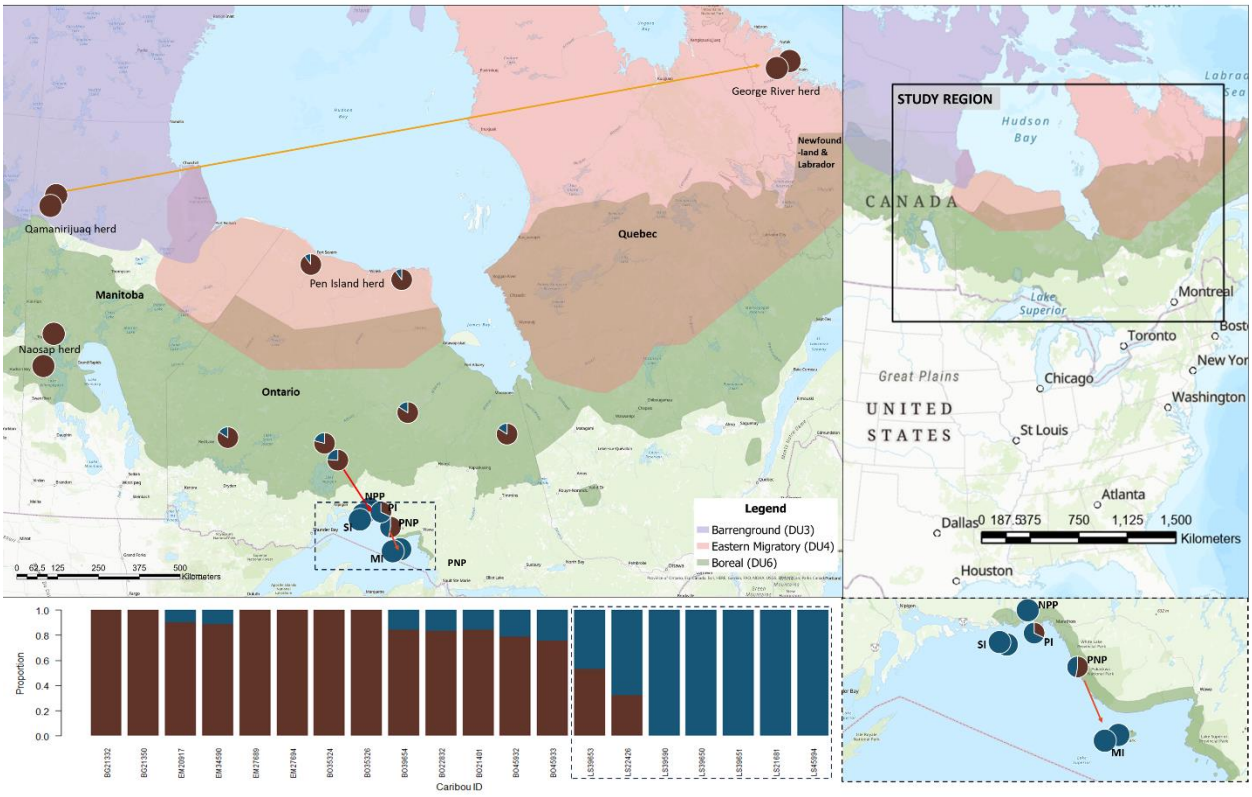

FIGURE 2 Sample sites of caribou (N = 20) in this study. Background colors show the ranges of three of the Canadian Designatable Units (DUs) included: barren- ground, eastern migratory, and boreal. Circles on map indicate sample locations and the colors indicate individual population assignment proportions under the best supported model of K = 2. The arrows on the map indicate migrations modeled by Treemix. The Lake Superior region is indicated by dashed borders; abbreviated labels indicate site names: Pukaskwa National Park (PNP), Neys Provincial Park (NPP), Pic Island (PI), The Slate Islands (SI), and Michipicoten Island (MI). X- axis label indicates individual IDs.

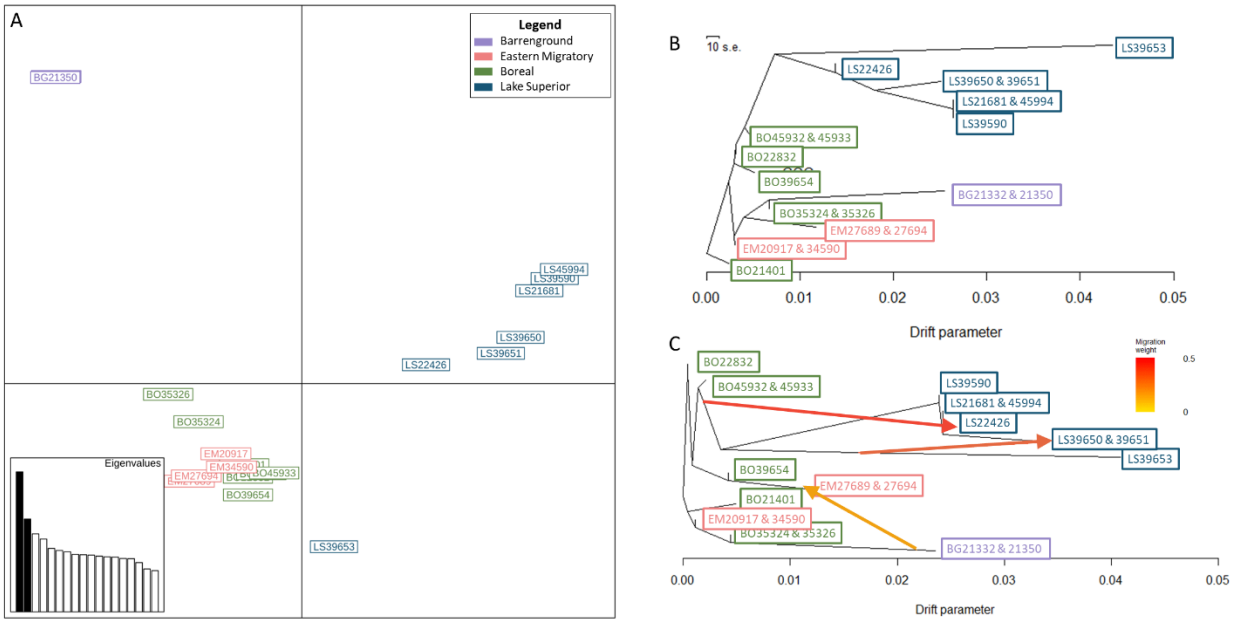

FIGURE 3 A) Principal Component Analysis (PCA) visualizing genomic variation among caribou (N=20). B) Evolutionary tree (Treemix) without migration showing evolutionary relationships (k = 1000). Branch lengths indicate drift estimates. C) Evolutionary tree (Treemix) migration model with the best support (k = 1000) depicting three migrations. Arrows indicate the direction of migration or gene flow; arrow colors indicate the strength of migration

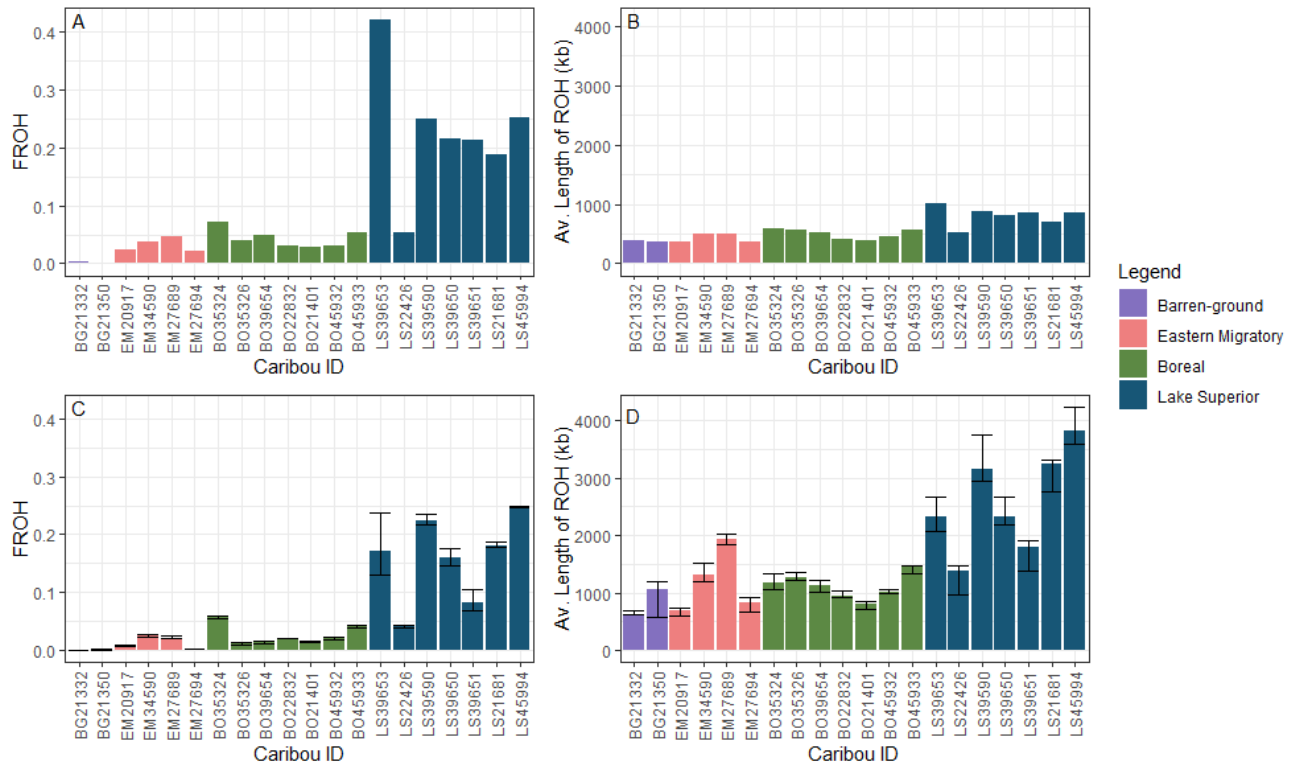

FIGURE 4 Inbreeding estimates based on Runs of Homozygosity (ROH) identified with PLINK (A,B) and ROHan (C,D) under the 250 kb size class. A,C) F_{ROH} indicates the proportion of the genome classified as ROH. B, D) Reflects the average length of ROH in kilobases, where shorter ROH indicate historical inbreeding and long ROH indicate recent inbreeding. Error bars (C,D) represent the minimum and maximum estimates produced by the Hidden Markov Model in ROHan.

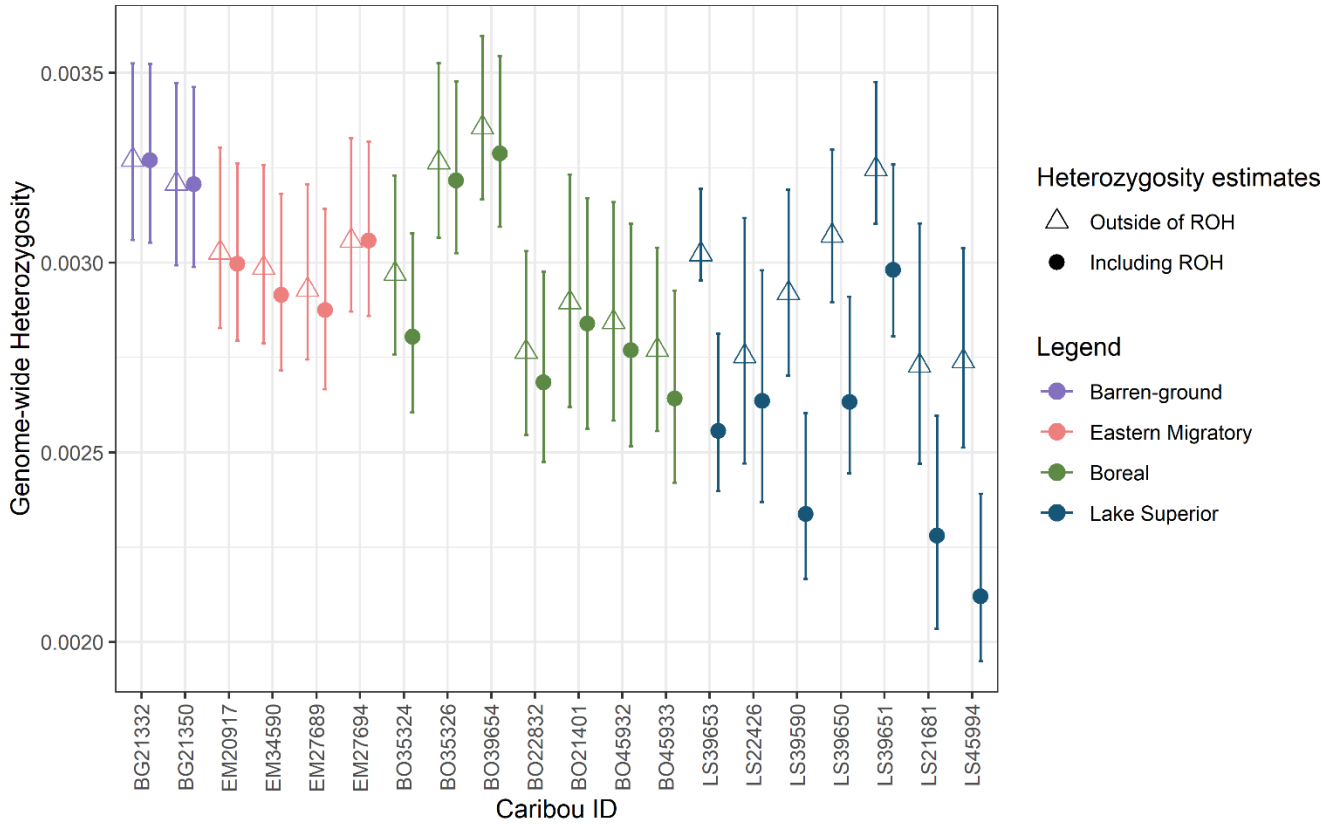

FIGURE 5 Individual genetic diversity calculated as genome-wide heterozygosity (Watterson's θ). Heterozygosity was calculated across the whole genome (including ROH; solid circles) and excluding regions in ROH (hollow triangles), using 250kb windows and allowing a heterozygosity rate of 0.0005 within ROH. Error bars indicate the minimum and maximum estimates of the Hidden Markov Model.

DATA ACCESSIBILITY

The raw reads are available at the National Centre for Biotechnology (NCBI) under the BioProject Accession numbers PRJNA 634908 and PRJNA 984705. Spatial layers for mapping the ranges of Designatable Units were provided by the Committee on the Status of Endangered Wildlife in Canada. Annotated code for all analyses can be found at: https://github.com/ksolmundson/PopGenomics.

AUTHOR CONTRIBUTIONS

K.S., J.B., P.J.W., and M.M. conceived and designed the study. J.B., P.J.W., and M.M. oversaw the research as Co-Principal Investigators. K.S. and S.K. performed bioinformatic analyses with guidance from R.S.T.. K.S. wrote the manuscript and J.B., R.S.T, M.M., and P.J.W. provided feedback and edited the manuscript.

SUPPLEMENTARY MATERIALS

Whole genomes reveal caribou population structure and inbreeding histories

Table S1. Sample ID reference numbers and genomic reference numbers to access raw sequence data used in this study.

Sample ID	Subspecies	Sample type	Genome accession number
BG21332	*R. t. groenlandicus*	Previously sequenced	PRJNA634908
BG21350	*R. t. groenlandicus*	Previously sequenced	PRJNA634908
EM20917	*R. t. caribou*	Previously sequenced	PRJNA634908
EM34590	*R. t. caribou*	Previously sequenced	PRJNA634908
EM27689	*R. t. caribou*	Previously sequenced	PRJNA634908
EM27694	*R. t. caribou*	Previously sequenced	PRJNA634908
BO35324	*R. t. caribou*	Previously sequenced	PRJNA634908
BO35326	*R. t. caribou*	Previously sequenced	PRJNA634908
BO39654	*R. t. caribou*	Previously sequenced	PRJNA634908
BO22832	*R. t. caribou*	Hair	PRJNA984705
BO21401	*R. t. caribou*	Fecal*	PRJNA984705
BO45932	*R. t. caribou*	Muscle	PRJNA984705
BO45933	*R. t. caribou*	Muscle	PRJNA984705
LS39653	*R. t. caribou*	Hide	PRJNA984705
LS22426	*R. t. caribou*	Fecal	PRJNA984705
LS39590	*R. t. caribou*	Previously sequenced	PRJNA634908
LS39650	*R. t. caribou*	Muscle	PRJNA984705
LS39651	*R. t. caribou*	Muscle	PRJNA984705
LS21681	*R. t. caribou*	Fecal*	PRJNA984705
LS45994	*R. t. caribou*	Antler	PRJNA984705

*****DNA was extracted & prepared for sequencing following the protocol described by Taylor et. al (2022).

Table S2. Model probabilities based on the mean and SD of log likelihood values produced by NGSAdmix across an array of 10 runs. The model with the highest probability (K=2) is the best supported value of K.

K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9
Inf	1385.0690	248.9362	308.2456	343.2897	140.3178	133.4855	111.7235

Table S3. Inbreeding estimates calculated using VCFtools, PLINK, and ROHan with a size class of 250 kb. We used the strictly filtered VCF file to calculate the inbreeding coefficient (F) in VCFtools and to identify ROH using PLINK with two sets of parameters (as specified in Methods). We also identified ROH in the individual BAM files using ROHan with two sets of parameters.

Sample ID	Mean Depth	F (VCFtools)	F_{ROH} (PLINK strict)	F_{ROH} (PLINK relaxed)	F_{ROH} (ROHan strict)	F_{ROH} (ROHan relaxed)
BG21332	38x	-0.03	0.002	0.003	0	0.001
BG21350	38x	-0.03	0.001	0.002	0	0.002
EM20917	36x	0	0.015	0.023	0	0.009
EM34590	37x	0.02	0.029	0.037	0	0.025
EM27689	38x	0.07	0.037	0.048	0	0.021
EM27694	39x	0.04	0.013	0.022	0	0.002
BO35324	35x	0.05	0.06	0.071	0	0.058
BO35326	38x	-0.04	0.034	0.041	0	0.013
BO39654	38x	0.05	0.039	0.048	0	0.014
BO22832	19x	0.01	0.022	0.03	0.002	0.021
BO21401	10x	-0.08	0.016	0.028	0.001	0.017
BO45932	13x	-0.03	0.022	0.03	0.001	0.022
BO45933	18x	0.04	0.043	0.053	0.009	0.043
LS39653	40x	0.41	0.399	0.42	0	0.171
LS22426	10x	-0.06	0.04	0.052	0.008	0.04
LS39590	35x	0.23	0.235	0.25	0	0.225
LS39650	37x	0.20	0.202	0.215	0	0.161
LS39651	40x	0.20	0.199	0.212	0	0.081
LS21681	10x	0.08	0.165	0.188	0.074	0.179
LS45994	17x	0.22	0.233	0.251	0.041	0.248

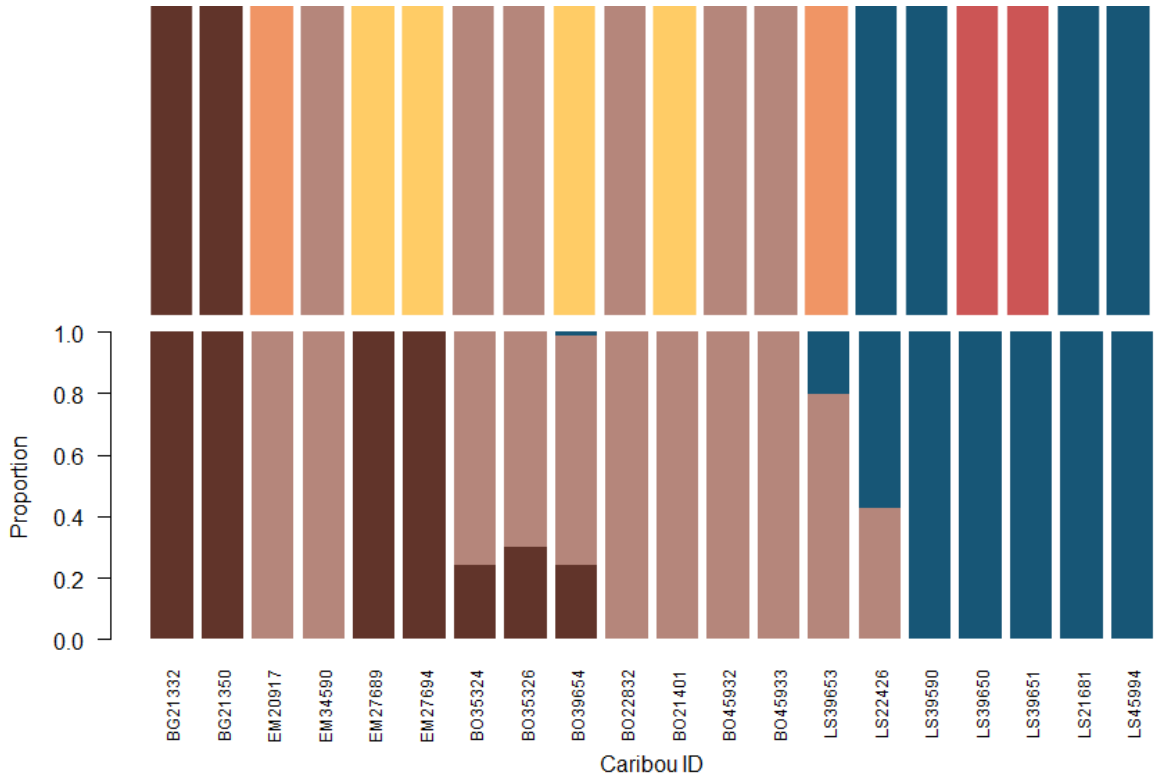

Figure S1. NGSAdmix plot for the next best supported K-values (3 and 6). Each bar represents an individual and colours represent population assignments.

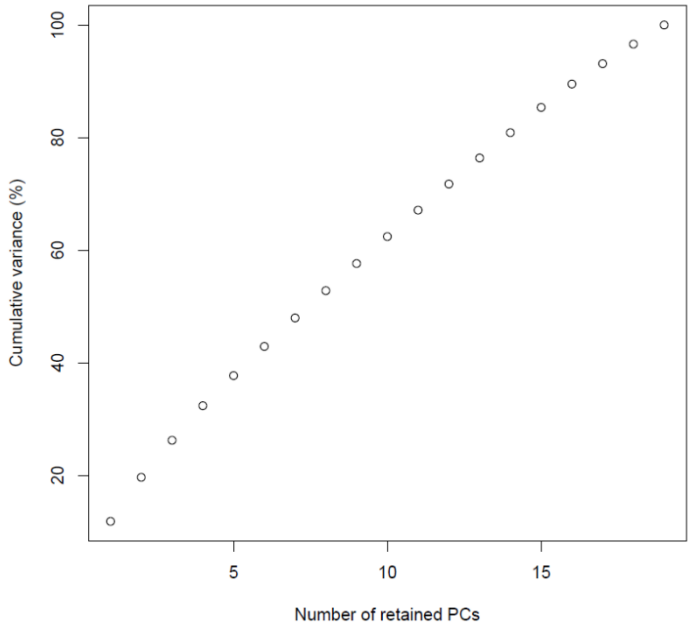

Figure S2. Cumulative variance for the 19 Principal Components of the PCA.

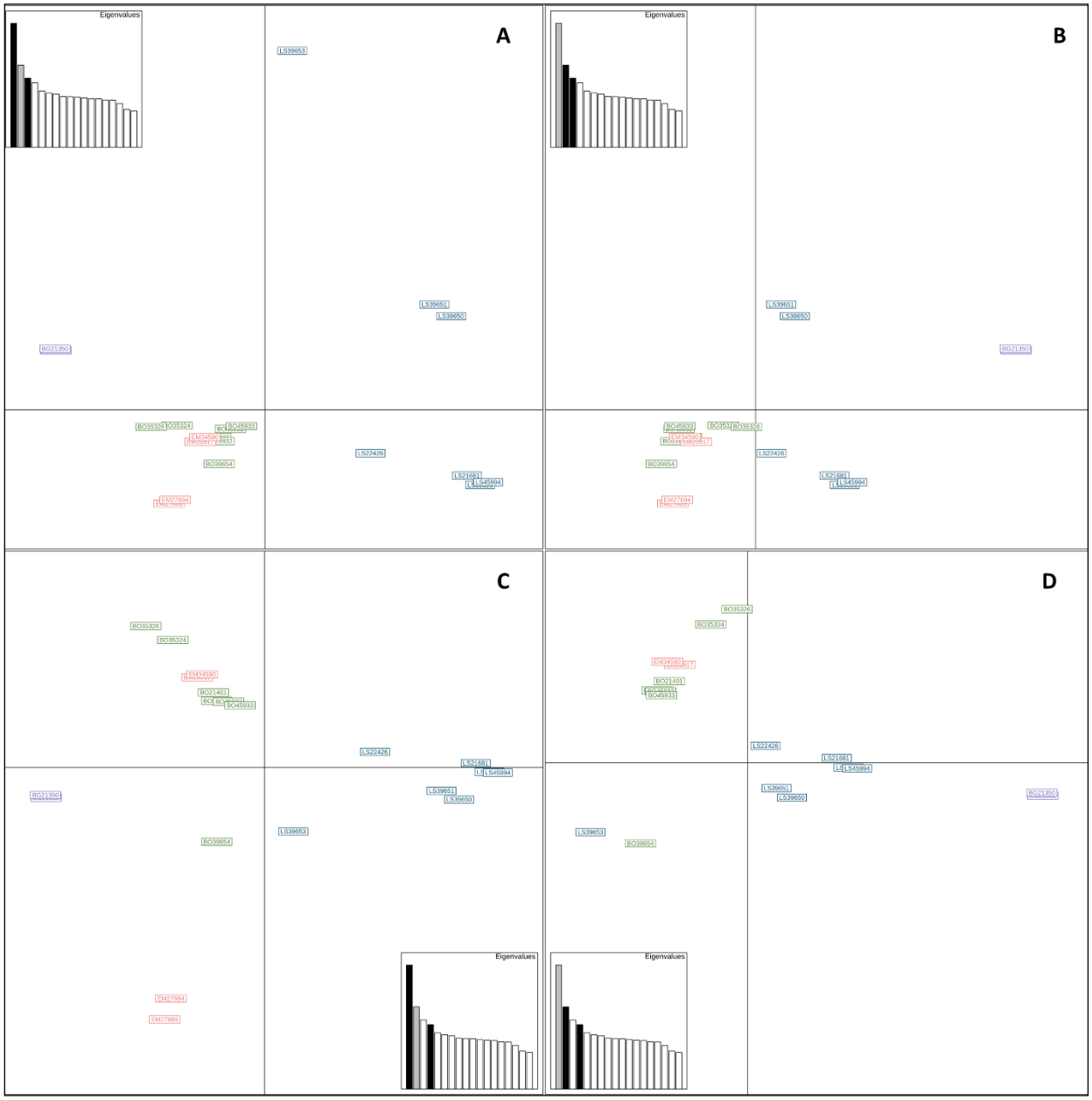

Figure S3. Additional Principal Component Axes: A) 1vs3 B) 2vs3 C) 1vs4 D) 2vs4

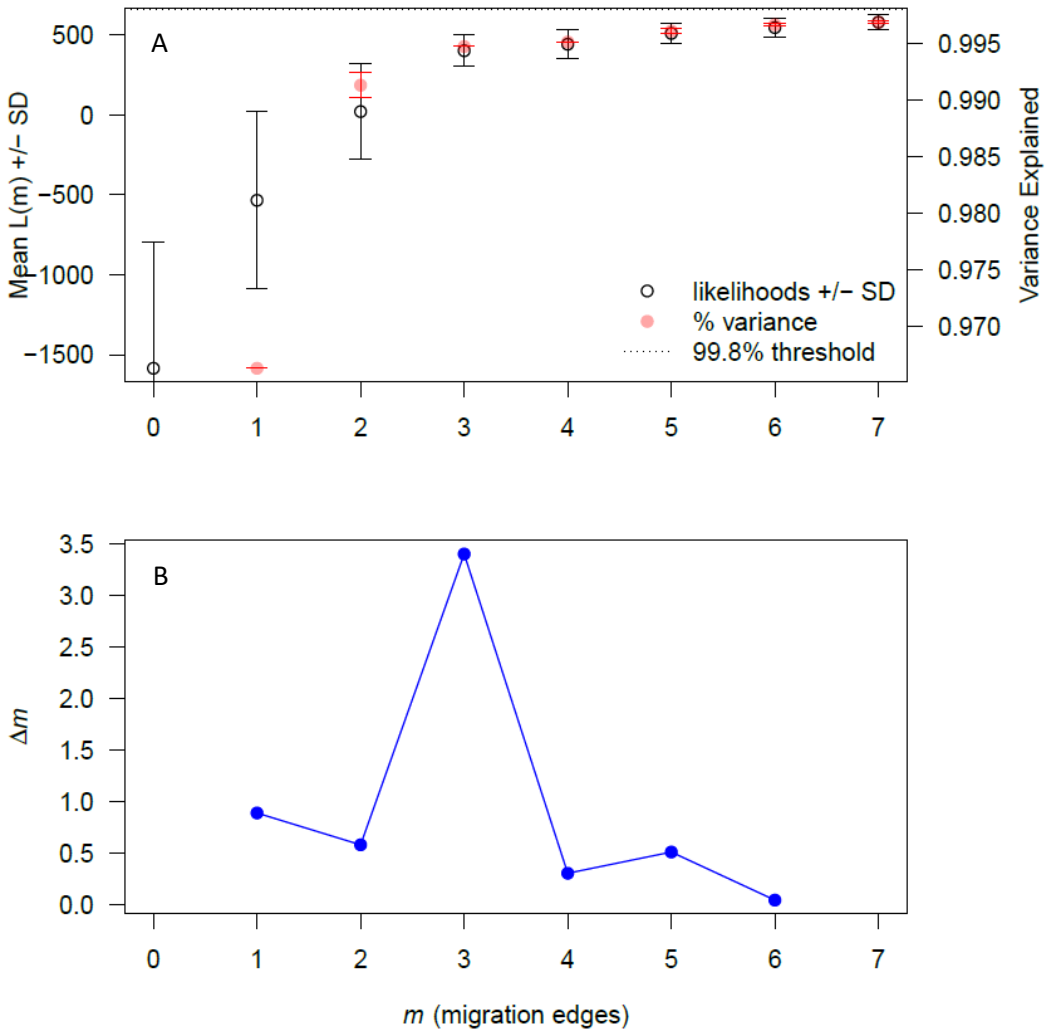

Figure S4. Comparison of Treemix migration models. We performed 10 iterations at k-values of 500, 1000, and 2000 for each migration model (m=0-7). The likelihood and SD values (A) and comparison of Δm values (B) indicate the model with 3 migrations has the best support.

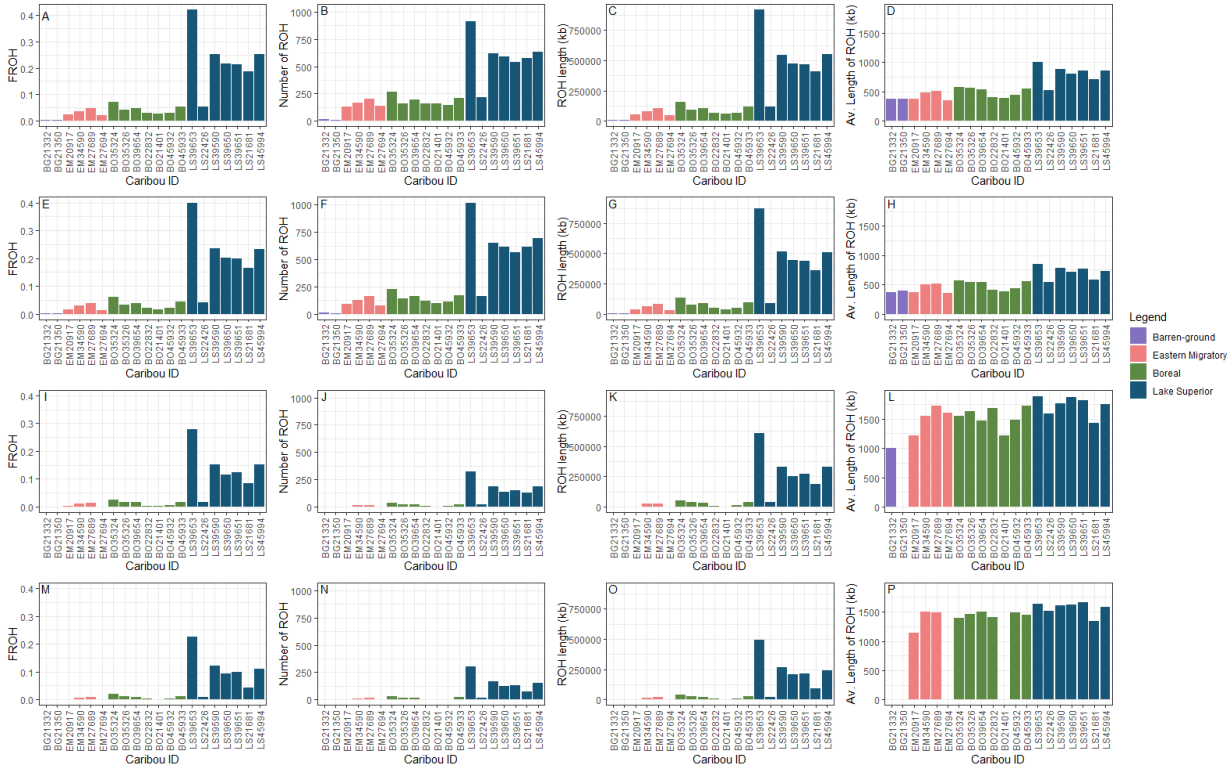

Figure S5. Runs of homozygosity identified with PLINK under different sets of parameters. A-D) relaxed parameters, >250kb ROH. E-H) strict parameters, >250kb ROH. I-L) relaxed parameters, >1Mb ROH. M-P) strict parameters, >1Mb ROH.

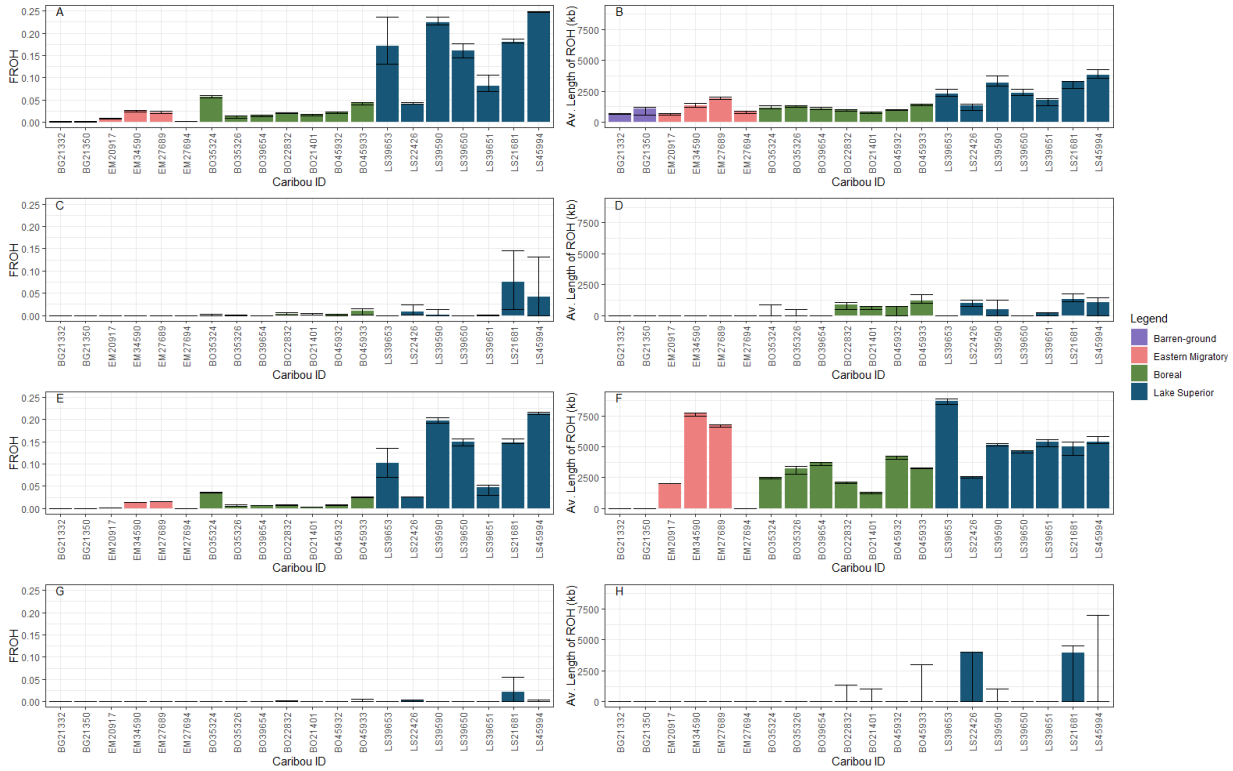

Figure S6. Runs of homozygosity identified with ROHan under different sets of parameters. We used window sizes of 250kb (A-D) and 1Mb (E-H) under relaxed (A,B,E,F) and more strict (C,D,G,H) parameters. Error bars indicate minimum and maximum estimates produced by the Hidden Markov Model.

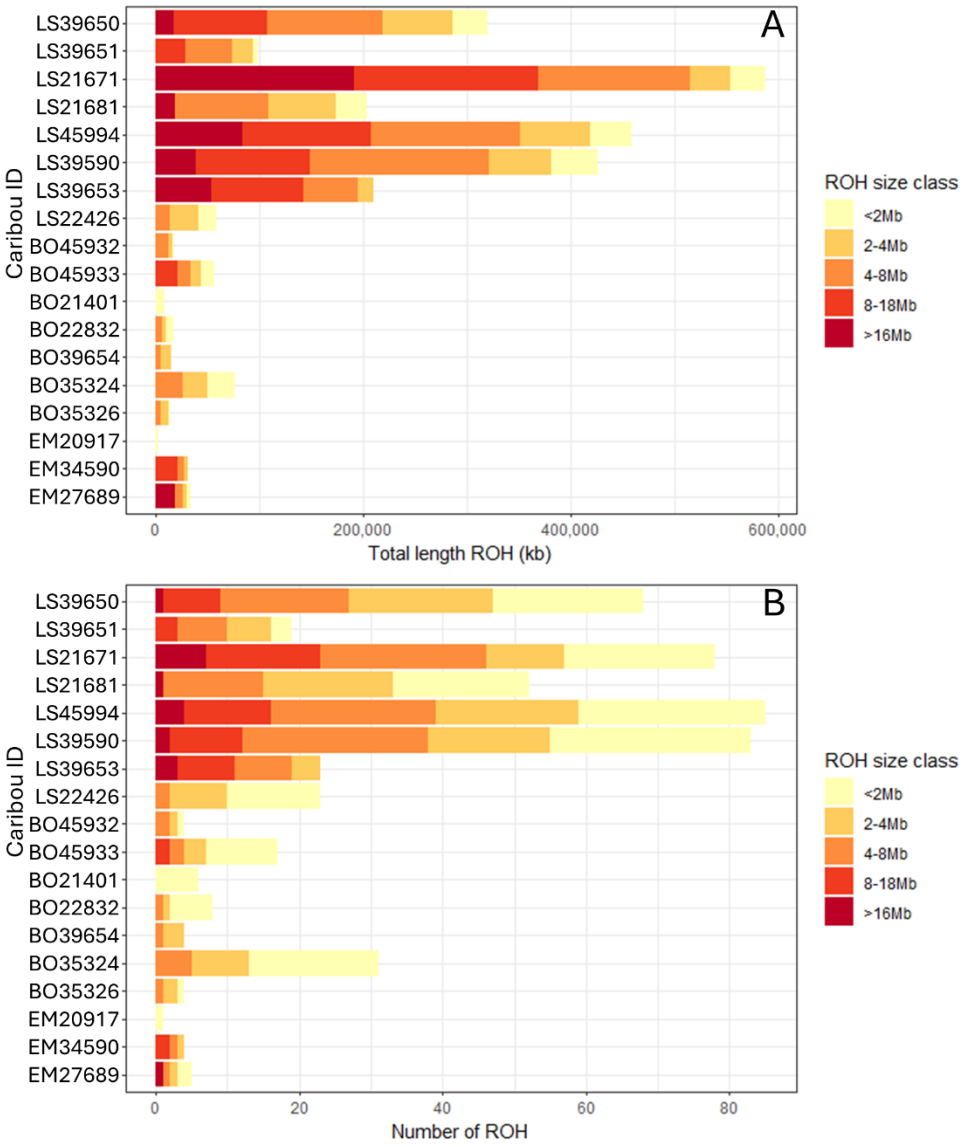

Figure S7. Size classes for Runs of Homozygosity (ROH) identified with ROHan under relaxed parameters with 250kb windows. Y-axis indicates individual caribou ID.

SUPPLEMENTARY REFERENCES

Taylor RS, Manseau M, Redquest B, et al (2022) Whole genome sequences from non-invasively collected caribou faecal samples. Conserv Genet Resour 14:53–68. https://doi.org/10.1007/s12686-021-01235-2

CHAPTER 3

Heterozygosity Rich Regions maintained in isolated caribou

A VERSION OF THIS ARTICLE HAS BEEN POSTED AS A PREPRINT

Solmundson, K., Bowman, J., Wilson, P. J., Taylor, R. S., Horn, R. L., Keobouasone, S., & Manseau, M. (2020). Genomic islands of heterozygosity maintained across caribou populations despite inbreeding. *BioRxiv*, 2020-12.

ABSTRACT

The caribou (*Rangifer tarandus*) is an iconic Canadian species facing conservation concerns due to anthropogenic factors. In particular, boreal caribou (*R. t. caribou*) have lost approximately half of their historic range, primarily attributed to habitat loss and fragmentation along the southern range edge. Nevertheless, small, isolated caribou populations have persisted in refugia along the coast and islands of Lake Superior, at the current southern range periphery of caribou. The persistence of caribou in coastal and island habitats in this region despite extirpation in surrounding areas could potentially indicate the presence of local adaptations. In this study, we used high-coverage whole genomes from 20 caribou representing different populations and ecotypes, to test for evidence of local adaptation via balancing selection in the Lake Superior range. We found several signatures of high genomic variation, which we classified as Heterozygosity Rich Regions (HRR), suggesting the presence of balancing selection. The HRR were maintained at higher rates in Lake Superior caribou than in other populations, which could indicate putative local adaptations. We then identified genes within these HRR and explored the associated molecular functions and biological processes. We found nine broader categories of molecular functions and six categories for biological processes, and further investigated additional subcategories for these molecular functions and biological processes. For instance, metabolic processes were associated with only two genes, but representing six subprocesses. The maintenance of high heterozygosity within these functional gene regions, especially in isolated populations, suggests the potential role of balancing selection in maintaining variation. This study demonstrates the preservation of genetic diversity in isolated caribou populations, providing evolutionary insights on wildlife threatened with extinction.

KEYWORDS: caribou, whole genomes, high-coverage, Heterozygosity Rich Regions

1| INTRODUCTION

The caribou (*Rangifer tarandus*) is a diverse species with several lineages (Festa-Bianchet, Ray, Boutin, Côté, & Gunn, 2011; Polfus et al., 2016). Caribou are diverse in terms of morphology, behaviour, and habitat type, demonstrating local adaptations to the diverse habitats occupied across their broad geographic range (Festa-Bianchet, Ray, Boutin, Côté, & Gunn, 2011; Polfus et al., 2016). This local adaptation is exemplified in Peary caribou (*R. t. pearyi*), a subspecies characterized by a relatively small body size and white coat, which are beneficial traits for living on arctic islands (COSEWIC, 2015). Beyond physical characteristics, there are also differences in adaptive behaviours among lineages. For instance, barren-ground caribou (*R. t. groenlandicus*) congregate on the tundra when calving, whereas boreal caribou (*R. t. caribou*) space out in densely forested areas to avoid predation (COSEWIC, 2014a, 2016).

Despite the diversity of caribou across Canada, all extant lineages are currently considered to be at risk of extinction by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC, 2014b, 2014a, 2015, 2016, 2017a, 2017b). Several factors have been implicated; however, the declining trends are predominantly attributable to anthropogenic changes, including climate change and habitat loss and fragmentation (Festa-Bianchet et al., 2011). Boreal caribou are particularly sensitive to habitat fragmentation; when dense forests are fragmented by linear corridors, caribou lose critical habitat for calving and face increased predation pressures (James & Stuart-Smith, 2000). The loss and fragmentation of habitat is especially evident along the southern edge of the species' range distribution; across the provinces of Canada, boreal caribou have lost approximately half of their historic range (COSEWIC, 2014a; Vors, Schaefer, Pond, Rodgers, & Patterson, 2007).

Despite this range loss, along the trailing range edge small groups of caribou remain in refugia on Lake Superior islands (Fig 1; Bergerud, Dalton, Butler, Camps, & Ferguson, 2007; Ontario Ministry of Natural Resources and Forestry, 2018). The Lake Superior region has a unique microclimate characterized by cooler summer temperatures, warmer winter temperatures, and greater amounts of winter precipitation in comparison to surrounding areas (Kopec, 1967; Scott & Huff, 1996). The survival of the Lake Superior caribou could perhaps be attributed the local climate or other habitat features, such as islands that allow escape from predators and generally low levels of human disturbance provided by protected areas such as Pukaskwa National Park, and Neys and the Slate Islands Provincial Parks (Schaefer, 2003). Nevertheless, the persistence of caribou in this region despite extirpation in surrounding areas could indicate that these caribou possess beneficial adaptations (Hampe & Petit, 2005). Caribou from the Lake Superior region may have adaptations that were selected for by the unique pressures associated with a southern distribution or the island and coastal ecosystem (Mallory & Hillis, 1996).

The field of genetics has provided many advances to wildlife conservation; however, historically, most genetic studies lacked the ability to detect signatures of selection because wildlife research has been limited to the use of very few markers and primarily focused on non-coding regions (e.g. microsatellites; Primmer, 2009). The field of genomics offers the ability to gain new insights into regions of adaptation (Mahony et al., 2020). Studies that use whole genomes or SNP panels often use outlier approaches to identify sites that may be under selection (Hohenlohe, Funk, & Rajora, 2021). This approach is rooted in a central principle of population genomics: neutral loci across the genome will be similarly affected by demography and the evolutionary history of populations, whereas loci under selection will often behave differently, revealing ‘outlier’ patterns of variation (Luikart, England, Tallmon, Jordan, & Taberlet, 2003).

For example, regions that have maintained a high level of diversity, measured as genomic heterozygosity, may be under the pressure of balancing selection (Selli et al., 2021).

Inbreeding depression is thought to be caused by two genetic effects: the increased expression of recessive deleterious alleles, and increased homozygosity at loci with heterozygote advantage (Charlesworth & Willis, 2009). The expression of inbreeding depression can be counterbalanced by balancing selection, which can maintain variation at sites with heterozygote advantage (Díez-del-Molino, Sánchez-Barreiro, Barnes, Gilbert, & Dalén, 2017). Thus, heterozygosity rich regions (HRR) can indicate regions that might be associated with heterozygote advantage (Selli et al., 2021).Using an annotated reference genome, it is possible to identify genes and the associated molecular functions and biological processes linked with these sites of high heterozygosity. A study of endangered brown bears (*Ursus arctos marsicanus*) demonstrated fixation of several partially deleterious alleles; however, high-variation was found in regions related to the immune system, olfactory signaling pathways, and digestion despite inbreeding, indicating that variation at important genes may be maintained by balancing selection (Benazzo et al., 2017).

Investigating patterns of hetero- and homozygosity requires a moderate to high depth of coverage due to the high uncertainty associated with low coverage sequences (Renaud, Hanghoj, Korneliussen, Willerslev, & Orland, 2019); however, high coverage data is costly to sequence, limiting research to a few representatives per populations. In this study, we used high-coverage whole genomes from 20 caribou (Figure 1) to test for evidence of local adaptation via balancing selection in the Lake Superior range. If caribou from the Lake Superior range possess local adaptations due to balancing selection, we predicted to find HRR that are maintained in Lake Superior caribou at higher rates than in caribou from other ranges. However, if caribou from the

Lake Superior range have lost adaptive variation in areas under balancing selection, due to isolation, bottlenecks, and inbreeding, we predicted to find HRR in caribou from other populations that are absent in caribou from the Lake Superior range.

2 | METHODS

2.1 | Sample Selection

We used moderately high coverage (≥10x) sequence data from 20 caribou, representing herds from different populations and ecotypes (Table S1). All sequence data used in this study are available through the National Center for Biotechnology Information (NCBI), BioProject accession numbers PRJNA 634908 and PRJNA 984705 (Solmundson et al., 2023; Taylor et al., 2020). From the boreal caribou ecotype, we included caribou from the continuous boreal range of Manitoba and Ontario, and from the Lake Superior range (Fig 1). Within the Lake Superior range, we sampled caribou from Michipicoten, Pic, and the Slate Islands, as well as mainland coastal caribou from Pukaskwa National Park and the area near Neys Provincial Park (Solmundson et al., 2023). We also included eastern migratory caribou from disjunct ranges in Ontario and Quebec, and barren-ground caribou from the Qamanirijuaq herd in Manitoba (Fig 1).

2.2 | Heterozygosity Rich Regions

We followed the same genome alignment and filtering methodology described in Solmundson et al. (2023), using the 35 largest scaffolds of a chromosome-level reference genome, which represents >99% of the caribou genome (Taylor et al., 2022). We then categorized regions of heterozygosity using a Markov Chain Monte Carlo (MCMC) based

approach that incorporates quality into the model considerations and requires a minimum average depth of 5x (Renaud et al., 2019). Specifically, we calculated Runs of Homozygosity (ROH) and genomic heterozygosity, measured as Waterson's θ with error bounds, using 1Mb windows across the entire genome of each individual, allowing a heterozygosity tolerance of $5x10^{-4}$ within ROH. We summarized the average genome with heterozygosity including and excluding regions in ROH for each caribou (Table S2). We categorized regions as heterozygosity rich (HRR) if the average heterozygosity across a 1MB window exceeded the top 0.999 SNPs of the percentile distribution of all heterozygosity estimates (Selli et al., 2021). We plotted the local heterozygosity estimates for all individuals across the scaffolds that contained HRR using ggplot2 (Wickham, 2016) in R (Fig 2).

2.3 | Gene Characterization

We extracted predicted genes within the HRR from an annotation file that corresponded to the caribou reference genome (Taylor et al., 2022). We then conducted a Gene Ontology (GO) analysis (Mi, Muruganujan, Ebert, Huang, & Thomas, 2019) to identify the molecular functions and biological processes associated with these genes. To select a model organism, we tested both *Homo sapiens* and *Bos taurus*. We excluded duplicate entries in cases where the same gene was identified multiple times. For each molecular function and biological pathway identified, we further explored the sub-functions (Fig S1) or sub-processes (Fig S2) within that category. We plotted a summary of the molecular functions (Fig 3A) and biological pathways (Fig 3B) in R.

3 | RESULTS

We found six HRR that were maintained in more than one individual from our sample of 20 caribou. All six of the HRR were maintained in at least one caribou from the Lake Superior range. In fact, for five of the six HRR, the highest heterozygosity estimate corresponded to a Lake Superior caribou (Table S3). Within the HRR, we found a total of 99 predicted genes (Table S2). We were able to identify 53 of these genes, while 46 represented proteins of unknown function. Several of the 53 identified genes were associated with repetitive gene regions; thus, in total we identified 27 unique genes (Table 1). The GO analysis with *H. sapiens* as a model organism resulted in corresponding database entries for 25 of these genes, whereas we only found 15 database entries when *B. taurus* was selected as the model organism. However, there was agreement among genes that were identified under both model organisms, thus we selected *H. sapiens* as the model. With the 25 genes that had GO database entries, we identified 39 associated molecular functions, which could be divided into nine categories; the most common functions were binding and transcription regulation (Fig 3A). We further explored the specific functions within these broader molecular functions (Fig S1). Within the binding category, which was associated with 14 genes, the most common functions were heterocyclic compound binding and organic cyclic compound binding, which were both associated with 12 of the genes. Conversely, we found the second most common function, transcription regulation, had only one sub-function; all ten transcription regulation genes were associated with DNA-binding transcription factor activity.

We also identified 33 biological processes representing six categories among the 25 genes with GO database entries; the most common were biological regulation and cellular process (Fig 3B). We found some of the associated biological processes represented a diversity of subprocesses (Fig S2). The most common process, biological regulation was represented by

twelve genes, and could be further subdivided into biological regulation, cellular process, and response to stimulus; all twelve genes contributed to the first process, but the two later processes only had one associated gene each. In contrast, metabolic process, represented by only two genes, could further be divided into six sub-processes: both of the genes were associated with cellular metabolic process, nitrogen compound metabolic process, organic substance metabolic process, and primary metabolic process, and the other sub-processes (small molecule metabolic process, biosynthetic process) were associated with one gene.

4 | DISCUSSION

Caribou in the Lake Superior range may face unique selection pressures introduced by several factors such as a trailing edge distribution, the local climate of the Lake Superior region, and the unique coastal and island ecosystems. In this study, we tested for evidence of balancing selection by investigating the maintenance of HRR. Balancing selection is an evolutionary mechanism that predictably results in increased heterozygosity, a measure of genomic variation (Díez-del-Molino et al., 2017). Thus, we characterized HRR in caribou populations using high-coverage, whole genomes to identify regions that might be associated with heterozygote advantage (Selli et al., 2021). We found several HRR that were maintained at higher rates in the caribou from the Lake Superior range than the other caribou populations we sampled. We then identified genes within these regions and the associated molecular pathways and predicted biological functions.

Previous research revealed inbreeding in caribou from the Lake Superior range, but relatively high heterozygosity was observed outside of ROH, which are genomic patterns caused by inbreeding (Solmundson et al., 2023). However, it was not clear if this discrepancy was

simply reflecting the difference between background genomic heterozygosity compared to areas in ROH, or if there were specific sites driving the observed differences in heterozygosity estimates. In this study, we identified specific genomic regions where high heterozygosity has been maintained and provided insights on genes within these regions. The high variation found within these genomic regions, especially even given relatively high inbreeding levels (Solmundson et al., 2023), suggests that heterozygosity may be evolutionarily maintained by balancing selection (Benazzo et al., 2017; Selli et al., 2021).

Caribou from the Lake Superior region may experience unique selection pressures associated with their southern distribution and island and coastal ecosystem (Mallory & Hillis, 1996). This area is characterized by low anthropogenic disturbance, low predation rates, low genetic connectivity to the continuous range, a unique microclimate, and alternative forage sources (Kopec, 1967; Schaefer, 2003; Scott & Huff, 1996). Although the islands in this region can support caribou, the typical source of forage for caribou, lichen, is often not available. On Michipicoten Island, which historically supported hundreds of caribou, the primary source of forage for caribou is yew trees (*Taxus* spp.; Mitchell et al., 2022), suggesting the possibility of dietary adaptations. Interestingly, we found several biological sub-processes that are associated with metabolic process, which could be linked to selective pressures related to climate (Borghi, Perez de Souza, Yoshida, & Fernie, 2019) or diet (Li et al., 2022). The potential presence of local adaptations could help to explain the persistence of caribou in the Lake Superior range despite extirpation from nearby regions (Hampe & Petit, 2005; Mallory & Hillis, 1996). Unfortunately, given recent rapid declines in the Lake Superior populations and a high uncertainty of future pressures, such as predation and industry (Ontario Ministry of Natural

Resources and Forestry, 2018), the presence of local adaptation alone is likely insufficient to ensure the caribou's long-term survival.

The maintenance of relatively high heterozygosity across populations could also suggest that the HRR represent ancestral polymorphisms that are maintained at selected sites, representing another form of balancing selection. However, other evolutionary processes, such as drift, can also cause genomic outliers. Previous research has demonstrated that caribou from the Lake Superior range are experiencing high rates of genomic drift, likely due to the recent isolation of this region from the continuous boreal caribou range (Drake et al., 2018; Solmundson et al., 2023). Drift predictably results in increasing homozygosity, ultimately resulting in allelic loss or fixation under extreme cases (Ezard & Travis, 2006). Thus, the HRR observed in this study are likely not the result of drift. Nevertheless, the HRR could be caused by some mechanism other than balancing selection. There are limited ways to confirm adaptation, such as knockout (Monroe et al., 2018) or common garden experiments (Fitzpatrick, Chhatre, Soolanayakanahally, & Keller, 2021; Mahony et al., 2020); however, these methods are typically not possible in the context of *in situ* conservation research of a threatened wildlife population. To gain further insights, future studies could explore fitness differences between caribou populations by estimating rates of survival and reproduction while considering climate or other ecological variables.

One caveat of comparative genomic analyses is that outlier sites can also be caused by errors during genome sequencing, assembly, or alignment. Notably, we found several repetitive gene regions within HRR. This finding is consistent with balancing selection, which selects for increased variation; for example, immune regions such as the MHC are commonly associated with repetitive genes under balancing selection (Aguilar *et al*. 2004). However, repetitive regions

are also prone to genome mapping and alignment errors, which can introduce biases (Treangen and Salzberg 2012; Ekblom and Wolf 2014). One of the HRR (Figure 2C) had notably higher maximum heterozygosity estimates than the others and was maintained across all individuals, which may underscore the importance of genomic variation in this region but could potentially be caused by an error in the genome reference assembly that all samples were subsequently mapped to. To reduce the chances that the observed genomic variation was caused by a random sequencing or alignment error, we used an approach that takes base quality and mapping quality into account, excluded results from the edges of scaffolds where there is higher uncertainty, and only considered HRR that were maintained in at least two individuals. Nevertheless, additional studies may aim to further investigate these putatively adaptive regions with additional samples, alternative approaches, or additional metrics of adaptation.

In this study, we used high-coverage whole genomes from different caribou populations to test for evidence of local adaptation via balancing selection. Comparisons of genome-wide diversity patterns revealed several putatively adaptive regions with high heterozygosity. This study demonstrates the maintenance of genomic variation in isolated caribou populations, providing important evolutionary insights on a species that is at risk of extinction (Festa-Bianchet et al., 2011; SARA, 2003).

REFERENCES

Aguilar A, Roemer G, Debenham S, et al (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc Natl Acad Sci U S A 101:3490–3494. https://doi.org/10.1073/PNAS.0306582101

Benazzo, A., Trucchi, E., Cahill, J. A., Maisano Delser, P., Mona, S., Fumagalli, M., … Bertorelle, G. (2017). Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers. Proceedings of the National Academy of Sciences, 114(45), 1–9. doi: 10.1073/pnas.1707279114

Bergerud, A. T., Dalton, W. J., Butler, H., Camps, L., & Ferguson, R. (2007). Woodland caribou persistence and extirpation in relic populations on Lake Superior. Rangifer, 27(17), 57–78.

Borghi, M., Perez de Souza, L., Yoshida, T., & Fernie, A. R. (2019). Flowers and climate change: a metabolic perspective. New Phytologist, 224(4), 1425–1441. doi: 10.1111/nph.16031

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. Nature Reviews Genetics, 10, 783–796.

COSEWIC. (2014a). COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Newfoundland population, Atlantic-Gaspésie population and Boreal population, in Canada. Retrieved from www.registrelep-sararegistry.gc.ca/default_e.cfm

COSEWIC. (2014b). COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Northern Mountain Population, Central Mountain Population, and Southern Mountain Population in Canada. Retrieved from http://www.sararegistry.gc.ca/virtual_sara/files/cosewic/sr_Caribou_Northern_Central_Southern_2014_e.pdf

COSEWIC. (2015). COSEWIC assessment and status report on the Peary caribou *Rangifer tarandus pearyi*, in Canada.

COSEWIC. (2016). COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Barren-Ground population, in Canada. Retrieved from http://www.registrelep-sararegistry.gc.ca/default.asp?lang=en&n=24F7211B-1

COSEWIC. (2017a). Caribou (*Rangifer tarandus*) some populations: COSEWIC assessment and status report. Retrieved from https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/cosewic-assessments-status-reports/caribou-some-populations-2017.html#_04_8

COSEWIC. (2017b). COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Dolphin and Union population, in Canada. Retrieved from https://wildlife-

species.canada.ca/species-risk-registry/virtual_sara/files/cosewic/srCaribouDolphinUnion2017e1.pdf

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2017). Quantifying Temporal Genomic Erosion in Endangered Species. Trends in Ecology and Evolution, xx, 1–10. doi: 10.1016/j.tree.2017.12.002

Drake, C. C., Manseau, M., Klütsch, C. F. C., Priadka, P., Wilson, P. J., Kingston, S., & Carr, N. (2018). Does connectivity exist for remnant boreal caribou (*Rangifer tarandus caribou*) along the Lake Superior Coastal Range? Options for landscape restoration. Rangifer, 38(1), 13. doi: 10.7557/2.38.1.4124

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. Evol Appl. https://doi.org/10.1111/eva.12178

Ezard, T. H. G., & Travis, J. M. J. (2006). The impact of habitat loss and fragmentation on genetic drift and fixation time. Oikos, 114(2), 367–375. doi: 10.1111/j.2006.0030-1299.14778.x

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(5), 419–434. doi: 10.1139/z11-025

Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y., & Keller, S. R. (2021). Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. Molecular Ecology Resources, 21(8), 2749–2765. doi: 10.1111/1755-0998.13374

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: The rear edge matters. Ecology Letters, 8(5), 461–467. doi: 10.1111/j.1461-0248.2005.00739.x

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. Molecular Ecology, 30(1), 62–82. doi: 10.1111/mec.15720

James, A. C. R., & Stuart-Smith, A. K. (2000). Distribution of Caribou and Wolves in Relation to Linear Corridors. The Journal of Wildlife Management, 64(1), 154–159.

Kopec, R. J. (1967). Areal patterns of seasonal temperature anomalies in the vicinity of the Great Lakes. Bulletin of the American Meteorological Society, 48(12), 884–889. doi: 10.1175/1520-0477-48.12.884

Li, L., Shen, F., Jie, X., Zhang, L., Yan, G., Wu, H., … Zhang, X. (2022). Comparative Transcriptomics and Methylomics Reveal Adaptive Responses of Digestive and Metabolic Genes to Dietary Shift in Giant and Red Pandas. Genes, 13(8). doi: 10.3390/genes13081446

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics, 4(12), 981–994. doi: 10.1038/nrg1226

Mahony, C. R., MacLachlan, I. R., Lind, B. M., Yoder, J. B., Wang, T., & Aitken, S. N. (2020). Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. Evolutionary Applications, 13(1), 116–131. doi: 10.1111/eva.12871

Mallory, F. F., & Hillis, T. L. (1996). Demographic characteristics of circumpolar caribou populations: ecotypes, ecological constraints, releases, and population dynamics. Rangifer, 10, 49–60. doi: 10.1139/z97-193

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research, 47(D1), D419–D426. doi: 10.1093/nar/gky1038

Mitchell, G., Wilson, P. J., Manseaua, M., Redquest, B., Patterson, B. R., & Rutledge, L. Y. (2022). DNA metabarcoding of faecal pellets reveals high consumption of yew (*Taxus* spp.) by caribou (*Rangifer tarandus*) in a lichen-poor environment. Facets, 7, 701–717. doi: 10.1139/facets-2021-0071

Monroe, J. G., Powell, T., Price, N., Mullen, J. L., Howard, A., Evans, K., … McKay, J. K. (2018). Drought adaptation in *Arabidopsis thaliana* by extensive genetic loss-of-function. ELife, 7, 1–18. doi: 10.7554/eLife.41038

Ontario Ministry of Natural Resources and Forestry. (2018). Seeking advice on the future of caribou in the Lake Superior coast range. Retrieved from https://ero.ontario.ca/notice/013-2587

Polfus, J. L., Manseau, M., Simmons, D., Neyelle, M., Bayha, W., Andrew, F., & Andrew, L. (2016). Łegha ́ gots ’ enete ́ (learning together): the importance of indigenous perspectives in the identification of biological variation. In Ecology and Society (Vol. 21). doi: 10.5751/ES-08284-210218

Primmer, C. R. (2009). From conservation genetics to conservation genomics. Annals of the New York Academy of Sciences, 1162, 357–368. doi: 10.1111/j.1749-6632.2009.04444.x

Renaud, G., Hanghoj, K., Korneliussen, T. S., Willerslev, E., & Orland, L. (2019). Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. Genetics, 212(July), 587–614.

SARA. (2003). Caribou (*Rangifer tarandus*), Boreal population - Species at risk registry. Retrieved July 8, 2022, from Government of Canada website: https://species-registry.canada.ca/index-en.html#/species/636-252%23species_summary

Schaefer, J. A. (2003). Long-term range recession and the persistence of caribou in the taiga. Conservation Biology, 17(5), 1435–1439. doi: 10.1046/j.1523-1739.2003.02288.x

Scott, R. W., & Huff, F. A. (1996). Impacts of the Great Lakes on regional climate conditions. Journal of Great Lakes Research, 22(4), 845–863. doi: 10.1016/S0380-1330(96)71006-7

Selli, A., Ventura, R. V, Fonseca, P. A. S., Buzanskas, M. E., Andrietta, L. T., Balieiro, J. C. C., & Brito, L. F. (2021). Detection and Visualization of Heterozygosity-Rich Regions and Runs of Homozygosity in Worldwide Sheep Populations. 11, 2696. doi: 10.3390/ani11092696

Solmundson, K., Bowman, J., Manseau, M., Taylor, R. S., Keobouasone, S., & Wilson, P. J. (2023). Genomic population structure and inbreeding history of Lake Superior caribou. Ecology and Evolution, 13(7), 1–14. doi: 10.1002/ece3.10278

Taylor, R. S., Manseau, M., Horn, R. L., Keobouasone, S., Golding, G. B., & Wilson, P. J. (2020). The role of introgression and ecotypic parallelism in delineating intra-specific conservation units. Molecular Ecology, (00), 1–17. doi: 10.1111/mec.15522

Taylor, R. S., Manseau, M., Keobouasone, S., Mastromonaco, G., Solmundson, K., Kelly, A., … Paul, J. (2022). Continent-wide phylogenomic framework reveals introgression as a driver of intra-specific diversity and enriched molecular pathways in caribou. BioRxiv. doi: 10.1101/2022.12.19.521038

Treangen TJ, Salzberg SL (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36–46. https://doi.org/10.1038/nrg3117

Vors, L. S., Schaefer, J. A., Pond, B. A., Rodgers, A. R., & Patterson, B. R. (2007). Woodland Caribou Extirpation and Anthropogenic Landscape Disturbance in Ontario. Journal of Wildlife Management, 71(4), 1249–1256. doi: 10.2193/2006-263

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York.: Springer-Verlag. Retrieved from https://ggplot2.tidyverse.org

Table 1. Heterozygosity Rich Regions identified in high-coverage whole genomes of caribou (N=20). Columns 1-3 indicate the locations of genetically diverse regions in BED format. Column 4 represents the number of predicted genes within each region, whereas column 5 indicates how many genes were identified, excluding duplicate entries.

Scaffold/Chromosome	Start	End	Predicted genes	Identified genes
Scaffold 1	27000000	28000000	17	7
Scaffold 2	81000000	82000000	10	3
Scaffold 7	62000000	63000000	16	3
Scaffold 19	11000000	12000000	25	17
Scaffold 27	28000000	29000000	7	0
Scaffold 29	33000000	34000000	24	1

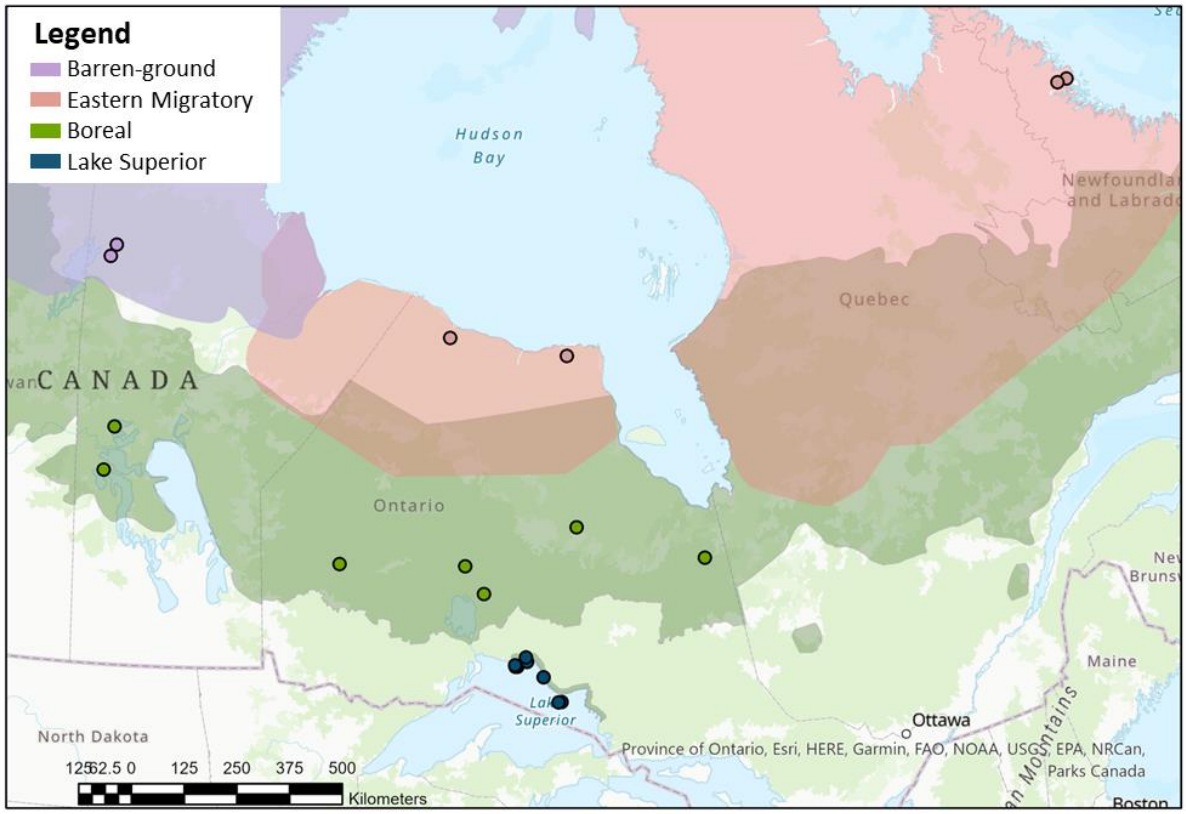

Figure 1. Map of caribou sample (N=20) locations. Background layers reflect the ranges of the Designable Units (DU) of caribou: barren-ground, eastern migratory, and boreal. Within the boreal DU, we sampled caribou from the continuous range and the Lake Superior range.

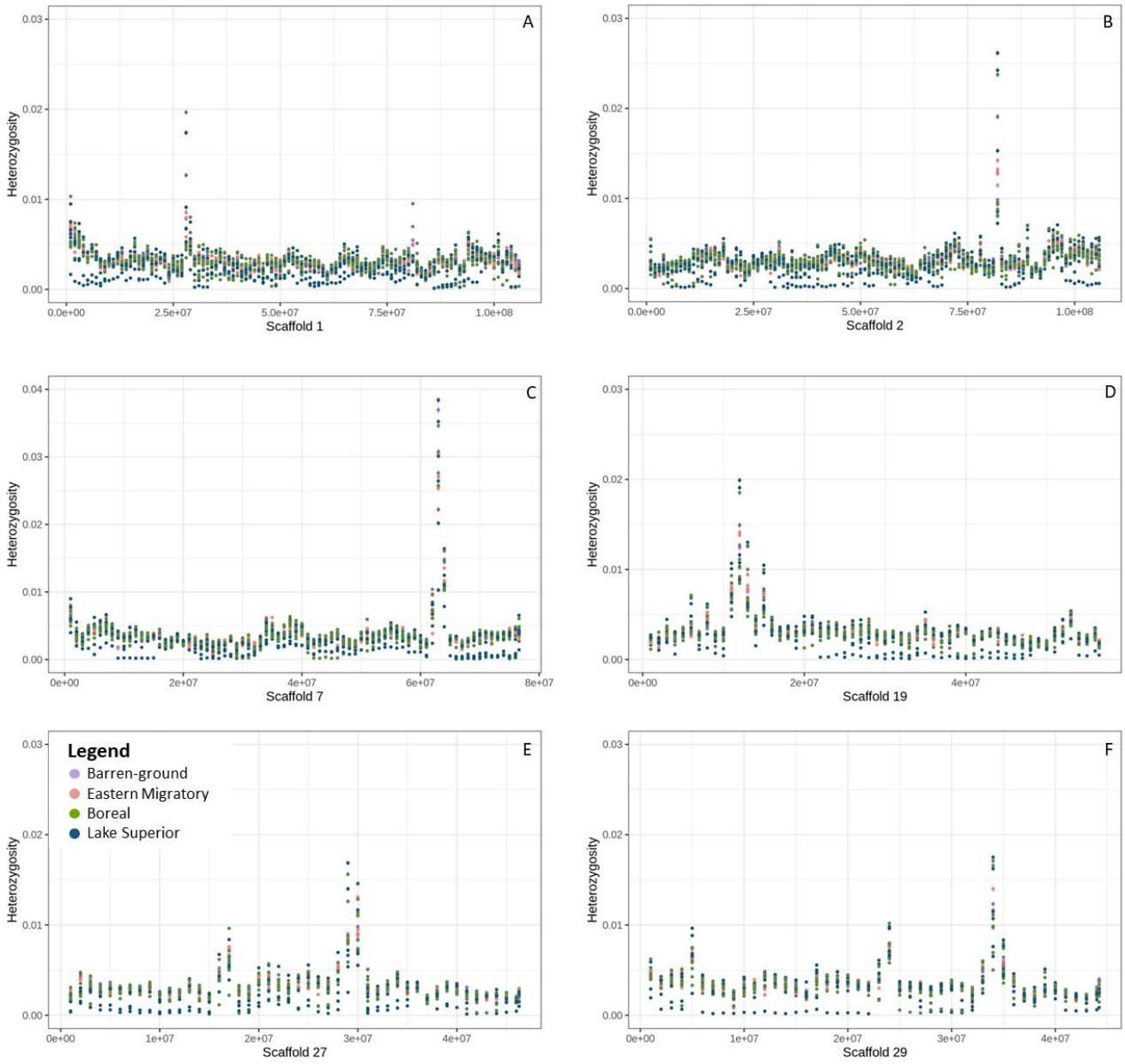

Figure 2. Heterozygosity Rich Regions (HRR) identified using high-coverage whole genomes of 20 caribou. Points indicate heterozygosity, estimated as Waterson's θ with 1Mb windows across each genome. Grey error bars indicate the error bounds of the Markov Chain Monte Carlo. Note the Y-axis of Scaffold 7 (plot C) has a maximum of 0.04, whereas all other plots have a maximum Y of 0.03.

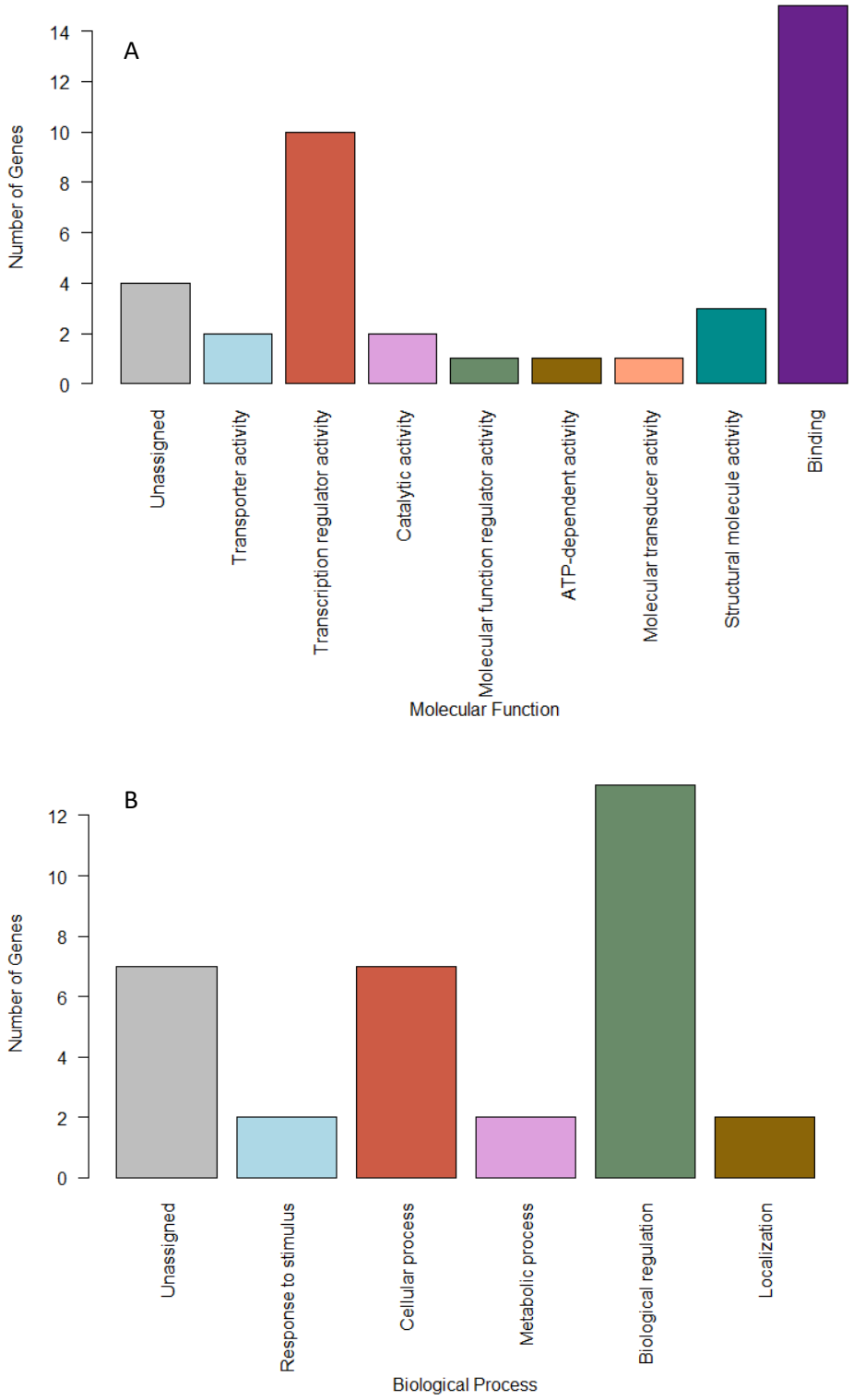

Figure 3. Summary of Gene Ontology analysis results for the A) molecular functions and B) biological processes associated with genes located in HRR.

DATA ACCESSIBILITY

The raw reads are available at the National Centre for Biotechnology (NCBI) under the BioProject Accession numbers PRJNA 634908 and PRJNA 984705. Spatial layers for mapping the ranges of Designatable Units were provided by the Committee on the Status of Endangered Wildlife in Canada. Annotated code can be found at: https://github.com/ksolmundson/HRR

AUTHOR CONTRIBUTIONS

K.S., J.B., P.J.W., and M.M. conceived and designed the study. J.B., P.J.W., and M.M. oversaw the research as Co-Principal Investigators. K.S. performed bioinformatic analyses with guidance from R.S.T. and S.K.. K.S. wrote the manuscript and J.B., R.S.T, M.M., and P.J.W. provided feedback and edited the manuscript.

ACKNOWLEDGMENTS

Funding was provided by the Species at Risk Stewardship Program from Ontario's Ministry of Environment, Conservation, and Parks; the Genomic Applications Partnership Program from Genome Canada and Environment and Climate Change Canada; and Discovery Grants from the Natural Sciences and Engineering Research Council of Canada. We thank Austin Thompson and Broderick Crosby for assisting with preparing and extracting DNA, and The Centre for Applied Genomics at the Hospital for Sick Children (Toronto, Ontario) for library preparation and sequencing. This research was conducted using high-performance computing services provided by the Digital Research Alliance of Canada (alliancecan.ca). We are grateful for the efforts of those who provided samples including the Government of Ontario and Environment and Climate Change Canada.

SUPPLEMENTARY MATERIALS

Heterozygosity Rich Regions maintained in isolated caribou

Table S1. Metadata for all high coverage whole caribou genomes included in this study (N=20).

Sample ID	Ecotype/Population	Depth	Heterozygosity (no ROH)	Heterozygosity (with ROH)	Genome accession number
BG21332	Barren-ground	38x	0.00303	0.00300	PRJNA634908
BG21350	Barren-ground	38x	0.00327	0.00327	PRJNA634908
EM20917	Eastern migratory	36x	0.00321	0.00321	PRJNA634908
EM34590	Eastern migratory	37x	0.00290	0.00284	PRJNA634908
EM27689	Eastern migratory	38x	0.00273	0.00228	PRJNA634908
EM27694	Eastern migratory	39x	0.00275	0.00264	PRJNA634908
BO35324	Boreal	35x	0.00276	0.00268	PRJNA634908
BO35326	Boreal	38x	0.00293	0.00287	PRJNA634908
BO39654	Boreal	38x	0.00306	0.00306	PRJNA634908
BO22832	Boreal	19x	0.00299	0.00291	PRJNA984705
BO21401	Boreal	10x	0.00297	0.00280	PRJNA984705
BO45932	Boreal	13x	0.00326	0.00322	PRJNA984705
BO45933	Boreal	18x	0.00292	0.00234	PRJNA984705
LS39653	Lake Superior	40x	0.00307	0.00263	PRJNA984705
LS22426	Lake Superior	10x	0.00325	0.00298	PRJNA984705
LS39590	Lake Superior	35x	0.00302	0.00256	PRJNA634908
LS39650	Lake Superior	37x	0.00336	0.00329	PRJNA984705
LS39651	Lake Superior	40x	0.00284	0.00277	PRJNA984705
LS21681	Lake Superior	10x	0.00277	0.00264	PRJNA984705
LS45994	Lake Superior	17x	0.00274	0.00212	PRJNA984705

Table S2. Predicted genes within Heterozygosity Rich Regions.

Scaffold/ Chromosome	Region type	Start	End	Predicted gene
Scaffold 1	gene	27002184	27037357	ID=ANN09425;Note=Similar to RAPGEF5: Rap guanine nucleotide exchange factor 5 (Homo sapiens OX%3D9606);
Scaffold 1	gene	27248332	27268198	ID=ANN09426;Note=Protein of unknown function;
Scaffold 1	gene	27302488	27316183	ID=ANN09427;Note=Protein of unknown function;
Scaffold 1	gene	27324219	27341770	ID=ANN09428;Note=Protein of unknown function;
Scaffold 1	gene	27329788	27333553	ID=ANN09429;Note=Protein of unknown function;
Scaffold 1	gene	27342186	27372186	ID=ANN09430;Note=Protein of unknown function;
Scaffold 1	gene	27372599	27376920	ID=ANN09431;Note=Protein of unknown function;
Scaffold 1	gene	27461768	27480851	ID=ANN09432;Note=Protein of unknown function;
Scaffold 1	gene	27481335	27494680	ID=ANN09433;Note=Similar to MAGEA4: Melanoma-associated antigen 4 (Homo sapiens OX%3D9606);
Scaffold 1	gene	27497777	27510748	ID=ANN09434;Note=Protein of unknown function;
Scaffold 1	gene	27552417	27563954	ID=ANN09435;Note=Protein of unknown function;
Scaffold 1	gene	27629335	27647784	ID=ANN09436;Note=Similar to F52C9.6: Putative uncharacterized transposon-derived protein F52C9.6 (Caenorhabditis elegans OX%3D6239);
Scaffold 1	gene	27675798	27679746	ID=ANN09437;Note=Similar to IL6: Interleukin-6 (Bos taurus OX%3D9913);
Scaffold 1	gene	27769618	27779162	ID=ANN09438;Note=Similar to slc40a1: Solute carrier family 40 member 1 (Danio rerio OX%3D7955);
Scaffold 1	gene	27952361	27953038	ID=ANN09439;Note=Similar to RPL5: 60S ribosomal protein L5 (Gallus gallus OX%3D9031);
Scaffold 1	gene	27956593	27968996	ID=ANN09440;Note=Protein of unknown function;
Scaffold 1	gene	27974402	28045723	ID=ANN09441;Note=Similar to NUP42: Nucleoporin NUP42 (Homo sapiens OX%3D9606);
Scaffold 2	gene	81008103	81011708	ID=ANN02918;Note=Protein of unknown function;
Scaffold 2	gene	81014743	81037887	ID=ANN02919;Note=Similar to AMPD1: AMP deaminase 1 (Homo sapiens OX%3D9606);
Scaffold 2	gene	81091199	81095937	ID=ANN02920;Note=Similar to SIKE1: Suppressor of IKBKE 1 (Bos taurus OX%3D9913);
Scaffold 2	gene	81126607	81148070	ID=ANN02921;Note=Protein of unknown function;
Scaffold 2	gene	81172693	81182010	ID=ANN02922;Note=Protein of unknown function;
Scaffold 2	gene	81200893	81233254	ID=ANN02923;Note=Similar to OR4C11: Olfactory receptor 4C11 (Homo sapiens OX%3D9606);
Scaffold 2	gene	81276534	81285930	ID=ANN02924;Note=Similar to OR4C11: Olfactory receptor 4C11 (Homo sapiens OX%3D9606);
Scaffold 2	gene	81706653	81732437	ID=ANN02925;Note=Protein of unknown function;
Scaffold 2	gene	81770638	81777004	ID=ANN02926;Note=Protein of unknown function;
Scaffold 2	gene	81814011	81814868	ID=ANN02927;Note=Protein of unknown function;
Scaffold 7	gene	62100233	62189276	ID=ANN25508;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62178835	62189349	ID=ANN25509;Note=Protein of unknown function;
Scaffold 7	gene	62250108	62250341	ID=ANN25510;Note=Protein of unknown function;
Scaffold 7	gene	62302115	62327739	ID=ANN25511;Note=Protein of unknown function;

Scaffold 7	gene	62343906	62365889	ID=ANN25512;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62436158	62436469	ID=ANN25513;Note=Similar to FAM136A: Protein FAM136A (Bos taurus OX%3D9913);
Scaffold 7	gene	62460107	62478496	ID=ANN25514;Note=Protein of unknown function;
Scaffold 7	gene	62473968	62492932	ID=ANN25515;Note=Protein of unknown function;
Scaffold 7	gene	62553535	62631648	ID=ANN25516;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62649071	62684561	ID=ANN25517;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62711241	62750709	ID=ANN25518;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62768442	62769281	ID=ANN25519;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62781581	62836432	ID=ANN25520;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62855815	62901973	ID=ANN25521;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62914510	62948707	ID=ANN25522;Note=Similar to ABCC4: Multidrug resistance-associated protein 4 (Homo sapiens OX%3D9606);
Scaffold 7	gene	62985964	62998002	ID=ANN25523;Note=Similar to Transposon TX1 uncharacterized 149 kDa protein (Xenopus laevis OX%3D8355);
Scaffold 19	gene	11018649	11019632	ID=ANN08013;Note=Similar to ZNF420: Zinc finger protein 420 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11037734	11038063	ID=ANN08014;Note=Similar to ZKSCAN8: Zinc finger protein with KRAB and SCAN domains 8 (Pan paniscus OX%3D9597);
Scaffold 19	gene	11093556	11098537	ID=ANN08015;Note=Protein of unknown function;
Scaffold 19	gene	11100528	11100764	ID=ANN08016;Note=Protein of unknown function;
Scaffold 19	gene	11100869	11101129	ID=ANN08017;Note=Protein of unknown function;
Scaffold 19	gene	11101234	11101554	ID=ANN08018;Note=Protein of unknown function;
Scaffold 19	gene	11138037	11147546	ID=ANN08019;Note=Protein of unknown function;
Scaffold 19	gene	11248949	11249634	ID=ANN08020;Note=Protein of unknown function;
Scaffold 19	gene	11329680	11344166	ID=ANN08021;Note=Similar to ZNF677: Zinc finger protein 677 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11344234	11357548	ID=ANN08022;Note=Similar to ZNF347: Zinc finger protein 347 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11376901	11391510	ID=ANN08023;Note=Similar to ZNF677: Zinc finger protein 677 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11396824	11397684	ID=ANN08024;Note=Similar to ZNF480: Zinc finger protein 480 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11464968	11472383	ID=ANN08025;Note=Protein of unknown function;

Scaffold 19	gene	11474750	11484267	ID=ANN08026;Note=Similar to ZNF665: Zinc finger protein 665 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11495423	11513416	ID=ANN08027;Note=Similar to ZNF845: Zinc finger protein 845 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11544316	11544798	ID=ANN08028;Note=Similar to RPL21: 60S ribosomal protein L21 (Sus scrofa OX%3D9823);
Scaffold 19	gene	11592659	11630358	ID=ANN08029;Note=Similar to ZNF91: Zinc finger protein 91 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11680602	11688823	ID=ANN08030;Note=Similar to ZNF160: Zinc finger protein 160 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11692250	11694706	ID=ANN08031;Note=Similar to ZNF665: Zinc finger protein 665 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11715094	11720471	ID=ANN08032;Note=Similar to ZNF695: Zinc finger protein 695 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11739721	11754695	ID=ANN08033;Note=Protein of unknown function;
Scaffold 19	gene	11828574	11828945	ID=ANN08034;Note=Similar to RPL35: 60S ribosomal protein L35 (Bos taurus OX%3D9913);
Scaffold 19	gene	11833226	11843931	ID=ANN08035;Note=Similar to ZNF665: Zinc finger protein 665 (Homo sapiens OX%3D9606);
Scaffold 19	gene	11860031	11870357	ID=ANN08036;Note=Protein of unknown function;
Scaffold 19	gene	11895685	11935959	ID=ANN08037;Note=Similar to GLRX: Glutaredoxin-1 (Sus scrofa OX%3D9823);
Scaffold 27	gene	28117569	28143039	ID=ANN20647;Note=Protein of unknown function;
Scaffold 27	gene	28128075	28146735	ID=ANN20648;Note=Protein of unknown function;
Scaffold 27	gene	28315519	28319356	ID=ANN20649;Note=Protein of unknown function;
Scaffold 27	gene	28459609	28478524	ID=ANN20650;Note=Protein of unknown function;
Scaffold 27	gene	28617386	28618333	ID=ANN20651;Note=Protein of unknown function;
Scaffold 27	gene	28904226	28924139	ID=ANN20652;Note=Protein of unknown function;
Scaffold 27	gene	28921306	28937975	ID=ANN20653;Note=Protein of unknown function;
Scaffold 29	gene	33041309	33059087	ID=ANN22352;Note=Protein of unknown function;
Scaffold 29	gene	33058086	33065025	ID=ANN22353;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33152644	33157853	ID=ANN22354;Note=Protein of unknown function;
Scaffold 29	gene	33198157	33200241	ID=ANN22355;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33208699	33216346	ID=ANN22356;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33251746	33268860	ID=ANN22357;Note=Protein of unknown function;
Scaffold 29	gene	33345430	33349125	ID=ANN22358;Note=Protein of unknown function;
Scaffold 29	gene	33411043	33419329	ID=ANN22359;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33481923	33488302	ID=ANN22360;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33514256	33517200	ID=ANN22361;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33518235	33520349	ID=ANN22362;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33601215	33607051	ID=ANN22363;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33631039	33634871	ID=ANN22364;Note=Protein of unknown function;

Scaffold 29	gene	33644467	33650568	ID=ANN22365;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33659928	33677556	ID=ANN22366;Note=Protein of unknown function;
Scaffold 29	gene	33704422	33722007	ID=ANN22367;Note=Protein of unknown function;
Scaffold 29	gene	33713619	33717909	ID=ANN22368;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33743412	33743831	ID=ANN22369;Note=Similar to Pregnancy-associated glycoprotein 1 (Ovis aries OX%3D9940);
Scaffold 29	gene	33771167	33794006	ID=ANN22370;Note=Protein of unknown function;
Scaffold 29	gene	33778481	33782868	ID=ANN22371;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33826799	33834860	ID=ANN22372;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33849544	33866300	ID=ANN22373;Note=Protein of unknown function;
Scaffold 29	gene	33873242	33881549	ID=ANN22374;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);
Scaffold 29	gene	33922122	33931062	ID=ANN22375;Note=Similar to Pregnancy-associated glycoprotein 1 (Bos taurus OX%3D9913);

Table S3. Heterozygosity estimates (Watterson's θ) within Heterozygosity Rich Regions for each caribou (N=20). Columns 3-5 indicate the location of HRR windows in BED format.

Sample ID	Scaffold/Chromosome	Start	End	Heterozygosity	Error
EM20917	Scaffold 1	27000000	28000000	0.0058	0.0002
BG21332	Scaffold 1	27000000	28000000	0.0091	0.0002
BG21350	Scaffold 1	27000000	28000000	0.0078	0.0002
BO21401	Scaffold 1	27000000	28000000	0.0048	0.0002
LS21681	Scaffold 1	27000000	28000000	0.0079	0.0006
LS22426	Scaffold 1	27000000	28000000	0.0068	0.0003
BO22832	Scaffold 1	27000000	28000000	0.0045	0.0002
EM27689	Scaffold 1	27000000	28000000	0.0080	0.0002
EM27694	Scaffold 1	27000000	28000000	0.0085	0.0002
EM34590	Scaffold 1	27000000	28000000	0.0086	0.0002
BO35324	Scaffold 1	27000000	28000000	0.0044	0.0002
BO35326	Scaffold 1	27000000	28000000	0.0127	0.0002
LS39590	Scaffold 1	27000000	28000000	0.0054	0.0002
LS39650	Scaffold 1	27000000	28000000	0.0091	0.0002
LS39651	Scaffold 1	27000000	28000000	0.0174	0.0003
LS39653	Scaffold 1	27000000	28000000	0.0174	0.0003
BO39654	Scaffold 1	27000000	28000000	0.0197	0.0003
BO45932	Scaffold 1	27000000	28000000	0.0051	0.0002
BO45933	Scaffold 1	27000000	28000000	0.0035	0.0002
LS45994	Scaffold 1	27000000	28000000	0.0053	0.0002
EM20917	Scaffold 2	81000000	82000000	0.0115	0.0003
BG21332	Scaffold 2	81000000	82000000	0.0129	0.0003
BG21350	Scaffold 2	81000000	82000000	0.0098	0.0003

BO21401	Scaffold 2	81000000	82000000	0.0087	0.0003
LS21681	Scaffold 2	81000000	82000000	0.0122	0.0007
LS22426	Scaffold 2	81000000	82000000	0.0081	0.0003
BO22832	Scaffold 2	81000000	82000000	0.0081	0.0002
EM27689	Scaffold 2	81000000	82000000	0.0127	0.0003
EM27694	Scaffold 2	81000000	82000000	0.0142	0.0003
EM34590	Scaffold 2	81000000	82000000	0.0132	0.0003
BO35324	Scaffold 2	81000000	82000000	0.0096	0.0003
BO35326	Scaffold 2	81000000	82000000	0.0191	0.0003
LS39590	Scaffold 2	81000000	82000000	0.0105	0.0003
LS39650	Scaffold 2	81000000	82000000	0.0153	0.0003
LS39651	Scaffold 2	81000000	82000000	0.0242	0.0003
LS39653	Scaffold 2	81000000	82000000	0.0261	0.0003
BO39654	Scaffold 2	81000000	82000000	0.0238	0.0003
BO45932	Scaffold 2	81000000	82000000	0.0094	0.0003
BO45933	Scaffold 2	81000000	82000000	0.0088	0.0003
LS45994	Scaffold 2	81000000	82000000	0.0072	0.0002
EM20917	Scaffold 7	62000000	63000000	0.0383	0.0004
BG21332	Scaffold 7	62000000	63000000	0.0306	0.0005
BG21350	Scaffold 7	62000000	63000000	0.0370	0.0005
BO21401	Scaffold 7	62000000	63000000	0.0254	0.0005
LS21681	Scaffold 7	62000000	63000000	0.0245	0.0010
LS22426	Scaffold 7	62000000	63000000	0.0264	0.0006
BO22832	Scaffold 7	62000000	63000000	0.0271	0.0004
EM27689	Scaffold 7	62000000	63000000	0.0270	0.0004
EM27694	Scaffold 7	62000000	63000000	0.0305	0.0004
EM34590	Scaffold 7	62000000	63000000	0.0255	0.0004
BO35324	Scaffold 7	62000000	63000000	0.0222	0.0004
BO35326	Scaffold 7	62000000	63000000	0.0276	0.0004
LS39590	Scaffold 7	62000000	63000000	0.0233	0.0004
LS39650	Scaffold 7	62000000	63000000	0.0301	0.0004
LS39651	Scaffold 7	62000000	63000000	0.0352	0.0004
LS39653	Scaffold 7	62000000	63000000	0.0384	0.0004
BO39654	Scaffold 7	62000000	63000000	0.0346	0.0004
BO45932	Scaffold 7	62000000	63000000	0.0257	0.0005
BO45933	Scaffold 7	62000000	63000000	0.0308	0.0005
LS45994	Scaffold 7	62000000	63000000	0.0202	0.0004
EM20917	Scaffold 19	11000000	12000000	0.0124	0.0003
BG21332	Scaffold 19	11000000	12000000	0.0127	0.0003
BG21350	Scaffold 19	11000000	12000000	0.0126	0.0003
BO21401	Scaffold 19	11000000	12000000	0.0087	0.0003
LS21681	Scaffold 19	11000000	12000000	0.0126	0.0007
LS22426	Scaffold 19	11000000	12000000	0.0089	0.0003

BO22832	Scaffold 19	11000000	12000000	0.0085	0.0002
EM27689	Scaffold 19	11000000	12000000	0.0149	0.0003
EM27694	Scaffold 19	11000000	12000000	0.0141	0.0003
EM34590	Scaffold 19	11000000	12000000	0.0138	0.0003
BO35324	Scaffold 19	11000000	12000000	0.0111	0.0003
BO35326	Scaffold 19	11000000	12000000	0.0149	0.0003
LS39590	Scaffold 19	11000000	12000000	0.0128	0.0003
LS39650	Scaffold 19	11000000	12000000	0.0107	0.0002
LS39651	Scaffold 19	11000000	12000000	0.0191	0.0003
LS39653	Scaffold 19	11000000	12000000	0.0199	0.0003
BO39654	Scaffold 19	11000000	12000000	0.0185	0.0003
BO45932	Scaffold 19	11000000	12000000	0.0092	0.0003
BO45933	Scaffold 19	11000000	12000000	0.0102	0.0003
LS45994	Scaffold 19	11000000	12000000	0.0116	0.0003
EM20917	Scaffold 27	28000000	29000000	0.0080	0.0002
BG21332	Scaffold 27	28000000	29000000	0.0084	0.0002
BG21350	Scaffold 27	28000000	29000000	0.0087	0.0002
BO21401	Scaffold 27	28000000	29000000	0.0079	0.0003
LS21681	Scaffold 27	28000000	29000000	0.0092	0.0006
LS22426	Scaffold 27	28000000	29000000	0.0072	0.0003
BO22832	Scaffold 27	28000000	29000000	0.0059	0.0002
EM27689	Scaffold 27	28000000	29000000	0.0090	0.0002
EM27694	Scaffold 27	28000000	29000000	0.0085	0.0002
EM34590	Scaffold 27	28000000	29000000	0.0081	0.0002
BO35324	Scaffold 27	28000000	29000000	0.0088	0.0002
BO35326	Scaffold 27	28000000	29000000	0.0126	0.0002
LS39590	Scaffold 27	28000000	29000000	0.0091	0.0002
LS39650	Scaffold 27	28000000	29000000	0.0066	0.0002
LS39651	Scaffold 27	28000000	29000000	0.0140	0.0002
LS39653	Scaffold 27	28000000	29000000	0.0169	0.0003
BO39654	Scaffold 27	28000000	29000000	0.0156	0.0003
BO45932	Scaffold 27	28000000	29000000	0.0080	0.0003
BO45933	Scaffold 27	28000000	29000000	0.0083	0.0002
LS45994	Scaffold 27	28000000	29000000	0.0054	0.0002
EM20917	Scaffold 29	33000000	33000000	0.0117	0.0003
BG21332	Scaffold 29	33000000	33000000	0.0116	0.0003
BG21350	Scaffold 29	33000000	33000000	0.0123	0.0003
BO21401	Scaffold 29	33000000	33000000	0.0097	0.0003
LS21681	Scaffold 29	33000000	33000000	0.0113	0.0007
LS22426	Scaffold 29	33000000	33000000	0.0107	0.0003
BO22832	Scaffold 29	33000000	33000000	0.0069	0.0002
EM27689	Scaffold 29	33000000	33000000	0.0111	0.0003
EM27694	Scaffold 29	33000000	33000000	0.0140	0.0003

EM34590	Scaffold 29	33000000	33000000	0.0111	0.0003
BO35324	Scaffold 29	33000000	33000000	0.0099	0.0002
BO35326	Scaffold 29	33000000	33000000	0.0166	0.0003
LS39590	Scaffold 29	33000000	33000000	0.0125	0.0003
LS39650	Scaffold 29	33000000	33000000	0.0115	0.0002
LS39651	Scaffold 29	33000000	33000000	0.0162	0.0003
LS39653	Scaffold 29	33000000	33000000	0.0175	0.0003
BO39654	Scaffold 29	33000000	33000000	0.0171	0.0003
BO45932	Scaffold 29	33000000	33000000	0.0050	0.0002
BO45933	Scaffold 29	33000000	33000000	0.0112	0.0003
LS45994	Scaffold 29	33000000	33000000	0.0076	0.0002

Table S4. Molecular sub-functions associated with genes in Heterozygosity Rich Regions

Molecular Function	**Subfunction**	**Number of genes**
Binding	Organic cyclic compound binding	12
Binding	Hormone binding	1
Binding	Heterocyclic compound binding	12
Binding	Protein binding	1
Binding	Amide binding	1
Catalytic activity	Hydrolase activity	1
Catalytic activity	Oxidoreductase activity	1
Molecular function regulator activity	Enzyme regulator activity	1
Molecular transducer activity	Singling receptor activity	1
Structural molecule activity	Structural constituent of ribosome	3
Transcription regulator activity	Dna-binding transcription factor activity	10
Transporter activity	Transmembrane transporter activity	2
ATP-dependent activity	ATP-dependent activity	1
NA*	NA	5

*No GO category assigned

Table S5. Biological sub-processes associated with genes in Heterozygosity Rich Regions

Biological Process	**Subprocess**	**Number of genes**
Biological regulation	Regulation of biological process	12
Biological regulation	Regulation of molecular function	1
Cellular process	Cell communication	1
Cellular process	Cellular component organization or biogensis	2
Cellular process	Celllular metabolic process	2
Cellular process	Cellular response to stimulus	1
Cellular process	Signal transduction	1
Cellular process	Transmembrane transport	2
Localization	Establishment of localization	2
Metabolic process	Biosynthetic process	1
Metabolic process	Cellular metabolic process	2
Metabolic process	Nitrogren compound metabolic process	2
Metabolic process	Organic substance metabolic process	2
Metabolic process	Primary metabolic process	2
Metabolic process	Small molecule metabolic process	1
Response to stimulus	Cellular response to stimulus	1
NA*	NA	8

*No GO category assigned

CHAPTER 4

Whole genomes reveal the demographic history of Ontario's boreal caribou

Kirsten Solmundson[1*], Jeff Bowman[1,2], Micheline Manseau[1,3], Rebecca S. Taylor[3], Paul J. Wilson[4]

[1]Environmental & Life Sciences Graduate Program, Trent University, Peterborough, ON, Canada

[2]Wildlife Research and Monitoring Section, Ontario Ministry of Natural Resources and Forestry, Peterborough, ON, Canada

[3]Landscape Science and Technology Division, Environment and Climate Change Canada, Ottawa, ON, Canada

[4]Biology Department, Trent University, Peterborough, ON, Canada

ABSTRACT

Boreal caribou (*Rangifer tarandus caribou*) have experienced significant declines and range contractions over the past century which has resulted in several small and isolated populations, raising conservation concerns about their long-term viability. In Ontario, most boreal caribou reside in the northern boreal continuous range. Along the southern edge, the range has been receding northward for over a century, resulting in isolated populations along the coast and nearshore islands of Lake Superior. In this study, we investigated the demographic histories of boreal caribou populations across Ontario, Canada. We extracted and sequenced DNA from dried blood and other archived samples, producing 197 chromosome-level re-sequenced whole genomes. We then tested for population structure and modelled changes in population sizes throughout historical and recent time. We found evidence of genetic structure throughout the study area; however, in general, all populations shared similar demographic histories. Following the late Pleistocene, the populations slowly expanded, plateaued, and declined at similar times. Notably, the ancestors of two population groupings from the current Lake Superior range had larger effective population sizes during the late Pleistocene and had the largest maximum historical population sizes out of all population groupings. Collectively, our results indicate that caribou from the Lake Superior and continuous boreal ranges of Ontario do not have divergent evolutionary histories, and the observed differences in population structure are likely due to recent bottleneck effects that occurred over the past century. This study delineates the population history of boreal caribou in Ontario, providing valuable insights for conservation and management.

KEYWORDS: genomics, caribou, effective population size, demographic history

1 | INTRODUCTION

Found in boreal forests across Canada, boreal caribou (*Rangifer tarandus caribou*) are iconic and ecologically important animals (COSEWIC 2014b). However, like many wildlife populations worldwide, boreal caribou have experienced significant declines and range contractions over the past century due to various anthropogenic factors, including habitat loss, fragmentation, and increased human activity (Schaefer, 2003; Vors, Schaefer, Pond, Rodgers, & Patterson, 2007). These declines have resulted in some small and isolated caribou populations, which raises concerns about their long-term viability given the potential for inbreeding depression and loss of genetic diversity (Festa-Bianchet, Ray, Boutin, Côté, & Gunn, 2011).

To effectively manage and conserve declining wildlife populations, it is important to understand their evolutionary history and genetic structure (COSEWIC 2020; Hohenlohe et al. 2021). The field of genomics has expanded beyond the limitations of traditional genetic studies, providing new insights deep into population history, shedding light on responses to past environmental changes and potential adaptability to future environmental changes (Supple and Shapiro 2018; Schweizer et al. 2021). Advances in genome sequencing and bioinformatics have provided valuable tools to investigate historical demographic processes and population dynamics in non-model species, allowing researchers to explore the effects of historical evolutionary events such as glacial cycles (Primmer, 2009).

Glaciation events during the Pleistocene significantly shaped the distribution and genetic diversity of many wildlife species globally, including caribou (Hofreiter & Stewart, 2009; Lister, 2004). As the glaciers receded at the end of the Pleistocene, caribou populations began to colonize their current habitats, resulting in rapid range expansions and introgression among

allopatrically-evolved lineages, which is reflected in the diversity patterns of the species (Klutsch, Manseau, & Wilson, 2012; Polfus, Manseau, Klütsch, Simmons, & Wilson, 2017; Taylor et al., 2022).

Boreal caribou occupy a vast range across Canada despite recent population declines and range contractions (SARA, 2012). In Ontario, the majority of boreal caribou reside in the northern boreal continuous range (henceforth BCR; Ontario Ministry of Natural Resources, 2009). Along the southern edge, range retraction has resulted in small and disconnected populations (Schaefer, 2003; Solmundson et al. 2023). Some caribou populations have managed to persist along the islands and coastal areas of Lake Superior, approximately 150 km south of the trailing edge of the BCR (Schaefer, 2003). The recent history of caribou in the Lake Superior range (henceforth LSR) is well-documented but the future of these populations remains precarious. Better understanding their evolutionary history can provide important insights for conservation and management.

The LSR has been the focus of recent caribou conservation efforts and is of particular interest due to its southern location and unique history. Historically, caribou were present on various islands within the LSR. In the 1980s, a population was established on Michipicoten Island, which grew to over 600 caribou by the 2010s (Ontario Ministry of Natural Resources and Forestry, 2018). However, in 2014, the arrival of wolves on the island led to a rapid decline in the population, prompting the relocation of the remaining caribou to other nearby islands without wolf populations in 2018 (Ontario Ministry of Natural Resources and Forestry, 2018). This history of population declines and relocations may have significant implications for the genetic structure and demographic dynamics of caribou in the LSR.

Previous whole-genome research indicated that caribou from the LSR form a consistent genetic group, differentiated from caribou in the BCR (Solmundson et al. 2023). However, the underlying causes of this genomic differentiation remain uncertain. It is not clear whether the differentiation is a result of genomic effects associated with recent island colonization (e.g., bottleneck effect caused by island colonization, followed by inbreeding and drift over subsequent generations) or if there are deeper differences in the evolutionary history between these two ranges. In this study, we re-sequenced 197 whole genomes of caribou from Ontario, Canada to investigate the demographic histories of Lake Superior caribou in relation to caribou from the boreal continuous range. Specifically, we tested for population structure and modelled changes in population sizes throughout historical and recent time.

We present two hypotheses for the demographic history of caribou in the LSR. First, if caribou from the Lake Superior range represent a recently diverged island population, we predicted all LSR and BCR populations would demonstrate similar historical effective population sizes. However, if caribou from the LSR have a divergent demographic history, we predicted to observe distinct patterns in historical effective population size among the population groupings of the LSR and predicted the BCR population groups would show similar trends to one another. In general, we expected that population groupings with relatively more abundant populations occupying vast areas would have larger effective population sizes than caribou that presently occur in small and isolated groupings, such as the LSR.

2 | MATERIALS & METHODS

2.1 | Population sampling

Samples were collected by Government of Ontario and Parks Canada field biologists and sent to Trent University for DNA extraction. We selected samples spanning the entire width of the province, providing a thorough representation of the boreal BCR (Fig. 1); we explicitly selected samples from the lower portion of the boreal range and did not include samples from northern regions where there is an overlap between boreal and eastern migratory caribou (Pond, Brown, Wilson, & Schaefer, 2016). Although we use the term BCR throughout this study, there are several recognized boreal caribou ranges within this larger region. Specifically, there are nine boreal caribou ranges within the province of Ontario that are delineated in the Species At Risk Act (SARA) Recovery Strategy for the Woodland Caribou (*Rangifer tarandus caribou*), Boreal population (Environment Canada 2011; SARA 2012). This study includes samples from all except for the most northerly range (ON9; SARA, 2012). From the LSR (boreal range ON6; SARA, 2012), we sequenced seven historical (1990s) samples from Pukaskwa NP, and 55 samples from Michipicoten Island, including several caribou that have recently been translocated to other Lake Superior Islands (OMNRF, 2018). The samples from the now extirpated, historical herd in Pukaskwa NP were provided by the Canadian Museum of Nature, Ottawa, ON. Notably, at least one caribou from the historical coastal Lake Superior population, Pukaskwa National Park (NP), was a founder of the Michipicoten Island population that was established in the 1980s (Bergerud, Dalton, Butler, Camps, & Ferguson, 2007), thus these two populations may display similar historical population trends reflecting their shared evolutionary history.

2.2 | Genome sequencing, assembly, and quality control

We extracted DNA from muscle tissue, whole blood, and dried blood cards using the Qiagen DNeasy kit, following the manufacturer's protocols (Qiagen, Hilden, Germany). Dried blood preserved on cards (e.g. Whatman® Classic FTA® cards [GE Healthcare Bio-Sciences, Marlborough, MA, USA]) has a limited amount of DNA, especially for mammalian samples due to the lack of nucleated blood cells. Thus, to increase the DNA yielded from the blood cards, prior to extraction, we incubated the samples at a high temperature (65°C) on a shaker overnight (18hrs) following the protocol described by Love Stowell *et al.* (2018). Following DNA extraction, we increased the final concentration of DNA with an Amicon centrifugal filter (Millipore Sigma, Darmstadt, Germany). We quantified the extracted DNA using a Qubit system (Thermo Fisher Scientific, MA, USA) to ensure all samples were above the minimum threshold required for next-generation sequencing and then normalized the concentrations of all samples (1.03-1.5 ng/μL).

The extracted DNA was then sent to The Centre for Applied Genomics (TCAG), at The Hospital for Sick Children (Toronto, ON). The samples were prepared using an Illumina library prep kit (Illumina, San Diego, CA, USA) and sequenced with the Illumina NovaSeq6000 system, yielding paired-end 150bp low coverage sequence reads.

We used secure computing resources from the Digital Research Alliance of Canada (alliancecan.ca) to conduct all analyses and created all figures in R (R Core Team, 2023). First, we used Trimmomatic v0.39 (Bolger, Lohse, & Usadel, 2014) to trim the sequence reads and bwa v0.7.17 (Heng Li & Durbin, 2009) to align them to a chromosome-level caribou reference genome (Taylor et al., 2022 [preprint]) retaining the 35 largest scaffolds/chromosomes,

corresponding to >99% of the genome and the full karyotype (2n=70) of this species (Gripenberg and Nieminen 1986; Proskuryakova et al. 2023). We carefully considered the filters applied at each stage through the genome assembly and population genotyping process to ensure we retained high-quality sites. We used Samtools v1.17 (Heng Li et al., 2009) to sort the mapped bam file and to remove sites with a mapping quality (q) lower than 20. We then removed duplicates and clipped overlapping regions with GATK v4.2.4.0 (DePristo et al. 2011) and Bamutils v1.0.14 respectively. We indexed the mapped bam files with Samtools and used GATK to perform indel realignment. Finally, we calculated the depth of coverage (Appendix S1) and checked the quality of all final mapped genomes using Samtools and FastQC (Andrews, 2010).

2.3 | Population genotyping

We used angsd (Korneliussen, Albrechtsen, & Nielsen, 2014) to combine the individual mapped genomes into a grouped file with genotype likelihood values, with the following filters -minMapQ 20 -minQ 20 -snp_pval 1e-6 -minmaf 0.05. We used PCAngsd (Meisner & Albrechtsen, 2018), to conduct Principal Component Analyses (PCA); this program incorporates genotype likelihood estimates, which is ideal for the low-coverage data used in this study. We further explored population structure with NGSAdmix (Skotte, Korneliussen, & Albrechtsen, 2013). We tested for population groupings of K=2-8, with 10 arrays for each value of K and used R to plot the results and calculate the probability for each K value based on the log-likelihoods produced by NGSAdmix (Appendix S2).

We used PCAngsd and NGSAdmix to identify sample groupings. We then used angsd to create BCF files for each sample grouping; we required data for all individuals within each population with the -minInd option and applied the following filters: -minMapQ 20 -minQ 20 -

snp_pval 0.01 -rmTriallelic 0.01 -C 50 -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1. We converted each grouped BCF to a VCF and indexed the files with BCFtools (Li, 2011), and then split the VCF files by chromosomes with R.

2.4 | Demographic history modeling

We used the program PopsizeABC (Boitard, Rodríguez, Jay, Mona, & Austerlitz, 2016) to model changes in population sizes over time using an Approximate Bayesian Computation (ABC) framework. This bioinformatic software can accurately estimate population sizes up to more recent time frames than other commonly used genomic demographic models such as PSMC, MSMC, or SMC++ (Nadachowska-Brzyska, Konczal and Babik, 2022). For each population grouping, we calculated statistics from empirical data and performed 500,000 simulations under two sets of time parameters. Specifically, we simulated 500,000 datasets containing 100 independent scaffolds that were 2MB in length, corresponding to the same length used for the linkage disequilibrium calculations. We used a lower bound of $0.1x10^{-8}$ and an upper bound of $1x10^{-8}$ for the per generation per bp recombination rate, with a per generation per bp mutation rate of $1x10^{-8}$. We specified a caribou generation time of 7 years (Polfus et al., 2017; Taylor et al., 2020). For the final ABC models, we included the calculated statistics for allele frequencies and linkage disequilibrium, using a neural network regression to retain the 0.1% of simulations with summary statistics that most closely matched the empirical data.

We used two sets of time parameters to model population sizes in both deep and more recent history. First, we used the default program settings of 21 windows beginning 130,000 generations ago. To gain insights on more recent changes in population size, we conducted additional analyses with the first window starting 1000 generations ago, while retaining a total of

21 windows to ensure a high resolution was maintained into recent history. Given a caribou generation time of 7 years, the first time window under the more recent parameters roughly corresponds to when the Laurentide Ice Sheet had retreated from the Lake Superior region (approximately 7kya), allowing caribou to move into the habitats they now occupy (Saarnisto 1974; Klutsch et al. 2012). Given the known history of isolation, we may see evidence of historical bottlenecks in caribou from the LSR but the models cannot detect very recent events (e.g. from within the past century).

3 | RESULTS

3.1 | Population structure

We explored population structure through PCA and population assignment tests (NGSAdmix; Skotte et al., 2013). The best supported NGSAdmix model (Appendix S2) revealed 2 broad population groupings, where Michipicoten Island in the LSR was assigned as one group, and all samples in the BCR represented the other group, with Pukaskwa NP showing mixed assignment to both groups (Fig. 1). The next best supported NGSAdmix model was K=5, where Michipicoten Island and Pukaskwa NP were each assigned as a distinct population, while samples from the BCR were assigned to the three remaining populations (Fig. 1). The PCA reflected similar trends, where Michipicoten Island formed a tight cluster, caribou from Pukaskwa NP also clustered together but showed more spread among samples, and the BCR separated into 3 clusters (Fig. 1). Considering the population assignment results collectively we separated the samples into five groups (Table 1) to perform population-specific genotyping for

subsequent analyses: MICH (ON6 under SARA, 2012), PUK (ON6), BCR1 (ON1-5), BCR2 (ON7), BCR3 (ON8).

3.2 | Historical effective population sizes

There was some variance in the historical effective population sizes among population groupings but generally, the groups followed the same trajectories, with considerable quantile overlap (Fig. 2). During the late Pleistocene (approximately 100kya), the ancestors of populations from the BCR appear to have existed in relatively low abundance, whereas the ancestors of population groupings from the LSR had historically higher effective population sizes (Fig. 2). All ancestral populations demonstrated a steady increase in effective population size following the end of the Pleistocene, with an onset slightly before 10,000 years ago. Approximately 1,000 years ago, the populations plateaued and subsequently began to decline, although the default parameters provide limited insights on recent history compared to the windows focused on the most recent 1000 generations, starting approximately 7000 years before present (Fig. 2).

Using time windows focused on the most recent 1000 generations provided insight into the population sizes from 1000 to 100 years before present (Fig. 2B). All populations demonstrated similar trends over their recent history, steadily declining over the past 1000 years. Notably, under the more recent set of parameters, the modelling of deeper evolutionary history is more uncertain, as indicated by the 5% and 95% quantiles (Fig. 2B).

Interestingly, the largest sample grouping, based on both number of individuals and range size, (BCR1) did not have a larger historical or recent effective population size than the other populations. Under both sets of parameters, the ABC models indicated all populations reached a

maximum size at approximately the same time, roughly 5000 to 1000 years before present; although in general, all population estimates under the more recent time parameters appear slightly shifted forward in time when compared to the default parameters (Fig. 2).

4 | DISCUSSION

We sequenced whole genomes from 197 caribou to investigate the population histories of boreal caribou in Ontario, Canada. Contrary to our prediction, we found the largest population, based on both number of individuals and range size, (BCR1) did not have a larger effective population size than the other populations. We hypothesized if caribou from the LSR represent a recently diverged population, they would display similar patterns in demographic history with the other population groupings from the BCR. Overall, we found all caribou population groupings had similar demographic histories; following the late Pleistocene, the populations slowly expanded, plateaued, and declined at similar times. Notably, the ancestors of the populations from the LSR had larger estimated population sizes during the late Pleistocene. However, in general, all populations followed similar trajectories, with a relatively high degree of uncertainty as indicated by the quantiles. Collectively, our results indicate that caribou from the LSR and BCR do not have divergent demographic histories, and the observed genetic differentiation is likely due to recent isolation effects that occurred over the past century. This finding corroborates a previous genetic study that suggested shared ancestry between the LSR and BCR based on mtDNA (Klutsch et al. 2012).

4.1 | Population structure

We found evidence of population structure across the study region that was subsequently divided into five population groupings: three from the BCR, and two from the LSR. We hypothesized caribou from Michipicoten Island and Pukaskwa NP might be assigned to the same population grouping or show similar demographic histories because an individual from Pukaskwa NP was among the founders when the population was established on Michipicoten Island in the 1980s (Bergerud et al. 2007). These two population groupings from the LSR shared similar demographic histories; however, they did not group together under our population assignment tests (Fig. 1), confirming the inferences of previous research based on a single genome from Pukaskwa NP (Solmundson et al. 2023). Using a single individual to represent an entire population is typically not an ideal approach, but it may be necessary in the context of conservation research; for example, in cases where there are very few individuals remaining or a local extinction has already occurred. Through the use of historical archives, we provided 197 new whole genome sequences, including seven from the now apparently extirpated population in Pukaskwa NP. We also note that the field of conservation genomics currently lacks a standard set of protocols for the assembly, filtering, and analysis of wildlife sequence data. Thus, further insights could potentially be gained by future studies that attempt replicate our findings using other forms of genomic data or alternative demographic modelling approaches.

4.2 | Demographic histories

We explored changes in the effective population sizes throughout ancient and recent windows of time. The default model parameters provided insights into the deeper histories of these populations, corresponding to the Pleistocene when caribou populations persisted in glacial

refugia (Polfus et al. 2017). As the glaciers retreated, caribou expanded from their refugia and colonized new habitats (Klütsch, Manseau, Anderson, Sinkins, & Wilson, 2017; Klutsch et al. 2012), resulting in a slow, steady increase in the effective population size for all populations (Fig. 2A). The expansion and diversification of populations following the late Pleistocene has been documented in a wide range of taxa across several continents, including marsh deer in South America (Márquez et al. 2006), pit vipers in Asia (Ding, Gan, He, & Zhao, 2011), and fish in Australia (Bishop, Hughes, & Schmidt, 2018; Faulks, Gilligan, & Beheregaray, 2010). However, this phenomenon is particularly prevalent in mammals from boreal North America because the boreal region represents an extreme on the latitudinal gradient of biotic diversity (Lessa, Cook, & Patton, 2003).

Despite similar demographic histories across all populations, the two population groupings from the LSR had notably larger estimated population sizes during the late Pleistocene (approximately 100,000 years before present) than the other population groupings (Fig. 2A). All of the populations in this study are from the North American Lineage (NAL) of caribou, which means these groups were located in refugia south of the Laurentide Ice Sheet during the Pleistocene (Klutsch et al. 2012; Polfus et al. 2017). The observed differences in effective population sizes between LSR and BCR caribou may indicate that the LSR caribou persisted as a separate population, or perhaps in a different refugium than the BCR caribou, prior to the recession of the Laurentide Ice Sheet. Future studies should further investigate these findings, perhaps through the use other demographic models that focus further back in time such as MSCM-IM, which can incorporate isolation and migration events (Wang et al. 2020).

The ABC models focused on more recent history suggested all populations declined at similar times and rates over the past few hundred years (Fig. 2B). Over recent decades, several

population bottlenecks have occurred in the LSR (Bergerud et al. 2007; Ontario Ministry of Natural Resources and Forestry, 2018) but we did not find evidence of additional historical bottlenecks in LSR population groupings, which have been detected in similar studies of small and fragmented populations (Bemmels et al. 2021). This suggests that the bottlenecks in this population have occurred in the most recent 100 years, beyond the limits of model inference. Notably, there was a high degree of uncertainty around some estimates, such as the population grouping from Pukaskwa NP. Despite obtaining several archive samples from the cryobank at the Canadian Museum of Nature, our inferences are still limited by the relatively small sample size, which is an unresolvable constraint as caribou have been extirpated from this area.

4.3 | Conclusions

This study examines the population history of caribou in Ontario, providing insights for informed conservation and management strategies to aid in the preservation of this iconic species. We found evidence of genetic structure throughout the study area; however, in general, all populations shared similar demographic histories, indicating these groupings do not represent divergent evolutionary lineages. Although all populations shared similar histories, the LSR caribou had larger effective population sizes during the late Pleistocene than the population groupings from the BCR. We also found the population groupings from the LSR had the largest maximum historical population sizes and did not find evidence of dramatic historical bottlenecks, despite a known history of inbreeding (Solmundson et al. 2023). Collectively, our results indicate that the LSR and BCR do not have divergent evolutionary histories, and the observed differences in population structure are likely due to recent bottleneck effects that occurred over the past century.

REFERENCES

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

Bemmels, J. B., Mikkelsen, E. K., Haddrath, O., Colbourne, R. M., Robertson, H. A., & Weir, J. T. (2021). Demographic decline and lineage-specific adaptations characterize New Zealand kiwi. Proceedings of the Royal Society B: Biological Sciences, 288(1965). https://doi.org/10.1098/rspb.2021.2362

Bergerud, A. T., Dalton, W. J., Butler, H., Camps, L., & Ferguson, R. (2007). Woodland caribou persistence and extirpation in relic populations on Lake Superior. Rangifer, 27(17), 57–78.

Bishop, C. R., Hughes, J. M., & Schmidt, D. J. (2018). Mitogenomic analysis of the Australian lungfish (*Neoceratodus forsteri*) reveals structuring of indigenous riverine populations and late Pleistocene movement between drainage basins. Conservation Genetics, 19(3), 587–597. https://doi.org/10.1007/s10592-017-1034-7

Boitard, S., Rodríguez, W., Jay, F., Mona, S., & Austerlitz, F. (2016). Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. PLOS Genetics, 12(3), e1005877. https://doi.org/10.1371/journal.pgen.1005877

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

DePristo, M. a., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next- generation DNA sequencing data. Nat Genet, 43(5), 491–498. https://doi.org/10.1038/ng.806.A

Ding, L., Gan, X. N., He, S. P., & Zhao, E. M. (2011). A phylogeographic, demographic and historical analysis of the short-tailed pit viper (*Gloydius brevicaudus*): Evidence for early divergence and late expansion during the Pleistocene. Molecular Ecology, 20(9), 1905–1922. https://doi.org/10.1111/j.1365-294X.2011.05060.x

Faulks, L. K., Gilligan, D. M., & Beheregaray, L. B. (2010). Evolution and maintenance of divergent lineages in an endangered freshwater fish, Macquaria australasica. Conservation Genetics, 11(3), 921–934. https://doi.org/10.1007/s10592-009-9936-7

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(5), 419–434. https://doi.org/10.1139/z11-025

Hofreiter, M., & Stewart, J. (2009). Ecological Change, Range Fluctuations and Population Dynamics during the Pleistocene. Current Biology, 19(14), R584–R594. https://doi.org/10.1016/j.cub.2009.06.030

Keillor, P., Grant, W. S., Aquatic Sciences Center, & University of Wisconsin-Madison. (2003). Glacial Geology. In E. White, Wisconsin Sea Grant, Aquatic Sciences Center, & University of Wisconsin-Madison (Eds.), Living on the Coast. Detroit: US Army Corp of Engineers, University of Wisconsin Sea Grant Institute. Retrieved from https://www.lre.usace.army.mil/Missions/Great-Lakes-Information/Coastal-Program/Coastal-Processes/

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(12). https://doi.org/10.1111/jbi.13090

Klutsch, 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(12), 1–10. https://doi.org/10.1371/journal.pone.0052661

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15(1), 1–13. https://doi.org/10.1186/s12859-014-0356-4

Lessa, E. P., Cook, J. A., & Patton, J. L. (2003). Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proceedings of the National Academy of Sciences of the United States of America, 100(18), 10331–10334. https://doi.org/10.1073/pnas.1730921100

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27, 2987–2993.

Li, Heng, & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, Heng, Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lister, A. M. (2004). The impact of Quaternary Ice Ages on mammalian evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 359(1442), 221–241. https://doi.org/10.1098/rstb.2003.1436

Love Stowell, S. M., Bentley, E. G., Gagne, R. B., Gustafson, K. D., Rutledge, L. Y., & Ernest, H. B. (2018). Optimal DNA extractions from blood on preservation paper limits conservation genomic but not conservation genetic applications. Journal for Nature Conservation, 46(February), 89–96. https://doi.org/10.1016/j.jnc.2018.09.004

Márquez, A., Maldonado, J. E., González, S., Beccaceci, M. D., Garcia, J. E., & Duarte, J. M. B. (2006). Phylogeography and Pleistocene demographic history of the endangered marsh

deer (Blastocerus dichotomus) from the Río de la Plata Basin. Conservation Genetics, 7(4), 563–575. https://doi.org/10.1007/s10592-005-9067-8

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low Depth NGS Data. Genetics, 1–14. https://doi.org/10.1534/genetics.XXX.XXXXXX

Nadachowska-Brzyska, K., Konczal, M., & Babik, W. (2022). Navigating the temporal continuum of effective population size. Methods in Ecology and Evolution, 13(1), 22–41. https://doi.org/10.1111/2041-210X.13740

Ontario Ministry of Natural Resources. (2009). Ontario's woodland caribou conservation plan, 28pp.

Ontario Ministry of Natural Resources and Forestry. (2018). Seeking advice on the future of caribou in the Lake Superior coast range. Retrieved from https://ero.ontario.ca/notice/013-2587

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(2), 386–396. https://doi.org/10.1111/jbi.12918

Pond, B. A., Brown, G. S., Wilson, K. S., & Schaefer, J. A. (2016). Drawing lines: Spatial behaviours reveal two ecotypes of woodland caribou. Biological Conservation, 194(January), 139–148. https://doi.org/10.1016/j.biocon.2015.12.005

Primmer, C. R. (2009). From conservation genetics to conservation genomics. Annals of the New York Academy of Sciences, 1162, 357–368. https://doi.org/10.1111/j.1749-6632.2009.04444.x

R Core Team, . (2023). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

SARA. (2012). Recovery Strategy for the Woodland Caribou (*Rangifer tarandus caribou*), Boreal population, in Canada. Retrieved from www.sararegistry.gc.ca

Schaefer, J. A. (2003). Long-term range recession and the persistence of caribou in the taiga. Conservation Biology, 17(5), 1435–1439. https://doi.org/10.1046/j.1523-1739.2003.02288.x

Schweizer, R. M., Saarman, N., Ramstad, K. M., Forester, B. R., Kelley, J. L., Hand, B. K., … Luikart, G. (2021). Big data in conservation genomics: boosting skills, hedging bets, and staying current in the field. Journal of Heredity. https://doi.org/10.1093/jhered/esab019

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. Genetics, 195(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Solmundson, K., Bowman, J., Manseau, M., Taylor, R. S., Keobouasone, S., & Wilson, P. J. (2023). Genomic population structure and inbreeding history of Lake Superior caribou. Ecology and Evolution, 13(7), 1–14. https://doi.org/10.1002/ece3.10278

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. Genome Biology, 19(1). https://doi.org/10.1186/s13059-018-1520-3

Taylor, R. S., Manseau, M., Horn, R. L., Keobouasone, S., Golding, G. B., & Wilson, P. J. (2020). The role of introgression and ecotypic parallelism in delineating intra-specific conservation units. Molecular Ecology, (00), 1–17. https://doi.org/10.1111/mec.15522

Taylor, R. S., Manseau, M., Keobouasone, S., Mastromonaco, G., Solmundson, K., Kelly, A., … Paul, J. (2022). Continent-wide phylogenomic framework reveals introgression as a driver of intra-specific diversity and enriched molecular pathways in caribou. BioRxiv. https://doi.org/10.1101/2022.12.19.521038

Vors, L. S., Schaefer, J. A., Pond, B. A., Roders, A. R., & Patterson, B. R. (2007). Woodland Caribou Extirpation and Anthropogenic Landscape Disturbance in Ontario. Journal of Wildlife Management, 71(4), 1249–1256. https://doi.org/10.2193/2006-263

Statements & Declarations

Funding

This work was supported by funding provided through the Ontario Species at Risk Stewardship Program to KS and RT, Genomic Applications Partnership Program of Genome Canada and Environment and Climate Change Canada to PJW and MM, and NSERC Discovery grants to JB and PJW.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Sample collection was coordinated by KS, JB, MM, and PJW. KS prepared samples, extracted DNA, assembled genomic data, conducted analyses, created the figures, and wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Data Availability

Sequence data will be made available through NCBI upon acceptance for publication. Archive samples from Pukaskwa National Park were provided by the Canadian Museum of Nature. Annotated code for this study is available on GitHub (https://github.com/ksolmundson/LowCovGenomes). Spatial layers for mapping the ranges of Designatable Units were provided by the Committee on the Status of Endangered Wildlife in Canada.

Acknowledgments

Funding was provided by the Species at Risk Stewardship Program from Ontario's Ministry of Environment, Conservation, and Parks; the Genomic Applications Partnership Program of Genome Canada and Environment and Climate Change Canada; and Discovery Grants from the Natural Sciences and Engineering Research Council of Canada. We are grateful to Austin Thompson and Broderick Crosby for assisting with preparing and extracting DNA, and The Centre for Applied Genomics at the Hospital for Sick Children (Toronto, Ontario) for library preparation and sequencing. This research was conducted using high-performance computing services provided by the Digital Research Alliance of Canada (alliancecan.ca). We are grateful for the efforts of those who provided samples including the Canadian Museum of Nature, Government of Ontario, and Environment and Climate Change Canada.

Table 1. Summary of sample groupings of boreal caribou, determined through comparative analyses of low coverage whole genomes (N=197).

Sample grouping	Number of sequenced individuals	Range area
Michipicoten Island (MICH)	55	Lake Superior
Pukaskwa National Park (PUK)	7	Lake Superior
Ontario West (BCR1)	83	Boreal continuous
Ontario Central (BCR2)	16	Boreal continuous
Ontario East (BCR3)	36	Boreal continuous

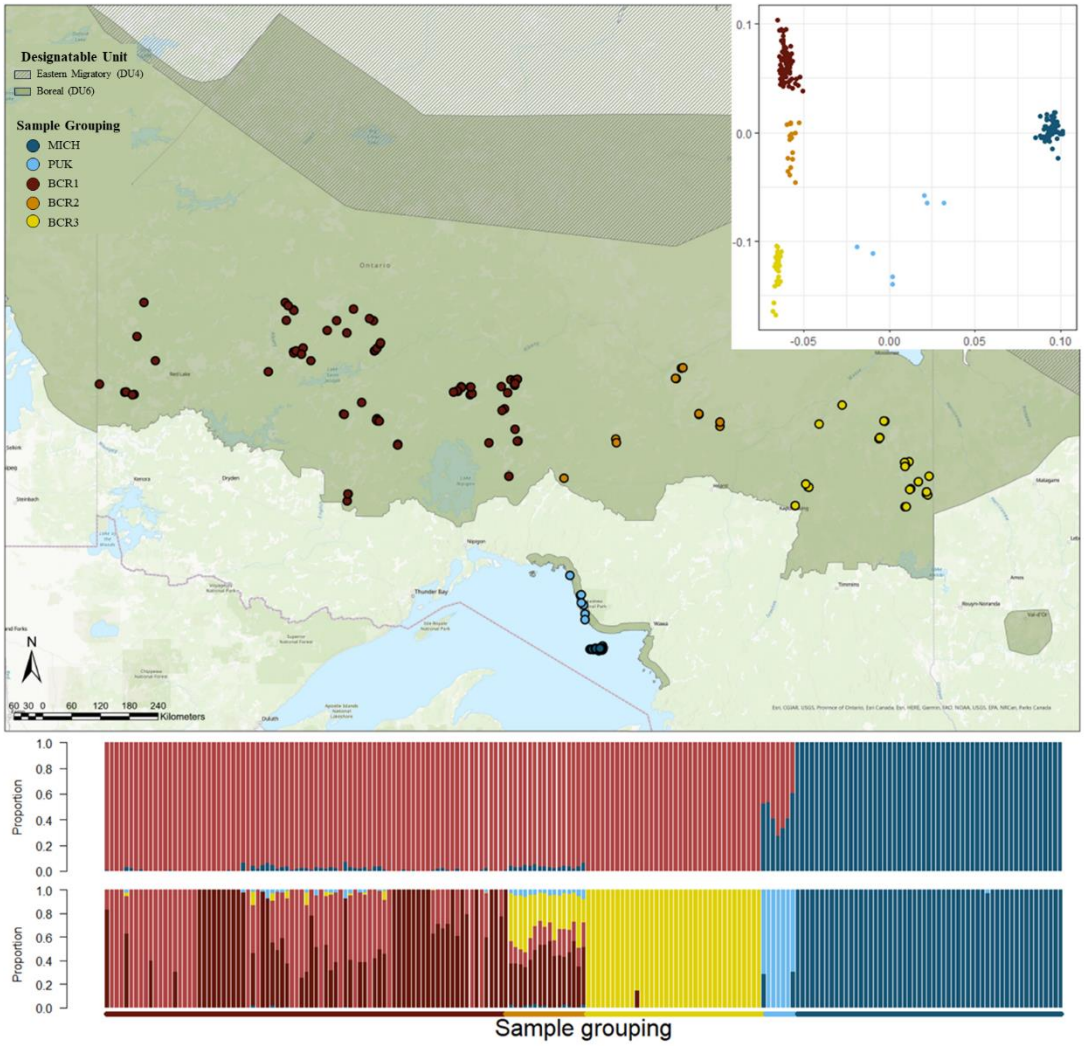

Fig. 1. Sample locations and population assignment results from whole genomes of boreal caribou from Ontario, Canada (N=197). The background layers of the map indicate the current range distributions. The colours of map location points, PCA points, and the bar beneath the NGSAdmix plots reflect the sample groupings used for grouped genotyping: MICH = Michipicoten Island; PUK = Pukaskwa National Park; BCR1, BCR2, BCR3 = boreal continuous range of caribou in Ontario.

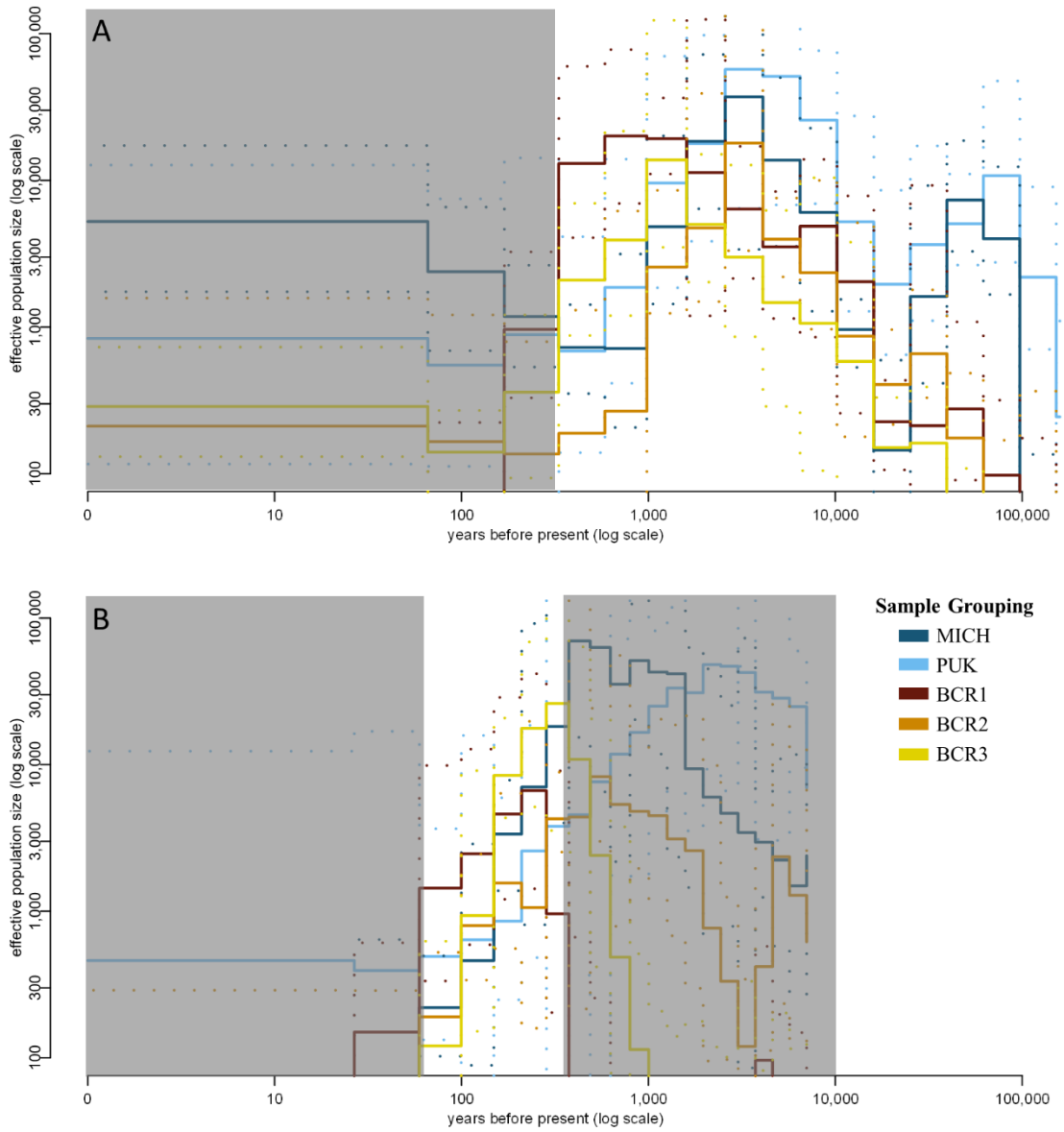

Fig. 2. Historical effective population sizes of boreal caribou groups modelled using an Approximate Bayesian Computation approach with 500,000 simulations. The upper plot (A) displays deeper evolutionary history with the model starting 130,000 generations ago; the plot reflects the most recent 100,000 years. The lower (B) used time windows focused on more recent history, corresponding to the most recent 1000 generations before present. Dotted lines indicate the 5% and 95% quantiles and shaded regions indicate areas of high model uncertainty.

SUPPLEMENTARY MATERIALS

Whole genomes reveal the demographic history of Ontario's boreal caribou

Table S1. Metadata for genomic samples: identifier, tissue type, collection year, range area (BCR vs LSR), and average genome-wide sequencing depth

Sample ID	Sample type	Collection year	Range area	Genome depth
22900	Whole blood	2010	Boreal continuous	0.96
22902	Whole blood	2010	Boreal continuous	1.01
22903	Whole blood	2010	Boreal continuous	1.11
22904	Whole blood	2010	Boreal continuous	1.03
22906	Whole blood	2010	Boreal continuous	1.08
22910	Whole blood	2010	Boreal continuous	1.03
22911	Whole blood	2010	Boreal continuous	1.04
22912	Whole blood	2010	Boreal continuous	0.87
22913	Whole blood	2010	Boreal continuous	1.05
22915	Whole blood	2010	Boreal continuous	0.92
22916	Whole blood	2010	Boreal continuous	1.15
22917	Whole blood	2010	Boreal continuous	0.96
22919	Whole blood	2010	Boreal continuous	1.03
22920	Whole blood	2010	Boreal continuous	1.05
22921	Whole blood	2010	Boreal continuous	1
22923	Whole blood	2010	Boreal continuous	0.93
22925	Whole blood	2010	Boreal continuous	1.03
22949	FTA card	2011	Boreal continuous	1.04
22950	FTA card	2011	Boreal continuous	0.91
22951	FTA card	2011	Boreal continuous	0.94
22952	FTA card	2011	Boreal continuous	1.06
22953	FTA card	2011	Boreal continuous	0.84

22954	FTA card	2011	Boreal continuous	1.07
22955	FTA card	2011	Boreal continuous	0.91
22956	FTA card	2011	Boreal continuous	0.97
22957	FTA card	2011	Boreal continuous	1.11
22958	FTA card	2011	Boreal continuous	0.98
22959	FTA card	2011	Boreal continuous	1.05
22960	FTA card	2011	Boreal continuous	0.97
22961	FTA card	2011	Boreal continuous	1.32
22963	FTA card	2011	Boreal continuous	1.23
22964	FTA card	2011	Boreal continuous	1.22
22965	FTA card	2011	Boreal continuous	1.06
22966	FTA card	2011	Boreal continuous	1.09
22967	FTA card	2011	Boreal continuous	1.15
22968	FTA card	2011	Boreal continuous	1.11
22969	FTA card	2011	Boreal continuous	1.13
22970	FTA card	2011	Boreal continuous	1.09
22971	FTA card	2011	Boreal continuous	1.18
22972	FTA card	2011	Boreal continuous	1.11
22973	FTA card	2011	Boreal continuous	1.15
22974	FTA card	2011	Boreal continuous	1.00
22976	FTA card	2011	Boreal continuous	1.19
22977	FTA card	2011	Boreal continuous	1.16
22978	FTA card	2011	Boreal continuous	1.00
22979	FTA card	2011	Boreal continuous	1.21
22980	FTA card	2011	Boreal continuous	1.11
22981	FTA card	2011	Boreal continuous	1.18
22982	FTA card	2011	Boreal continuous	1.07

22983	FTA card	2011	Boreal continuous	1.11
22984	FTA card	2011	Boreal continuous	1.11
22985	FTA card	2011	Boreal continuous	0.98
22986	FTA card	2011	Boreal continuous	0.96
23044	FTA card	2012	Boreal continuous	1.04
23046	FTA card	2012	Boreal continuous	1.18
23056	FTA card	2012	Boreal continuous	1.25
23060	FTA card	2012	Boreal continuous	0.94
23068	FTA card	2012	Boreal continuous	1.1
23069	FTA card	2012	Boreal continuous	1.08
23073	FTA card	2012	Boreal continuous	1.01
23080	FTA card	2012	Boreal continuous	1.03
23087	FTA card	2012	Boreal continuous	1.03
23200	FTA card	2011	Boreal continuous	1.05
23203	FTA card	2011	Boreal continuous	1.08
23204	FTA card	2011	Boreal continuous	1.00
23208	FTA card	2011	Boreal continuous	1.11
23210	FTA card	2011	Boreal continuous	1.21
23213	FTA card	2011	Boreal continuous	1.22
23214	FTA card	2011	Boreal continuous	1.2
23215	FTA card	2011	Boreal continuous	1.17
23216	FTA card	2011	Boreal continuous	1.24
23217	FTA card	2011	Boreal continuous	1.14
23224	FTA card	2011	Boreal continuous	1.1
23227	FTA card	2011	Boreal continuous	1.09
23228	FTA card	2011	Boreal continuous	1.02
23230	FTA card	2011	Boreal continuous	1.05

23231	FTA card	2011	Boreal continuous	0.97
23232	FTA card	2011	Boreal continuous	0.97
23233	FTA card	2012	Boreal continuous	1.14
23238	FTA card	2010	Boreal continuous	1.26
23239	FTA card	2010	Boreal continuous	0.98
23240	FTA card	2010	Boreal continuous	1.13
23241	FTA card	2010	Boreal continuous	1.06
23242	FTA card	2011	Boreal continuous	1.1
23243	FTA card	2011	Boreal continuous	1.18
23244	FTA card	2011	Boreal continuous	1.06
23245	FTA card	2011	Boreal continuous	1.12
23246	FTA card	2011	Boreal continuous	1.27
23250	FTA card	2011	Boreal continuous	0.93
23251	FTA card	2011	Boreal continuous	0.99
23252	FTA card	2011	Boreal continuous	1.03
23256	FTA card	2011	Boreal continuous	1.07
23257	FTA card	2011	Boreal continuous	1.23
23261	FTA card	2011	Boreal continuous	1.11
23262	FTA card	2011	Boreal continuous	1.13
23263	FTA card	2011	Boreal continuous	1.15
23265	FTA card	2011	Boreal continuous	1.2
23266	FTA card	2011	Boreal continuous	1.29
23268	FTA card	2011	Boreal continuous	1.11
23269	FTA card	2011	Boreal continuous	1.22
23271	FTA card	2011	Boreal continuous	1.02
23272	FTA card	2011	Boreal continuous	1.08
23273	FTA card	2012	Boreal continuous	1.33

23278	FTA card	2011	Boreal continuous	1.13
23280	FTA card	2011	Boreal continuous	1.03
23281	FTA card	2011	Boreal continuous	1.25
23283	FTA card	2011	Boreal continuous	1.1
23286	FTA card	2012	Boreal continuous	1.13
23289	FTA card	2012	Boreal continuous	1.15
23290	FTA card	2012	Boreal continuous	0.99
23293	FTA card	2012	Boreal continuous	1.14
23298	FTA card	2010	Boreal continuous	1.06
23303	FTA card	2010	Boreal continuous	1.02
23305	FTA card	2010	Boreal continuous	1.09
23307	FTA card	2010	Boreal continuous	1.12
23308	FTA card	2010	Boreal continuous	1.23
23312	FTA card	2010	Boreal continuous	1.05
23313	FTA card	2010	Boreal continuous	0.94
23316	FTA card	2010	Boreal continuous	1.11
23317	FTA card	2010	Boreal continuous	1.14
23319	FTA card	2010	Boreal continuous	1.39
23320	FTA card	2010	Boreal continuous	1.36
23322	FTA card	2010	Boreal continuous	1.08
23325	FTA card	2010	Boreal continuous	1.25
23327	FTA card	2010	Boreal continuous	1.11
23335	FTA card	2010	Boreal continuous	1.16
23336	FTA card	2010	Boreal continuous	1.08
23346	FTA card	2011	Boreal continuous	1.11
23348	FTA card	2011	Boreal continuous	1.37
23349	FTA card	2011	Boreal continuous	1.1

23352	FTA card	2011	Boreal continuous	1.15
23353	FTA card	2011	Boreal continuous	0.99
23356	FTA card	2012	Boreal continuous	1.25
23357	FTA card	2012	Boreal continuous	0.95
23358	FTA card	2012	Boreal continuous	1.68
44894	FTA card	2018	Lake Superior	2.13
44895	FTA card	2018	Lake Superior	1.97
50247	Cyrobank (whole blood)	1996	Lake Superior	2.09
50248	Cyrobank (whole blood)	1996	Lake Superior	1.76
50250	Cyrobank (extracted DNA)	1990	Lake Superior	1.83
50251	Cyrobank (extracted DNA)	1992	Lake Superior	1.79
50252	Cyrobank (extracted DNA)	1992	Lake Superior	1.99
50253	Cyrobank (extracted DNA)	1992	Lake Superior	1.62
50254	Cyrobank (extracted DNA)	1992	Lake Superior	1.84
50260	FTA card	2018	Lake Superior	1.41
50261	FTA card	2018	Lake Superior	1.55
50262	FTA card	2018	Lake Superior	1.44
50263	FTA card	2018	Lake Superior	1.35
50264	FTA card	2016	Lake Superior	1.92
50265	FTA card	2016	Lake Superior	1.94
50266	FTA card	2016	Lake Superior	2.09
50267	FTA card	2016	Lake Superior	1.9
50268	FTA card	2016	Lake Superior	2
50269	FTA card	2016	Lake Superior	1.74
50270	FTA card	2016	Lake Superior	1.87
50271	FTA card	2016	Lake Superior	1.85
50274	FTA card	2016	Lake Superior	1.73

50275	FTA card	2016	Lake Superior	1.84
50276	FTA card	2016	Lake Superior	1.85
50277	FTA card	2016	Lake Superior	1.73
50278	FTA card	2016	Lake Superior	2.08
50279	FTA card	2016	Lake Superior	2.04
50280	FTA card	2016	Lake Superior	2
50281	FTA card	2016	Lake Superior	2.09
50282	FTA card	2016	Lake Superior	1.95
50283	FTA card	2016	Lake Superior	2.06
50285	FTA card	2016	Lake Superior	2.08
50286	FTA card	2016	Lake Superior	2.03
50287	FTA card	2016	Lake Superior	2.09
50288	FTA card	2016	Lake Superior	2.14
50289	FTA card	2016	Lake Superior	2.13
50290	FTA card	2016	Lake Superior	1.95
50291	FTA card	2016	Lake Superior	1.75
50292	FTA card	2016	Lake Superior	1.92
50293	FTA card	2016	Lake Superior	1.95
50294	FTA card	2017	Lake Superior	1.87
50295	FTA card	2017	Lake Superior	1.97
50296	FTA card	2017	Lake Superior	2.01
50298	FTA card	2017	Lake Superior	1.99
50299	FTA card	2017	Lake Superior	2.25
50300	FTA card	2017	Lake Superior	1.99
50301	FTA card	2017	Lake Superior	1.99
50302	FTA card	2017	Lake Superior	2.12
50303	FTA card	2017	Lake Superior	1.92

50304	FTA card	2017	Lake Superior	2.05
50305	FTA card	2017	Lake Superior	2.26
50306	FTA card	2017	Lake Superior	1.95
50307	FTA card	2017	Lake Superior	2.12
50308	FTA card	2018	Lake Superior	1.99
50309	FTA card	2018	Lake Superior	1.85
50310	FTA card	2018	Lake Superior	2.06
50311	FTA card	2018	Lake Superior	2.4
50312	FTA card	2018	Lake Superior	2.08
50313	FTA card	2018	Lake Superior	1.99
50314	FTA card	2018	Lake Superior	1.91
50315	Hide	2015	Lake Superior	2.05
50316	Hide	2016	Lake Superior	1.97

Table S2. Probability values for NGSAdmix models of K=2-8. Values were calculated from the log likelihood results from 10 arrays for each value of K.

K=2	K=3	K=4	K=5	K=6	K=7	K=8
9.68×10^{9}	1.06×10^{3}	1.79×10^{3}	4.22×10^{3}	1.91×10^{3}	2.01×10^{3}	2.71×10^{3}

CHAPTER 5

CONCLUSIONS

Conservation Genomics

Globally many wildlife species are declining in range and abundance, necessitating conservation management to ensure their survival (Steffen et al. 2007). However, properly conserving wildlife requires knowledge of the evolutionary history of a species to make informed management decisions (COSEWIC 2020). Past evolutionary events, including bottlenecks, admixture, inbreeding, and selective pressures, influence the genetic composition of populations, resulting in variation among lineages (Coates et al. 2018). The caribou (*Rangifer tarandus*) is an iconic species with remarkable variation among lineages, attributable to local adaptation coupled with a complex history of evolutionary events such as bottlenecks and introgression (Polfus et al. 2017; Taylor et al. 2022). In species with complex evolutionary histories, such as the caribou, it can be difficult to categorize variation to inform conservation and management decisions such as the designation of management units. The field of genomics is shedding new light however, on the variation within and between species (Primmer 2009). Next-generation-sequencing (NGS) approaches allow us to gain new insights on recent and historic evolutionary events such as inbreeding and introgression (Hohenlohe et al. 2021). Unfortunately, the field of genomics is advancing so rapidly that there is an increasing gap between academics and wildlife managers, reducing the applicability to wildlife conservation (Britt et al. 2018). My dissertation research, conducted in collaboration with government wildlife managers, bridges the gap between genomics and conservation management, demonstrating how advances in DNA sequencing can be used to gain new insights on wildlife populations that are facing conservation risks.

Recommendations for Researchers

Given the precarious state of species at risk, we encourage researchers to use minimally invasive techniques, where possible, to conduct genomic studies. New advances allow WGS from non-optimal sources of DNA, such as fecal matter, hair, shed antlers, and dried blood (Love Stowell et al. 2018; Taylor et al. 2022b; Solmundson et al. 2023). Researchers can increase the yield of DNA from non-optimal sources by using high temperature incubation (Love Stowell et al. 2018) and concentrator columns (e.g. Amicon Ultra Centrifugal Filter Devices), prior to and post extraction respectively. If high quality tissue is required, researchers could opportunistically collect samples from roadkill or predation sites if conditions allow (e.g. the sample is relatively fresh or the environmental conditions are below freezing). Kill sites can also provide insights into phenotypic data, which can otherwise be collected non-invasively using trail cams (Peltier et al. 2023). Furthermore, we encourage researchers to make use of previously collected, archived data including museum databases, cryobanks, government archives, and DNA sequence archives.

We encourage genomic researchers to collaborate with non-academic partners to ensure projects are valuable to multiple interest groups. To ensure that research findings are applicable for conservation management, academic researchers should strive to partner with government agencies that are involved in wildlife management decisions (Britt et al. 2018). Collaborating with partners from diverse backgrounds, such as conservation agencies and other wildlife research organizations, can provide new insights and potential funding opportunities. Importantly, researchers should ensure the outcomes of the project are communicated in a way that is understandable to the relevant interest groups (Kadykalo et al. 2020).

Given the rapid advances and changes in the field of genomics, best practices include providing transparent, reproducible, publicly available data. We recommend that all raw read

DNA sequence data should be posted in a public repository (e.g. The National Center For Biotechnology Information Sequence Read Archive, NCBI SRA). Additionally, annotated code for analyses should be publicly available in a repository (e.g. GitHub) unless restricted by copyright. Finally, we encourage genomic researchers to share their findings in open-source publications or publicly available reports to increase the potential reach and allow greater uptake.

Translating Genomics to Wildlife Management

Wildlife populations are typically divided into management units that reflect their evolutionary lineages, which are sometimes referred to as Evolutionarily Significant Units or Distinct Population Segments (Coates et al. 2018; Funk et al. 2019). In Canada, wildlife populations that are considered to be both distinct and evolutionarily significant are categorized as Designatable Units (DUs) by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC 2020). Genomic data can help to delineate distinct populations through tests of population structure and genomic differentiation, such as principal component analyses (PCA) or population assignment tests like NGSAdmix (Skotte et al. 2013; Lou et al. 2020). Insights into evolutionary significance can be gained from phylogenomic trees and tests for adaptive differentiation (Supple and Shapiro 2018).

Whole genome sequencing (WGS) allows researchers to further explore the demographic histories of populations by estimating the effective population size (Ne) throughout time (Nadachowska-Brzyska et al. 2022). For species with complex evolutionary histories, WGS can provide insights on historical divergence, admixture, and introgression among lineages (Supple and Shapiro 2018). Categorization of genome-wide inbreeding patterns, or Runs of Homozygosity (ROH), can be a powerful tool for conservation, revealing not only the extent of inbreeding, but how recently inbreeding occurred based on the length of the ROH (Ceballos et al.

2018; Meyermans et al. 2020). Advances in bioinformatics have resulted in analysis tools that can incorporate evolutionary events such as isolation or migration, which are likely to be present in wild populations (Pickrell and Pritchard 2012; Wang et al. 2020).

The Lake Superior Caribou

Conservation management increasingly requires human intervention to ensure the long-term success of wild populations (Nielsen et al. 2022). Boreal caribou, a threatened DU (COSEWIC 2014), have lost approximately half of their historical range, resulting in small and isolated populations that persist on islands in the Lake Superior Range (LSR), over 100 km south of the boreal continuous range (BCR; Vors, Schaefer, Pond, Rodgers, & Patterson, 2007). Caribou historically inhabited the Lake Superior islands; however, by the 1980s, only one bull was sighted on Michipicoten Island when nine caribou were translocated from the Slate Islands to Michipicoten Island (Ontario Ministry of Natural Resources and Forestry 2018). In the absence of predation, this population grew to approximately 680 caribou by 2010 (Ontario Ministry of Natural Resources and Forestry 2018). However, in the winter of 2014, an ice bridge formed, allowing wolves to colonize the island. The Michipicoten Island caribou population quickly collapsed under the new predation pressure, prompting a translocation of some of the few remaining caribou to the Slate and Caribou Islands in early 2018 (Ontario Ministry of Natural Resources and Forestry, 2018).

The Slate Islands once had the highest density caribou population in North America (Bergerud et al. 2007); however, over the past decade the population had also collapsed. At the time of translocation in 2018, there appeared to be only two resident bulls remaining on the Slate Islands, indicating the population was functionally extirpated (Ontario Ministry of Natural Resources and Forestry 2018). Caribou Island is a small island, located relatively far from the

shore, potentially representing greater safety from predators but limited by a relatively small carrying capacity (Ontario Ministry of Natural Resources and Forestry 2018). Given their complex history and precarious future, the Lake Superior caribou populations provide a compelling subject for conservation genomics research.

In this dissertation, I used WGS from caribou in Manitoba, Ontario, and Quebec to address conservation questions in the LSR. In Chapters 2 and 3, I used 20 high-coverage (>10x) genomes that were sequenced from a variety of tissue types due to minimally invasive, opportunistic sample collection. Whole genomes were sequenced from hair, fecal pellets, a shed antler, and tissue samples collected from hunters, road-kill sites, and predation sites (Solmundson et al. 2023). I explored the evolutionary history of the LSR caribou in comparison to other boreal populations and other caribou ecotypes (specifically, eastern migratory and barren-ground). Chapter 4 provided a more comprehensive examination of the genomic population structure and demographic history of boreal caribou in Ontario, using 197 low-coverage (~1.4x) whole genomes. To generate the dataset, I extracted DNA from dried blood stored on FTA cards (Flinders Technology Associates, Whatman™), successfully using a data source that was previously thought to be impractical for WGS (Love Stowell et al. 2018).

Informing Caribou Management

Wildlife management actions can present collateral risks. For example, wildlife translocations can introduce parasites, disease, and outbreeding depression; however, translocating animals can successfully rescue genetically imperilled populations or re-populate areas of range loss (Banes et al. 2016; Robinson et al. 2020). In western Canada, a study examining the survival of translocated and native mountain caribou suggested translocation programs can stabilize or increase small and declining populations of species at risk (Grant et al.

2019). Genomics can provide insights into the individuals who are best suited for translocation or the most likely to survive. For instance, a recent study of translocated desert tortoises revealed that individual heterozygosity predicted tortoise survival, whereas translocation distance or geographic unit of origin did not (Scott et al. 2020).

WGS of the Lake Superior caribou revealed high levels of inbreeding in the LSR but shared ancestry and historical gene flow with BCR (Solmundson et al. 2023), and similar demographic histories within the LSR and BCR (Ch 3. Figure 2). These results confirm inferences from genetic data that suggested the BCR and LSR caribou share ancestry (Klutsch et al. 2012; Drake et al. 2018) and corroborate the current management designation as one DU (COSEWIC 2014). Analysis of 197 whole genomes of caribou from the BCR and LSR revealed genomic substructure within the BCR that is consistent with management range boundaries (Ch 3. Figure 1). Using genomic population assignment tests, we found the samples could be grouped into five populations, which predominantly reflected the recognized caribou ranges in Ontario, although the largest population grouping from the boreal continuous range encompassed several of the ranges (ON1-5) recognized under the Recovery Strategy for the Woodland Caribou (*Rangifer tarandus caribou*), Boreal population (Environment Canada 2011; SARA 2012).

The genomic results confirm the BCR and LSR have recently become disjunct due to anthropogenic range loss (Schaefer 2003), preventing gene flow that historically occurred. Thus, the LSR populations could potentially be stabilized by augmented gene flow from the BCR to boost population sizes and reduce inbreeding levels. If establishing future island populations, managers could include caribou from both the BCR and LSR; however, BCR alleles should generally represent less than half of the established population to reduce the risk of genetically swamping the recipient population (Frankham 2015). This balanced approach could help ensure

that any beneficial alleles that the LSR caribou possess (Solmundson et al. 2020) are introduced to the new island populations, while decreasing levels of inbreeding and genetic load. Furthermore, recurring translocations from the BCR to the island populations over multiple years could reduce future risks presented by isolation, recent population bottlenecks, and high inbreeding load (Morris et al. 2021). Additionally, when managing caribou from the LSR, individual inbreeding (Table S1) and relatedness levels (Table S2) should be considered given the known history of inbreeding and recent declines (Solmundson et al. 2023).

Conclusion

As the earth faces its sixth mass extinction event (Barnosky et al. 2011), genomic sequencing presents a viable tool for conservation researchers. Advances in sequencing technology now allow researchers to delve deep into the evolutionary history of each individual, providing insights into adaptation, admixture, and inbreeding (Supple and Shapiro 2018). Genomic data can offer new insights to inform conservation status listings and recovery decisions at a critical juncture (Funk et al. 2019). By broadening the datasets to encompass the whole genome, researchers can pose and address entirely new questions, yielding valuable insights that have direct applications to conservation. Although genetic approaches have a lengthy history in conservation, the transition to genomic technologies is still in its nascent stages; the potential benefits of conservation genomics are limited by a gap between genomic researchers and wildlife managers. In this dissertation, I used WGS to gain new insights on isolated caribou populations, a species at risk of extinction, demonstrating how genomics can be applied to inform conservation management.

REFERENCES

Banes GL, Galdikas BMF, Vigilant L (2016) Reintroduction of confiscated and displaced mammals risks outbreeding and introgression in natural populations, as evidenced by orang-utans of divergent subspecies. Sci Reports 2016 61 6:1–9. https://doi.org/10.1038/srep22026

Bergerud AT, Dalton WJ, Butler H, et al (2007) Woodland caribou persistence and extirpation in relic populations on Lake Superior. Rangifer 27:57–78

Britt M, Haworth SE, Johnson JB, et al (2018) The importance of non-academic coauthors in bridging the conservation genetics gap. Biol Conserv 218:118–123. https://doi.org/10.1016/j.biocon.2017.12.019

Ceballos FC, Joshi PK, Clark DW, et al (2018) Runs of homozygosity: Windows into population history and trait architecture. Nat Rev Genet 19:220–234. https://doi.org/10.1038/nrg.2017.109

Coates DJ, Byrne M, Moritz C (2018) Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. Front Ecol Evol 6:1–13. https://doi.org/10.3389/fevo.2018.00165

COSEWIC (2020) COSEWIC guidelines for recognizing designatable units. In: Gov. Canada. https://cosewic.ca/index.php/en-ca/reports/preparing-status-reports/guidelines-recognizing-designatable-units.html. Accessed 20 Feb 2023

COSEWIC (2014) COSEWIC assessment and status report on the caribou *Rangifer tarandus*, Newfoundland population, Atlantic-Gaspésie population and Boreal population, in Canada

Grant L, Johnson C, Thiessen C (2019) Evaluating the efficacy of translocation: maintaining habitat key to long-term success for an imperiled population of an at-risk species. Biodivers Conserv 28:2727–2743. https://doi.org/10.1007/s10531-019-01789-6

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

Lou RN, Jacobs A, Wilder A, Therkildsen NO (2020) A beginner's guide to low-coverage whole genome sequencing for population genomics. Authorea Prepr 1–68

Meyermans R, Gorssen W, Buys N, Janssens S (2020) How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. BMC Genomics 21:94. https://doi.org/10.1186/s12864-020-6463-x

Nadachowska-Brzyska K, Konczal M, Babik W (2022) Navigating the temporal continuum of effective population size. Methods Ecol Evol 13:22–41. https://doi.org/10.1111/2041-210X.13740

Nielsen ES, Hanson JO, Carvalho SB, Beger M (2022) Ecology & Evolution Molecular ecology meets systematic conservation planning. Trends Ecol Evol xx:1–13. https://doi.org/10.1016/j.tree.2022.09.006

Ontario Ministry of Natural Resources and Forestry (2018) Seeking advice on the future of caribou in the Lake Superior coast range

Pickrell JK, Pritchard JK (2012) Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet 8:. https://doi.org/10.1371/journal.pgen.1002967

Polfus JL, Manseau M, Klütsch CFC, et al (2017) Ancient diversification in glacial refugia leads to intraspecific diversity in a Holarctic mammal. J Biogeogr 44:386–396. https://doi.org/10.1111/jbi.12918

Primmer CR (2009) From conservation genetics to conservation genomics. Ann N Y Acad Sci 1162:357–368. https://doi.org/10.1111/j.1749-6632.2009.04444.x

Robinson ZL, Bell DA, Dhendup T, et al (2020) Evaluating the outcomes of genetic rescue attempts. Conserv Biol 35:666–677. https://doi.org/10.1111/cobi.13596

Scott PA, Allison LJ, Field KJ, et al (2020) Individual heterozygosity predicts translocation success in threatened desert tortoises. Science (80-) 370:1086–1089. https://doi.org/10.1126/science.abb0421

Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. Genetics 195:693–702. https://doi.org/10.1534/genetics.113.154138

Solmundson K, Bowman J, Manseau M, et al (2023) Genomic population structure and inbreeding history of Lake Superior caribou. Ecol Evol 13:1–14. https://doi.org/10.1002/ece3.10278

Steffen W, Crutzen PJ, McNeill JR (2007) The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature. AMBIO A J Hum Environ 36:614–621. https://doi.org/10.1579/0044-7447(2007)36[614:TAAHNO]2.0.CO;2

Supple MA, Shapiro B (2018) Conservation of biodiversity in the genomics era. Genome Biol 19:. https://doi.org/10.1186/s13059-018-1520-3

Taylor RS, Manseau M, Keobouasone S, et al (2022) Continent-wide phylogenomic framework reveals introgression as a driver of intra-specific diversity and enriched molecular pathways in caribou. bioRxiv. https://doi.org/10.1101/2022.12.19.521038

Vors LS, Schaefer JA, Pond BA, et al (2007) Woodland Caribou Extirpation and Anthropogenic Landscape Disturbance in Ontario. J Wildl Manage 71:1249–1256. https://doi.org/10.2193/2006-263

Wang K, Mathieson I, O'Connell J, Schiffels S (2020) Tracking human population structure through time from whole genome sequences. PLoS Genet 16:1–24. https://doi.org/10.1371/journal.pgen.1008552

SUPPLEMENTARY MATERIALS

CONCLUSIONS

We created a grouped population file using 55 low coverage genomes collected from Michipicoten Island following the protocol as described in Ch 3. Methods. The average depth of coverage across all 55 samples was 1.95x (range=1.35x - 2.4x; Ch 3. Table S1). The Michipicoten Island grouping includes caribou that were historically sampled on Michipicoten Island prior to 2018 (N=42), as well as some of the caribou that were relocated to the Slate (N=9) and Caribou (N=4) Islands in early 2018 (Ontario Ministry of Natural Resources and Forestry 2018).

We calculated individual inbreeding estimates (Table S1) using ngsF, which is recommended for low coverage genomic data because it incorporates genotype likelihood values into the calculations (Fumagalli et al. 2014). We compared relatedness levels (Table S2) with NGSRelate (Hanghøj et al. 2019), a program created for low coverage data that takes inbreeding into consideration, to calculate relatedness as R_{AB} (originally published as R_{XY}; Hedrick et al. 2015). A recent review comparing six genome-based relatedness estimators concluded R_{AB} produced the most consistent results when compared to known pedigree relatedness (Hauser et al. 2022). Under this metric, the expectation is 1 for an individual compared to itself and 0 for individuals within a population. The R_{AB} metric calculates relatedness based on genomic segments that share Identity by Descent (IBD); this is preferable to methods that calculate relatedness from Identity by State, which can overestimate relatedness under inbreeding (Hanghøj et al., 2019).

SUPPLEMENTARY REFERENCES

Fumagalli M, Vieira FG, Linderoth T, Nielsen R (2014) NgsTools: Methods for population genetics analyses from next-generation sequencing data. Bioinformatics 30:1486–1487. https://doi.org/10.1093/bioinformatics/btu041

Hanghøj K, Moltke I, Andersen PA, et al (2019) Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. Gigascience 8:1–9. https://doi.org/10.1093/gigascience/giz034

Hauser SS, Galla SJ, Putnam AS, et al (2022) Comparing genome-based estimates of relatedness for use in pedigree-based conservation management. Mol Ecol Resour 22:2546–2558. https://doi.org/10.1111/1755-0998.13630

Hedrick PPW, Lacy RC, Baker CS (2015) Measuring relatedness between inbred individuals. J Hered 106:20–25. https://doi.org/10.1093/jhered/esu072

Ontario Ministry of Natural Resources and Forestry (2018) Seeking advice on the future of caribou in the Lake Superior coast range

Table S1. Inbreeding estimates of caribou from Michipicoten Island. Samples were collected on the island prior to 2018 (Population group = Michipicoten Island) and during the translocation of caribou to the Slate and Caribou Islands.

Archive ID	PCID	Population group	Inbreeding estimate (F)
CMI001	50264	Michipicoten Island	0.0027
CMI003	50265	Michipicoten Island	0.0015
CMI004	50266	Michipicoten Island	0.0016
CMI005	50267	Michipicoten Island	0.003
CMI006	50268	Michipicoten Island	0.0021
CMI007	50269	Michipicoten Island	0.0024
CMI008	50270	Michipicoten Island	0.0024
CMI009	50271	Michipicoten Island	0.003
CMI011	50274	Michipicoten Island	0.0031
CMI012	50275	Michipicoten Island	0.0021
CMI013	50276	Michipicoten Island	0.0023
CMI014	50277	Michipicoten Island	0.0032
CMI015	50278	Michipicoten Island	0.0015
CMI016	50279	Michipicoten Island	0.002
CMI017	50280	Michipicoten Island	0.0023
CMI018	50281	Michipicoten Island	0.0016
CMI019	50282	Michipicoten Island	0.0027
CMI020	50283	Michipicoten Island	0.0024
CMI022	50285	Michipicoten Island	0.0018
CMI023	50286	Michipicoten Island	0.0023
CMI024	50287	Michipicoten Island	0.0022
CMI025	50288	Michipicoten Island	0.0015
CMI026	50289	Michipicoten Island	0.0019
CMI027	50290	Michipicoten Island	0.0018
CMI028	50291	Michipicoten Island	0.0019
CMI029	50292	Michipicoten Island	0.002
CMI030	50293	Michipicoten Island	0.0017
CMI031	50294	Michipicoten Island	0.0018
CMI032	50295	Michipicoten Island	0.0024
CMI033	50296	Michipicoten Island	0.0024
CMI035	50298	Michipicoten Island	0.0021
CMI036	50299	Michipicoten Island	0.0021
CMI037	50300	Michipicoten Island	0.002
CMI038	50301	Michipicoten Island	0.0019
CMI039	50302	Michipicoten Island	0.0021
CMI040	50303	Michipicoten Island	0.0023
CMI041	50304	Michipicoten Island	0.0016

CMI042	50305	Michipicoten Island	0.0019
CMI044	50306	Michipicoten Island	0.0015
CMI045	50307	Michipicoten Island	0.0015
MIH001	50315	Michipicoten Island	0.0018
MIH002	50316	Michipicoten Island	0.0019
CSI001	44894	Slate Islands	0.0013
CSI005	44895	Slate Islands	0.0022
CSI006	50308	Slate Islands	0.0015
CSI007	50309	Slate Islands	0.002
CSI008	50310	Slate Islands	0.0019
CSI009	50311	Slate Islands	0.0012
CSI010	50312	Slate Islands	0.0014
CSI001	50313	Slate Islands	0.0017
CSI005	50314	Slate Islands	0.0029
CSI006	50260	Caribou Island	0.0033
CSI007	50261	Caribou Island	0.0032
CSI008	50262	Caribou Island	0.0033
CSI009	50263	Caribou Island	0.0038

Table S2. Pairwise relatedness (R_{AB}) estimates of caribou from Michipicoten Island. PCID indicates the individual genomic sample identifier.

PCID	44895	50260	50261	50262	50263	50264	50265	50266	50267	50268	50269	50270	50271	50274	50275	50276	50277	50278	50279	50280	50281	50282	50283	50285	50286	50287	50288	50289	50290	50291	50292	50293	50294	50295	50296	50298	50299	50300	50301	50302	50303	50304	50305	50306	50307	50308	50309	50310	50311	50312	50313	50314	50315	50316
44894	0.07	0.01	0.01	0.00	0.00	0.04	0.01	0.00	0.03	0.00	0.02	0.00	0.00	0.03	0.07	0.03	0.03	0.01	0.04	0.03	0.03	0.02	0.04	0.05	0.01	0.00	0.01	0.00	0.06	0.00	0.02	0.03	0.00	0.00	0.00	0.08	0.01	0.01	0.05	0.00	0.02	0.03	0.02	0.00	0.05	0.04	0.02	0.05	0.01	0.00	0.00	0.02	0.02	0.01
44895		0.04	0.00	0.02	0.03	0.00	0.00	0.02	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.03	0.03	0.00	0.05	0.04	0.00	0.05	0.01	0.03	0.00	0.03	0.00	0.00	0.06	0.06	0.00	0.00	0.03	0.00	0.01	0.01	0.05	0.04	0.06	0.00	0.01	0.01	0.06	0.06	0.02	0.03	0.05	0.08	0.00	0.01	0.00	0.00	0.02	0.00
50260			0.00	0.00	0.00	0.00	0.03	0.02	0.00	0.05	0.03	0.03	0.01	0.02	0.00	0.04	0.03	0.00	0.09	0.01	0.02	0.05	0.02	0.01	0.00	0.00	0.02	0.06	0.00	0.02	0.00	0.01	0.03	0.02	0.00	0.03	0.01	0.00	0.02	0.06	0.00	0.00	0.00	0.02	0.10	0.04	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.02
50261				0.08	0.02	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.00	0.01	0.06	0.08	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.04	0.04	0.00	0.04	0.00	0.02	0.06	0.02	0.17	0.04	0.02	0.00	0.03	0.04	0.01	0.03	0.04	0.01	0.00	0.02	0.00	0.00
50262					0.01	0.00	0.01	0.00	0.04	0.05	0.01	0.01	0.00	0.03	0.00	0.02	0.00	0.06	0.01	0.00	0.00	0.00	0.03	0.00	0.02	0.06	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.01	0.00	0.02	0.00	0.01	0.03	0.05	0.36	0.03	0.00	0.00	0.04	0.00	0.01	0.02	0.00	0.02	0.00	0.00	0.02	0.02
50263						0.00	0.01	0.00	0.00	0.02	0.03	0.04	0.03	0.00	0.11	0.04	0.00	0.01	0.00	0.05	0.03	0.01	0.02	0.00	0.03	0.05	0.00	0.00	0.01	0.00	0.02	0.05	0.02	0.02	0.01	0.03	0.00	0.05	0.00	0.01	0.00	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.05	0.07	0.03	0.00	0.02	0.06
50264							0.05	0.03	0.05	0.00	0.02	0.03	0.05	0.03	0.03	0.00	0.00	0.03	0.00	0.06	0.05	0.00	0.03	0.02	0.02	0.00	0.05	0.02	0.00	0.00	0.05	0.00	0.00	0.06	0.07	0.01	0.06	0.02	0.00	0.00	0.02	0.02	0.04	0.00	0.01	0.02	0.00	0.01	0.04	0.02	0.05	0.08	0.02	0.00
50265								0.05	0.11	0.07	0.03	0.06	0.03	0.03	0.00	0.02	0.00	0.01	0.00	0.00	0.09	0.01	0.05	0.01	0.03	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.09	0.03	0.05	0.04	0.00	0.01	0.09	0.01	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.00	0.03	0.03
50266									0.06	0.01	0.05	0.01	0.01	0.01	0.01	0.03	0.01	0.00	0.03	0.00	0.09	0.04	0.00	0.01	0.00	0.01	0.06	0.01	0.00	0.00	0.03	0.05	0.06	0.03	0.02	0.03	0.07	0.00	0.05	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.04	0.01	0.01	0.00	0.01	0.03	0.01	0.02
50267										0.03	0.05	0.00	0.20	0.03	0.00	0.03	0.04	0.02	0.03	0.00	0.05	0.04	0.04	0.07	0.00	0.00	0.01	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.04	0.04	0.02	0.03	0.00	0.04	0.15	0.03	0.00	0.00
50268											0.02	0.01	0.03	0.05	0.03	0.09	0.04	0.00	0.08	0.00	0.01	0.00	0.09	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.02	0.04	0.05	0.01	0.08	0.00	0.00	0.06	0.01	0.01	0.00	0.01	0.05	0.01	0.01	0.01	0.04	0.00	0.02	0.00	0.03	0.01
50269												0.00	0.04	0.01	0.03	0.09	0.00	0.01	0.06	0.00	0.03	0.00	0.00	0.00	0.02	0.06	0.03	0.02	0.00	0.00	0.01	0.01	0.00	0.06	0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.03	0.04	0.00	0.02	0.01	0.02	0.03	0.03	0.01
50270													0.09	0.00	0.00	0.01	0.00	0.01	0.02	0.01	0.00	0.00	0.01	0.00	0.03	0.01	0.02	0.01	0.00	0.04	0.04	0.05	0.00	0.01	0.00	0.04	0.03	0.04	0.02	0.04	0.00	0.01	0.03	0.04	0.04	0.00	0.01	0.00	0.03	0.07	0.03	0.00	0.02	0.06
50271														0.03	0.00	0.02	0.05	0.00	0.04	0.00	0.00	0.01	0.03	0.04	0.00	0.00	0.00	0.11	0.00	0.00	0.01	0.00	0.00	0.04	0.01	0.00	0.08	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.01	0.02	0.00	0.01	0.16	0.00	0.00	0.00
50274															0.00	0.00	0.06	0.05	0.04	0.00	0.00	0.00	0.03	0.05	0.01	0.00	0.00	0.02	0.00	0.01	0.01	0.00	0.03	0.05	0.05	0.01	0.03	0.00	0.00	0.00	0.07	0.01	0.00	0.01	0.02	0.00	0.02	0.05	0.03	0.00	0.01	0.02	0.05	0.04
50275																0.03	0.10	0.01	0.00	0.02	0.01	0.05	0.01	0.01	0.06	0.04	0.01	0.02	0.05	0.02	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.05	0.01	0.00	0.03	0.02	0.00	0.02	0.00	0.02	0.02	0.03	0.03
50276																	0.00	0.00	0.02	0.03	0.04	0.03	0.02	0.00	0.00	0.10	0.03	0.01	0.00	0.00	0.02	0.01	0.00	0.02	0.00	0.05	0.03	0.00	0.00	0.11	0.00	0.05	0.01	0.00	0.03	0.00	0.01	0.01	0.00	0.01	0.00	0.02	0.00	0.00
50277																		0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.03	0.04	0.00	0.00	0.04	0.04	0.01	0.00	0.06	0.00	0.00	0.00	0.00	0.01	0.00	0.04	0.00	0.06	0.01	0.04	0.05	0.00	0.00	0.01	0.04	0.05
50278																			0.02	0.00	0.01	0.00	0.00	0.05	0.02	0.05	0.05	0.03	0.04	0.00	0.01	0.00	0.02	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.03	0.05	0.00	0.00	0.01	0.00	0.00	0.04	0.04	0.00	0.01	0.14	0.03	0.02
50279																				0.02	0.07	0.00	0.02	0.03	0.00	0.03	0.02	0.04	0.01	0.02	0.00	0.00	0.05	0.03	0.04	0.01	0.04	0.00	0.02	0.05	0.03	0.01	0.00	0.02	0.02	0.05	0.00	0.01	0.01	0.00	0.01	0.02	0.01	0.00
50280																					0.01	0.05	0.00	0.02	0.02	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.03	0.00	0.00	0.04	0.00	0.02	0.02	0.01	0.01	0.01	0.04	0.01	0.05	0.02	0.02	0.09	0.01	0.14	0.02	0.02	0.00	0.01
50281																						0.05	0.01	0.01	0.00	0.01	0.05	0.03	0.00	0.00	0.02	0.05	0.02	0.04	0.00	0.06	0.02	0.00	0.02	0.00	0.02	0.06	0.01	0.00	0.03	0.01	0.01	0.00	0.00	0.04	0.02	0.00	0.00	0.03
50282																							0.00	0.02	0.00	0.03	0.03	0.02	0.02	0.02	0.04	0.02	0.04	0.00	0.00	0.03	0.00	0.02	0.04	0.03	0.01	0.03	0.02	0.02	0.03	0.05	0.04	0.01	0.00	0.04	0.00	0.01	0.00	0.04
50283																								0.05	0.01	0.00	0.02	0.01	0.03	0.00	0.03	0.02	0.03	0.01	0.00	0.00	0.04	0.01	0.02	0.04	0.03	0.02	0.01	0.02	0.19	0.00	0.02	0.03	0.02	0.00	0.02
50285																									0.01	0.00	0.02	0.00	0.03	0.01	0.02	0.00	0.04	0.00	0.00	0.00	0.11	0.01	0.01	0.02	0.01	0.08	0.00	0.01	0.00	0.02	0.02	0.05	0.02	0.03	0.01	0.03	0.00	0.00
50286																										0.05	0.00	0.00	0.02	0.01	0.07	0.04	0.01	0.05	0.00	0.03	0.00	0.21	0.00	0.07	0.03	0.06	0.04	0.01	0.01	0.02	0.03	0.01	0.03	0.01	0.00	0.01	0.12	0.00
50287																											0.01	0.00	0.00	0.00	0.02	0.05	0.03	0.00	0.00	0.06	0.00	0.03	0.02	0.01	0.09	0.03	0.00	0.00	0.04	0.02	0.00	0.00	0.02	0.01	0.03	0.01	0.03	0.01
50288																												0.03	0.05	0.00	0.01	0.03	0.02	0.01	0.04	0.03	0.02	0.00	0.01	0.00	0.00	0.08	0.00	0.00	0.03	0.08	0.00	0.02	0.04	0.04	0.08	0.00	0.00
50289																													0.02	0.01	0.01	0.00	0.03	0.04	0.05	0.00	0.02	0.00	0.04	0.00	0.00	0.00	0.01	0.01	0.02	0.01	0.02	0.01	0.03	0.00	0.03	0.02	0.02	0.06
50290																														0.05	0.01	0.00	0.00	0.00	0.04	0.00	0.04	0.03	0.00	0.00	0.00	0.06	0.08	0.05	0.00	0.01	0.05	0.04	0.09	0.07	0.01	0.00	0.03	0.00
50291																															0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.05	0.00	0.00	0.01	0.00	0.69	0.02	0.00	0.03	0.00	0.01	0.03	0.00	0.00	0.00	0.05
50292																																0.18	0.00	0.01	0.01	0.04	0.01	0.01	0.00	0.05	0.03	0.04	0.00	0.00	0.02	0.04	0.00	0.00	0.02	0.06	0.00	0.02	0.06	0.03
50293																																	0.02	0.03	0.02	0.03	0.00	0.02	0.01	0.04	0.02	0.03	0.01	0.01	0.01	0.03	0.00	0.02	0.00	0.00	0.02	0.00	0.09	0.05
50294																																		0.00	0.00	0.03	0.06	0.06	0.06	0.02	0.02	0.04	0.00	0.01	0.01	0.00	0.00	0.05	0.00	0.02	0.00	0.05	0.04	0.03
50295																																			0.08	0.02	0.01	0.04	0.00	0.03	0.00	0.00	0.05	0.00	0.01	0.04	0.02	0.00	0.04	0.00	0.01	0.00	0.06	0.00
50296																																				0.00	0.08	0.04	0.00	0.00	0.00	0.00	0.05	0.00	0.02	0.04	0.03	0.00	0.05	0.05	0.09	0.01	0.04	0.00
50298																																					0.00	0.00	0.02	0.02	0.04	0.04	0.04	0.01	0.06	0.02	0.00	0.00	0.00	0.02	0.01	0.02	0.00	0.04
50299																																						0.00	0.02	0.00	0.00	0.01	0.00	0.01	0.00	0.03	0.00	0.02	0.02	0.02	0.00	0.00	0.00	0.00
50300																																							0.00	0.03	0.02	0.01	0.05	0.00	0.03	0.00	0.01	0.01	0.02	0.03	0.10	0.00	0.03	0.02
50301																																								0.01	0.05	0.05	0.01	0.05	0.01	0.04	0.02	0.02	0.06	0.01	0.00	0.00	0.04	0.03
50302																																									0.04	0.03	0.00	0.00	0.01	0.01	0.03	0.00	0.00	0.01	0.02	0.03	0.02	0.02
50303																																										0.05	0.01	0.00	0.09	0.03	0.01	0.01	0.00	0.02	0.00	0.01	0.01	0.03
50304																																											0.02	0.02	0.01	0.00	0.00	0.03	0.00	0.01	0.00	0.03	0.01	0.02
50305																																												0.00	0.00	0.00	0.04	0.03	0.08	0.01	0.04	0.01	0.00	0.01
50306																																													0.02	0.00	0.03	0.00	0.01	0.03	0.00	0.00	0.00	0.05
50307																																														0.03	0.00	0.03	0.00	0.03	0.02	0.00	0.01	0.04
50308																																															0.04	0.04	0.03	0.01	0.00	0.04	0.03	0.00
50309																																																0.04	0.03	0.02	0.02	0.01	0.02	0.00
50310																																																	0.00	0.01	0.00	0.00	0.01	0.00
50311																																																		0.02	0.03	0.04	0.03	0.02
50312																																																			0.04	0.00	0.00	0.05
50313																																																				0.03	0.00	0.03
50314																																																						0.00	0.03
50315																																																							0.02

Appendix