Article

Current Biology

High genetic load without purging in caribou, a diverse species at risk

Highlights

- Caribou are genetically diverse and form nine lineages across North America
- Levels of inbreeding vary, with some highly inbred caribou in some locations
- Regardless of inbreeding level, high genetic load is observed across all lineages
- Genetically diverse species may be especially vulnerable to inbreeding depression

Authors

Rebecca S. Taylor, Micheline Manseau, Sonesinh Keobouasone, ..., Dave Hervieux, Deborah Simmons, Paul J. Wilson

Correspondence

rebecca.taylor@ec.gc.ca

In brief

Taylor et al. use whole-genome sequences of caribou in North America, a genetically diverse yet declining species, and find nine phylogenomic lineages discordant from conservation units. Results show evidence for historical purging of deleterious alleles but no increased purging even with extremely high inbreeding in some populations.

Taylor et al., 2024, Current Biology *34*, 1–13
March 25, 2024
https://doi.org/10.1016/j.cub.2024.02.002

CellPress

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

Current Biology

CellPress
OPEN ACCESS

Article

High genetic load without purging in caribou, a diverse species at risk

Rebecca S. Taylor,[1,12,13,*] Micheline Manseau,[1] Sonesinh Keobouasone,[1] Peng Liu,[1] Gabriela Mastromonaco,[2] Kirsten Solmundson,[3] Allicia Kelly,[4] Nicholas C. Larter,[4] Mary Gamberg,[5] Helen Schwantje,[6] Caeley Thacker,[6] Jean Polfus,[7] Leon Andrew,[8] Dave Hervieux,[9] Deborah Simmons,[8,11] and Paul J. Wilson[10]

[1]Landscape Science and Technology, Environment and Climate Change Canada, Colonel By Drive, Ottawa, ON K1S 5B6, Canada
[2]Wildlife Science, Toronto Zoo, Meadowvale Road, Toronto, ON M1B 5K7, Canada
[3]Environmental & Life Sciences Graduate Program, Trent University, Peterborough, ON K9L 1Z8, Canada
[4]Department of Environment and Natural Resources, Government of Northwest Territories, PO Box 900, Fort Smith, NT X0E 0P0, Canada
[5]Gamberg Consulting, Jarvis Street, Whitehorse, YK Y1A 2J2, Canada
[6]British Columbia Ministry of Forest, Lands, Natural Resource Operations, and Rural Development, Labieux Road, Nanaimo, BC V9T 6E9, Canada
[7]Canadian Wildlife Service - Pacific Region, Environment and Climate Change Canada, 1238 Discovery Avenue, Kelowna, BC V1V 1V9, Canada
[8]ʔehdzo Got'ıné Gots'ę Nákedı (Sahtú Renewable Resources Board), P.O. Box 134, Tulít'a, NT X0E 0K0, Canada
[9]Alberta Ministry of Environment and Protected Areas, Government of Alberta, 10320-99 Street, Grande Prairie, AB T8V 6J4, Canada
[10]Biology Department, Trent University, East Bank Drive, Peterborough, ON K9L 1Z8, Canada
[11]Deceased
[12]X (formerly Twitter): @BeckySTaylor
[13]Lead contact
*Correspondence: rebecca.taylor@ec.gc.ca
https://doi.org/10.1016/j.cub.2024.02.002

SUMMARY

High intra-specific genetic diversity is associated with adaptive potential, which is key for resilience to global change. However, high variation may also support deleterious alleles through genetic load, thereby increasing the risk of inbreeding depression if population sizes decrease. Purging of deleterious variation has been demonstrated in some threatened species. However, less is known about the costs of declines and inbreeding in species with large population sizes and high genetic diversity even though this encompasses many species globally that are expected to undergo population declines. Caribou is a species of ecological and cultural significance in North America with a wide distribution supporting extensive phenotypic variation but with some populations undergoing significant declines resulting in their at-risk status in Canada. We assessed intra-specific genetic variation, adaptive divergence, inbreeding, and genetic load across populations with different demographic histories using an annotated chromosome-scale reference genome and 66 whole-genome sequences. We found high genetic diversity and nine phylogenomic lineages across the continent with adaptive diversification of genes, but also high genetic load among lineages. We found highly divergent levels of inbreeding across individuals, including the loss of alleles by drift but not increased purging in inbred individuals, which had more homozygous deleterious alleles. We also found comparable frequencies of homozygous deleterious alleles between lineages regardless of nucleotide diversity. Thus, further inbreeding may need to be mitigated through conservation efforts. Our results highlight the "double-edged sword" of genetic diversity that may be representative of other species atrisk affected by anthropogenic activities.

INTRODUCTION

Intra-specific diversity is crucial for adaptive potential and resilience of species under environmental changes.[1–4] Therefore, understanding the drivers of intra-specific genetic variation and its interplay with adaptive divergence is essential to understanding how current species respond to environmental variations.[5–7] Conversely, there is growing evidence suggesting that populations with high genetic diversity have a larger inbreeding or masked load which is defined as the reduction of fitness caused by deleterious recessive mutations which are unmasked by inbreeding, a category of overall genetic load which is the actual or potential loss of fitness due to deleterious alleles.[8–10] Recent research on threatened populations or species with low genetic diversity has demonstrated the purging of some putatively deleterious genetic variation, for example, in the Sumatran rhinoceros,[11] the kākāpō,[12] Alpine ibex,[13] and Indian tigers.[14] Some threatened species nevertheless developed inbreeding depression likely due to historical demography and a constant influx of deleterious mutations with selection not sufficiently purging high enough

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

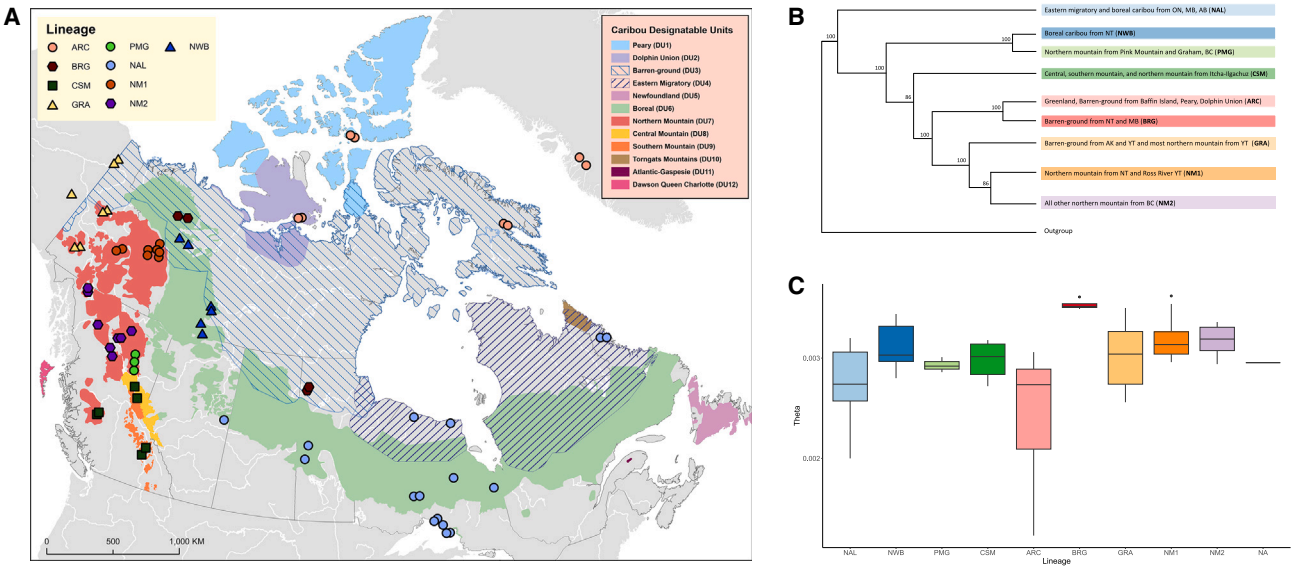


Figure 1. Map and intra-specific diversity of sampled individuals

(A) A map showing the sample location of each of the individuals sequenced. The background shading indicates the Designatable Unit (DU) ranges, and the colors of the points indicate which phylogenomic lineage the individual belongs to.

(B) Simplified version of the whole-genome phylogenomic reconstruction showing the caribou in each lineage. ON, MB, and AB refer to Canadian provinces (Ontario, Manitoba, and Alberta, respectively).

(C) Genetic diversity (θ) for each lineage. See also Figures S1 and S2 and Tables S1 and S2.

proportions, as well as the interplay between the level of deleteriousness and the dominance of deleterious alleles.[15,16]

Less is known about the costs of population declines and inbreeding in species with large population sizes and high genetic diversity,[17] even though this encompasses many species globally that will likely undergo rapid declines and fragmentation into isolated populations due to anthropogenic impacts. Such species with large historical effective populations sizes are expected to have high inbreeding load and thus may be at high risk of inbreeding depression upon a population bottleneck when compared with populations that have maintained small sizes over time.[8,9] We investigate these processes in an example of such a species, caribou (*Rangifer tarandus*), a widespread and diverse species at risk.

Caribou (known as reindeer in Eurasia) is a highly mobile species with a wide distribution, ranging from the high Arctic to the boreal forests, and spanning from the east to the west coast of North America[18] (Figure 1A). Across its range, caribou have a large amount of phenotypic and genetic variation and have been divided into 12 conservation units, known as Designatable Units (DUs), by the Committee on the Status of Endangered Wildlife in Canada[18] (Figure 1A). Caribou DUs face threats including habitat destruction and climate change,[19–21] with nine DUs currently listed as endangered or threatened, two as special concern, and one that is extinct.[18,22] Globally, in 2015 the species changed from least concern to vulnerable on the International Union for Conservation of Nature (IUCN) Red List due to the species undergoing a 40% decline over three generations.[23]

We investigated intra-specific lineage diversity, genetic variation, adaptive diversification, inbreeding extent, and genetic load in caribou across North America and Greenland using whole-genome sequencing thus undertaking a comprehensive reconstruction of intra-specific caribou diversity. We first assess intra-specific diversity and the processes that may have led to high genetic variation, as well as adaptive diversification. We then characterize inbreeding and compare genetic load in individuals with different demographic histories (high vs low inbreeding) to understand the potential impact of population declines on genetic diversity as well as on deleterious variation in a genetically and phenotypically diverse species.

RESULTS AND DISCUSSION

Phylogenomic, genetic, and demographic diversity

We assembled a new caribou reference genome with a contig N50 of 32.82 KB, scaffold N50 of 64.42 MB, and an L50 of 14, with 99.5% of the assembly being on 36 scaffolds. As the chromosome number is 70 for the species[24] (34 autosome pairs plus the sex chromosomes), this likely represents a chromosome-scale assembly. We used RNA-seq data to perform a high-quality annotation of the genome, which identified the locations of 34,407 protein-coding genes. Using 66 re-sequenced genomes from across North America and Greenland representing eight DUs and 33 subpopulations (Figure 1A; Tables S1 and S2), phylogenomic reconstruction using two different methods was generally consistent and separated caribou into nine major lineages (reciprocally monophyletic groups of individuals). Lineages were not concordant with DU designations (Figures 1B and S1), likely due to their designation based on life history, morphology, as well as some genetic data.[18] Our whole-genome phylogenies reconstructed the two major lineages known from previous studies using mitochondrial DNA[25]; the North American lineage (NAL) and the larger and more diverse Beringian-Eurasian

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

lineage (BEL), which contains eight lineages in our results (Figures 1B and S1). The only discordance between our two reconstructions was that the NWB lineage was basal to the PMG lineage instead of a sister group, and that the BRG lineage individuals were within (although on the outside) of the ARC group instead of a sister lineage in the SNP based reconstruction (Figure S1B). We show the results based on the more powerful method using full sequence data (Figure 1A), especially given the large differences between BRG and ARC individuals in further analyses. The principal component analysis (PCA) is also concordant with the phylogenomic results separating into three major clusters on PC1: the NAL individuals, NWB, PMG, and CSM individuals, and the three major northern mountain lineages (GRA, NM1, and NM2; Figure S2A). On PC2, the BRG and ARC lineages separate (Figure S2A). ADMIXTURE similarly separates the NAL and BEL at K = 2 (both K = 1 and K = 2 had the lowest and very similar cross validation error; 0.287 and 0.289, respectively) and then separates the ARC and BRG at K = 3 (Figures S2B and S2C). However, the analysis loses power at higher values of K likely due to small and uneven sample sizes for this kind of analysis (Figures S2D and S2E). F_{st} measured between lineages was highest between the NAL lineage and all others (between 0.09 and 0.15), with next highest values between the ARC lineage and most others (between 0.15 and 0.04), and the rest generally following the branching order from the phylogeny (Table S3).

We added to previously published demographic history reconstructions[26] using the pairwise sequentially Markovian coalescent[27] (PSMC) and completed the analysis for newly sequenced individuals and compiled all results (Figure S3). Historical trends are generally similar and show large increases in effective population size (N_e) beginning around 100,000 years ago, with N_e varying between populations but always very high at the peak sizes—into the hundreds of thousands at least (Figure S3). These patterns have been shown to be related to glacial cycles occurring at the time and represent very large N_e values.[26]

As well as exploring lineage diversity and demographic history, we calculated individual genetic diversity, θ, an approximation of heterozygosity under the infinite sites model.[11,28,29] We found overall high heterozygosity in caribou, although with some variation among individuals (overall mean of 0.0030, range of 0.0012 to 0.0036; Figures 1C and S2F). Some individuals from within the NAL and ARC lineages had lower diversity than the others, with the ARC lineage mean θ at 0.0024 and the NAL at 0.0028 compared with the mean of all others at 0.0031 (Figures 1C and S2F). When compared with other mammal species where genome-wide heterozygosity has been calculated, our mean is around some of the highest heterozygosity values (see Figure 3 in Morin et al.[30]), demonstrating a high genetic diversity, as well as high phylogenetic lineage diversity, in caribou. We also calculated nucleotide diversity of each lineage and found it to range between 0.0025 in the ARC and NAL lineages to 0.0028 in the BRG lineage, (all others = 0.0027) also demonstrating high diversity when compared across mammal species.[31] The lower diversity seen in the ARC and NAL lineages may, at least in part, be explained by lower historical N_e. Although the historical peak in N_e is always high, it is lowest in the NAL individuals when compared with all other lineages, and with the most recent observable N_e using the PSMC analysis being lowest in the ARC individuals (Figure S3). These potential signatures of historical drift, when compared with other lineages, may help explain their elevated genetic differentiation (Table S3; Figures S2A–S2C) as well as lower heterozygosity and nucleotide diversity.

To measure introgression among the reconstructed lineages, we used D and f4-ratio tests, which control for incomplete lineage sorting. These statistics gave many significant signatures of introgression between groups (Data S1). We then calculated the f-branch statistic, which accounts for many correlated signatures of introgression using the f4-ratio statistics to show when along the phylogeny introgression occurred, and whether the gene flow was into an ancestral group. The f-branch results generally indicated widespread introgression between lineages; however, some lineages show only a small amount or even a lack of gene flow (with the caveat that sister groups cannot be tested for gene flow using ABBA BABA tests). For example, no lineage shows gene flow into the ARC, GRA, or NM1 caribou lineages (Figure 2A). To explore potential introgression between sister lineages, we visualized "admixture graphs" in SplitsTree. Unsurprisingly these graphs show some phylogenetic uncertainty, putatively due to gene flow, within lineages (Figure S4). We also see potential signals of introgression between the NWB, PMG, and CSM lineages, which may help to explain the placement of the PMG individuals next to the NWB lineage in the phylogenetic reconstruction, as well as some potential introgression between the three major northern mountain lineages (GRA, NM1, and NM2).

Altogether our results point toward a high level of intra-specific diversity in caribou, with some strong signals of introgression among many of the lineages. Our results build on previous studies showing high genetic diversity in caribou, for example, large numbers of mitochondrial haplotypes[25,26,32] and high diversity in microsatellite loci.[33–35] The reasons behind the high diversity and number of intra-specific lineages are likely multifaceted. The large Beringian refugium, where the individuals from the BEL lineages spent the glacial cycles of the Quaternary, harbored high levels of genetic diversity for some, particularly cold adapted, species such as caribou,[36,37] and is reflected in the large diversity in the BEL vs the NAL caribou (Figure 1B). Post-glacial expansion out of refugia can lead to genetic bottlenecks and low diversity further away from the refugial populations.[38] However, repeated secondary contact and admixture between glacial lineages can increase genetic diversity,[39,40] in a similar mechanism to the "glacial pulse model," which describes how lineage fusion during glacial cycles can be a source of intra-specific lineage diversity.[41] Indeed, well-known determinants of diversification during adaptive radiations include high standing variation, gene flow, and habitat to diversify into,[42] all of which are true for caribou post-glacial recolonizations. Historical demographic reconstruction supports very large N_e, in the hundreds of thousands, during times of post-glacial expansion, and so the high diversity we see today was likely, at least in part, driven by these processes.[26]

Adaptive genetic diversity

The same processes leading to high genetic variation will likely also have increased standing adaptive variation in caribou. We used the branch model approach in the codeml module of PAML[43] to calculate the ratio of synonymous to non-synonymous mutations (dN/dS ratio) within genes. Positive selection is

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

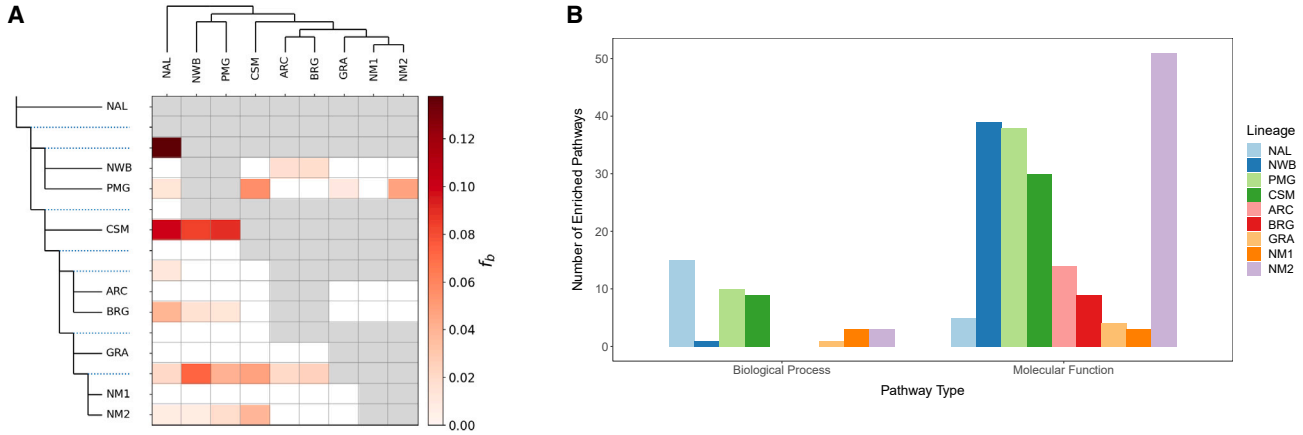


Figure 2. Levels of introgression between lineages and their number of enriched functional pathways

(A) Heatmap showing the f-Branch statistics alongside the phylogeny. The gene flow is from the lineage indicated in the top phylogeny (x axis) going into the phylogeny represented on the y axis. Dotted lines indicate where gene flow is going into an ancestral group on the phylogeny. Grayed out squares indicate tests that could not be made with this statistic, and white squares indicate where no gene flow was detected.

(B) The number of enriched functional pathways within each lineage, under the categories biological processes and molecular function, from their rapidly evolving genes. See also Figure S4, Table S3, Data S1, and S2S–S2AA.

indicated by a gene having a signature of a higher non-synonymous substitution rate.[44] To test this, the program performs likelihood ratio tests to elucidate whether the "focal" branch has a significantly different ratio from the rest of the phylogenetic tree and is thus putatively a rapidly evolving gene within that branch when compared with the overall phylogeny. We ran each of the nine major linages from our whole-genome phylogeny as the focal branch to find those rapidly evolving genes significant to each, thus potentially involved in the adaptive diversification of the lineage.

We found a number of rapidly evolving genes within each lineage (Table S3; Data S2A–S2I). However, contrary to our expectations to find more lineage specific genes in those with lower levels of introgression (for example in the ARC lineage), we found similar numbers of statistically significant genes after Bonferroni correction for each both overall ($X^2 = 3.058$, df = 8, p = 0.931) and for the genes unique to each lineage ($X^2 = 8.142$, df = 8, p = 0.420; Table S3). This could be due to the relatively short time scale of the diversification[26] (all within ~120,000 years) limiting the number of genes within each lineage with a high dN/dS ratio. There was overlap in the genes that pulled out as significant for each lineage (Data S2A–S2I), which may be indicative of the diversification occurring from a large pool of standing genetic variation, which is known to be a driver of diversification during adaptive radiations.[42] It is important to note that it is possible that some of the resulting genes are false positives, for example, due to complex substitutions (e.g., in multiple nucleotides), or if a gene falls into a region of high recombination. However, using a Bonferroni correction has been shown to be a very conservative filter with the CodeML approach.[44]

We performed gene ontology (GO) analyses to assign functional categories to the genes under both "molecular function" (molecular-level processes or activities carried out by gene products) and "biological process" (the larger biological objective which encompasses a number of molecular-level processes[45]) and found a number of processes represented in the significant genes such as immune processes, stress responses, carbohydrate binding, among many others (Data S2J–S2R). We then performed enrichment analyses for each lineage, also under both "biological process" and "molecular function," in order to find specific pathways containing multiple genes with signatures of rapid evolution. We found that caribou lineages had different numbers of enriched pathways, with some lineages showing a significantly larger number compared with the others ($X^2 = 112.71$, df = 8, p = 2.2e−16; Figure 2B; Data S2S–S2AA). In particular, the CSM, NWB, and PMG lineages all show high numbers of enriched pathways, and these lineages also show some of the highest signatures of introgression (Figure 2A and S4). Gene flow may further enhance the variability of functional pathways by exchanging gene variants among lineages creating new combinations, thereby increasing adaptive potential. These combinational pathway changes may well facilitate expression levels and timing prompting adaptation to the range of ecozones inhabited by caribou and the larger *Rangifer* range.[46] It is known that gene flow can facilitate adaptive diversification, as well as inflating standing genetic variation as a whole[42,47,48] and may have been a driver of adaptive diversification in these caribou lineages.

Inbreeding and genetic load

Despite their abundance as well as their high phenotypic, lineage, and genetic diversity (Figure 1) with overall high introgression (Figure 2A), and differential adaptive genetic diversity of caribou (Figure 2B; Table S3), some populations have undergone dramatic declines in recent years. For example, the range of boreal caribou in Ontario (NAL lineage) has become disjunct and the populations along the southern edge of the distribution (Lake Superior) have declined to very small numbers of individuals and have already been shown to have elevated signatures of inbreeding.[49] Other populations, for example, the Qamanirjuaq barren-ground caribou (BRG lineage) are decreasing but are still in large numbers[22] (~250,000 individuals for the Qamanirjuaq caribou), while some, for example, northern mountain

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

CellPress
OPEN ACCESS

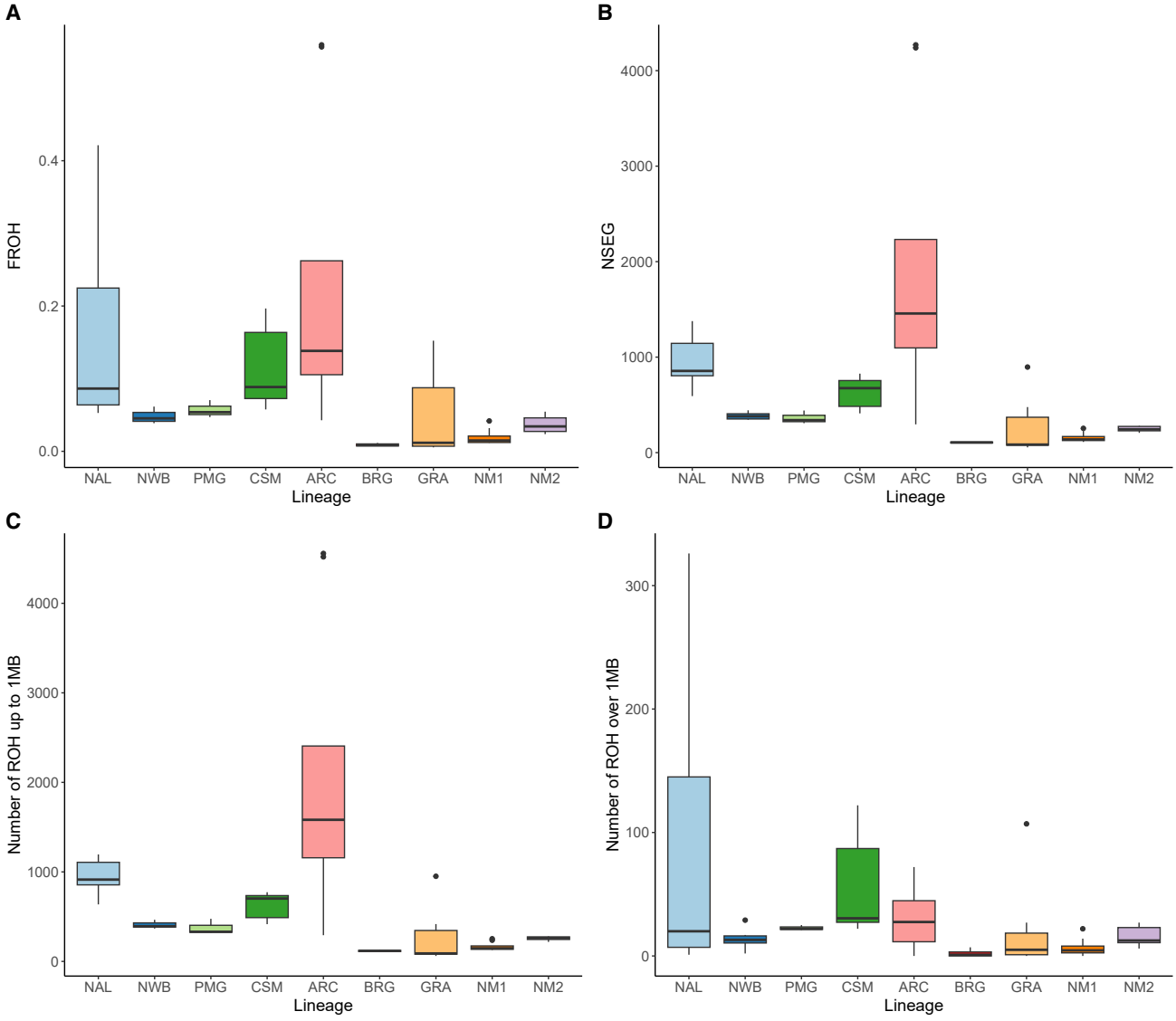


Figure 3. Inbreeding extent in the caribou lineages

(A) F_{ROH}, or the proportion of the genome in runs of homozygosity, (B) the number of runs of homozygosity in total, (C) the number of runs of homozygosity up to one million base pairs, and (D) the number of runs of homozygosity over one million base pairs. The outlier dots for the ARC lineage (A–C) are the Kangerlussuaq caribou from Greenland, for the GRA lineage (B–D) are Aishihik, for NWB (D) is one individual from South Slave, and for NM1 (A–D) is one individual from Frog. See also Figure S5; Data S3J.

caribou from the Redstone (NM1 lineage) and Aishihik (GRA lineage) have remained stable or are increasing[22] (see STAR Methods for more detail on what is known about the effective and census population sizes of each of the sampled caribou subpopulations). We measured the effect of demography on signatures of inbreeding using runs of homozygosity (ROH) estimation and found varied levels of inbreeding across individuals ranging from F_{ROH} (proportion of the genome in ROH) of 1% or less in the barren-ground caribou in the BRG and GRA lineages, up to around 56% in caribou from Kangerlussuaq in Greenland (ARC lineage; Figures 3 and S5; Data S3J), a number comparable to some of the most endangered species such as southern resident killer whales,[15] kākāpō,[12] and Indian tigers.[14]

The most inbred caribou are generally from the most northern or southern portions of the distribution where genetic erosion due an extreme environment (north) or anthropogenic disturbance (south) are the strongest[49] except for the Aishihik caribou that are known to have been introgressed with introduced reindeer.[26] The caribou with the highest F_{ROH} are also those with lower heterozygosity values (Figures 1C and 3), although the magnitude of the variance in F_{ROH} is generally higher. However, wider variation in F_{ROH} does align with results found across mammal species.[50]

We calculated when the longest ROHs originated in the highly inbred individuals. For the boreal caribou from the disjunct part of the range in Ontario, which were estimated at 492 mature

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

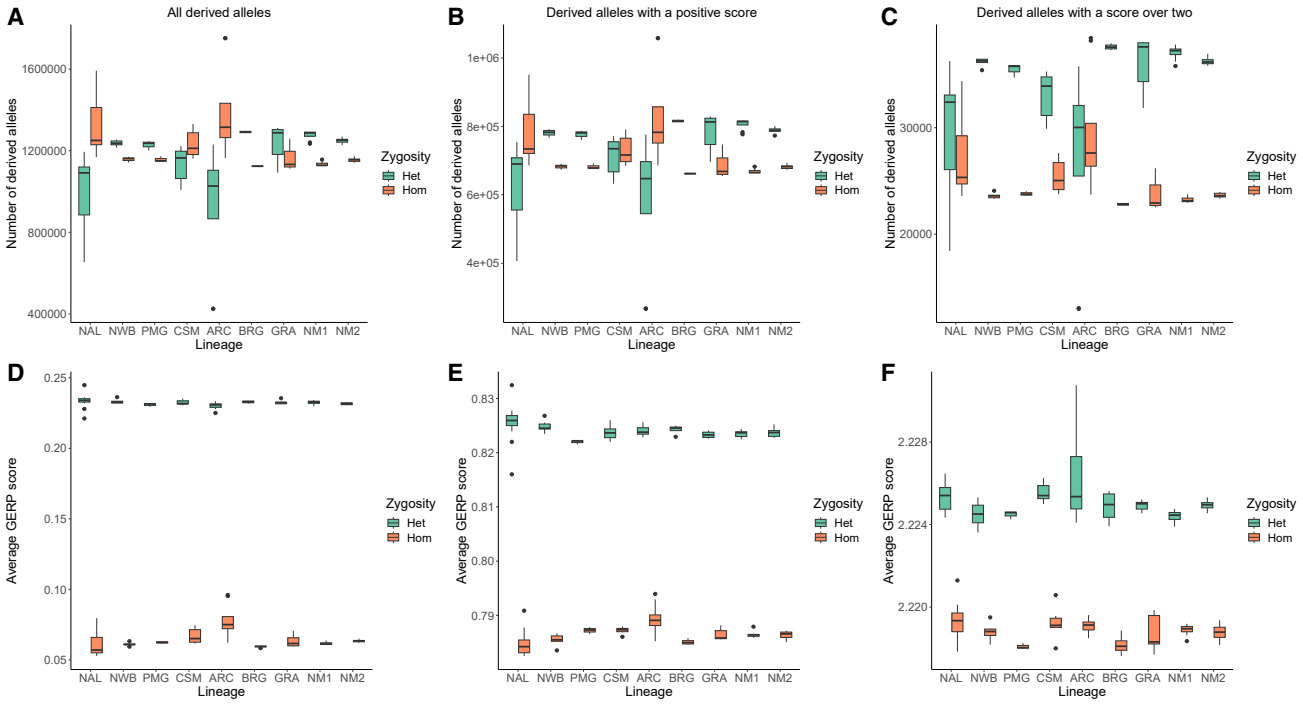


Figure 4. Genetic load as profiled using genomic evolutionary rate profiling (GERP)

(A) The number of all derived alleles, (B) the number of derived alleles with a positive GERP score for each lineage, (C) the number of derived alleles with a GERP score over two for each lineage, (D) the average GERP score for all derived alleles for each lineage, (E) the average GERP score for the derived alleles with positive scores for each lineage, and (F) the average GERP score for the derived alleles with scores over two for each lineage. For all plots, the number of alleles which are heterozygous (green) and homozygous (orange) are shown separately. See also Data S4A.

adults as of 2012,[22] the ROHs arise from shared parental ancestors around the late 1950s–early 1960s, likely reflecting the period when the range became fragmented and discontinuous over the past century.[51] For the Itcha-Ilgachuz caribou, the longest ROHs arise from shared parental ancestors in the early 1900s, coinciding to when declines were seen potentially due to hunting.[52] Exact historical population sizes are not known at that time, but the population recovered from 350 animals in the 1970s up to a peak of 2,800 in 2003,[52,53] although, is once again declining and was estimated at 1,220 adults in 2012.[18,22] The Kangerlussuaq caribou in Greenland are known to have undergone a strong and sustained decline, maintaining low population sizes for over a hundred years between 1845 and the 1950s.[54] Historical population sizes are not known however preceding the decline (between 1845 and 1850) around 6,000 caribou were hunted each year in the region suggesting a large population size. Following the decline, caribou were reported to be scarce with only around 300 caribou harvested between 1910–1920.[54] The longest ROH in the Greenland caribou arose during this period (1884 and 1929, respectively). Caribou have subsequently increased steadily peaking in the 1970s. In recent years, they have declined but were still estimated over 60,000 animals in 2018.[54]

Overall, our results demonstrate that different demographic histories have had a dramatic impact on levels of inbreeding and so we investigated proxies of genetic load, which is the actual or potential reduction in fitness due to mutation, with the highly deleterious alleles expected to be at least partially recessive.[8,10,55] We measured whether there is evidence of purging of deleterious variation, or the removal of highly deleterious alleles due to inbreeding causing homozygosity and thus allowing selection to act on these mutations,[13] in highly inbred vs non-inbred individuals. Purging is dependent on the degree of deleteriousness as well as dominance interactions of alleles, and thus population bottlenecks have been shown to remove highly deleterious recessive mutations, however mildly deleterious mutations can increase due to genetic drift and the reduced efficacy of selection,[13,14] and so we have measured multiple categories of putatively deleterious alleles at both the individual and lineage level using two different methods.

The first method we used was multispecies comparisons and genomic evolutionary rate profiling[56] (GERP), where we calculated the number of derived alleles which were homozygous or present in a heterozygous state per individual, as well as their overall average GERP scores, in three categories of putative deleteriousness; all derived alleles regardless of GERP score, those with a positive GERP score and so at positions demonstrating some level of evolutionary conservation across species, and those with a GERP score over two (at the top end of the score range in our dataset) and so at positions showing a high level of evolutionary conservation across species and thus putatively more highly deleterious. Overall, for the derived alleles with a score over two, the number of homozygous loci per individual is generally lower than the number of heterozygous ones (Figure 4C; Data S4A) when compared with the other two categories of deleteriousness (Figures 4A and 4B; Data S4A). However,

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

CellPress
OPEN ACCESS

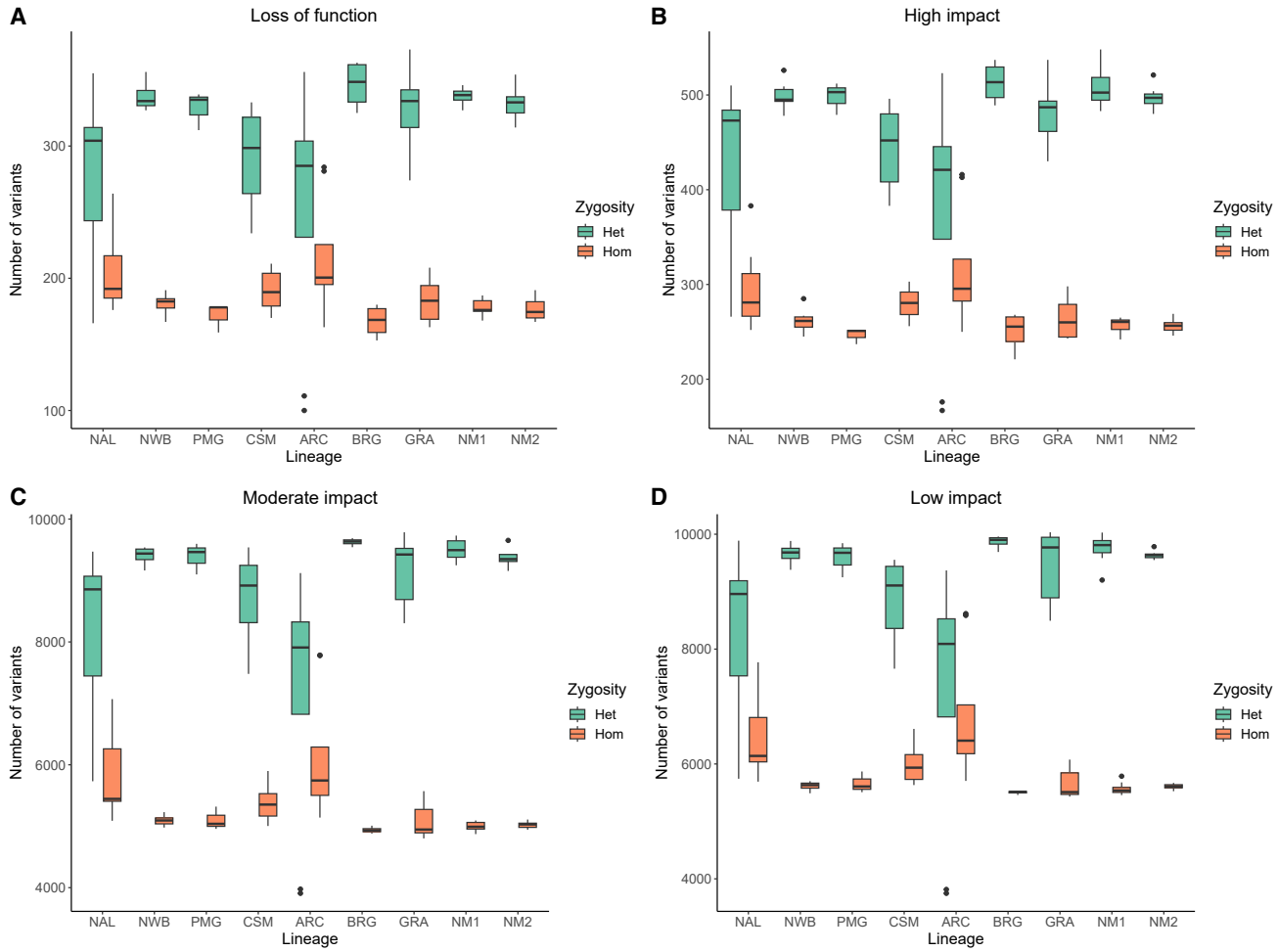


Figure 5. Genetic load as profiled using the genome annotation

The number of derived loss-of-function, high-impact, moderate-impact, and low-impact alleles in each lineage (A–D, respectively). For all plots, the number of alleles which are heterozygous (green) and homozygous (orange) are shown separately. See also Data S4B.

contrary to the pattern expected with purging of deleterious alleles, this is less so in the most inbred individuals (in the ARC and NAL lineages) where the numbers of homozygous-derived alleles with scores over two is higher and the number of derived alleles present as heterozygous is lower than other individuals (Figure 4; Data S4A). Similarly, when looking at the average GERP scores for derived alleles over a score of two, we found the average score in the inbred individuals to be similar though slightly higher than non-inbred individuals (Figure 4F; Data S4A).

In line with the GERP analysis, using our new genome annotation, we found fewer heterozygous and more homozygous derived loss-of-function (LOF) and high-impact alleles in the more inbred individuals, as well as for moderate- and low-impact alleles (Figure 5; Data S4B). This is in contrast to similar analyses in other species showing a decline in the number of alleles in the most deleterious categories that are homozygous, e.g., LOF, for the most inbred individuals, for example, in the Alpine ibex,[13] Indian tigers,[14] and Iberian lynx.[57] Our results also follow the "drift only" pattern described in Dussex et al.[58] and thus indicate genomic erosion and loss of overall diversity through drift without preferential purging of deleterious variation due to inbreeding.

In all categories, the numbers of derived alleles that are homozygous was proportional to F_{ROH} both across all individuals (Figure S6) and when only considering lineages with relatively larger sample sizes (at least eight individuals), two with signatures of inbreeding and higher F_{ROH} values (NAL and ARC), and two with low inbreeding (NM1 and NM2), although the relationship appears to be slightly weaker in the lineages with low F_{ROH} values (Figure 6). This is similar to patterns seen in species such as Indian tigers and suggests that higher inbreeding may have a potential fitness cost[14] and is in contrast to patterns seen in some other species such as killer whales where historical population processes appeared to also have an impact on homozygous genetic load.[15]

In addition to the individual based results, we calculated the proportions of derived alleles which are present as heterozygous in all individuals, fixed homozygous and segregating homozygous in each deleteriousness category for all lineages (Figures 7A–7F; Data S4C and S4D). We focus again on those lineages with sample sizes of at least eight caribou, although

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

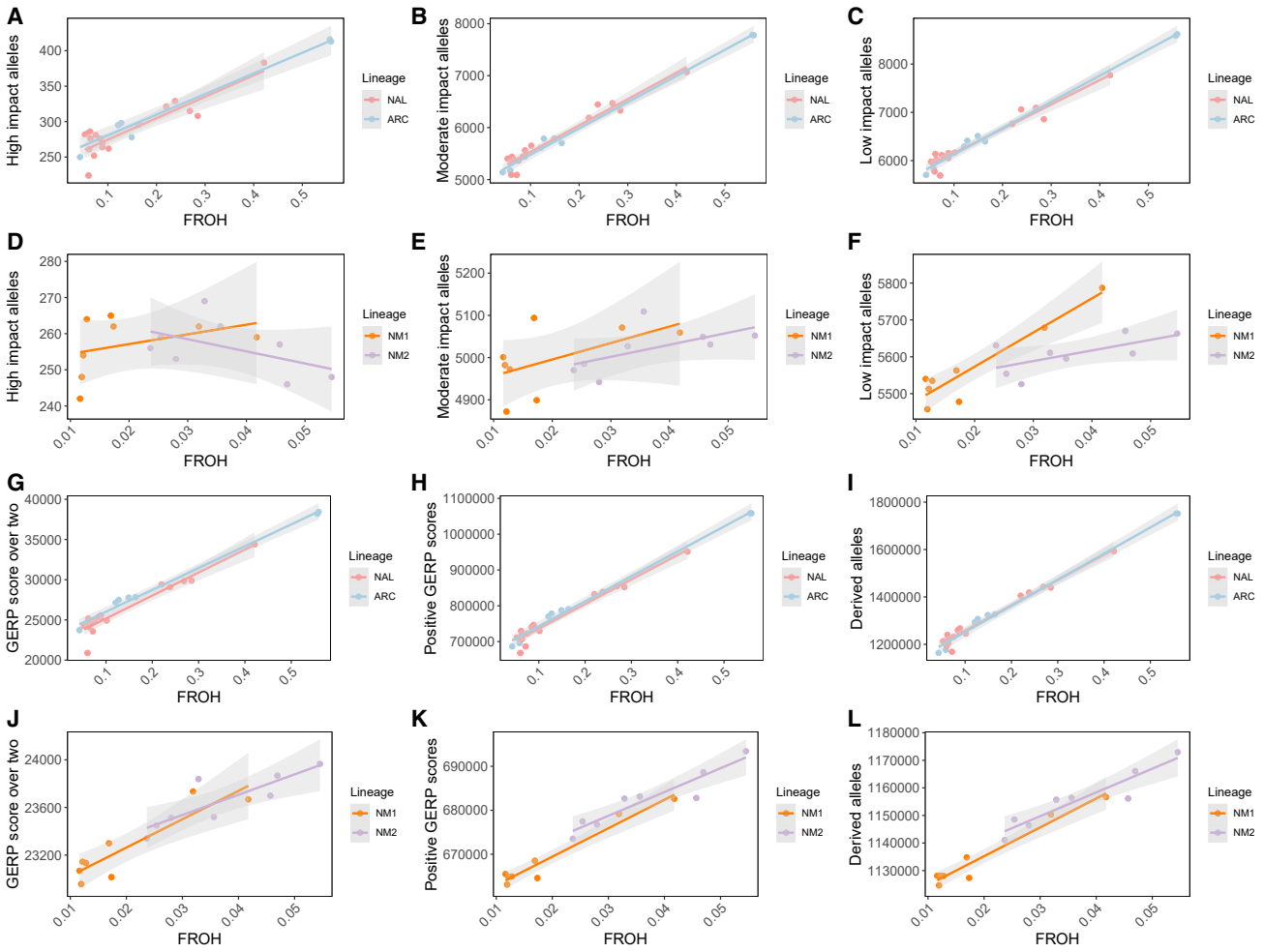


Figure 6. Homozygous genetic load versus F_{ROH}

Numbers of derived alleles that are homozygous representing different deleterious categories plotted against FROH for lineages with at least eight individuals (NAL and ARC representing higher F_{ROH} values, and NM1 and NM2 representing lower F_{ROH}) including high impact (which includes loss-of-function alleles; A and D), moderate impact (B and E), low impact (C and F), high GERP scores (G and J), positive GERP scores (H and K), and all derived sites (I and L). Linear regression lines were added for each lineage. See also Figure S6 and Data S4.

results were generally consistent across all lineages aside from a potential skew in the PMG and BRG lineages with three and four individuals, respectively (Figure S7). The proportions (to control for the lower number of higher impact alleles) of derived alleles that are homozygous and segregating were similar across deleteriousness categories and lineages (Figures 7C and 7F); however, some interesting patterns emerge in the proportions of alleles which are fixed homozygous and sites where all individuals were heterozygous. For alleles which are fixed homozygous, those that are high impact, LOF, and with the highest GERP scores, representing the most deleterious alleles, are found in lower proportions than alleles in the lower impact categories across all lineages apart from NM1, likely indicating purifying selection against these alleles (Figures 7B and 7E). The overall proportions of alleles that are homozygous and fixed were higher in the ARC lineage but lower in the NAL lineage when compared with the non-inbred ones. For the ARC lineage, this may be due to the overall increase in homozygosity of alleles as we see in the individual level results and again may indicate increased risk of inbreeding depression. That the NAL lineage shows lower proportions was more surprising and may be due to that lineage containing individuals from the continuous range which may be less likely to share the same fixed homozygous alleles as caribou from the discontinuous Lake Superior range given that they are now geographically isolated.[49] Either way, the ARC and NAL lineages, with higher inbreeding and lower nucleotide diversity, do not have lower proportions of putatively highly deleterious alleles compared with lower impact categories than do other non-inbred lineages (Figures 7B and 7E), indicating no increased purging due to inbreeding in those lineages.

Proportions of derived alleles that are present as heterozygous in all individuals within a lineage, which are not as frequently investigated, revealed an increase in the proportion of moderate-impact alleles and those with a GERP score over two in all lineages (Figures 7A and 7D). This intriguing pattern may be caused by the differences in the level of deleteriousness across

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

CellPress
OPEN ACCESS

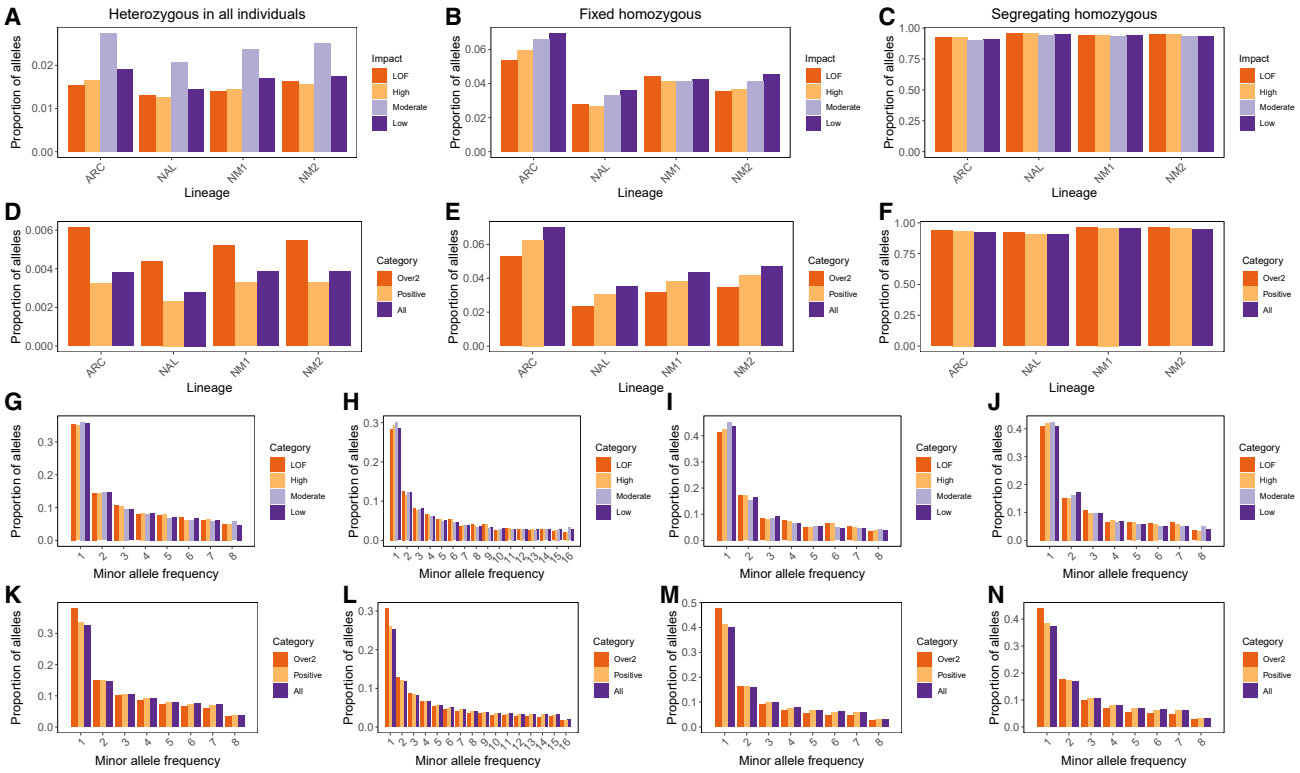


Figure 7. Per lineage proportions and site frequency spectra of genetic load

Proportions of derived alleles that are heterozygous in all individuals (A and D), fixed homozygous (B and E), and segregating homozygous (C and F) in all categories of deleteriousness, as well as the site frequency spectra of all categories for each lineage (G–N). Only lineages with at least eight individuals (G and K represent the ARC lineage, H and L the NAL lineage, I and M the NM1 lineage, and J and N the NM2 lineage) are shown. See also Figure S7 and Data S4C and S4D.

the categories. The moderate-impact category, which encompasses alleles such as non-synonymous mutations, are likely to be of milder deleteriousness than the very-high-impact, LOF alleles.[14] Although we see evidence of some selection against them being homozygous (Figures 7B and 7E), this is likely not nearly as strong as for high-impact mutations. GERP analysis is not as able to discriminate weakly from highly deleterious alleles,[13] therefore the "scores over two" category likely contains a combination of both potentially explaining this pattern also being seen in alleles at positions with high conservation scores. Given that the build-up of mildly deleterious alleles due to drift has been cited as the reason for inbreeding depression occurring, even with the purging of highly deleterious alleles in those species (e.g., Grossen et al.[13]), that these putatively mildly deleterious alleles are present in slightly higher proportions as heterozygous across individuals could indicate some vulnerability to this phenomenon in caribou. However, the overall proportion of the derived alleles present as heterozygous across individuals within a lineage is still very low when compared with those alleles that are fixed or segregating homozygous (Figure 7). These results highlight the complexity of genetic load and the different potential patterns due to dominance effects, levels of deleteriousness, as well as demographic history differences between populations and species.[13,15]

As another way to test for purging of the derived putatively highly deleterious alleles, we calculated the site frequency spectrum (SFS) for each lineage for each category to see if there are shifts in the spectrum for more deleterious alleles (e.g., as seen in Kardos et al.,[15] Grossen et al.,[13] and Kahn et al.[14]). Once again, patterns are the same across lineages (Figures 7G–7N and S7). For the genome annotation categories, there is no strong shift in the spectrum apart from perhaps a slight shift in the moderate-impact alleles (Figures 7G–7I). The GERP plots show a shift in the spectrum of the mutations in the most highly conserved genomic positions with scores over two (Figures 7K–7N), which may indicate selection against these derived alleles overall, but is again seen across all lineages regardless of inbreeding or nucleotide diversity. That we only see this pattern in the GERP results could indicate that most of the deleterious alleles in the genome are non-coding and thus only detectable via an analysis such as GERP. Alternatively, it could mean that annotation category is a bad predictor of actual fitness effects (selection coefficient) in this system.

Overall, our analyses exploring genetic load at both the individual and lineage levels show evidence for selection against deleterious alleles but not more so due to inbreeding. In fact, our results show increased numbers of derived alleles that are homozygous in inbred individuals at all categories of deleteriousness (Figures 4 and 5; Data S4A–S4D), a relationship between F_{ROH} and deleterious alleles that are homozygous (Figure 6), potentially higher proportions of deleterious alleles that are fixed in one of the most inbred lineages with lower nucleotide diversity

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

(ARC lineage), and high proportions of putatively mildly deleterious alleles which are heterozygous (Figure 7). Our results may indicate that, even despite extremely high levels of inbreeding in some individuals (e.g., 22%–42% of the genome in ROH in disjunct Ontario boreal caribou and 56% in Kangerlussuaq caribou), the increased homozygosity and inbreeding does not appear, as of yet, to have led to purging of highly deleterious alleles. Given that we see a shift in the SFS regardless of inbreeding, there does appear to have been some purging of deleterious alleles, but this likely occurred over a longer evolutionary history and is thus relatively uniform across populations. Recent population declines may have been over too short a timescale to substantially impact the distribution of allele frequencies for putatively highly deleterious versus neutral alleles. However, the analyses we have used cannot infer how deleterious alleles are, and it is thus possible that some very highly deleterious alleles have been purged recently due to strong inbreeding, but this effect is not visible with these types of analyses.

It is difficult to compare load across studies due to differences in genome annotations and data filtering, as well as the multispecies alignment used for GERP analysis. However, our data indicate hundreds of high-impact alleles which are homozygous in each individual (between 221 and 416) as well as thousands of moderate-impact alleles that are homozygous (Data S4B), representing a large overall genetic load in caribou. This is not surprising given their high historical effective population sizes,[26] high phenotypic diversity,[22] and high genetic and intra-specific lineage diversity and gene flow we reconstructed here. It is, however, important to note that the genetic load analyses we have conducted cannot directly measure fitness and rely on commonly held assumptions that these inferred deleterious alleles, on average, would have deleterious impacts on fitness.[14]

Preserving the high genetic diversity of caribou may be important for their persistence and ability to adapt to environmental changes.[2,3] Given that small populations have not purged deleterious variation and show increased homozygosity, maintaining connectivity between populations and lineages is important as introgression appears to be a driver of increased genetic variation (Figure 2), allowing flow of adaptive genes[59] and preventing an increase in homozygous load as has been recommended in other species (e.g., Smeds and Ellegren[60]). As some caribou populations have recently declined to small census sizes, particular attention should be put in assessing the potential impact of inbreeding on current and future trends.

Conservation implications

Much has been published recently in understanding the impacts of low genetic variation and genetic load associated with inbreeding in threatened and generally genetically depauperate species.[11–14,57,61,62] However, high genetic load is likely supported in many widespread and diverse species with similar demographic histories to caribou[9] that are perhaps not yet threatened but that have started or will inevitably be impacted by anthropogenic activities such as habitat loss and climate change into the future. As caribou have already begun to be impacted and undergo declines in some areas, the genetic erosion and lack of purging even with high inbreeding levels might foreshadow what will occur in these other taxa.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Caribou reference genome assembly
  - Genome annotation
  - Caribou samples background
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Re-sequenced whole genome filtering
  - Whole genome phylogenomic reconstruction
  - Genetic diversity and demographic history
  - Rapidly evolving genes and gene ontology analysis
  - Runs of homozygosity (ROH) estimation
  - Genetic load

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.02.002.

ACKNOWLEDGMENTS

We would like to thank the many field collectors and hunters who supplied samples for this work, as well as Bridget Redquest and Austin Thompson for technical support in the laboratory. We would also like to thank Brody Crosby for support with data management and other members of the Ecogenomics team (https://www.ecogenomicscanada.ca/) for the many insightful conversations. We are also thankful to the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca), Compute Canada, and Amazon Cloud Computing for high-performance computing services. Funding from this research was provided by the Genomic Applications Partnership Program of Genome Canada, Environment and Climate Change Canada, the Government of Canada's Genomics Research and Development Initiative (GRDI), the Species-at-Risk Stewardship Program (SARSP), and the WCS Garfield Weston Fellowship. We would also like to thank Marty Kardos, Lukas Keller, and three anonymous reviewers for their feedback and improvement on the manuscript.

AUTHOR CONTRIBUTIONS

R.S.T. helped to conceive the study, did the bioinformatics, and wrote the manuscript. M.M. helped to conceive the study, secured funding, and edited the manuscript. S.K. helped with the bioinformatics and edited the manuscript, P.L. wrote scripts for some bioinformatic analyses, and G.M. did the laboratory work to produce the fibroblast cells for the reference genome and annotation. K.S. secured funding and support with bioinformatics, A.K., N.C.L., M.G., H.S., C.T., J.P., L.A., D.H., and D.S. coordinated or collected samples and edited the manuscript. P.J.W. helped to conceive the study, secured funding, and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 5, 2023
Revised: November 17, 2023
Accepted: February 1, 2024
Published: February 27, 2024

REFERENCES

1. Hoban, S., Archer, F.I., Bertola, L.D., Bragg, J.G., Breed, M.F., Bruford, M.W., Coleman, M.A., Ekblom, R., Funk, W.C., Grueber, C.E., et al. (2022). Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. Biol. Rev. Camb. Philos. Soc. *97*, 1511–1538.
2. Andrello, M., D'Aloia, C.D., Dalongeville, A., Escalante, M.A., Guerrero, J., Perrier, C., Torres-Florez, J.P., Xuereb, A., and Manel, S. (2022). Evolving spatial conservation prioritization with intraspecific genetic data. Trends Ecol. Evol. *37*, 553–564.
3. Carvalho, S.B., Velo-Antón, G., Tarroso, P., Portela, A.P., Barata, M., Carranza, S., Moritz, C., and Possingham, H.P. (2017). Spatial conservation prioritization of biodiversity spanning the evolutionary continuum. Nat. Ecol. Evol. *1*, 151.
4. O'Brien, D., Laikre, L., Hoban, S., Bruford, M.W., Ekblom, R., Fischer, M.C., Hall, J., Hvilsom, C., Hollingsworth, P.M., Kershaw, F., et al. (2022). Bringing together approaches to reporting on within species genetic diversity. J. Appl. Ecol. *59*, 2227–2233.
5. Des Roches, S., Pendleton, L.H., Shapiro, B., and Palkovacs, E.P. (2021). Conserving intraspecific variation for nature's contributions to people. Nat. Ecol. Evol. *5*, 574–582.
6. Leigh, D.M., van Rees, C.B., Millette, K.L., Breed, M.F., Schmidt, C., Bertola, L.D., Hand, B.K., Hunter, M.E., Jensen, E.L., Kershaw, F., et al. (2021). Opportunities and challenges of macrogenetic studies. Nat. Rev. Genet. *22*, 791–807.
7. Yiming, L., Siqi, W., Chaoyuan, C., Jiaqi, Z., Supen, W., Xianglei, H., Xuan, L., Xuejiao, Y., and Xianping, L. (2021). Latitudinal gradients in genetic diversity and natural selection at a highly adaptive gene in terrestrial mammals. Ecography *44*, 206–218.
8. Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H.E., and van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. Nat. Rev. Genet. *23*, 492–503.
9. van Oosterhout, C., Speak, S.A., Thomas, B., Chiara, B., Lawrence, P.-.A., Lara, H.U., Jim, J.G., Gernot, S., and Hernán, E.M. (2022). Genomic erosion in the assessment of species extinction risk and recovery potential. bioRxiv. https://doi.org/10.1101/2022.09.13.507768.
10. Hedrick, P.W., and Garcia-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. Trends Ecol Evol *31*, 94–952.
11. von Seth, J., Dussex, N., Díez-Del-Molino, D., van der Valk, T., Kutschera, V.E., Kierczak, M., Steiner, C.C., Liu, S., Gilbert, M.T.P., Sinding, M.S., et al. (2021). Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. Nat. Commun. *12*, 2393.
12. Dussex, N., van der Valk, T., Morales, H.E., Wheat, C.W., Díez-Del-Molino, D., von Seth, J., Foster, Y., Kutschera, V.E., Guschanski, K., Rhie, A., et al. (2021). Population genomics of the critically endangered kākāpō. Cell Genom. *1*, 100002.
13. Grossen, C., Guillaume, F., Keller, L.F., and Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine Ibex. Nat. Commun. *11*, 1001.
14. Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., Zachariah, A., Jhala, Y.V., Kardos, M., et al. (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proc. Natl. Acad. Sci. USA *118*, e2023018118.
15. Kardos, M., Zhang, Y., Parsons, K.M., A, Y., Kang, H., Xu, X., Liu, X., Matkin, C.O., Zhang, P., Ward, E.J., et al. (2023). Inbreeding depression explains killer whale population dynamics. Nat. Ecol. Evol. *7*, 675–686.
16. Stoffel, M.A., Johnston, S.E., Pilkington, J.G., and Pemberton, J.M. (2021). Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. Nat. Commun. *12*, 2972.
17. Fraimout, A., Rastas, P., Lv, L., and Merilä, J. (2023). Inbreeding depression in an outbred stickleback population. Mol. Ecol. *32*, 3440–3449.
18. COSEWIC (2011). Designatable Units for Caribou (*Rangifer tarandus*) in Canada (Committee on the Status of Endangered Wildlife in Canada).
19. Festa-Bianchet, M., Ray, J.C., Boutin, S., Côté, S.D., and Gunn, A. (2011). Conservation of caribou (*Rangifer tarandus*) in Canada: an uncertain future 1 This review is part of the virtual symposium "Flagship Species – Flagship Problems" that deals with ecology, biodiversity and management issues, and climate impacts on species at risk and of Canadian importance, including the polar bear (Ursus maritimus), Atlantic cod (Gadus morhua), Piping Plover (Charadrius melodus), and caribou (Rangifer tarandus). Can. J. Zool. *89*, 419–434.
20. Vors, L.S., and Boyce, M.S. (2009). Global declines of caribou and reindeer. Glob. Change Biol. *15*, 2626–2633.
21. Weckworth, B.V., Hebblewhite, M., Mariani, S., and Musiani, M. (2018). Lines on a map: conservation units, meta-population dynamics, and recovery of woodland caribou in Canada. Ecosphere *9*, e02323.
22. Government of Canada (2022). Species at Risk Act: COSEWIC assessments and status reports. https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/cosewic-assessments-status-reports.html.
23. Gunn, A. (2016). Rangifer tarandus. The IUCN Red List of Threatened Species 2016: e.T29742A22167140. https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T29742A22167140.en.
24. Gripenberg, U., and Nieminen, M. (1986). The chromosomes of reindeer (*Rangifer tarandus*). Rangifer *6*, 109.
25. Weckworth, B.V., Musiani, M., McDevitt, A.D., Hebblewhite, M., and Mariani, S. (2012). Reconstruction of caribou evolutionary history in Western North America and its implications for conservation. Mol. Ecol. *21*, 3610–3624.
26. Taylor, R.S., Manseau, M., Klütsch, C.F.C., Polfus, J.L., Steedman, A., Hervieux, D., Kelly, A., Larter, N.C., Gamberg, M., Schwantje, H., et al. (2021). Population dynamics of caribou shaped by glacial cycles before the Last Glacial Maximum. Mol. Ecol. *30*, 6121–6143.
27. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.
28. Haubold, B., Pfaffelhuber, P., and Lynch, M. (2010). mlRho – a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. Mol. Ecol. *19* (*Supplement 1*), 277–284.
29. Foote, A.D., Hooper, R., Alexander, A., Baird, R.W., Baker, C.S., Ballance, L., Barlow, J., Brownlow, A., Collins, T., Constantine, R., et al. (2021). Runs of homozygosity in killer whale genomes provide a global record of demographic histories. Mol. Ecol. *30*, 6162–6177.
30. Morin, P.A., Archer, F.I., Avila, C.D., Balacco, J.R., Bukhman, Y.V., Chow, W., Fedrigo, O., Formenti, G., Fronczek, J.A., Fungtammasan, A., et al. (2021). Reference genome and demographic history of the most endangered marine mammal, the vaquita. Mol. Ecol. Resour. *21*, 1008–1020.
31. Teixeira, J.C., and Huber, C.D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. Proc. Natl. Acad. Sci. USA *118*, e2015096118.
32. Polfus, J.L., Manseau, M., Klütsch, C.F.C., Simmons, D., and Wilson, P.J. (2017). Ancient diversification in glacial refugia leads to intraspecific diversity in a Holarctic mammal. J. Biogeog. *44*, 386–396.
33. Boulet, M., Couturier, S., Côté, S.D., Otto, R.D., and Bernatchez, L. (2007). Integrative use of spatial, genetic, and demographic analyses for investigating genetic connectivity between migratory, montane, and sedentary caribou herds. Mol. Ecol. *16*, 4223–4240.
34. McLoughlin, P.D., Paetkau, D., Duda, M., and Boutin, S. (2004). Genetic diversity and relatedness of boreal caribou populations in western Canada. Biol. Conserv. *118*, 593–598.

35. Zittlau, K., Coffin, J., Farnell, R., Kuzyk, G., and Strobeck, C. (1998). Genetic relationships of three Yukon caribou herds determined by DNA typing. Rangifer *20*, 59–62.

36. Galbreath, K.E., Cook, J.A., Eddingsaas, A.A., and DeChaine, E.G. (2011). Diversity and demography in Beringia: multilocus tests of paleodistribution models reveal the complex history of Arctic ground squirrels. Evolution *65*, 1879–1896.

37. Dussex, N., Alberti, F., Heino, M.T., Olsen, R.A., van der Valk, T., Ryman, N., Laikre, L., Ahlgren, H., Askeyev, I.V., Askeyev, O.V., et al. (2020). Moose genomes reveal past glacial demography and the origin of modern lineages. BMC Genomics *21*, 854.

38. Roberts, D.R., and Hamann, A. (2015). Glacial refugia and modern genetic diversity of 22 western North American tree species. Proc. Biol. Sci. *282*, 20142903.

39. Petit, R.J., Aguinagalde, I., de Beaulieu, J.L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. Science *300*, 1563–1565.

40. Alcala, N., and Vuilleumier, S. (2014). Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. Proc. Biol. Sci. *281*, 20141369.

41. Maier, P.A., Vandergast, A.G., Ostoja, S.M., Aguilar, A., and Bohonak, A.J. (2019). Pleistocene glacial cycles drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). Evolution *73*, 2476–2496.

42. Berner, D., and Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. *31*, 491–499.

43. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

44. Álvarez-Carretero, S., Kapli, P., and Yang, Z. (2023). Beginner's guide on the use of PAML to detect positive selection. Mol. Biol. Evol. *40*, msad041.

45. Dessimoz, C., and Skunca, N. (2017). The Gene Ontology Handbook (Humana Press).

46. Weldenegodguad, M., Pokharel, K., Niiranen, L., Soppela, P., Ammosov, I., Honkatukia, M., Lindeberg, H., Peippo, J., Reilas, T., Mazzullo, N., et al. (2021). Adipose gene expression profiles reveal insights into the adaptation of northern Eurasian semi-domestic reindeer (Rangifer tarandus). Commun. Biol. *4*, 1170.

47. Streicher, J.W., Devitt, T.J., Goldberg, C.S., Malone, J.H., Blackmon, H., and Fujita, M.K. (2014). Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. Mol. Ecol. *23*, 3273–3291.

48. Lexer, C., Marthaler, F., Humbert, S., Barbará, T., de la Harpe, M., Bossolini, E., Paris, M., Martinelli, G., and Versieux, L.M. (2016). Gene flow and diversification in a species complex of *Alcantarea* inselberg bromeliads. Bot. J. Linn. Soc. *181*, 505–520.

49. Solmundson, K., Bowman, J., Manseau, M., Taylor, R.S., Keobouasone, S., and Wilson, P.J. (2023). Genomic population structure and inbreeding history of Lake Superior caribou. Ecol. Evol. *13*, e10278.

50. Brüniche-Olsen, A., Kellner, K.F., Anderson, C.J., and DeWoody, J.A. (2018). Runs of homozygosity have utility in mammalian conservation and evolutionary studies. Conserv. Genet. *19*, 1295–1307.

51. Drake, C.C., Manseau, M., Klütsch, C.F.C., Priadka, P., Wilson, P.J., Kingston, S., and 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*, 13–26.

52. Seip, D.R., and Cichowski, D.B. (1996). Population Ecology of caribou in British Columbia. Rangifer *16*, 73–80.

53. Ministry; Forests; Lands; Natural; Resource; Operations, and Rural Development (2018). Itcha-Ilgachuz and Rainbow Caribou Herd Population and Habitat Information. https://a100.gov.bc.ca/pub/acat/documents/r54951/ItchaandRainbowcaribouherdinformation_1541116628869_1116541210.pdf.

54. Cuyler, C., Rosing, M., Linnell, J.D.C., Loison, A., Ingerslev, T., and Landa, A. (2002). Status of the Kangerlussuaq-Sisimiut Caribou Population (*Rangifer tarandus groenlandicus*) in 2000, West Greenland. In Greenland Institute of Natural Resources. Technical Report No. 42. https://doi.org/10.13140/2.1.1322.8164.

55. Agrawal, A.F., and Whitlock, M.C. (2012). Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. Annu. Rev. Ecol. Evol. Syst. *43*, 115–135.

56. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. *6*, e1001025.

57. Kleinman-Ruiz, D., Lucena-Perez, M., Villanueva, B., Fernández, J., Saveljev, A.P., Ratkiewicz, M., Schmidt, K., Galtier, N., García-Dorado, A., and Godoy, J.A. (2022). Purging of deleterious burden in the endangered Iberian lynx. Proc. Natl. Acad. Sci. USA *119*, e2110614119.

58. Dussex, N., Morales, H.E., Grossen, C., Dalén, L., and van Oosterhout, C. (2023). Purging and accumulation of genetic load in conservation. Trends Ecol. Evol. *38*, 961–969.

59. Hanson, J.O., Fuller, R.A., and Rhodes, J.R. (2019). Conventional methods for enhancing connectivity in conservation planning do not always maintain gene flow. J. Appl. Ecol. *56*, 913–922.

60. Smeds, L., and Ellegren, H. (2023). From high masked to high realized genetic load in inbred Scandinavian wolves. Mol. Ecol. *32*, 1567–1580.

61. Mathur, S., Tomeček, J.M., Tarango-Arámbula, L.A., Perez, R.M., and DeWoody, J.A. (2023). An evolutionary perspective on genetic load in small, isolated populations as informed by whole genome resequencing and forward-time simulations. Evolution *77*, 690–704.

62. Xie, H.X., Liang, X.X., Chen, Z.Q., Li, W.M., Mi, C.R., Li, M., Wu, Z.J., Zhou, X.M., and Du, W.G. (2022). Ancient demographics determine the effectiveness of genetic purging in endangered lizards. Mol. Biol. Evol. *39*, msab359.

63. Taylor, R.S., Manseau, M., Horn, R.L., Keobouasone, S., Golding, G.B., and Wilson, P.J. (2020). The role of introgression and ecotypic parallelism in delineating intraspecific conservation units. Mol. Ecol. *29*, 2793–2809.

64. Taylor, R.S., Manseau, M., Redquest, B., Keobouasone, S., Gagné, P., Martineau, C., and Wilson, P.J. (2022). Whole genome sequences from non-invasively collected caribou faecal samples. Conserv. Genet. Resour. *14*, 53–68.

65. Bolger, A.M., Lohse, M., and Usadel, B. (2014). trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

66. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with bowtie 2. Nat. Methods *9*, 357–359.

67. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

68. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

69. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

70. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

71. Vieira, F.G., Lassalle, F., Korneliussen, T.S., and Fumagalli, M. (2016). Improving the estimation of genetic distances from Next-Generation Sequencing data. Biol. J. Linn. Soc. *117*, 139–149.

72. Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: A comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. Mol. Biol. Evol. *32*, 2798–2800.

73. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455.

74. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575.

75. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

76. Korunes, K.L., and Samuk, K. (2021). pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol. Ecol. Resour. *21*, 1359–1368.

77. Malinsky, M., Matschiner, M., and Svardal, H. (2021). dsuite – fast D statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. *21*, 584–595.

78. Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267.

79. Macías, L.G., Barrio, E., and Toft, C. (2020). GWideCodeML: a python package for testing evolutionary hypotheses at the genome-wide level. G3 (Bethesda) *10*, 4369–4372.

80. Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. F1000Research *9*. ISCB Comm J-304.

81. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics *36*, 2628–2629.

82. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly *6*, 80–92.

83. Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge—accurate paired shotgun read merging via overlap. PLoS One *12*, e0185056.

84. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/10.48550/arXiv.1303.3997.

85. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. *5*, e1000695.

86. Taylor, R.S., Horn, R.L., Zhang, X., Golding, G.B., Manseau, M., and Wilson, P.J. (2019). The caribou (*Rangifer tarandus*) genome. Genes *10*, 540.

87. Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. *26*, 342–350.

88. Yamaguchi, K., Kadota, M., Nishimura, O., Ohishi, Y., Naito, Y., and Kuraku, S. (2021). Technical considerations in Hi-C scaffolding and evaluation of chromosome-scale genome assemblies. Mol. Ecol. *30*, 5923–5934.

89. Li, Z., et al. (2017). Draft genome of the reindeer (*Rangifer tarandus*). GigaScience *6*, gix102.

90. Weldenegodguad, M., Pokharel, K., Ming, Y., Honkatukia, M., Peippo, J., Peilas, T., Røed, K.H., and Kantanen, J. (2020). Genome sequence and comparative analysis of reindeer (*Rangifer tarandus*) in northern Eurasia. Sci. Rep. *10*, 8980.

91. Galloway, S.M., Hanrahan, V., Dodds, K.G., Potts, M.D., Crawford, A.M., and Hill, D.F. (1996). A linkage map of the ovine X chromosome. Genome Res. *6*, 667–677.

92. Liu, R., Low, W.Y., Tearle, R., Koren, S., Ghurye, J., Rhie, A., Phillippy, A.M., Rosen, B.D., Bickhart, D.M., Smith, T.P.L., et al. (2019). New insights into mammalian sex chromosome structure and evolution using high-quality sequences from bovine X and Y chromosomes. BMC Genomics *20*, 1000.

93. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.

94. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.

95. Meyermans, R., Gorssen, W., Buys, N., and 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 Genom. *21*, 94.

96. Thompson, E.A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. Genetics *194*, 301–326.

97. Johnston, S.E., Huisman, J., Ellis, P.A., and Pemberton, J.M. (2017). A high-density linkage map reveals sexual dimorphism in recombination landscapes in red deer (*Cervus elaphus*). G3 (Bethesda) *7*, 2859–2870.

98. Huber, C.D., Kim, B.Y., and Lohmueller, K.E. (2020). Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. PLOS Genet. *16*, e1008827.

99. Wootton, E., Robert, C., Taillon, J., Côté, S.D., and Shafer, A.B.A. (2023). Genomic health is dependent on long-term population demographic history. Mol. Ecol. *32*, 1943–1954.

100. Kumar, S., and Subramanian, S. (2002). Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. USA *99*, 803–808.

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
16 samples were used from federal or provincial tissue archives in Canada	This study	N/A
Critical commercial assays		
Qiagen DNAeasy tissue extraction kit	Qiagen, Hilden, German	Cat#69504
Qubit fluorometer High Sensitivity Assay Kit	Thermo Fisher Scientific	*Cat#Q32851*
Illumina HiSeq X.	Illumina, San Diego, California, USA	N/A
Deposited data		
Raw reads for new whole genome sequences	This study	NCBI: PRJNA1040806
Raw reads for whole genome sequences	Taylor et al.[63]	NCBI: PRJNA634908
Raw reads for whole genome sequences	Taylor et al.[26]	NCBI: PRJNA754521
Raw reads for whole genome sequences	Taylor et al.[64]	NCBI: PRJNA694662
Raw reads for whole genome sequences	Solmundson et al.[49]	NCBI: PRJNA984705
Raw reads for whole genome sequences	Weldenegodguad et al.[46]	European Nucleotide Archive: PRJEB37216
Software and algorithms		
Trimmomatic 0.38	Bolger et al.[65]	https://github.com/usadellab/Trimmomatic
Bowtie2 2.3.0	Langmead & Salzberg[66]	N/A
Samtools 1.5	Li et al.[67]	https://github.com/samtools/samtools
GATK4	McKenna et al.[68]	https://github.com/broadinstitute/gatk
VCFtools 0.1.16	Danecek et al.[69]	https://vcftools.github.io/index.html
IQtree 1.6.12	Nguyen et al.[70]	https://github.com/Cibiv/IQ-TREE
ngsDist	Vieira et al.[71]	https://github.com/fgvieira/ngsDist
FASTME 2.1.6.2	Lefort et al.[72]	https://gite.lirmm.fr/atgc/FastME/
RAxML-ng 1.0.1	Kozlov et al.[73]	https://github.com/amkozlov/raxml-ng
Plink 1.9	Purcell et al.[74]	https://www.cog-genomics.org/plink/
ADMIXTURE 1.3.0	Alexander et al.[75]	https://dalexander.github.io/admixture/
mlRho 2.9	Haubold et al.[28]	http://guanine.evolbio.mpg.de/mlRho/
pixy 1.2.7	Korunes and Samuk[76]	https://pixy.readthedocs.io/en/latest/
Dsuite 0.5	Malinsky et al.[77]	https://github.com/millanek/Dsuite
SplitsTree	Hudson and Bryant[78]	https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/algorithms-in-bioinformatics/software/splitstree/
pairwise sequentially Markovian coalescent (PSMC)	Li and Durbin[27]	https://github.com/lh3/psmc
GWideCodeML	Macías et al.[79]	https://github.com/lauguma/GWideCodeML
PAML	Yang[43]	http://abacus.gene.ucl.ac.uk/software/
Gffread 0.12.3	Pertea & Pertea[80]	https://github.com/gpertea/gffread
ShinyGo 0.76.2	Ge et al.[81]	http://bioinformatics.sdstate.edu/go74/
genomic evolutionary rate profiling (GERP)	Davydov et al.[56]	https://github.com/BeckySTaylor/Phylogenomic_Analyses/tree/main

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SnpEff	Cingolani et al.[82]	http://pcingola.github.io/SnpEff/
BBmap 38.86	Bushnell et al.[83]	https://github.com/BioInfoTools/BBMap
BWA-MEM	Li[84]	https://github.com/lh3/bwa
htsbox	N/A	https://github.com/lh3/htsbox
easySFS	Gutenkunst et al.[85]	https://github.com/isaacovercast/easySFS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Rebecca S. Taylor (rebecca.taylor@ec.gc.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All raw reads for re-sequenced genomes have been deposited to the NCBI Sequence Read Archive (SRA): PRJNA634908, PRJNA694662, PRJNA754521, PRJNA984705, PRJNA1040806; all data are publicly available as of the date of publication.
- Assembled reference genome and annotation are available at: https://www.caribougenome.ca/.
- All code used to run programs is accessible through https://github.com/BeckySTaylor/Phylogenomic_Analyses/.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

For the caribou reference genome, fibroblast cells were used from captive caribou at Toronto Zoo. The 16 new re-sequenced genomes were sequenced from blood or tissue obtained from provincial or federal tissue archives.

METHOD DETAILS

Caribou reference genome assembly

To ensure high quality, contiguous DNA for chromosome-scale reference genome assembly, fibroblast cells were taken from caribou at Toronto Zoo and cultured in T-75 flasks. Firstly, we pre-warmed DMEM1, DMEM3 and trypsin to 37-38 °C, then discarded media and rinsed each T-75 flask with 5 ml of DMEM1. We discarded the media, and added 3 ml of trypsin to each flask and incubate at 38 °C for 2 min. We checked to see that cells had lifted, and then added 9 ml of DMEM3 to each flask, rinsed the flask growing surface to retrieve as many cells as possible and transferred the entire volume to a 15 ml tube, leaving 4 x 15 ml tubes, 2 for each animal. We then centrifuged at 200 x g for 5 min to pellet cells, discarded supernatant and re-suspend pellet in 0.8 ml of PBS. We combined pellets for each individual together in a 2 ml tube, centrifuged in microcentrifuge at 200 x g for 5 min, and discarded the supernatant. The samples were then flash frozen using liquid nitrogen and transferred to a dry shipper and shipped to Dovetail Genomics as we wanted to improve our previous assembly also sequenced by Dovetail Genomics,[86] by further scaffolding using Omni-C libraries.[87,88] For each Dovetail Omni-C library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNAse I, chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed, and the DNA purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to produce ~30x sequence coverage. Then HiRise used MQ>50 reads for scaffolding.

The input de novo assembly and Dovetail OmniC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies.[87] Dovetail OmniC library sequences were aligned to the draft input assembly using bwa (https://github.com/lh3/bwa). The separations of Dovetail OmniC read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold. We used the stats.sh script in the program BBMap version 38.42[83] to calculate the assembly statistics.

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

Genome annotation

Cells were cultured as above and shipped to Genewiz (Azenta Life Sciences) for RNA sequencing for the annotation. Total RNA extraction was done using the QIAGEN RNeasy Plus Kit following manufacturer protocols. Total RNA was quantified using Qubit RNA Assay and TapeStation 4200. Prior to library prep, we performed Dnase treatment followed by AMPure bead clean up and QIAGEN FastSelect HMR rRNA depletion. Library preparation was done with the NEBNext Ultra II RNA Library Prep Kit following manufacturer protocols. Then these libraries were run on the NovaSeq6000 platform in 2 x 150 bp configuration. The annotation was also performed by Dovetail Genomics. Repeat families found in the genome assemblies of caribou were identified de novo and classified using the software package RepeatModeler (version 2.0.1). RepeatModeler depends on the programs RECON (version 1.08) and RepeatScout (version 1.0.6) for the de novo identification of repeats within the genome. The custom repeat library obtained from RepeatModeler were used to discover, identify, and mask the repeats in the assembly file using RepeatMasker (Version 4.1.0). Coding sequences from *Bos taurus*, caribou (www.caribougenome.ca) and reindeer[89] were used to train the initial ab initio model for caribou using the AUGUSTUS software (version 2.5.5). Six rounds of prediction optimisation were done with the software package provided by AUGUSTUS. The same coding sequences were also used to train a separate ab initio model for caribou using SNAP (version 2006-07-28). RNAseq reads were mapped onto the genome using the STAR aligner software (version 2.7) and intron hints generated with the bam2hints tools within the AUGUSTUS software. MAKER, SNAP and AUGUSTUS (with intron-exon boundary hints provided from RNA-Seq) were then used to predict for genes in the repeat-masked reference genome. To help guide the prediction process, Swiss-Prot peptide sequences from the UniProt database were downloaded and used in conjunction with the protein sequences from *Bos taurus*, caribou (www.caribougenome.ca), and reindeer to generate peptide evidence in the Maker pipeline. Only genes that were predicted by both SNAP and AUGUSTUS software were retained in the final gene sets. To help assess the quality of the gene prediction, AED scores were generated for each of the predicted genes as part of the MAKER pipeline. Genes were further characterised for their putative function by performing a BLAST search of the peptide sequences against the UniProt database, and tRNAs were predicted using the software tRNAscan-SE (version 2.05).

Caribou samples background

Whole genome sequences of 50 individuals used in this study are available on the National Centre for Biotechnology (NCBI) under BioProject Accession numbers NCBI: PRJNA634908, PRJNA694662, PRJNA754521, PRJNA984705[26,63,64] (Table S2). For this study, we sequenced 16 new genomes (Figure 1; Table S2) using the same protocols as before.[26,63] In total, we have sequenced the genomes of caribou representing 33 different subpopulations across North America and Greenland. For many of the subpopulations we have some information on census sizes (Table S1) and population trends which we outline here. All of the information we provide here, unless otherwise cited, comes from the COSEWIC status reports.[22]

Peary caribou (DU1) are listed as threatened under the species-at-risk-act (SARA) and sits within the ARC lineage. The first counts for Peary caribou were made in the 1960s, and there were an estimated 50,000 caribou at this time. In the late 1980's the population was at an estimated 22,000, but there was a mass die-off reducing the total population to just 5,400 mature individuals by 1996. The most recent estimated census population is 13,200 mature individuals (2015). There are four subpopulations of Peary caribou, and we include two samples from one of those, the Western Queen Elizabeth subpopulation which appears to have been increasing since the 1990s.

The Dolphin Union caribou (DU2), from which we sampled two individuals, are listed as endangered under the SARA and sit within the ARC lineage. This herd underwent a strong decline starting in the early 1900s, eventually starting to increase again in the 1970s with migration returning to the mainland. By 1993 up to 7,000 were once again migrating annually across Coronation Gulf and Dease Strait and in 1997 the population was estimated at 34,558 individuals. However, data suggests a decline of around 50-60% over the last 18 years (as of 2017).

The barren-ground caribou (DU3) are listed as threatened under the SARA, with 14-15 subpopulations. The overall census size for adult barren-ground caribou is estimated at 800,000 (as of 2016), down from 2 million in the 1990s. Barren-ground caribou are thought to fluctuate in abundance. Most subpopulations are declining, and here we use samples two samples each from four subpopulations: Bluenose-west, Qamanirjuaq (both BRG lineage), Porcupine (in the GRA lineage), and Baffin Island (in the ARC lineage). Bluenose-West have declined by over 80% and Qamanirjuaq by ~40% during the last three generations. Porcupine caribou have increased in recent generations, whereas Baffin Island caribou have undergone a dramatic decline of 98% between 1991 and 2014 and there is a large degree of uncertainty in population estimates, and especially low abundance was found on North Baffin.

Northern mountain caribou (DU7) is listed as special concern under the SARA and is estimated to contain a total of 43,000-48,000 mature individuals as of 2014. We find the northern mountain caribou to be split between four phylogenomic lineages; CSM, GRA, NM1, and NM2. There are 45 subpopulations, and we have one to two genome sequences from each of 12 subpopulations: Aishihik, Atlin, Chase, Frog, Graham, Hart River, Itcha-Ilgachuz, Muskwa, Pink Mountain, Spatzizi, Tay, and Tsenaglode. We also have six genome sequences from a 13th subpopulation, the larger Redstone subpopulation. The population trends for sampled populations are largely unknown, apart from Aishihik which is increasing, Atlin and Redstone which are stable, and Itcha-Ilgachuz which is decreasing.

Central mountain caribou (DU8) are listed as endangered under the SARA and sit within the CSM lineage, estimated at 469 mature individuals as of 2014. This DU has undergone a large decline of at least 64% over the last three generations. There are 10 subpopulations all of which contain fewer than 250 individuals with two now extirpated. All subpopulations have been undergoing long term declines. We have one genome from each of two subpopulations, Kennedy Siding and Quintette.

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

Southern mountain caribou (DU9) are listed as endangered under the SARA and sit within the CSM lineage, estimated at 1,395 mature individuals as of 2014, and have undergone a 45% decline in the last three generations. There are 15 subpopulations all but one of which have been undergoing declines for the last 27 years and contain fewer than 500 individuals, with two subpopulations now extirpated. We have two genomes from one subpopulation, Columbia North.

The boreal caribou (DU6), listed as threatened under the SARA, covers and extremely large range across Canada, however previous genetic evidence, as well as our current results, indicate that boreal caribou belong to two phylogenomic lineages and the phenotype evolved in parallel.[32] The boreal caribou from the Northwest Territories are from the NWB lineage and the rest sit within the NAL lineage in our results. In total, it was estimated that there are between 24,722 and 30,513 boreal DU individuals as of 2014, and there are currently 51 subpopulations, most of which are in decline. We have six genomes from the Northwest Territories boreal caribou, and 12 genomes from another six of the subpopulations: Coastal, Nipigon, Far North, Kesagami, Naosap, and Cold Lake.

The eastern migratory caribou (DU5) sit within the NAL lineage and is recommended for listing as endangered by COSEWIC. In total the eastern migratory DU was estimated to contain 170,636 mature individuals as of 2017, with an overall decline of 80% recorded over three generations. There are four subpopulations, and we have two genome sequences each from two of them; The eastern migratory caribou from the Southern Hudson Bay subpopulation and the George River subpopulation – a subpopulation which has undergone a particularly dramatic decline of 99% in three generations. It is known that eastern migratory subpopulations have historically fluctuated, however the George River subpopulation, which used the be the largest-sized subpopulation, is now lower than ever recorded and it is unclear if it will increase due to new threats.

We have two genomes from the Kangerlussuaq-Sisimiut caribou population on Greenland. Historically, the herd was in high numbers between 1815 and around 1845, however there was subsequently a huge decline in numbers, remaining low until the 1950s when numbers began to increase steadily again peaking in the 1970s.[54] The herd has been increasing once again in recent years.[54]

QUANTIFICATION AND STATISTICAL ANALYSIS

Re-sequenced whole genome filtering

Samples were extracted using a Qiagen DNAeasy tissue extraction kit following the manufacturer's instructions (Qiagen). Samples were run on a Qubit fluorometer (Thermo Fisher Scientific) using the High Sensitivity Assay Kit and normalized to 20 ng/μl at a final volume of 50 μl. The DNA was shipped to The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children (Toronto, Ontario) for library preparation and sequencing. The samples were sequenced two per lane on an Illumina HiSeq X.

All code used to filter and map re-sequenced genomes, as well as for downstream analyses, can be found on GitHub (https://github.com/BeckySTaylor/Phylogenomic_Analyses). Raw reads for all 66 individuals were cleaned using Trimmomatic version 0.38[65] using a sliding window of 4 base pairs to trim once phred score dropped below 15. We aligned all trimmed reads to the new reference genome, which we first indexed using Bowtie2 version 2.3.0.[66] We converted the SAM files to BAM files and sorted them using Samtools version 1.5,[67] and then added read group information using GATK4.[68] Using GATK4, we removed duplicates and used 'HaplotypeCaller' to call variants and produce a variant calling format (VCF) file. We used the 'CombineGVCFs' function, followed by 'GenotypeGVCFs' to produce a VCF file containing all individuals. We did two rounds of filtering on the VCF file using VCFtools version 0.1.16.[69] We removed indels, and any site low-quality genotype calls (minGQ) and low-quality sites (minQ), with scores below 20, as well as any site with a depth of less than five or more than double the mean depth of all genomes, filtering to remove sites with a depth of more than 55. For the second round of filtering, we made two VCF files; one made using a more 'stringent' filter to remove all missing data, and a 'less stringent' filter to removed sites with more than 5% missing data, resulting in 17,595,673 and 41,321,354 SNPs respectively.

We also downloaded the raw reads for five Fennoscandian Wild Tundra reindeer genomes to use as outgroups for phylogenomic analyses (ID numbers NMBU 38-42 from Weldenegodguad et al.,[90] European Nucleotide Archive: PRJEB37216). We mapped and filtered the reads as above, as well as producing a VCF file containing these and the 66 caribou genomes. We filtered the VCF file in VCFtools as above, this time removing sites with a depth over 48 (double the mean of this data set) and removing sites with more than 5% missing data, we had 16,119,954 SNPs.

We chose downstream analyses, outlined below, which are appropriate for our sampling across the very large caribou range, which included one or two samples from each of 33 different caribou subpopulations, representing eight DUs. Future work will aim to increase sample size within each subpopulation to enable analyses such as recent Ne reconstruction which need more than 1-2 samples at a fine scale to run and cannot be done grouping samples with genetic differentiation between them (e.g., using GONE or StairwayPlot2). Given the extremely large range of caribou and high number of subpopulations, this will require a huge sequencing effort and likely need to be done at a more regional scale. Many of the results we present here are, however, plotted grouped by lineage for clarity, but all statistics for each individual are given in the supplementary materials.

Whole genome phylogenomic reconstruction

For the phylogenomic reconstruction, we used IQtree version 1.6.12.[70] First, we made a consensus fasta file for each individual from the VCF file which included the reindeer using the 'consensus' command in BCFtools. We found that running IQtree on the full genome sequences required too much computational power, so we split the genome into seven sections of close to 300 million base pairs, made a phylogeny with each, and then made a consensus tree as follows. Firstly, we used the 'CSplit' command to split

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

each individual fasta file into one file per scaffold, retaining the files for scaffolds 1-36 (which contains ~99% of the reference genome, see results). We concatenated all caribou individuals together for each scaffold, and ran each scaffold from 1-36 in Model Finder in IQtree. Using the Bayesian Information Criterion, Model Finder gave the model TVM+F+I+G4 for all scaffolds apart from 33 where it selected GTR+F+I+G4. For all scaffolds the scores for these two models were close, and for scaffold 33 the likelihood score for the two models was similar (111,917,190.538 and 111,917,210.388), and so when concatenating the scaffolds, we used TVM+F+I+G4 for the full phylogenomic run.

To run the phylogenomic analysis, we then concatenated scaffolds together for each individual into seven fasta files of roughly 300 million base pairs (scaffolds 1-3, 4-7, 8-11, 12-16, 17-21, 22-27, 28-35), excluding scaffold 36 which is putatively part of the X chromosome based on the presence of known X chromosome genes on that scaffold.[91,92] We then reformatted each so that the sequence was on one line using 'awk' and 'grep' commands, and then concatenated all individuals together into one file, including the reindeer, for each of the seven sections so we had one fasta file with all individuals for each of the ~300 million base pair regions. We then ran IQtree using 100,000 bootstraps to obtain branch supports[93] (-bb command) to produce the phylogenies. We then made a consensus phylogeny from the seven using the IQtree '-con' command.

We also reconstructed an unrooted phylogeny following the protocol from von Seth et al.[11] We used ngsDist[71] to estimate a genetic distance matrix with 1000 bootstrap replicates from the genotypes allowing no missing data. We then used FASTME v2.1.6.2[72] to reconstruct the phylogeny, adding bootstrap support to the nodes using RAxML-ng v1.0.1.[73]

Genetic diversity and demographic history

We used Plink version 1.9[74] to convert the VCF file with the 66 caribou and no missing data into a BED file. We then pruned the dataset to remove sites with a correlation co-efficient of 0.1 or above in sliding windows of 50 SNPs, leaving 3,916,295 putatively unlinked SNPs, and then ran a PCA also in plink, and plotted in R studio version 1.2.5041. We also performed an assignment test using ADMIXTURE version 1.3.0[75] using the same dataset testing K1 to 10 and performing ten iterations of each and plotted the outputs in R studio (Figure S2).

To estimate individual genetic diversity, we used mlRho v2.9[28] to calculate heterozygosity for each individual from the bam files. mlRho calculates θ, an estimator of the population mutation rate which approximates heterozygosity under the infinite sites model,[11,28,29] which can be calculated from whole genomes as heterozygosity using $\theta = 4N_e\mu$[50] (where μ is the mutation rate) The files were first filtered using Samtools to remove bases with a mapping quality below 30, sites with a base quality below 30, and with a depth over 10X the average for the dataset. To estimate nucleotide diversity for each lineage, we used BCFtools to create a VCF file with all sites (variant and invariant) from the bam files, again filtering low-quality genotype calls (minGQ) and low-quality sites (minQ), as well as any site with a depth of less more than double the mean depth of all genomes as above. We then used pixy v1.2.7[76] to estimate nucleotide diversity of the 35 putative autosomes, testing two window sizes (50kb and 10kb). Results were consistent so we present results for 50kb windows.

To assess genetic distance between lineages we calculated Wier and Cockerham weighted F_{st}[94] between each pair of lineages using the VCF file with no missing data in VCFtools using a window size of 50kb and a step size of 25 kb (Table S3). We also measured introgression between each pair of lineages using ABBA BABA tests to control for incomplete lineage sorting. We used Dsuite version 0.5[77] to run the 'Dtrios' function to calculate D and f4-ratio statistics, grouping our individuals by the lineages uncovered in our phylogenomic analysis (see results), using the phylogeny as input using the '-t' command. When groups share branches on a phylogeny, many elevated D and f4-ratio statistics can occur, however these correlated statistics can be informative to uncover the relative time of the gene flow events across the phylogeny and to discover whether the gene flow occurred on internal branches by using the f-branch statistic.[77] We calculated the f-branch statistics using the output from Dtrios, and then plotted alongside the phylogeny using the 'dtools.py' script included with DSuite, setting the p-value to 0.05. As these statistics are unable to measure gene flow between sister groups, we used SplitsTree[78] to visualize the phylogenetic network as an 'admixture graph', using the seven files SplitsTree output by IQtree (one for each of the 300 million base pair phylogenomic analyses; Figure S4).

Reconstruction of historical effective population sizes has already been performed using the pairwise sequentially Markovian coalescent[27] (PSMC) for most of the individuals.[26] We analyzed the rest of the individuals in PSMC using the same filtering and parameters as used previously (Figure S3).[26]

Rapidly evolving genes and gene ontology analysis

We used GWideCodeML,[79] a python package to run the codeml function of PAML[43] in a computationally efficient way using genome-wide data. We used our annotation file to extract all genes from the genomes of our individuals to use in GWideCodeML. To do this, we made a consensus fasta file for each individual in the VCF file with our 66 caribou filtering to remove sites with more than 5% missing data, as described above. We then used the '-x' function in Gffread version 0.12.3[80] which pulls out the coding sequence for each gene as indicated in the annotation file and splices them together (to remove introns), to create one fasta file per individual with all genes. We reformatted the files so each gene sequence is on one line using 'awk' commands, and then renamed the header line of each gene to include the ID of each individual (in addition to the gene ID from the annotation) using 'sed' commands. We used the 'CSplit' command to split the files into one file per gene, and then concatenated the files for each gene – resulting in one file per gene containing the sequence for all 66 individuals.

We first ran the genes using an unrooted version of the tree as required by codeml. We removed the outgroup reindeer and then used the 'ape' package in R studio to transform the tree into an unrooted version with a trifurcation at the root, the format needed

by codeml. We used the branch model, which uses a Likelihood Ratio Test (LRT) to test whether the genes have a significantly different dN/dS ratio on the focal branch, compared to all other branches of the tree. We tested this for each of the nine major lineages uncovered in our phylogenomic analysis, excluding one individual, the boreal caribou from Alberta, which is a hybrid between the NAL and BEL lineages as indicated by the PCA analysis (Figure S2). We then performed a Bonferroni multiple testing correction on the Likelihood ratio results, adjusting the significance threshold to account for running the model over nine lineages. We then took those sites where the focal branch was putatively under positive selection (larger dN/dS ratio), and as some genes can be significant in multiple branches, we also calculated how many genes were unique to each branch. We then used a chi-squared test in R studio to determine if there were a significantly different number of positively selected genes across the different lineages.

To assign putative functions to the significant genes (after Bonferroni correction), we used ShinyGo v0.76.2[81] and assigned functions based on the GO Biological Process and GO Molecular Function databases. The enrichment analysis outputs any biological pathways over-represented in the list of genes with signatures of positive selection and was performed for the significant rapidly evolving genes for each of the nine major lineages separately.

Runs of homozygosity (ROH) estimation

To estimate the proportion of the genome in ROH we used Plink from the VCF file with no missing data and not LD pruned, and only using scaffolds 1-35 to ensure removing sex chromosomes as above. To test the impact of the key settings (homozyg-snp, homozyg-density, homozyg-gap, homozyg-window-snp, homozyg-window-het, homozyg-het) on the resulting data, we ran 11 different combinations to optimize the runs (Data S3A). Due to our high coverage (over 15X as recommended for this analysis in Plink) and very high SNP density dataset (an average of ~1 SNP every 125 bp) many of the settings did not affect the results and for those that did we chose a conservative approach (results for all runs available Data S3B–S3K) and landed on final settings of homozyg-snp 100, homozyg-density 20, homozyg-gap 1000, homozyg-window-snp 100, homozyg-window-het 1, homozyg-window-missing 5, and homozyg-het 3. Homozyg-kb and homozyg-window-threshold were set using recommendations from Meyermans et al.,[95] so using a homozyg-kb set the same as the scanning window size (100) and using their formula setting homozyg-window-threshold to 0.05. To calculate when the longest ROHs originated, we used $g = 100/(2rL)$[96], with L being the length of ROH in MB, r representing the recombination rate for which we used 1.04 cM/Mb as calculated for red deer,[97] and g the number of generations ago.

Genetic load

We used two approaches to estimate genetic load, one annotation free method and one using our new annotation, to ensure concordance of our results using different approaches and in case of any bias arising from the annotation. For this, we used both genomic evolutionary rate profiling (GERP) analysis[56] and SnpEff[82] which used our new annotation. For the GERP analysis we largely followed the protocol from von Seth et al.[11] Firstly, we generated a TimeTree phylogeny (http://www.timetree.org/) of 48 mammal species representing those with available genomes from the even-toed and odd-toed ungulates due to turnover of constrained sites over larger phylogenetic distances.[98] We downloaded the reference genomes for each of the 48 species and converted to fastq format using BBmap v38.86,[83] and then aligned to the caribou reference genomes using BWA-MEM.[84] We converted the resulting alignment files to BAM format and filtered them to remove reads aligning to more than one location as well as supplementary reads and sorted the resulting file. We then used htsbox (https://github.com/lh3/htsbox) with quality filters (-R -q 30 -Q 30 -l 35 -s 1) to convert into fasta format, and then split each file to make one file per scaffold for the first 36 scaffolds (~99% of the genome assembly) which were then concatenated together to make one fasta alignment file for each scaffold with all species.

The resulting alignment files were run through a modified version of the gerpcol function in GERP (tar file available here: https://github.com/BeckySTaylor/Phylogenomic_Analyses). Because it can lead to biases[99] the focal species, here caribou, should not be included in the GERP analysis. However, this leads to missing data in the alignment which makes it difficult to interpret the output files which don't print which site the score pertains to. We modified the code for the gerpcol function to print out the position for each score, as well as the allele for three specified sister species, here the white-tailed deer, the moose, and the red deer, which are used to determine the ancestral alleles for each site. This was run using a Ts/Tv ratio of 2.06 as calculated in BCFtools for our caribou dataset. Additionally, as the TimeTree phylogeny outputs the branch lengths in millions of years but gercol requires substitutions per site, we used a tree scaling factor (-s) of 0.0022 reflecting the number of mutations per million years on average per site based on the average mammal mutation rate of $2.2 \times 10-9$[100]. We then wrote a custom R script (available here: https://github.com/BeckySTaylor/Phylogenomic_Analyses) to automate taking the output and extracting the derived alleles (based on the three outgroup species) at all positions in the genome, all positions with positive scores, as well as all positions with a score over 2 (representing the top portion of the possible score range which is a maximum of 2.46 for our dataset and therefore the most highly constrained sites) from our 66 caribou genomes using the VCF file with no missing data.

To get another measure of genetic load, we ran SnpEff using our new annotation and then pulled out SNPs labelled as loss of function (LOF), high impact (which includes the LOF), moderate impact (e.g. missense variants), and low impact (e.g. synonymous variants). We extracted derived alleles only using three outgroup species using the same custom R script as above from the GERP analysis. For the results from both GERP and SnpEff, we compiled the number of heterozygous and homozygous variants for each individual for each category. We also calculated how many fixed homozygous alleles, alleles which were heterozygous in all

Please cite this article in press as: Taylor et al., High genetic load without purging in caribou, a diverse species at risk, Current Biology (2024), https://doi.org/10.1016/j.cub.2024.02.002

individuals, and segregating homozygous alleles there are for each lineage by extracting all positions for each lineage where at least one individual has the allele of interest, and calculated the proportions of each to allow comparisons across the load categories (as there are many more derived alleles in the lower impact categories). All numbers are presented in Data S4C and S4D. We also used this data to calculate the site frequency spectrum for each category for each lineage using easySFS[85] (available: https://github.com/isaacovercast/easySFS) using default settings, and again plotted the proportions of derived alleles in each category with all numbers available in Data S4C and S4D. We removed one individual from the in-text results (PCID 35326, NAL lineage) due a slight skew in the genetic load results likely due this individual also being used for the reference genome, although results for this individual is included in Data S4A and S4B.

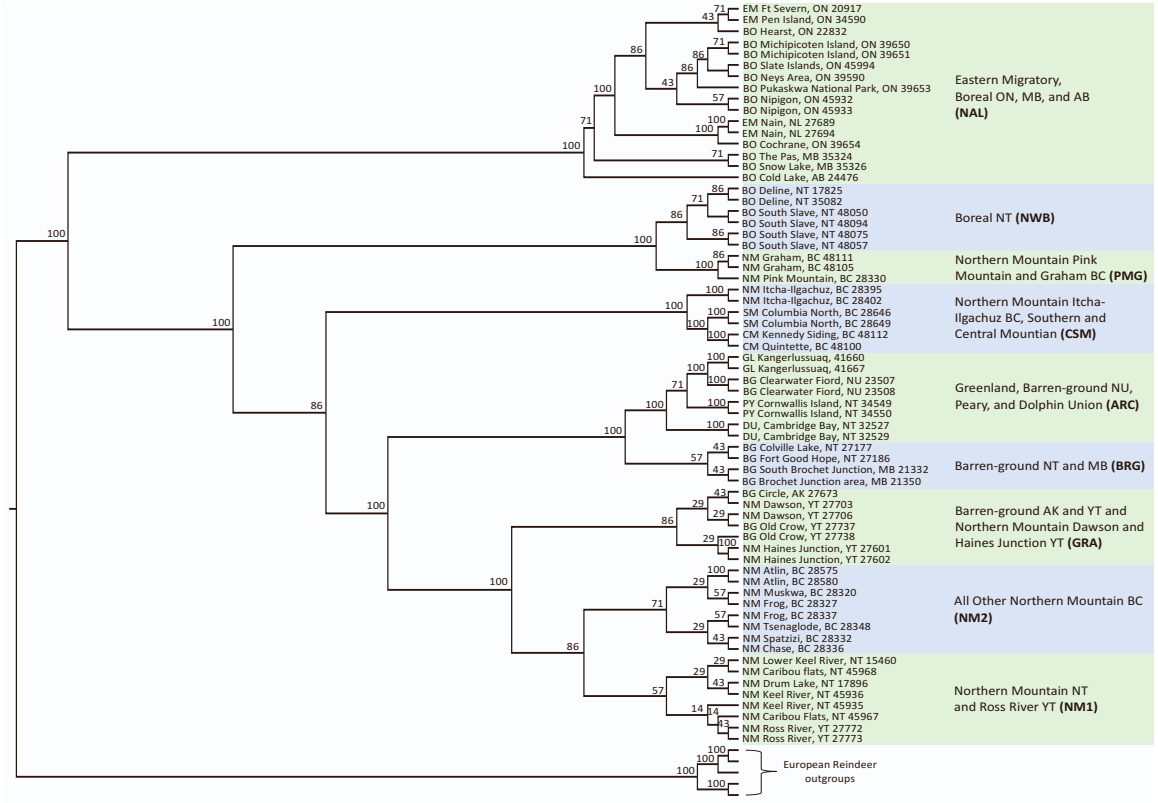
Current Biology, Volume *34*

Supplemental Information

High genetic load without purging in caribou, a diverse species at risk

Rebecca S. Taylor, Micheline Manseau, Sonesinh Keobouasone, Peng Liu, Gabriela Mastromonaco, Kirsten Solmundson, Allicia Kelly, Nicholas C. Larter, Mary Gamberg, Helen Schwantje, Caeley Thacker, Jean Polfus, Leon Andrew, Dave Hervieux, Deborah Simmons, and Paul J. Wilson

A



B

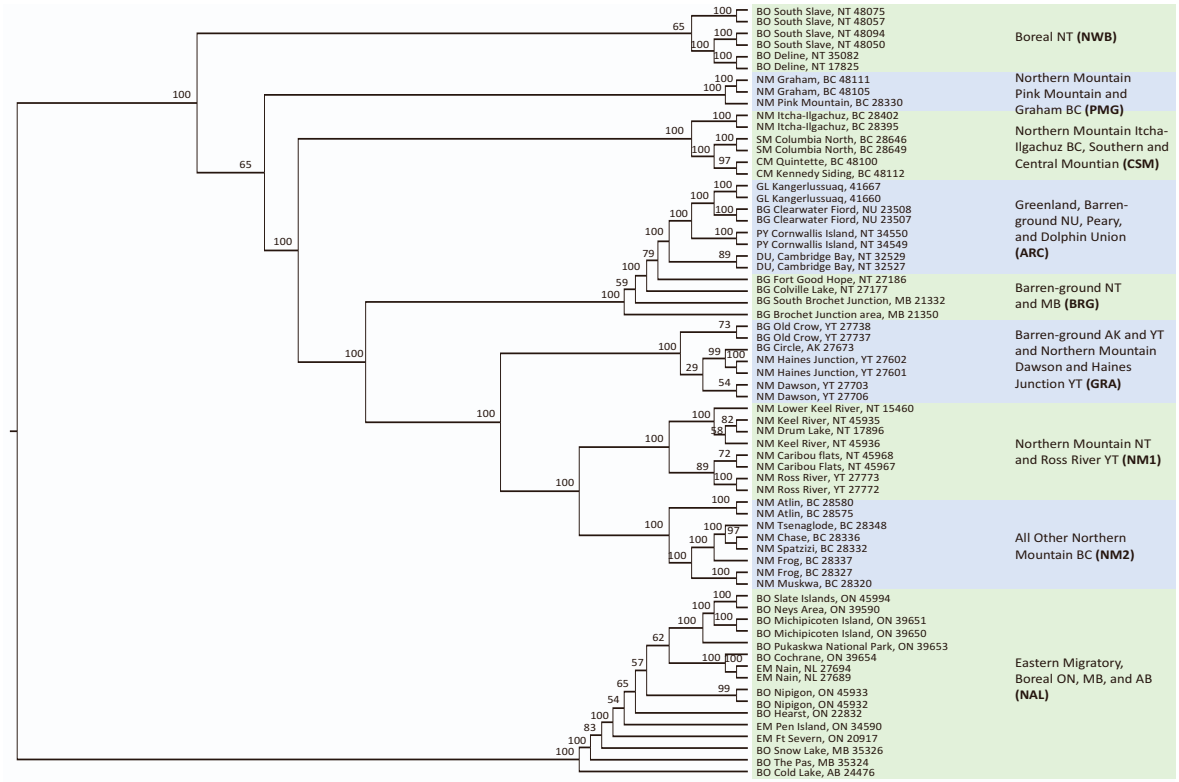


Figure S1. Phylogenomic reconstructions of the caribou genomes, related to Figure 1. Maximum likelihood phylogenomic reconstruction using the whole genome sequences and five reindeer genomes as outgroups (A) and a SNP based phylogenomic reconstruction (B) of the 66 caribou. Bootstrap supports are shown on the nodes. Each individual is identified on the tree with their PCID number (Table S2) and their Designatable Unit; BG – barren-ground, BO – boreal, CM – central mountain, DU – Dolphin Union, EM – eastern migratory, GL – Greenland, NM – northern mountain, PY – Peary, and SM – southern mountain.

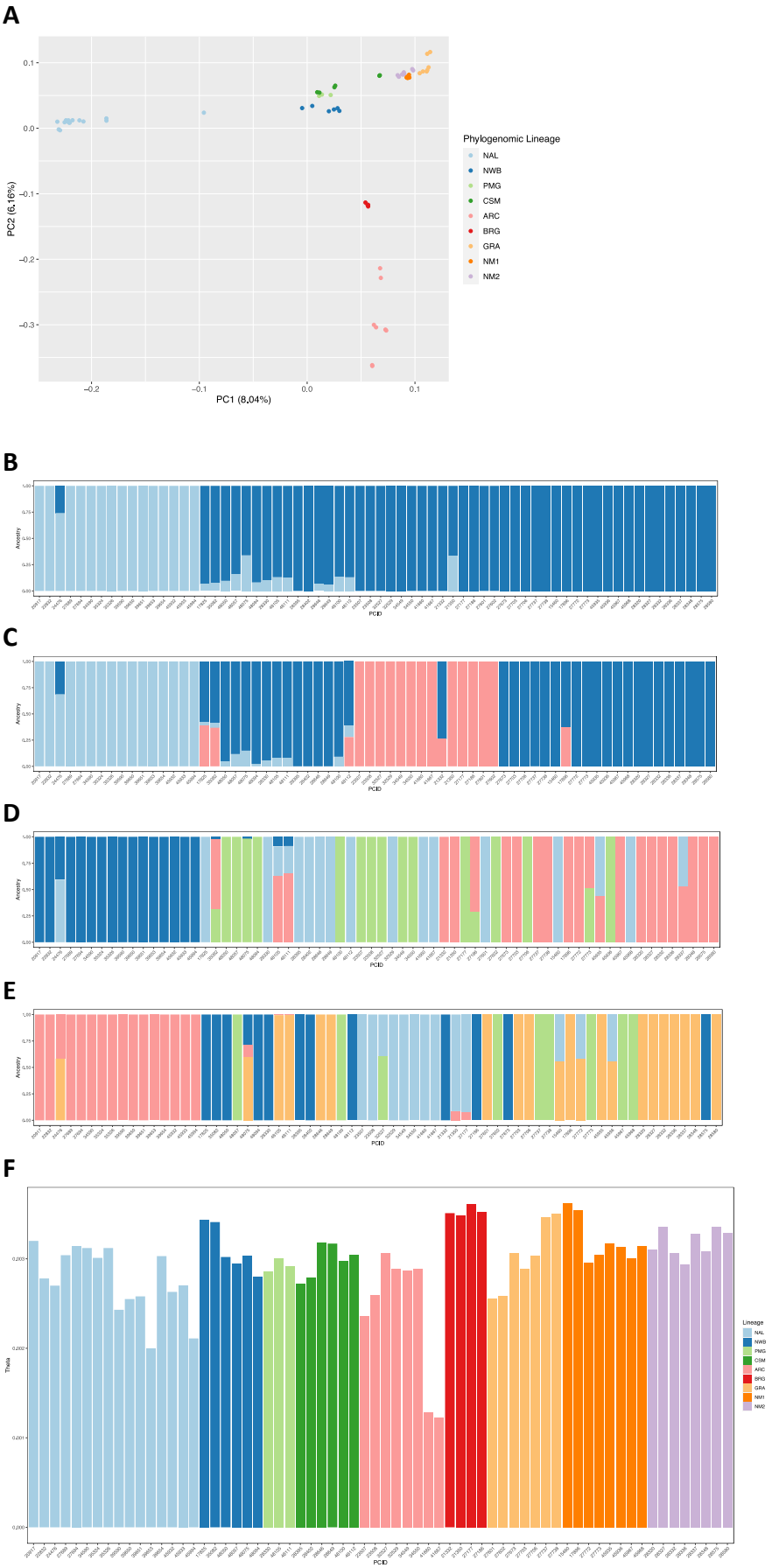


Figure S2. Genetic structure and diversity, related to Figure 1 and STAR Methods. PCA with all 66 caribou genomes, with the points coloured by the lineage within which it sits on the phylogenomic reconstruction (A), and results from an ADMIXTURE analysis showing K=2 (B), K=3 (C), K=4 (D), and K=5 (E) and Individual genetic diversity, θ, an approximation of heterozygosity (F).

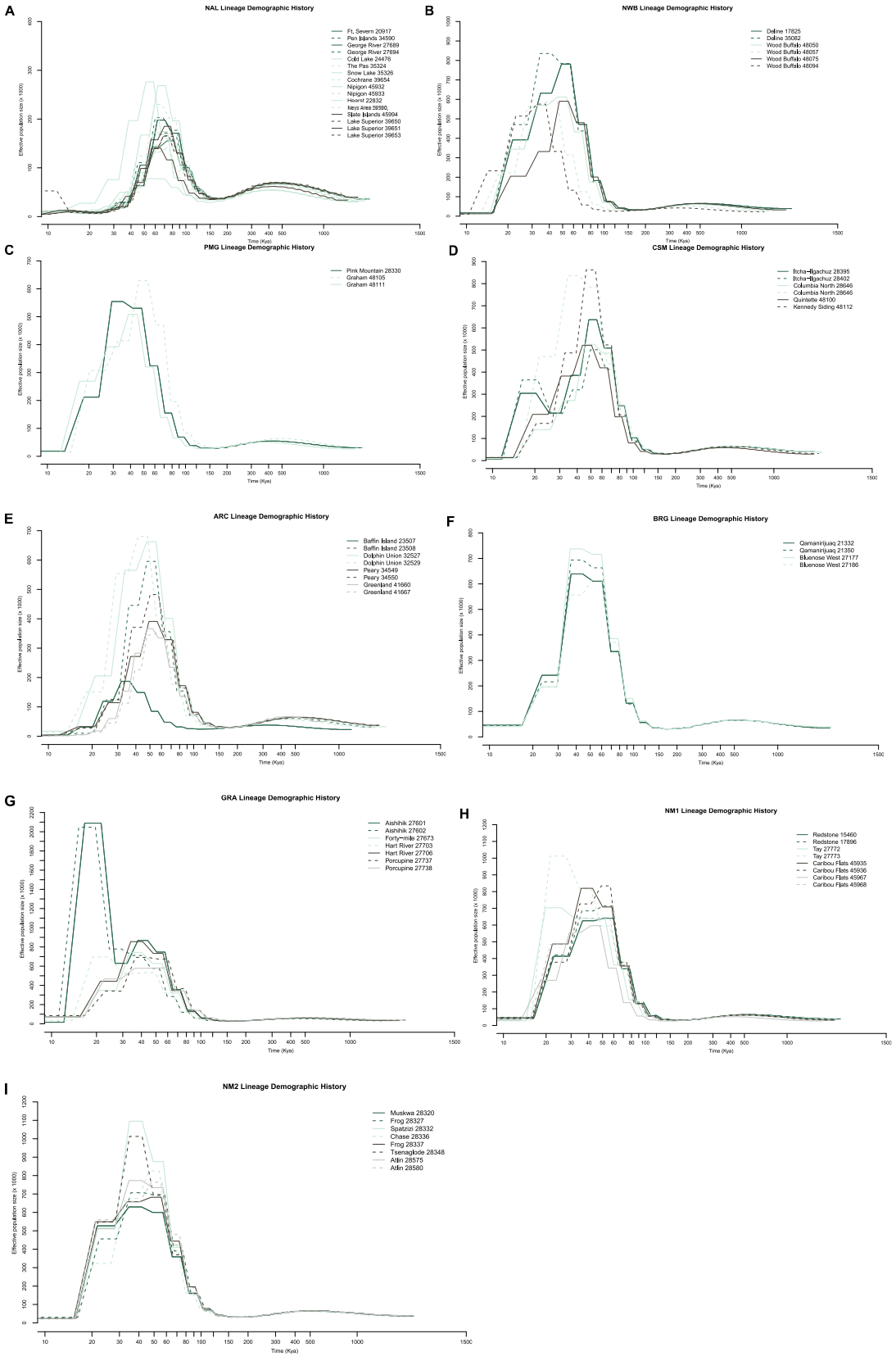


Figure S3. Reconstruction of historical effective population sizes (Ne) using the pairwise sequentially Markovian coalescent (PSMC), related to STAR Methods. Panels include the individuals in the NAL (A), NWB (B), PMG (C), CSM (D), ARC (E), BRG (F), GRA (G), NM1 (H), and NM2 (I) lineages.

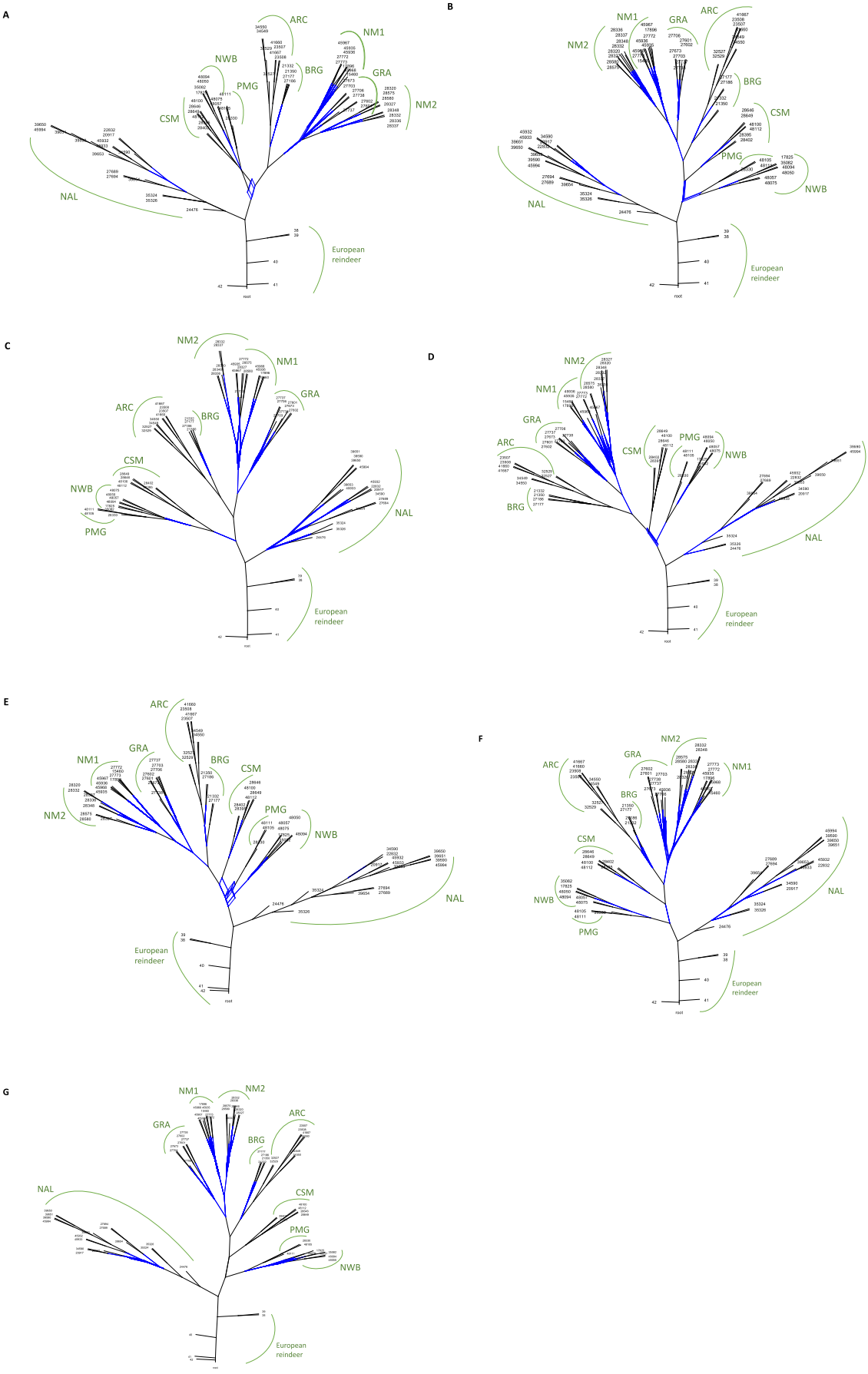


Figure S4. Admixture graphs resulting from the phylogenomic reconstruction of 66 caribou, related to Figure 2 and STAR Methods. Panels show the graphs using scaffolds 1-3 (A), 4-7 (B), 8-11 (C), 12-16 (D), 17-21 (E), 22-27 (F), and 28-35 (G). Blue lines represent phylogenetic uncertainty putatively resulting from gene flow. Green labels show the lineage to which the individuals belong.

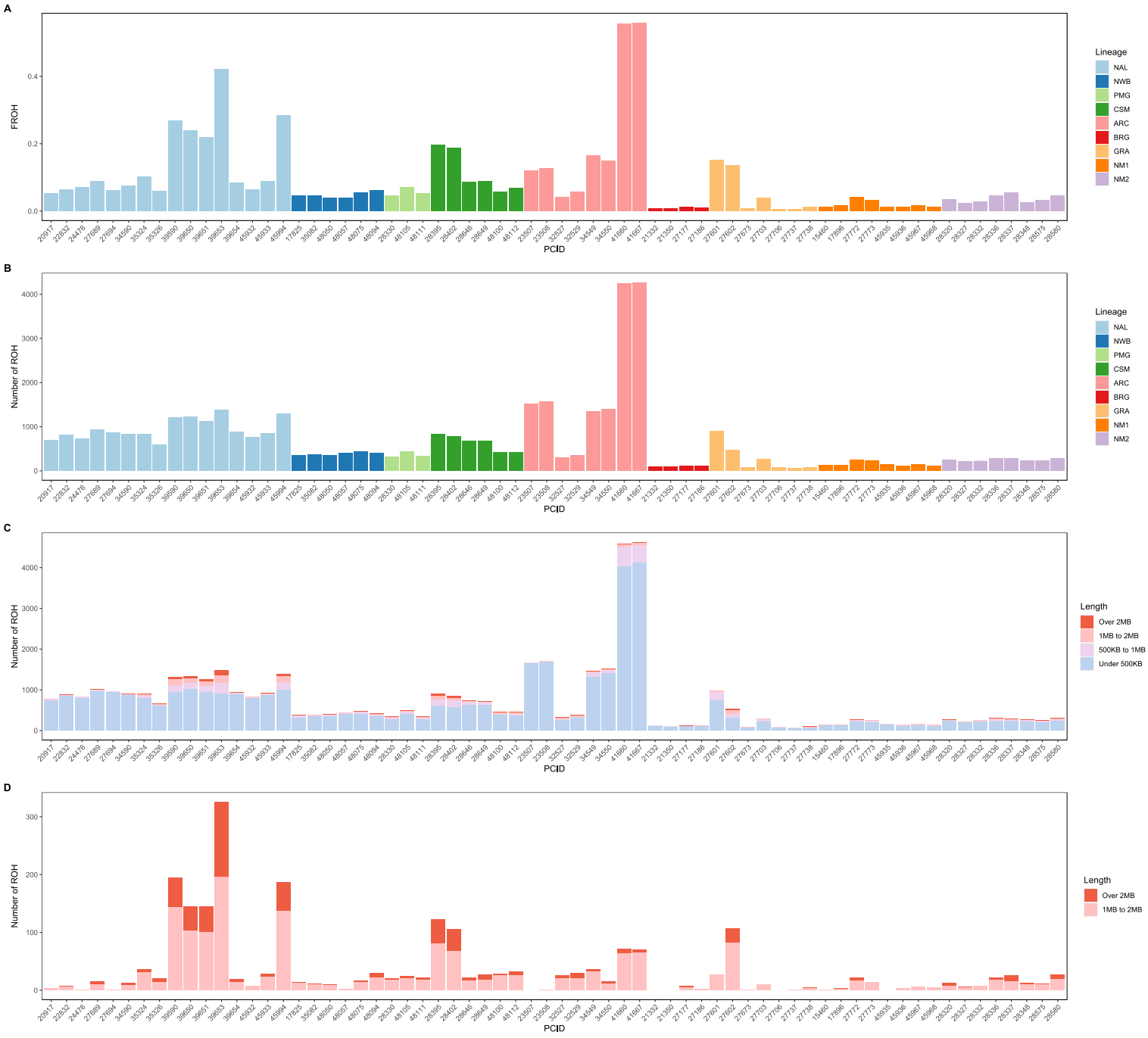


Figure S5. Runs of Homozygosity estimation, related to Figure 3. The proportion of the genome in runs of homozygosity (F_{ROH}; A), the number of runs of homozygosity for each individual (B) the number of runs of homozygosity in each individual under different length classes (C) and the number of runs of homozygosity over 1 million base pairs (1MB) for each individual (D).

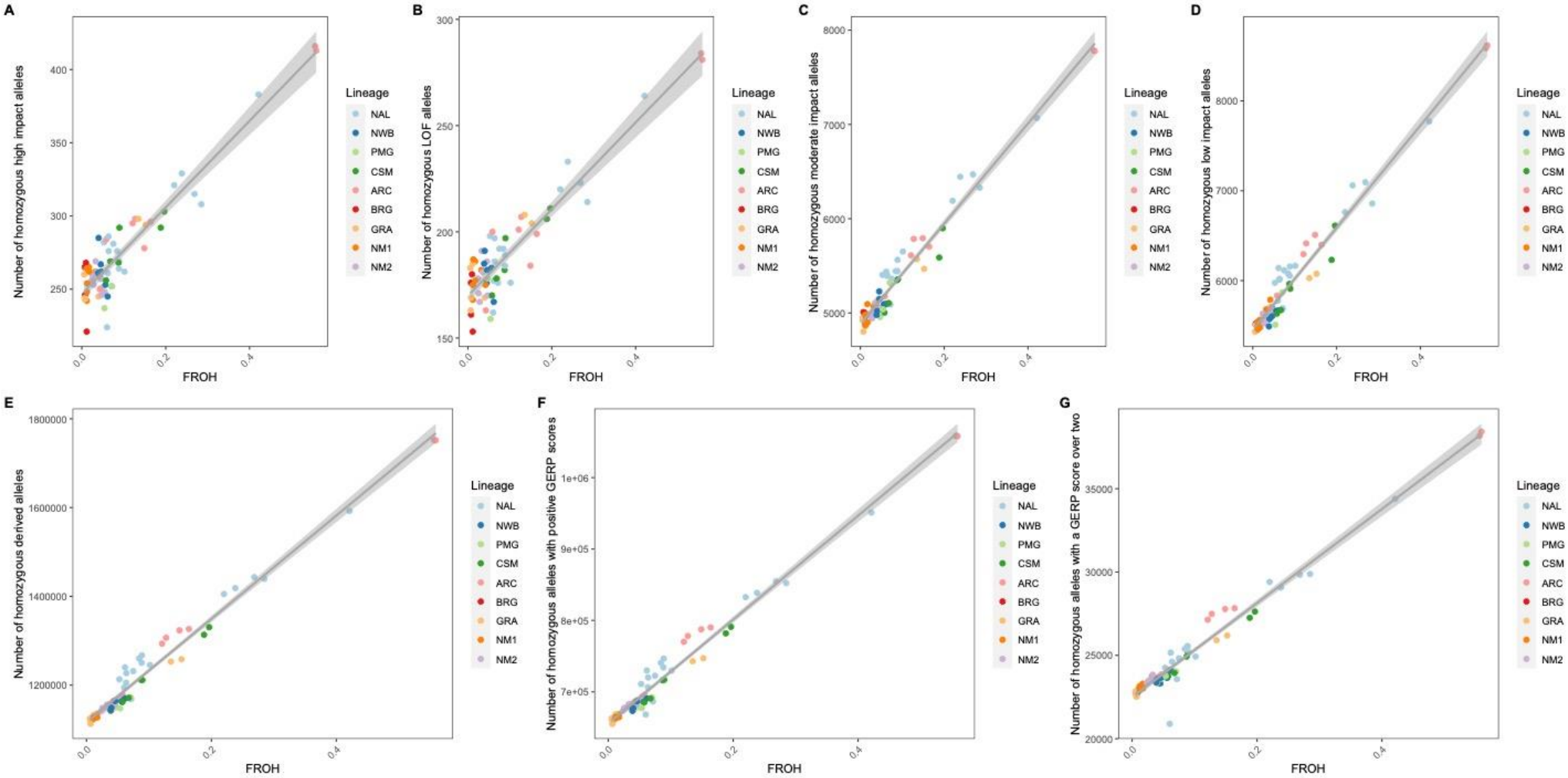


Figure S6. F_{ROH} versus homozygous genetic load, related to Figure 6. Plots show numbers of derived homozygous alleles representing different deleterious categories, including high impact (A), loss of function (B) moderate impact (C), low impact (D), all derived sites (E), positive GERP scores (F), and GERP scores over two (G). Linear regression lines were added including all individuals.

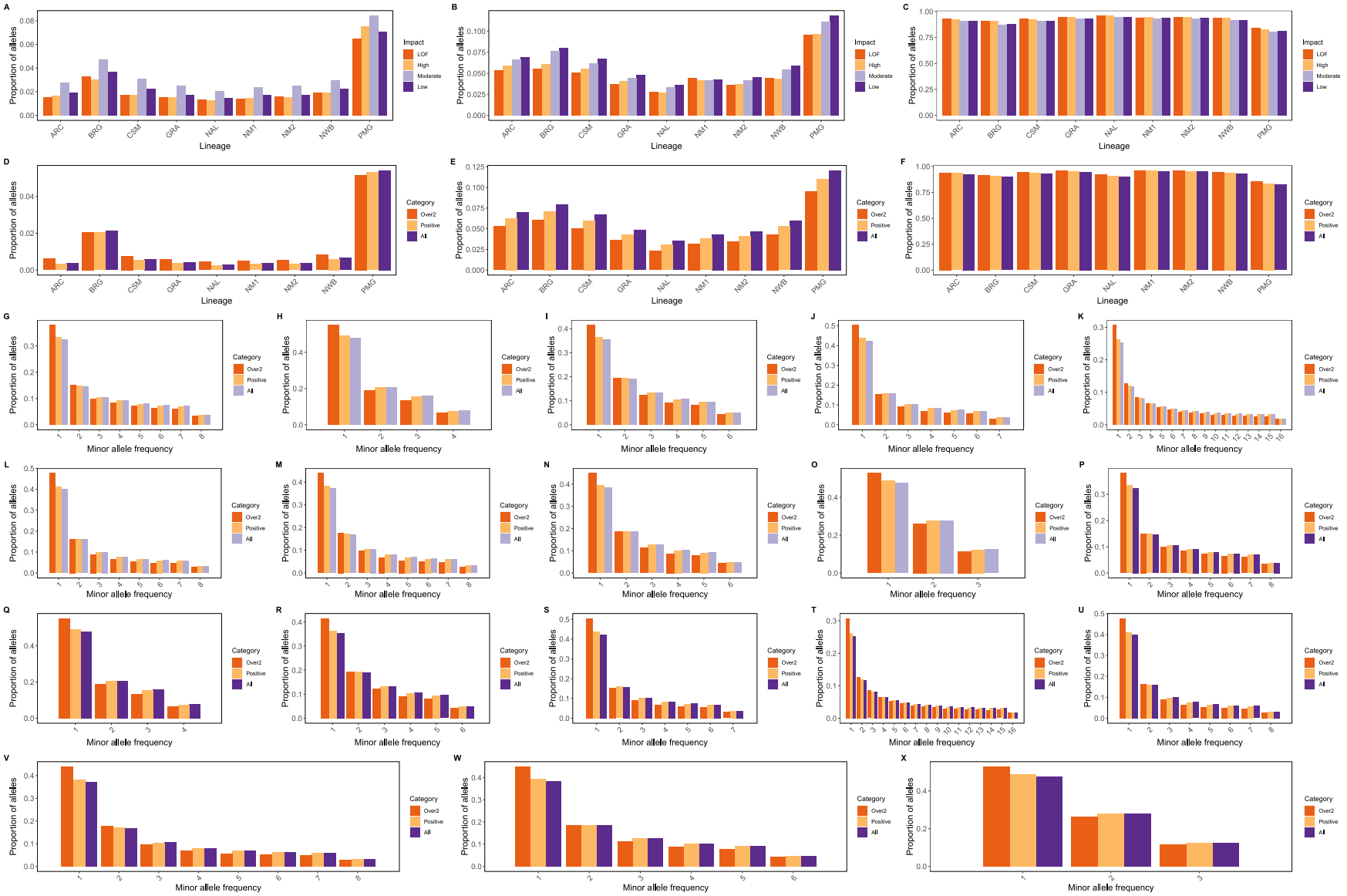


Figure S7. Per lineage proportions and site frequency spectra of genetic load, related to Figure 7. Panels show the proportions of alleles which are heterozygous in all individuals (A and D), fixed homozygous (B and E). and segregating homozygous (C and F) in all categories of deleteriousness, as well as the site frequency spectra of all categories for each lineage (G and P represent the ARC lineage, H and Q represent the BRG lineage, I and R represent the CSM lineage, J and S represent the GRA lineage, K and T represent the NAL lineage, L and U represent the NM1 lineage, M and V represent the NM2 lineage, N and W represent the NWB lineage, and O and X represent the PMG lineage).

DU	Subpopulation	Most recent population estimate	Year of most recent estimate	Year(s) of sampling	Closest population estimate to year sampled	Year of closest estimate
Peary	Western Queen Elizabeth	7,300	2012-2013	1993	Unknown exactly but there was a catastrophic die off in 1994-5 when the Bathurst and adjacent islands crashed from 3,155 to 542	1994-5
Dolphin Union	Dolphin Union	18,000	2015	1986	34,558	1997
Barrenground	Bluenose West	15,268	2015	2013	20,465	2012
Barrenground	Qamanirjuaq	264,718	2014	2007/8	348,661	2008
Barrenground	Porcupine	197,000	2013	2001	123,000	2001
Barrenground	Baffin Island	4,856	2014	1995	235,000	1991
Barrenground	Fortymile	40,000	2022	1994	22,000	1990
Northern mountain	Aishihik	1,813	2009	2002	889	1997
Northern mountain	Atlin	514-857	2007	2006	514-857	2007
Northern mountain	Chase	404	2009	2004	301	2002
Northern mountain	Frog	199	2001	2002/3	199	2001
Northern mountain	Graham	637	2009	2021	637	2009
Northern mountain	Hart River	1853	2006	1999/2000	1853	2006
Northern mountain	Itcha-Ilgachuz	1220	2012	2006	1547	2007
Northern mountain	Muskwa	828	2007	2006	828	2007
Northern mountain	Pink Mountain	1145	1993	2004	1145	1993
Northern mountain	Spatzizi	2258	1994	2003	2258	1994
Northern mountain	Tay	2907	1991	2002/5	2907	1991
Northern mountain	Tsenaglode	85-340	2008	2006	85-340	2008
Northern mountain	Redstone	>7300	2012	2013/4 and 2019	>7300	2012
Central mountain	Kennedy Siding	29	2014	2018	29	2014
Central mountain	Quintette	87	2014	2021	87	2014
Southern mountain	Columbia North	157	2013	2014	157	2013
Boreal	Northwest Territories	6,500	2012	2013/5 and 2019/20	6,500	2012
Boreal	Coastal	492	2012	1999, 2011-2017	492	2012
Boreal	Nipigon	300	2012	2011/12	300	2012
Boreal	Far North	Unknown	2012	2009	Unknown	2012
Boreal	Kesagami	492	2012	2009	492	2012
Boreal	Naosap	100-200	2012	2008/9	100-200	2012
Boreal	Cold Lake	150	2012	2014	150	2012
Eastern migratory	Southern Hudson Bay	12,479	2011	1992	At least 10,798	1994
Eastern migratory	George River	6,704	2016	2008	27,600	2012
Greenland	Kangerlussuaq-Sisimiut	60,469	2018	2009	98,300	2010

Table S1. **Caribou subpopulations and census sizes, related to Figure 1**. Information on census sizes of the sampled subpopulations, giving the population estimates as per COSEWIC status reports (for Canadian subpopulations) in numbers of mature individuals, both for the most recent estimate and the estimate closest to the sampling time.

Caribou PCID	Sampling location	Latitude/longitude	Designatable Unit	Canadian Sub-population	Lineage	Sample type	Mean depth
15460	Lower Keel River, NT	64.725884, -127.051600	Northern Mountain	Redstone	NM1	Tissue	38 (38)
17825	Deline, NT	65.496308, -122.825531	Boreal	Northwest Territories	NWB	Tissue	37 (37)
17896	Drum Lake, NT	63.891177, -126.276909	Northern Mountain	Redstone	NM1	Tissue	36 (36)
20917	Ft. Severn, ON	55.802900, -87.779700	Eastern migratory	Southern Hudson Bay	NAL	Tissue	35 (34)
21332	South Brochet Junction, MB	57.755310, -101.063060	Barren-ground	Qamanirijuaq	BRG	Tissue	36 (36)
21350	Brochet Junction area, MB	58.010560, -100.865000	Barren-ground	Qamanirijuaq	BRG	Tissue	35 (35)
22832	Hearst, ON	51.372380, -84.307190	Boreal	Far North	NAL	Hair	19 (19)
23507	Clearwater Fiord, NU	66.566670, -67.450000	Barren-ground	Baffin Island	ARC	Tissue	14 (13)
23508	Clearwater Fiord, NU	66.566670, -67.450000	Barren-ground	Baffin Island	ARC	Tissue	21 (21)
24476	Cold Lake, AB	54.994044, -110,695547	Boreal	Cold Lake	NAL	Fecal	14 (14)
27177	Colville Lake, NT	67.075993, -124.668217	Barren-ground	Bluenose West	BRG	Tissue	38 (38)
27186	Fort Good Hope, NT	66.924900, -126.385300	Barren-ground	Bluenose West	BRG	Tissue	36 (36)
27601	Haines Junction, YT	61.610434, -137.911647	Northern Mountain	Aishihik	GRA	Tissue	17 (17)
27602	Haines Junction, YT	61.610434, -137.911647	Northern Mountain	Aishihik	GRA	Tissue	16 (15)
27673	Circle, AK	64.285646, -143.215339	Barren-ground	NA, Fortymile herd	GRA	Tissue	17 (17)
27689	Nain, NL	56.913347, -61.716601	Eastern migratory	George River	NAL	Tissue	36 (36)
27694	Nain, NL	56.913347, -61.716601	Eastern migratory	George River	NAL	Tissue	36 (36)
27703	Dawson, YT	64.613165, -137.538322	Northern Mountain	Hart River	GRA	Tissue	16 (15)
27706	Dawson, YT	64.613165, -137.538322	Northern Mountain	Hart River	GRA	Tissue	17 (16)
27737	Old Crow, YT	67.660545, -140.955926	Barren-ground	Porcupine	GRA	Tissue	37 (37)
27738	Old Crow, YT	67.660545, -140.955926	Barren-ground	Porcupine	GRA	Tissue	37 (37)
27772	Ross River, YT	62.918412, -132.461189	Northern Mountain	Tay	NM1	Tissue	18 (18)
27773	Ross River, YT	62.918412, -132.461189	Northern Mountain	Tay	NM1	Tissue	22 (22)
28320	Muskwa, BC	58.665390, -125.28550	Northern Mountain	Muskwa	NM2	Tissue	25 (25)
28327	Frog, BC	57.842710, -126.473040	Northern Mountain	Frog	NM2	Tissue	34 (34)
28330	Pink Mountain, BC	57.345540, -123.581560	Northern Mountain	Pink Mountain	PMG	Tissue	15 (14)
28332	Spatzizi, BC	57.040520, -126.824510	Northern Mountain	Spatzizi	NM2	Tissue	22 (22)
28336	Chase, BC	56.585820, -126.117780	Northern Mountain	Chase	NM2	Tissue	19 (18)
28337	Frog, BC	57.930390, -126.365520	Northern Mountain	Frog	NM2	Tissue	35 (35)
28348	Tsenaglode, BC	58.028320, -129.558730	Northern Mountain	Tsenaglode	NM2	Tissue	23 (23)
28395	Itcha-Ilgachuz, BC	52.614600, -124.846290	Northern Mountain	Itcha-Ilgachuz	CSM	Tissue	32 (32)
28402	Itcha-Ilgachuz, BC	52.796100, -124.735300	Northern Mountain	Itcha-Ilgachuz	CSM	Tissue	34 (33)
28575	Atlin, BC	59.657170, -132.959110	Northern Mountain	Atlin	NM2	Tissue	35 (35)
28580	Atlin, BC	59.812640, -133.145770	Northern Mountain	Atlin	NM2	Tissue	33 (33)
28646	Columbia North, BC	51.750000, -118.430000	Southern Mountain	Columbia North	CSM	Tissue	35 (35)
28649	Columbia North, BC	51.750000, -118.430000	Southern Mountain	Columbia North	CSM	Tissue	34 (34)
32527	Cambridge Bay, NT	69.163000, -105.228000	Dolphin Union	Dolphin Union	ARC	Tissue	22 (22)
32529	Cambridge Bay, NT	69.163000, -105.228000	Dolphin Union	Dolphin Union	ARC	Tissue	18 (17)
34549	Cornwallis Island, NT	75.080000, -95.000000	Peary	Western Queen Elizabeth	ARC	Tissue	37 (37)
34550	Cornwallis Island, NT	75.080000, -95.000000	Peary	Western Queen Elizabeth	ARC	Tissue	37 (37)
34590	Pen Island, ON	55.024000, -83.537000	Eastern migratory	Southern Hudson Bay	NAL	Tissue	35 (35)
35082	Deline, NT	65.661384, -124.565284	Boreal	Northwest Territories	NWB	Tissue	37 (36)
35324	The Pas, MB	53.156200, -100.733400	Boreal	Naosap	NAL	Tissue	34 (33)
35326	Snow Lake, MB	54.093300, -100.449700	Boreal	Naosap	NAL	Tissue	35 (35)

39590	Neys Area, ON	48.8019, -86.6660	Boreal	coastal	NAL	Tissue	33 (33)
39650	Michipicoten Island, ON	47.751900, -85.745700	Boreal	Coastal	NAL	Tissue	35 (35)
39651	Michipicoten Island, ON	47.752000, -85.838500	Boreal	Coastal	NAL	Tissue	37 (37)
39653	Pukaskwa National Park, ON	48.326400, -86.182900	Boreal	Coastal	NAL	Tissue	37 (37)
39654	Cochrane, ON	50.168700, -80.289700	Boreal	Kesagami	NAL	Tissue	36 (36)
41660	Kangerlussuaq, Greenland	67.247700, -50.289200	Greenland	NA	ARC	Tissue	34 (33)
41667	Kangerlussuaq, Greenland	67.942700, -50.340300	Greenland	NA	ARC	Tissue	30 (29)
45932	Nipigon, ON	50.908, -88.34	Boreal	Nipigon	NAL	Tissue	16 (15)
45933	Nipigon, ON	50.252, -87.83	Boreal	Nipigon	NAL	Tissue	18 (18)
45935	Keel River, NT	64.210245, -126.514643	Northern Mountain	Redstone	NM1	Tissue	21 (21)
45936	Keel River, NT	64.210245, -126.514643	Northern Mountain	Redstone	NM1	Tissue	19 (19)
45967	Caribou Flats, NT	63.669889, -127.969174	Northern Mountain	Redstone	NM1	Tissue	16 (16)
45968	Caribou Flats, NT	63.669889, -127.969174	Northern Mountain	Redstone	NM1	Tissue	20 (20)
45994	Slate Islands, ON	48.667988, -87.025123	Boreal	Coastal	NAL	Antler	18 (17)
48050	South Slave, NT	62.14911, -116.21731	Boreal	Northwest Territories	NWB	Blood	19 (19)
48057	South Slave, NT	60.85677, -116.86862	Boreal	Northwest Territories	NWB	Blood	16 (16)
48075	South Slave, NT	60.218901, -116.222633	Boreal	Northwest Territories	NWB	Blood	22 (22)
48094	South Slave, NT	61.870213, -115.970767	Boreal	Northwest Territories	NWB	Blood	15 (14)
48100	Quintette, BC	54.67372, -121.30541	Central Mountain	Quintette	CSM	Blood	19 (19)
48105	Graham, BC	56.32065, -122.93355	Northern Mountain	Graham	PMG	Blood	22 (22)
48111	Graham, BC	56.85783, -123.30222	Northern Mountain	Graham	PMG	Blood	16 (16)
48112	Kennedy Siding, BC	55.3311, -122.0964	Central Mountain	Kennedy Siding	CSM	Blood	23 (23)

Table S2. **Metadata for samples used for whole genome re-sequencing, related to Figure 1 and STAR Methods**. Information includes ID numbers (PCID), sampling location including the Canadian province (AB = Alberta, BC = British Columbia, MB = Manitoba, NL = Newfoundland and Labrador, NT = Northwest Territories, NU = Nunavut, ON = Ontario, YT = Yukon Territory) or U.S. state (AK = Alaska), the Designatable Unit to which they belong, the sample type used for DNA extraction, and the average depth in the VCF files used for analyses with no missing data and when allowing 5% missing data shown in brackets.

	Number of genes with a significant dN/dS ratio	Number of unique genes with a significant dN/dS ratio	NAL	NWB	PMG	CSM	ARC	BRG	GRA	NM2
NAL	764	245								
NWB	758	250	0.089							
PMG	779	287	0.0907	0.024						
CSM	787	270	0.1043	0.0369	0.0355					
ARC	791	282	0.1494	0.0758	0.0838	0.0856				
BRG	777	276	0.0922	0.0216	0.0314	0.0318	0.0378			
GRA	749	271	0.1083	0.0296	0.0345	0.0352	0.0653	0.0119		
NM2	777	265	0.1111	0.0301	0.0338	0.0345	0.072	0.0201	0.0162	
NM1	742	278	0.1071	0.025	0.0318	0.0334	0.0661	0.0142	0.0095	0.01

Table S3. Rapidly evolving genes and $F_{st,}$ related to Figure 2 and STAR Methods. The overall number of significant rapidly evolving genes as well as the number of unique genes for each lineage, and Wier and Cockerham weighted F_{st} between each pair of lineages. The red values may need to be taken with caution due to very low sample sizes for those lineages.