Accurate runs of homozygosity estimation from low coverage genome sequences in non-model species

Rebecca Taylor[1], Micheline Manseau[1], and Paul Wilson[2]

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

November 15, 2024

Abstract

Runs of homozygosity (ROH) are increasingly being analyzed using whole genome sequences in non-model species as a measure of inbreeding and to assess demographic history, thus providing useful information for conservation. However, most studies have used Plink for ROH inference which has been shown to perform poorly when sequencing depth is below 10X, often underestimating the true proportion of the genome in ROH, which could lead to erroneous status assessment and management decisions. We use whole genome sequences from caribou, a non-model species at risk, subsampled to sequencing depths ranging from 1X to 15X, to assess the performance of ROHan, a program developed to enable ROH estimation using lower coverage sequences but so far only optimized for human data. We use 22 individuals with varying extent of inbreeding to assess the effects of sequencing depth, input parameters, and demographic history on the inference of ROH. We found that accurate estimation of the percentage of the genome and lengths of ROH can be achieved down to depths as low as 3-5X. However, input parameters and the demographic history of the individual can have a dramatic effect on results. Using our optimized settings, we then re-analyze low coverage sequences from a small and isolated caribou population and demonstrate high levels of inbreeding which had previously been missed. We provide recommendations for thorough optimization of parameters including the need for multiple runs as well as careful interpretation of outputs to enable robust ROH inference using low coverage whole genome sequences in wildlife species.

Accurate runs of homozygosity estimation from low coverage genome sequences in non-model species

Rebecca S. Taylor[1] | Micheline Manseau[1,2] | Paul J. Wilson[2]

[1] Landscape Science and Technology, Environment and Climate Change Canada, Ottawa, Ontario, Canada

[2]Biology Department, Trent University, Peterborough, Ontario, Canada

Correspondence

Rebecca S. Taylor, Landscape Science and Technology, Environment and Climate Change Canada, Ottawa, Ontario, Canada.

Email: rebecca.taylor@ec.gc.ca

Funding information

Funding from this research was provided by the Government of Canada's Genomics Research and Development Initiative (GRDI).

Abstract

Runs of homozygosity (ROH) are increasingly being analyzed using whole genome sequences in non-model species as a measure of inbreeding and to assess demographic history, thus providing useful information for conservation. However, most studies have used Plink for ROH inference which has been shown to perform poorly when sequencing depth is below 10X, often underestimating the true proportion of the genome in ROH, which could lead to erroneous status assessment and management decisions. We use whole genome sequences from caribou, a non-model species at risk, subsampled to sequencing depths ranging from 1X to 15X, to assess the performance of ROHan, a program developed to enable ROH estimation using lower coverage sequences but so far only optimized for human data. We use 22 individuals with varying extent of inbreeding to assess the effects of sequencing depth, input parameters, and demographic history on the inference of ROH. We found that accurate estimation of the percentage of the genome and lengths of ROH can be achieved down to depths as low as 3-5X. However, input parameters and the demographic history of the individual can have a dramatic effect on results. Using our optimized settings, we then re-analyze low coverage sequences from a small and isolated caribou population and demonstrate high levels of inbreeding which had previously been missed. We provide recommendations for thorough optimization of parameters including the need for multiple runs as well as careful interpretation of outputs to enable robust ROH inference using low coverage whole genome sequences in wildlife species.

KEYWORDS

Conservation genomics applications, homozygous-by-descent, inbreeding, low coverage sequencing, runs of homozygosity

1 | INTRODUCTION

Runs of homozygosity (ROH) are stretches of the genome which are entirely homozygous due to inheriting identical haplotypes from both parents, also known as autozygosity, and are increasingly being studied in non-model organisms in the field of conservation biology (Ceballos et al., 2018a; Ceballos et al., 2018b; de Assis Diniz Sobrinho et al., 2024; Duntsch et al., 2021). The inherited segments are identical by decent (IBD) when due to breeding between relatives (Brüniche-Olsen et al., 2018; Kardos, et al., 2017) or can be identical by state (IBS) due to demographic processes (Bruniche-Olsen et al., 2018). Very long ROH are more likely due to inbreeding and inheriting segments which are IBD due to recent common ancestors, whereas shorter ROH can be from more distantly related individuals and background relatedness within a population (Ceballos et al., 2018b; Martin et al., 2023). Over time with outbreeding ROH break down and so the length distribution is determined by numerous processes including both the recombination and mutation rates, the generation time, changes in effective population size, migration rate, as well as life history strategies (Bruniche-Olsen et al. 2018; Foote et al. 2021; Kardos et al. 2017). Measuring the proportion of ROH in the genome (F_{ROH}) as well as the length distributions of ROH can thus inform upon levels of inbreeding and the demographic history of the population (Allendorf, 2017; Bruniche-Olsen et al., 2018; Foote et al., 2021; Kardos et al., 2017).

Measuring ROH has given insight into the processes of genetic diversity loss and the reduction of fitness in the offspring of related individuals, known as inbreeding depression (Ceballos et al., 2018b; Duntsch et al., 2021). Studies in humans have demonstrated ROH to be enriched for homozygous deleterious variants, particularly for rare variants (Ceballos et al., 2018b), and increased investigation of ROH in non-model organisms is likely to substantially improve our understanding of the frequency and causal mechanisms of inbreeding depression in wild populations (Allendorf, 2017; Duntsch et al., 2021; Kardos et al., 2017; Silva et al., 2024). Similarly, heterozygosity has long been used to measure genetic diversity in natural populations, and can also be estimated from whole genome data using the scaled mutation rate theta (Bruniche-Olsen et al., 2018; Foote et al., 2021). As such, accurate measures of ROH, alongside heterozygosity, are highly informative metrics for conservation studies. However, the cost of producing high quality genomic data has been a limiting factor for whole genome assessments across more wildlife taxa (Lou et al., 2021).

One solution to bring down the costs of sequencing has been to reduce the depth of coverage per individual, however accurately measuring ROH with low coverage whole genome data can be difficult and lead to

incorrect results (Duntsch et al., 2021; Kardos et al., 2024; Silva et al., 2024). The large majority of studies use the 'homozyg' function in the software Plink to measure ROH likely due to its speed and simplicity (de Assis Diniz Sobrinho et al., 2024; Duntsch et al., 2021; Meyermans et al., 2020), however Plink uses an observational approach to scan windows for stretches of homozygous SNPs (Ceballos et al., 2018b; Silva et al., 2024) and so relies upon high quality genotypes. Lower coverage sequencing data may contain sites which are erroneously called as homozygous due to not covering both chromosomes during sequencing. Along with the potentially uneven distribution of called genotypes across the genome, this could lead to the overestimation of ROH in some regions (Ceballos et al., 2018a; Duntsch et al., 2021). Low coverage sequencing also has high error rates which can introduce false heterozygous SNPs and so conversely could lead to the underestimation of ROH (Ceballos et al., 2018a; Duntsch et al., 2021; Kardos et al., 2024). Due to these issues, it has been recommended by some to use at least 15X coverage to accurately estimate ROH, particularly when grouping individuals of different depths (Kardos et al., 2024), with a detailed investigation by Duntsch et al. (2021) finding that even with careful input parameter optimization in Plink, reliable estimation was not achieved when including sites below 8X coverage. Similarly, thorough exploration of Plink and BCFtools recommended 10X coverage for accurate ROH inference (Silva et al., 2024).

A program, named ROHan (Renaud et al., 2019), has been developed for the estimation of both ROH and theta from lower coverage genomes and ancient DNA, however it has not yet been as commonly applied to studies of wildlife. ROHan is a probabilistic program which runs in three steps; firstly, it estimates sequencing coverage. The calculated coverage of the sites is used in the next step by comparing to the genome-wide coverage to weight the contribution to the likelihood function. Secondly, ROHan uses a maximum weighted likelihood method to estimate local rates of heterozygosity within a set window size, and then thirdly runs a 2-state Hidden Markov Model (HMM) to find regions within ROH as well as to calculate genome-wide theta both including and excluding the ROH regions. To do this, the program calculates the expected value of segregating sites within a window given the local heterozygosity estimate and then computes the probability of a specific state generating a particular observation. Each state has a parameter which corresponds to theta and this is used to calculate the probability of a specific number of segregating sites within a window (Renaud et al., 2019). ROHan gives confidence intervals around estimates by running the Markov chain Monte Carlo (MCMC) three times using the range of heterozygosity estimates (lower bound, upper bound, and point estimates). The minimum and maximum values, once the three chains converge, are the confidence bounds (Renaud et al., 2019).

Renaud et al. (2019) used simulated data for human chromosome 1 from populations of differing effective population size (Ne) to extensively explore the effects of depth and ancient DNA damage, however the latter we will not discuss here as our focus is on modern genomes. They found that global estimates of theta were underestimated at low depths, with very wide confidence intervals for 1-2X coverage but with estimates stabilizing with smaller confidence intervals ~5X. With increasing Ne of the simulated data, the confidence intervals become larger though with good estimates still around 5X (Renaud et al., 2019). For ROH estimates, when using a larger window size, and so estimating long ROH, results were accurate with reasonable confidence intervals again around 5X. However, with smaller window sizes down to 100Kb, and so including the estimates of shorter ROH, confidence intervals were stable at 8-10X.

While Renault et al. (2019) mainly focused on the simulated human data, they did include calculations in two wildlife species, chimpanzees (*Pan troglodytes*) and Przewalkski's horses (*Equus ferus przewalskii*). However, both species were sequenced at high coverage (an average of 24.6X and 18.0-23.4X respectively; Der Sarkissian et al., 2015; Renaud et al., 2019), and so the effects of sequencing depth on non-model organisms from natural populations with complex demographic histories have not been fully explored. Further, the manuscript does not mention the impact of a key input parameter 'rohmu', or the heterozygosity rate which is tolerated within ROH regions, an important parameter to account for both mutation and sequencing errors within the data. The only guidance on choosing the parameter comes from the instruction page (https://github.com/grenaud/ROHan), which suggests that if you see regions which show reductions in heterozygosity but are not labelled as ROH, to increase the value of rohmu to, for example, 5e-5.

Where ROHan has been used on non-model species in the literature there is generally little discussion as to the input parameter choices made, with many using the default value of rohmu (or not specifying and so presumably using the default parameter) of 1e-5 (e.g., Chattopadhyay et al., 2021; Cui et al., 2023; Escoda & Castresana, 2021; Pečnerová et al., 2024). Some use the value of 2e-5 suggested on the manual page (e.g., Iannucci et al., 2021; Rasmussen et al., 2023) with fewer studies using a couple of different rohmu parameters and reporting the effect (e.g., Cerca et al., 2022). However, no full exploration of the impact of the rohmu parameter, especially in combination with different depth of sequencing and window size, has been carried out.

Given the emerging importance of ROH estimation for conservation genetics investigations of threatened species, we aim to build on the work by Renaud et al. (2019) and provide recommendations for estimating ROH with lower coverage data in non-model organisms and, importantly, show the implications of differing and complex demographic histories in natural populations as well as the impact of the rohmu parameter setting. Our specific aims are: 1) to assess the impact of the rohmu parameter on estimated of ROH and theta, 2) assess the impact of demographic history in a non-model wildlife species on the estimation of ROH and theta, and 3) demonstrate the impact of sequencing depth, down to as low as 1X, as well as window size on estimates of ROH and theta as well as their interplay with demographic history and the rohmu parameter. We then give recommendations based on our findings to allow for more accurate ROH and theta estimation in studies of natural populations.

To achieve this, we use high coverage whole genome sequences from a non-model wildlife species-at-risk, caribou (*Rangifer tarandus*), including individuals with highly varied demographic histories. Caribou are widespread across Canada with 11 extant intraspecific units, known as designatable units (DUs), and represent nine phylogenomic lineages (COSEWIC, 2011; Taylor et al., 2024). Due to various threats and/or population declines, all DUs are listed as either special concern, threatened, or endangered (COSEWIC, 2011). Alongside the large amount of phenotypic and genetic variation across their vast range, the demographic history of populations varies widely, from large and outbred, to those which have undergone strong bottlenecks but since recovered, to populations which have undergone more contemporary strong declines, all of which was reflected in a recent investigation of ROH and heterozygosity (Taylor et al., 2024). We subsampled 22 individuals representing different histories (Table 1) and with known F_{ROH} and theta values from this high coverage data to various lower depths as input to ROHan analyses to achieve our aims.

2 | MATERIALS AND METHODS

2.1 | Whole genome sequencing and filtering

We used 20 previously published genomes, 18 from across Canada and two from Greenland, all of which are 15X or higher sequencing depth and have been analyzed for ROH (Taylor et al., 2024), plus two new genomes from Newfoundland, an Island off the east coast of Canada, produced using the same laboratory and sequencing methods (Taylor et al., 2024). Individuals were chosen to represent a range of demographic histories, from highly diverse with no inbreeding, to very high F_{ROH} and relatively recent inbreeding (likely occurring population recovery (Taylor et al., 2024; Table 1). Bam files were created as per Taylor et al. (2024: all code available on Github: https://github.com/BeckySTaylor/Phylogenomic_Analyses), including trimming, mapping to the reference genome, and duplicate removal. We then used Samtools v1.1 (Li et al., 2009) to subsample the files down to multiple depths: 15X, 10X, 8X, 6X, 5X, 4X, 3X, 2X, and 1X.

2.2 | Plink analysis

We made individual VCF files for the 15X dataset, and then performed joint genotyping to produce a VCF containing all 22 individuals in GATK4, using the same protocol as Taylor et al. (2024). The files were subsequently filtered in VCFtools v0.1.16 (Danecek et al., 2011) to only keep the largest 35 scaffolds representing over 99% of the genome assembly and excluding the sex chromosomes, removing indels and sites with low-quality genotype calls (minGQ) and low-quality sites (minQ) with scores below 20. We also filtered the VCF by removing sites with less than half the average depth rounded down (so less than 7X) and removed sites over twice the average depth (more than 30X). We then did a second round of filtering

removing sites with any missing data.

We then used the Plink v2 (Purcell et al., 2007) homozyg function for ROH analysis. We have previously optimized the settings for the caribou dataset (Taylor et al., 2024) and so ran the same settings here, using: homozyg-snp 100, homozyg-density 20, homozyg-gap 1000, homozyg-window-snp 100, homozyg-window-het 1, homozyg-window-missing 5, homozyg-window-threshold 0.05, homozyg-het 3. We ran two different settings for the 'homozyg-kb' to enable direct comparisons to the two window sizes used in ROHan (below), and so this was ran using both 100 and 1000 (and so using windows of 100Kb and 1Mb). We calculated the percentage of the genome in ROH for each individual from each run as the percent ROH is also output by ROHan. This can easily be converted into the commonly used F_{ROH} by dividing by 100. We plotted the number of ROH (NROH) against the average size of ROH (SROH) using the 100Kb output to inform upon the demographic histories of the individuals, as per Ceballos et al. (2018b).

2.3 | ROHan analyses

For each Bam file, we clipped overlapping regions using BamUtil (Jun et al., 2015), and then indel realignment was performed using GATK v3.8 (McKenna et al., 2010). Bam files were not filtered for mapping or base quality as these sites are informative for the ROHan model, as per the instructions in the documentation (https://github.com/grenaud/ROHan). ROHan was then run for each bam file at each depth, always specifying the 35 autosomes as with Plink and using a transition:transversion ratio of 2.06 which was calculated from the VCF file using BCFtools v1.19 (Danecek et al., 2021). We investigated the effect of the rohmu parameter by running a range of different thresholds chosen based on a few preliminary runs: 2e-3, 8e-4, 5e-4, 2e-4, 8e-5, and 5e-5. We also investigated the effect of window size by running all depths and rohmu parameters with the default window size (1Mb) and a window size of 100Kb, for a total of 2,376 ROHan runs.

We also downloaded 10 low coverage genomes from the Gaspésie caribou population which were published previously (Dedato et al., 2022). The Gaspésie caribou are a small and isolated population which was thought to number fewer than 120 individuals as of 2013, and as such is listed as endangered under the Committee on the Status of Endangered Wildlife in Canada (COSEWIC; COSEWIC, 2014). Despite this, ROH analysis in Plink estimated extremely low F_{ROH}, averaging just 0.011 (or 1.1% of the genome). We downloaded the raw reads from the NCBI (SRR19545211- SRR19545216 and SRR1954518- SRR1954521) and aligned them to our caribou reference genome, including duplicate removal, using the same pipeline as before (Taylor et al., 2024). Resulting bam files were clipped with indels realigned as above and we calculated the depth using Samtools. We used ROHan to calculate the percentage of the genome in ROH and theta from the bam files using 'optimized' settings which we found to give the most comparable results to high coverage analysis in Plink at both window sizes (see Results), and so using a rohmu parameter of 2e-4 for the default window setting and 8e-4 for the 100Kb window setting.

3 | RESULTS AND DISCUSSION

3.2 | Impacts of sequence depth and demographic history at a smaller window size

We show the results for the rohmu parameters which gave the closest values to Plink for the 15X dataset (Figures 3-6), however it is possible that Plink is overestimating long ROH, as was found by Silva et al. (2024). However, we ran Plink to set a limit on the total number of heterozygous sites allowed in a ROH (3), as well as how many were allowed per window (1) which should reduce the amount that long ROH are overestimated, in contrast to Silva et al. (2024) who left the number of heterozygous sites allowed in a ROH at the default value (unlimited). Either way, we present the full results for all rohmu parameters in the supplementary material (Figures S1-20) and discuss the relevant patterns for the different settings.

For the smaller window size, and so including the measurement of shorter ROH in the genome, the results for a rohmu of 8e-4 (Figure 3) and 5e-4 (Figures S3) were similar. The inferred percentage of the genome in ROH and the average length of ROH were reasonably stable down to depths as low as 3X, although for percentage of the genome in ROH the confidence intervals were large at very low depth, stabilizing ~5X (Figure 3a and

b). The exception to this was for the bottlenecked individuals from Greenland and Newfoundland where the confidence intervals for percentage ROH were large at higher depths, and also had higher proportions of their genomes being unclassified by ROHan (Figure 3a and c). There was also a stronger impact of depth on the more inbred individuals, particularly on the inferred average length of ROH, which was consistently smaller at lower depths (Figure 3b). For the very lenient rohmu parameter (2e-3) confidence intervals were very wide and ROH clearly overestimated across the board (Figures S1). At more stringent settings the confidence intervals become much larger at 1X, particularly for the average length of ROH (Figures S5-S10). Interestingly, the percentage of the genome which was labelled as unclassified went down as the stringency of the rohmu parameter went up, although with the inbred and bottlenecked individuals always having the highest values (Figure 3c, Figures S1-10).

ROHan outputs theta both excluding and including ROH, and even though for conservation purposes the theta including ROH (and thus reflecting the loss of diversity due to inbreeding) is likely more informative, looking at both can be revealing about the performance of program settings. For both measurements of theta, we find the opposite pattern to Renaud et al. (2019) as for most of our inferences theta dropped as sequencing depth went up (Figure 4). This suggests that sequencing errors are increasing inferred heterozygosity at lower depth in our data whereas for Renaud et al. (2019) the low depth may have been causing false homozygous sites. Renaud et al. (2019) used simulated data for most of the program performance testing and so the different results may be due to our use of an empirical dataset. Lower coverage data has been shown to underestimate ROH in Plink, also likely due to sequencing errors creating false heterozygous sites, in other empirical datasets (Duntsch et al., 2021; Kardos et al., 2024).

At different rohmu parameters, the inferred theta including ROH was very similar (Figure 4b, Figures S1-S10). However, the theta excluding ROH changes dramatically, with the very lenient setting (2e-3) having such large portions of the genome in ROH or unclassified that theta excluding ROH became unrealistically low in inbred and bottlenecked individuals (Figure S2). For the most stringent settings of 2e-4, 8e-5, and 5e-5 we also found that theta excluding ROH started to drop likely due to 'true' ROH not being classified as such and thus falsely being included in the calculation (Figures S6-S10).

3.3 | Impacts of sequence depth and demographic history at a larger window size

For the larger window size of 1Mb, and so scanning for longer ROH likely as a result of recent inbreeding, the confidence intervals for the percentage of the genome in ROH are generally smaller than for 100Kb window size and stabilize at a lower depth of ~3X (Figure 5a; Figures S11 to S20). The exception to this was for the northern mountain caribou from Itcha-Ilgachuz which are skewed towards longer ROH and so showing signatures of inbreeding, though not as strong (or as recent) as for the Lake Superior caribou (Figure 1; Table 1). The long ROH for these two individuals were only detected at higher depths of coverage, starting around 5-6X (Figure 5a and b). The reasons for this are unclear but it may suggest that when scanning for long ROH using a larger window size, there is a threshold size of ROH below which the ROH are not easily detected at lower depth. Note that there was not an increase in the percentage of the genome labelled as unclassified for those individuals at lower depths either (Figure 5c).

As with the smaller window size, the percentage of the genome which is unclassified goes down with increasing stringency of the rohmu parameter (Figure 5c; Figures S11 to S20). At more lenient parameters the percentage unclassified is much higher for the highly bottlenecked Greenland individuals than for any others, although as the setting becomes more stringent the individuals with stronger recent inbreeding start to have the highest amount of the genome unclassified (Figures S11 to S20). For the rohmu parameter which was closest to the Plink results, 2e-4, the difference between the bottlenecked and recently inbred individuals is lower, however, and appears to be a 'middle ground' between the two (Figure 5c).

As with the smaller window size, estimates of theta go down with increasing depth (Figure 6; Figures S12 to S20). Theta including ROH was similar to those from the 100Kb window size as well, although with generally smaller confidence intervals (Figures 4b and 6b). The theta excluding ROH is always very low for the bottlenecked individuals from Greenland and Newfoundland, being similar to the theta including ROH

in those individuals (Figure 6a; Figures S12 to S20). When using a larger window size, seeing this pattern can thus indicate the presence of shorter ROH, such as seen in bottlenecked populations, which are not being detected.

not-yet-known not-yet-known

not-yet-known

unknown

AUTHOR CONTRIBUTIONS Rebecca S. Taylor conceptualized the study, performed all analyses, and wrote the manuscript. Micheline Manseau and Paul J. Wilson conceptualized the study and edited the manuscript.

ACKNOWLEDGEMENTS

We would like to thank Bridget Redquest and Austin Thompson for technical support in the laboratory. We are grateful to the National Biodiversity Cryobank of Canada at the Canadian Museum of Nature for providing the two samples from Newfoundland. We would also like to thank the Digital Research Alliance of Canada for their high-performance computing resources. Funding from this research was provided by the Government of Canada's Genomics Research and Development Initiative (GRDI).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw reads for 20 high coverage genomes used in this study have been deposited to the NCBI Sequence Read Archive (SRA): PRJNA634908, PRJNA694662, PRJNA754521, PRJNA984705, PRJNA1040806. The two new genomes used in this study will be made available upon acceptance. All scripts used to align and filter the data have been made available on GitHub: https://github.com/BeckySTaylor/Phylogenomic_Analyses/.

REFERENCES

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology, 26* , 420-430. https://doi.org/10.1111/mec.13948

Brüniche-Olsen, A., Kellner, K. F., Anderson, C. J., & DeWoody, J. A. (2018). Runs of homozygosity have utility in mammalian conservation and evolutionary studies. *Conservation Genetics* , *19* , 1295–1307. https://doi.org/10.1007/s1059 2-018-1099-y.

Ceballos, F. C., Hazelhurst, S., & Ramsay, M. (2018a). Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics, 19* , 106. https://doi.org/10.1186/s12864-018-4489-0

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018b). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics* , *19* , 220–234. https://doi.org/10.1038/nrg.2017.109

Cerca, J., Westbury, M. V., Heide-Jorgensen, M. P., Kovacs, K. M., Lorenzen, E. D., Lydersen, C., Shpak, O. V., Wiig, O., & Bachmann, L. (2022). High genomic diversity in the endangered East Greenland Svalbard Barents Sea stock of bowhead whales (*Balaena mysticetus*).*Scientific Reports, 12* , 6118. https://doi.org/10.1038/s41598-022-09868-5

Chattopadhyay, B., Forcina, G., Garg, K. M., Irestedt, M., Guerrini, M., Barbanera, F., & Rheindt, F. E. (2021). Novel genome reveals susceptibility of popular gamebird, the red-legged partridge (*Alectoris rufa* , Phasianidae), to climate change. *Genomics, 113* , 3430-3438. https://doi.org/10.1016/j.ygeno.2021.08.010

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

COSEWIC. (2014). COSEWIC Assessment and Status Report on the Caribou Rangifer tarandus, Newfoundland population, Atlantic-Gaspesie population, Boreal population in Canada.

Cui, R., Wu, J., Yan, K., Luo, S., Hu, Y., Feng, W., Lu, B., & Wang, J. (2023). Phased genome assemblies reveal haplotype-specific genetic load in the critically endangered Chinese *Bahaba* (Teleostei, Sciaenidae). *Molecular Ecology, 33* , e17250. https://doi.org/10.1111/mec.17250

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., McVean, G., & Durbin, R.: 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics 27* , 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., Li, H. (2021). Twelve years of SAMtools and BCFtools. Gigascience, 10, giab008. https://doi.org/10.1093/gigascience/giab008.

De Assis Diniz Sobrinho, F., Coelho, R. C., Moura, J. D. O., Bajay, M. M., Carvalho, L. C. B., Britto, F. B., Sarmento, J. L. R., Mastrangelo, S., & de Araujo, A. M. Genetic diversity and runs of homozygosity (ROH): A portrait of the quantitative academic publication dynamic and scientific metadata. Available at: https://www.researchsquare.com/article/rs-4202560/v1.

De Cara, M, A, R., Villanueva, B., Toro, M. A., & Fernandez, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology, 22* , 6091-6099. https://doi.org/10.1111/mec.12560

Dedato, M. N., Robert, C, Taillon, J., Shafer, A. B. A., & Cote, S. D. (2022). Demographic history and conservation genomics of caribou (*Rangifer tarandus*) in Quebec. *Evolutionary Applications, 15* , 2043-2053. https://doi.org/10.1111/eva.13495

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., Jonsson, H., Bar-Gal, G. K., Albrechtsen, A., Vieira, F. G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner, C., ... Orlando, L. (2015). Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Current Biology, 25* , 2577-2583.

Duntsch, L., Whibley, A., Brekke, P., Ewen, J. G., & Santure, A. W. (2021). Genomic data of different resolutions reveal consistent inbreeding estimates but contrasting homozygosity landscapes for the threatened Aotearoa New Zealand hihi. *Molecular Ecology ,30* , 6006–6020. https://doi.org/10.1111/mec.16068

Escoda, L., & Castresana, J. (2021). The genome of the Pyrenean desman and the effects of bottlenecks and inbreeding on the genomic landscape of an endangered species. *Evolutionary Applications, 14* , 1898-1913. https://doi.org/10.1111/eva.13249

Foote, A. D., Hooper, R., Alexander, A., Baird, R. W., Baker, C. S., Ballance, L., Barlow, J., Brownlow, A., Collins, T., Constantine, R., Rosa, L. D., Davison, N. J., Durban, J. W., Esteban, R., Excoffier, L., Fordyce Martin, S. L., Forney, K. A., Gerrodette, T., Gilbert, M. T. P., ... Morin, P. A. (2021). Runs of homozygosity in killer whale genomes provide a global record of demographic histories.*Molecular Ecology* , *30* , 6162–6177. https://doi.org/10.1111/mec.16137

Iannucci, A Benazzo, A., Natali, C., Arida, E. A., Zein. M. S. A., Jessop, T. S., Bertorelle G., Ciofi, C. (2021). Population structure, genomic diversity and demographic history of Komodo dragons inferred from whole-genome sequencing. Molecular Ecology, 30, 6309-6324. https://doi.org/10.1111/mec.16121

Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population scale DNA sequence data. *Genome Research* . https://doi.org/10.1101/gr.176552.114

Kardos, M., Qvarnstrom, A., & Ellegren, H. (2017). Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in *Ficedula* Flycatcher Genome Sequences. *Genetics, 205,*

1319-1334. https://doi.org/10.1534/genetics.116.198861

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

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology, 30* , 5966-5993. https://doi.org/10.1111/mec.16077

Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging populations of an island-colonizing bird.*Molecular Ecology, 32* , 1972-1989. https://doi.org/10.1111/mec.16865

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.*Genome Research, 20* , 1297–1303. https://doi.org/10.1101/gr.107524.110

Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics, 21* , 94. https://doi.org/10.1186/s1286 4-020-6463-x

Pečnerová, P., Lord, E., Garcia-Erill, G., Hanghøj, K., Rasmussen, M. S., Meisner, J., Liu, X., van der Valk, T., Santander, C. G., Quinn, L., Lin, L., Liu, S., Carøe, C., Dalerum, F., Götherström, A., Måsviken, J., Vartanyan, S., Raundrup, K., Al-Chaer, A., ... Siegismund, H. R. (2024). Population genomics of the muskox' resilience in the near absence of genetic variation. *Molecular Ecology, 33,* e17205. https://doi.org/10.1111/mec.17205

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., & Sham, P. K. (2007). PLINK: a Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics, 81* , 559–575. https://doi.org/10.1086/519795

Rasmussen, L., Fontsere, C., Soto-Calderón, I. D., Guillen, R., Savage, A., Hansen, A. J., Hvilsom, C., & Gilbert, M. T. P. (2023). Assessing the genetic composition of cotton-top Tamarins (*Saguinus oedipus*) before sweeping anthropogenic impact. *Molecular Ecology, 32* , 5514-5527. https://doi.org/10.1111/mec.17130

Renaud, G., Hanghøj, K., Korneliussen, T. S., Willerslev, E., & Orlando, L. (2019). Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics* ,*212* , 587–614. https://doi.org/ 10. 1534/ genetics.119.302057

Silva, G. A. A., Harder, A. M., Kirksey, K. B., Mathur, S., & Willoughby, J. R. (2024). Detectability of runs of homozygosity is influenced by analysis parameters and population-specific demographic history. *PLoS Computational Biology, 20* , e1012566. https://doi.org/10.1371/journal.pcbi.1012566

Taylor R. S., Manseau, M., Keobouasone, S., Lui, P., Mastromonaco, G., Solmundson, K., Kelly, A., Larter, N. C., Gamberg, M., Schwantje, H., Thacker, C., Polfus, J. L., Andrew, L., Hervieux, D., Simmons, D., & Wilson, P. J. (2024). High genetic load without purging in caribou, a diverse species-at-risk. *Current Biology, 34,* 1234-1246. DOI: 10.1016/j.cub.2024.02.002

TABLE 1 Location, demographic history, and the percent of the genome in ROH at 15X for each caribou using Plink and 'optimal' Rohan settings of 8e-4 at 100kb window size and 2e-4 at 1Mb window size.

Caribou PCID	**Location and Designatable Unit**	**Demographic history sum**
20917	Ft. Severn, Ontario, Eastern Migratory	Low signature of inbreeding

Caribou PCID	Location and Designatable Unit	Demographic history sum
21332	Qamanirijuaq, Manitoba, Barren-ground	Highly genetically diverse, re
21350	Qamanirijuaq, Manitoba, Barren-ground	Highly genetically diverse, re
22832	Hearst, Ontario, Boreal	Low signature of inbreeding
23507	Baffin Island, Nunavut, Barren-ground	Some inbreeding likely due to
23508	Baffin Island, Nunavut, Barren-ground	Some inbreeding likely due to
27689	George River, Newfoundland and Labrador, Eastern Migratory	Low signature of inbreeding
27694	George River, Newfoundland and Labrador, Eastern Migratory	Low signature of inbreeding
28395	Itcha-Ilgachuz, British Columbia, Northern Mountain	Signature of inbreeding being
28402	Itcha-Ilgachuz, British Columbia, Northern Mountain	Signature of inbreeding being
34590	Pen Islands, Ontario, Eastern Migratory	Low signature of inbreeding
39590	Neys Area, Ontario, Boreal	Strong signature of inbreedin
39650	Lake Superior, Michipicoten Island, Ontario, Boreal	Strong signature of inbreedin
39651	Lake Superior, Michipicoten Island, Ontario, Boreal	Strong signature of inbreedin
39653	Lake Superior, Pukaskwa National Park, Ontario, Boreal	Strong signature of inbreedin
39654	Cochrane, Ontario, Boreal	Low signature of inbreeding
41660	NK Kangerlussuaq, Greenland	Highly bottlenecked with mu
41667	NK Kangerlussuaq, Greenland	Highly bottlenecked with mu
45932	Nipigon (Nakina), Ontario, Boreal	Low signature of inbreeding
45933	Nipigon (Nakina), Ontario, Boreal	Low signature of inbreeding
50219	Terra Nova National Park, Newfoundland and Labrador, Newfoundland	Bottlenecked, skewed towards
50234	Gros Morne National Park, Newfoundland and Labrador, Newfoundland	Bottlenecked, skewed towards