Continent-wide phylogenomic framework reveals introgression as a driver of intra-specific diversity and enriched molecular pathways in caribou

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

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

[2]Toronto Zoo, Toronto, Canada

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

[4]Department of Environment and Natural Resources, Government of Northwest Territories, Fort Smith, Northwest Territories, Canada

[5]Retired, formally Department of Environment and Natural Resources, Government of Northwest Territories, Fort Simpson, Northwest Territories, Canada

[6]Gamberg Consulting, Whitehorse, Yukon, Canada

[7]Retired, formally British Columbia ministry of Forest, Lands, Natural Resource Operations, and Rural Development, Nanaimo, British Columbia, Canada

[8]British Columbia ministry of Forest, Lands, Natural Resource Operations, and Rural Development, Nanaimo, British Columbia, Canada

[9]Canadian Wildlife Service - Pacific Region, Environment and Climate Change Canada, Kelowna, British Columbia, Canada

[10]ʔehdzo Got’ı̨nęGots’ę́Nákedı (Sahtú Renewable Resources Board), Tulít'a, Northwest Territories, Canada

[11]Department of Environment and Parks, Government of Alberta, Grande Prairie, Alberta, Canada

Author for correspondence:

Rebecca S. Taylor

e-mail: rebecca.taylor@ec.gc.ca

Intra-specific diversification in the northern hemisphere is a recent phenomenon, particularly for vertebrates, due to climatic oscillations during the Quaternary. Caribou (*Rangifer tarandus*) are a recently diversified, keystone species in North America with a continental-wide distribution and a large amount of phenotypic variation. We assess patterns of intra-specific genetic variation and adaptive divergence, which are of critical interest to the fields of evolutionary biology and conservation. We characterized the relationship among caribou ecotypes and rapidly evolving genes among intra-specific lineages using phylogenomic analyses and a newly assembled chromosome-scale reference genome and 66 high-coverage genomes. Nine phylogenomic lineages showed different levels of introgression, with some lineages highly introgressed. Despite this, dN/dS ratios of all genes in the caribou genome (>30,000) identified 700-800 genes with significant signatures of positive selection in each lineage. These genes function in diverse pathways ranging from immune responses to circulatory systems, with the most enriched molecular pathways found in lineages with high levels of introgression. Glacial cycles and repeated introgression events between caribou lineages likely inflated both neutral and adaptive genetic variation: findings that are particularly timely given the importance of intra-specific diversity and adaptive variation in wild species for resilience to global change.

Key words:

adaptive diversification, intra-specific variation, phylogenetic history, positively selected genes, whole-genome sequences

1. Background

Dramatic climate oscillations during the Quaternary strongly affected the diversification of vertebrate species in the northern hemisphere leading to younger intra-specific lineages than in the south [1].

66 Unstable environments and contraction of populations into refugia during this time led to a loss of
67 genetic diversity in some species [1,2]. However, repeated cycles of secondary contact after periods of
68 isolation can create an excess of genetic diversity [3-5], and so the impacts of the glacial cycles on intra-
69 specific diversity is complex and likely impacted by the life-history of the species [5]. Further, although
70 many studies have investigated the impacts of these range shifts and refugia on neutral genetic
71 structure and diversity [e.g. 2, 6-8], little is known about the impact of climate cycles on the generation
72 of adaptive genetic variation [9]. Given how crucial intra-specific diversity is for ecosystem function [10-
73 12], a deeper understanding of the drivers of intra-specific genetic variation and its interplay with
74 adaptive divergence is essential to understanding how current species respond to environmental
75 variations [13-15]. Not only is this of critical interest to the fields of evolutionary biology and
76 conservation [16,17], it is also timely given the current recognition of the importance of intra-specific
77 diversity for species resilience [18-20].

78 We aim to address these questions in caribou (*Rangifer tarandus,* also known as reindeer) using
79 a whole genome phylogenomic framework. Caribou recently diversified in North America during the
80 glacial cycles of the Quaternary, and are thought to have undergone divergence between ecotypes
81 starting around 120,000 years ago [21-23]. Caribou are a highly mobile keystone species with a
82 continental-wide distribution, ranging from the high Arctic to the boreal forests, and spanning from the
83 east to the west coast of North America [24] (figure 1). Across its range caribou have a large amount of
84 phenotypic and genetic variation and have been divided into 12 conservation units, known as
85 Designatable Units (DUs), by the Committee on the Status of Endangered Wildlife in Canada [24] (figure
86 1). Caribou DUs face a number of threats including habitat destruction and climate change [25-27], with
87 eight DUs currently listed as endangered or threatened, two as special concern, and one that is extinct
88 [24,28-34].

Previous studies have demonstrated that caribou form two major mitochondrial genetic lineages due to separation in two major refugia during climate cycles - the Beringian-Eurasian (BEL) and the North American (NAL) lineages [22,35-39] - as well as further genetic sub-structuring due to micro-refugia [40-42]. Previous work also indicates that since their divergence 120,000 years ago to the present, caribou experienced potential repeated secondary contact as ice sheets contracted [23,37,42]. As such, caribou represents an interesting species to study the influence of the Quaternary on the diversification of both their neutral and adaptive genetic variation.

To this end, we assembled a chromosome-scale reference genome using Dovetail Omni-C technology and used 66 re-sequenced genomes to: 1) assess the levels of intra-specific diversity by reconstructing whole-genome phylogenomic history of caribou from across the continent for the first time, 2) measure introgression between reconstructed lineages of this highly mobile species, and 3) investigate signatures of rapidly evolving genes potentially involved with the diversification of caribou and their relationship to phylogenetic divergence and introgressive history, as a first step in understanding continent wide patterns of adaptive diversification. Given previous work on caribou, we hypothesized we would uncover divergent intra-specific lineages with some discordance from DU delineations, for example it is known that Boreal caribou are divided between the two major lineages (NAL and BEL) and that the phenotype likely evolved in parallel [43,44]. We also hypothesized that we would find a number of rapidly evolving genes, with more unique adaptive genes uncovered in lineages with lower levels of introgression due to the potentially homogenizing effect of gene flow between lineages.

2. Materials and Methods

(a) Caribou chromosome scale reference genome assembly and annotation

112 To ensure high quality, contiguous DNA for chromosome-scale reference genome assembly, fibroblast
113 cells were taken from caribou at Toronto Zoo and cultured in T-75 flasks. Firstly, we pre-warmed
114 DMEM1, DMEM3 and trypsin to 37-38 °C, then discarded media and rinsed each T-75 flask with 5 ml of
115 DMEM1. We discarded the media, and added 3 ml of trypsin to each flask and incubate at 38 °C for 2
116 min. We checked to see that cells had lifted, and then added 9 ml of DMEM3 to each flask, rinsed the
117 flask growing surface to retrieve as many cells as possible and transferred the entire volume to a 15 ml
118 tube, leaving 4 x 15 ml tubes, 2 for each animal. We then centrifuged at 200 x g for 5 min to pellet cells,
119 discarded supernatant and re-suspend pellet in 0.8 ml of PBS. We combined pellets for each individual
120 together in a 2 ml tube, centrifuged in microcentrifuge at 200 x g for 5 min, and discarded the
121 supernatant. The samples were then flash frozen using liquid nitrogen and transferred to a dry shipper
122 and shipped to Dovetail Genomics as we wanted to improve our previous assembly also sequenced by
123 Dovetail Genomics [45], by further scaffolding using Omni-C libraries [46,47]. Cells were cultured as
124 above and shipped to Genewiz (Azenta Life Sciences) for RNA sequencing for the annotation. Full details
125 of the sequencing, assembly, and annotation as performed by Dovetail Genomics are available in the
126 electronic supplementary material. We used the stats.sh script in the program BBMap version 38.42 [48]
127 to calculate the assembly statistics.
128

129 **(b) Re-sequenced whole genome sequences**

130 Whole genome sequences of 46 individual used in this study have been previously published and are
131 available on the National Centre for Biotechnology (NCBI) under BioProject Accession numbers
132 PRJNA634908, PRJNA694662, and PRJNA754521 (electronic supplementary material, table S1;
133 23,44,50,51]. For this study, we sequenced 20 new genomes (electronic supplementary material, table
134 S1; figure 1) using the same protocols as before (23,44). Briefly, samples were extracted using a Qiagen
135 DNAeasy tissue extraction kit following the manufacturer's instructions (Qiagen). Samples were run on a

136 Qubit fluorometer (Thermo Fisher Scientific) using the High Sensitivity Assay Kit and normalized to
137 20 ng/µl at a final volume of 50 µl. The DNA was shipped to The Centre for Applied Genomics (TCAG) at
138 the Hospital for Sick Children (Toronto, Ontario) for library preparation and sequencing. The samples
139 were sequenced two per lane on an Illumina HiSeq X. The raw reads for the new genomes will be made
140 available on the NCBI upon acceptance.

141 All code used to filter and map re-sequenced genomes, as well as for downstream analyses, can
142 be found on GitHub (https://github.com/BeckySTaylor/Phylogenomic_Analyses). Raw reads for all 66
143 individuals were cleaned using Trimmomatic version 0.38 [52] using a sliding window of 4 base pairs to
144 trim once phred score dropped below 15. We aligned all trimmed reads to the new reference genome,
145 which we first indexed using Bowtie2 version 2.3.0 [53]. We converted the SAM files to BAM files and
146 sorted them using Samtools version 1.5 [54], and then added read group information using GATK4 [55].
147 Using GATK4, we removed duplicates and used 'HaplotypeCaller' to call variants and produce a variant
148 calling format (VCF) file. We used the 'CombineGVCFs' function, followed by 'GenotypeGVCFs' to
149 produce a VCF file containing all individuals. We did two rounds of filtering on the VCF file using
150 VCFtools version 0.1.16 [56]. We removed indels, and any site low-quality genotype calls (minGQ) and
151 low-quality sites (minQ), with scores below 20, as well as any site with a depth of less than five or more
152 than double the mean depth of all genomes, filtering to remove sites with a depth of more than 55. For
153 the second round of filtering, we made two VCF files; one made using a more 'stringent' filter to remove
154 all missing data, and a 'less stringent' filter to removed sites with more than 5% missing data, resulting
155 in 17,595,673 and 41,321,354 SNPs respectively.

156 We also downloaded the raw reads for five Fennoscandian Wild Tundra reindeer genomes to
157 use as outgroups for phylogenomic analyses (ID numbers NMBU 38-42 [from 57], European Nucleotide
158 Archive accession PRJEB37216). We mapped and filtered the reads as above, as well as producing a VCF
159 file containing these and the 66 caribou genomes. We filtered the VCF file in VCFtools as above, this

160 time removing sites with a depth over 48 (double the mean of this data set) and removing sites with
161 more than 5% missing data, we had 16,119,954 SNPs.
162

163 **(c) Whole genome phylogenomic reconstruction**

164 For the phylogenomic reconstruction, we used IQtree version 1.6.12 [58]. First, we made a consensus
165 fasta file for each individual from the VCF file which included the reindeer using the 'consensus'
166 command in BCFtools. We found that running IQtree on the full genome sequences required too much
167 computational power, so we split the genome into seven sections of close to 300 million base pairs,
168 made a phylogeny with each, and then made a consensus tree as follows. Firstly, we used the 'CSplit'
169 command to split each individual fasta file into one file per scaffold, retaining the files for scaffolds 1-36
170 (which contains ~99% of the reference genome, see results). We concatenated all caribou individuals
171 together for each scaffold, and ran each scaffold from 1-36 in Model Finder in IQtree. Using the Bayesian
172 Information Criterion, Model Finder gave the model TVM+F+I+G4 for all scaffolds apart from 33 where it
173 selected GTR+F+I+G4. For all scaffolds the scores for these two models were close, and for scaffold 33
174 the likelihood score for the two models was similar (111,917,190.538 and 111,917,210.388), and so
175 when concatenating the scaffolds, we used TVM+F+I+G4 for the full phylogenomic run.
176 To run the phylogenomic analysis, we then concatenated scaffolds together for each individual
177 into seven fasta files of roughly 300 million base pairs (scaffolds 1-3, 4-7, 8-11, 12-16, 17-21, 22-27, 28-
178 35), excluding scaffold 36 which is putatively part of the X chromosome based on the presence of known
179 X chromosome genes on that scaffold [59,60]. We then reformatted each so that the sequence was on
180 one line using 'awk' and 'grep' commands, and then concatenated all individuals together into one file,
181 including the reindeer, for each of the seven sections so we had one fasta file with all individuals for
182 each of the ~300 million base pair regions. We then ran IQtree using 100,000 bootstraps to obtain

183 branch supports (-bb command) [61] to produce the phylogenies. We then made a consensus phylogeny
184 from the seven using the IQtree '-con' command.
185

186 **(d) ADMIXTURE and principal component analysis**

187 We performed an assignment test using ADMIXTURE version 1.3.0 [62]. We used Plink version 1.9 [63]
188 to convert the VCF file with the 66 caribou and no missing data into a BED file. We then pruned the
189 dataset to remove sites with a correlation co-efficient of 0.1 or above in sliding windows of 50 SNPs,
190 leaving 3,916,295 putatively unlinked SNPs, and then ran ADMIXTURE from K1 to 10. We plotted the
191 outputs in R studio version 1.2.5041 [64,65]. We used the pruned data set to run a PCA also in plink, and
192 plotted in R studio using GGplot2 [66]. To produce PCAs of subsets of individuals, we used the 'view -S'
193 function in BCFtools version 1.13 [67], and then ran plink and plotted in R studio as above.
194

195 **(e) Introgression measurements**

196 We measured introgression using ABBA BABA tests to control for incomplete lineage sorting. We used
197 Dsuite version 0.5 [68] to run the 'Dtrios' function to calculate D and f4-ratio statistics, grouping our
198 individuals by the lineages uncovered in our phylogenomic analysis (see results), using the phylogeny as
199 input using the '-t' command. When groups share branches on a phylogeny, many elevated D and f4-
200 ratio statistics can occur, however these correlated statistics can be informative to uncover the relative
201 time of the gene flow events across the phylogeny and to discover whether the gene flow occurred on
202 internal branches by using the f-branch statistic [68]. We calculated the f-branch statistics using the
203 output from Dtrios, and then plotted alongside the phylogeny using the 'dtools.py' script included with
204 DSuite, setting the p-value to 0.05.

205 As these statistics are unable to measure gene flow between sister groups, we used SplitsTree
206 [69] to visualize the phylogenetic network as an 'admixture graph', using the seven files SplitsTree
207 output by IQtree (one for each of the 300 million base pair phylogenomic analyses).
208

209 **(f) Rapidly evolving genes and gene ontology analysis**

210 We used GWideCodeML [70], a python package to run the codeml function of PAML [71] in a
211 computationally efficient way using genome-wide data. We used our annotation file to extract all genes
212 from the genomes of our individuals to use in GWideCodeML. To do this, we made a consensus fasta file
213 for each individual in the VCF file with our 66 caribou filtering to remove sites with more than 5%
214 missing data, as described above. We then used the '-x' function in Gffread version 0.12.3 [72] which
215 pulls out the coding sequence for each gene as indicated in the annotation file and splices them
216 together (to remove introns), to create one fasta file per individual with all genes. We reformatted the
217 files so each gene sequence is on one line using 'awk' commands, and then renamed the header line of
218 each gene to include the ID of each individual (in addition to the gene ID from the annotation) using
219 'sed' commands. We used the 'CSplit' command to split the files into one file per gene, and then
220 concatenated the files for each gene – resulting in one file per gene containing the sequence for all 66
221 individuals.

222 We first ran the genes using an unrooted version of the tree as required by codeml. We
223 removed the outgroup reindeer and then used the 'ape' package in R studio to transform the tree into
224 an unrooted version with a trifurcation at the root, the format needed by codeml. We used the branch
225 model, which uses a Likelihood Ratio Test (LRT) to test whether the genes have a significantly different
226 dN/dS ratio on the focal branch, compared to all other branches of the tree. We tested this for each of
227 the nine major lineages uncovered in our phylogenomic analysis (see results), excluding one individual,
228 the boreal caribou from Alberta, which is a hybrid between the two major clades as indicated by the PCA

229 analysis (see results). We then performed a Bonferroni multiple testing correction on the Likelihood
230 ratio results, adjusting the significance threshold to account for running the model over nine lineages.
231 We then took those sites where the focal branch was putatively under positive selection (larger dN/dS
232 ratio), and as some genes can be significant in multiple branches, we also calculated how many genes
233 were unique to each branch. We then used a chi-squared test in R studio to determine if there were a
234 significantly different number of positively selected genes across the different lineages.
235 To assign putative functions to the significant genes (after Bonferroni correction), we used
236 ShinyGo v0.76.2 [73] and assigned functions based on the GO Biological Process and GO Molecular
237 Function databases. The enrichment analysis outputs any biological pathways over-represented in the
238 list of genes with signatures of positive selection, and was performed for the significant rapidly evolving
239 genes for each of the nine major lineages separately.
240

241 ## 3. Results and Discussion

242 **(a) Chromosome scale assembly and re-sequenced genomes**

243 Our new caribou reference genome has a scaffold N50 of 64.42 MB, and an L50 of 14, with 99.5% of the
244 assembly being on 36 scaffolds. As the chromosome number is 70 for the species (34 autosome pairs
245 plus the sex chromosomes) [74], this represents a chromosome-scale assembly. We then used RNA-seq
246 data to perform a high-quality annotation of the genome, which identified the locations of 34,407
247 protein-coding genes.
248

249 **(b) Genetic structure, introgression, and high intra-specific diversity**

250 We reconstructed seven phylogenies, each using ~300 million base pairs of sequence from the 66 re-
251 sequenced genomes, and then made a final consensus phylogeny from the seven (figure 2). The final
252 phylogeny separated caribou into nine major well-supported lineages, which were not concordant with

253 DU designations but reconstructing the expected major lineages; the NAL and the larger and more
254 diverse BEL [39]. The 'NAL' branches off as the most basal group, and contains Eastern Migratory
255 caribou and Boreal caribou from Ontario and Manitoba. Boreal caribou from Alberta also belong to this
256 group but as an outside branch (figure 2). The other lineages are from the BEL, the first of which
257 contains the Boreal caribou from the Northwest Territories, which likely reflects the parallel evolution of
258 Boreal caribou as we have seen from previous results [43,44]. Surprisingly, our results indicate that the
259 Pink Mountain and Graham herds of the Northern Mountain caribou are a sister lineage (PMG) to the
260 Boreal caribou from the Northwest Territories (NWB lineage). We also found that caribou from the
261 Central Mountain and Southern Mountain groups form a lineage with the Northern Mountain caribou
262 from Itcha-Ilgachuz of British Columbia (CSM; figure 2). This correlates more closely with the current
263 population structuring under Schedule 1 of the Species at Risk Act (SARA) which groups the Itcha-
264 Ilgachuz caribou with animals from Central and Southern Mountain within the nationally significant
265 population: Woodland Caribou (Southern Mountain population) [75]. The Arctic caribou form a lineage
266 (ARC) with the mainland Barren-ground caribou from Manitoba and the Northwest Territories forming a
267 sister lineage (BRG). The Clearwater Fiord (Baffin Island) Barren-ground caribou sit most closely to
268 Greenland caribou and not alongside the mainland Barren-ground individuals (figure 2), which contrasts
269 some previous genetic results [76]. The last three lineages consist of most of the Northern Mountain
270 individuals, and Barren-ground caribou from the Yukon and Alaska (GRA, NM1 and NM2; figure 2). Our
271 finding that the Northern Mountain caribou form multiple distinct lineages (along with some Barren-
272 ground caribou individuals) aligns with previous genetic results that also showed sub-structuring of
273 Northern Mountain caribou [23,77] and that their distribution covers a large area with diverse habitats
274 [24].

275 For ADMIXTURE, the K value with the lowest cross-validation error was 1 (0.286), followed
276 closely by K=2 (0.289). Both the ADMIXTURE and PCA results are largely concordant with the

277 phylogenomic reconstruction (electronic supplementary material, figureS1 and figure S2). To measure
278 introgression between the reconstructed lineages, we used D and f4-ratio tests, which control for
279 incomplete lineage sorting. These statistics gave many significant signatures of introgression between
280 groups (electronic supplementary material). We then calculated the f-branch statistic, which accounts
281 for many correlated signatures of introgression using the f4-ratio statistics to show when along the
282 phylogeny introgression occurred, and whether the gene flow event was into an ancestral group. The f-
283 branch results indicated that the strongest signature of introgression was from the NAL lineage into the
284 ancestor of the NWB and PMG lineages (figure 3). We also found some introgression from the NAL into
285 the PMG group since they split from the NWB group. We also found relatively strong gene flow from the
286 NAL, NWB, and PMG lineages into the CSM lineage. The same three lineages (NAL, NWB, and PMG) also
287 show some gene flow going into the BRG lineage, with the NAL lineage having some gene flow into the
288 ancestor of the BRG and ARC groups. Most lineages showed gene flow into the ancestor of two main
289 Northern Mountain groups NM1 and NM2, with the NAL, NWB, PMG, and CSM lineages also having
290 gene flow into the NM2 lineage alone. Both the ARC and BRG lineages show gene flow into the NWB
291 lineage. Additionally, the GRA lineage, the CSM lineage, and the NM2 lineage show gene flow into the
292 PMG caribou lineage (figure 3). In contrast, some lineages show only a small amount or even a lack of
293 gene flow (with the caveat that sister groups cannot be tested for gene flow using ABBA BABA tests). For
294 example, no lineage shows gene flow into the ARC, GRA, or NM1 caribou lineages, with the ARC showing
295 some gene flow into NWB and the ancestral group of NM1 and NM2 and GRA showing some gene flow
296 into PMG (figure 3).

297 To try to explore potential introgression between sister lineages, we visualized 'admixture
298 graphs' in SplitsTree. Unsurprisingly these graphs show some phylogenetic uncertainty, putatively due to
299 gene flow, within lineages (electronic supplementary material, figure S3). We also see potential signals
300 of introgression between the NWB, PMG, and CSM lineages, which may help to explain the placement of

301 the PMG individuals next to the NWB lineage in the phylogenetic reconstruction (electronic
302 supplementary material, figure S3). We also see some potential introgression between the three major
303 northern mountain lineages (GRA, NM1, and NM2).

304 Altogether our results, reconstructing continent-wide whole genome phylogenomic history for
305 the first time, point towards a high level of intra-specific diversity and genetic structuring in caribou,
306 with some strong signals of introgression between many of the lineages. Our results build on previous
307 studies showing high genetic diversity in caribou, for example large numbers of mitochondrial
308 haplotypes [39,43,23] and high diversity in microsatellite loci [78-80]. The reasons behind the high
309 diversity and number of intra-specific lineages are likely multi-faceted. The large Beringian refugium
310 harboured high levels of genetic diversity for some, particularly cold adapted, species such as caribou
311 [81,82], and is reflected in the large diversity of caribou in the BEL vs the NAL lineage (figure 2). Post-
312 glacial expansion out of refugia can lead to genetic bottlenecks and low diversity further away from the
313 refugial populations [2]. However, repeated secondary contact and admixture between glacial lineages
314 can increase genetic diversity [3,5], in a similar mechanism to the 'glacial pulse model' which describes
315 how lineage fusion during glacial cycles can be a source of intra-specific lineage diversity [83]. Indeed,
316 well-known determinants of diversification during adaptive radiations include high standing variation,
317 gene flow, and habitat to diversify into [84], all of which are true for caribou post-glacial recolonizations.

318 The life history of caribou is also particularly conducive to the formation of high intra-specific
319 variation. Deer species have a high intrinsic rate of increase and dispersal capabilities [40]. High vagility
320 means caribou were able to undergo range shifts and quickly recolonize habitat once ice sheets recede,
321 or use ice-free corridors [40,82,23], facilitating repeated refugial lineage contact potentially increasing
322 genetic diversity as described above.

323

324 **(c) Rapidly evolving genes in caribou lineages**

325 The same processes leading to high genetic variation will likely have also increased standing adaptive
326 variation in caribou. Isolation in refugia may have promoted local adaptation, but the role of Quaternary
327 glacial cycles on adaptive diversification and its effects on biodiversity are not well known [9]. We used
328 the branch model approach in the codeml module of PAML [71] to calculate the ratio of synonymous to
329 non-synonymous mutations (dN/dS ratio) within genes. The program then performs likelihood ratio
330 tests to elucidate whether the 'focal' branch has a significantly different ratio from the rest of the
331 phylogenetic tree, and is thus putatively a rapidly evolving gene within that branch when compared to
332 the phylogeny as a whole. We ran each of the nine major linages from our whole genome phylogeny as
333 the focal branch (figure 2) to find those rapidly evolving genes significant to each, thus potentially
334 involved in the adaptive diversification of the lineage.

335 We found a number of rapidly evolving genes within each lineage (table 1, electronic
336 supplementary material). However, contrary to our hypothesis where we expected to find more lineage
337 specific genes in those with lower levels of introgression (as for example in the ARC lineage), we found
338 similar numbers of statistically significant genes after Bonferroni correction for each both overall (X^2 =
339 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 1).
340 This could be due to the relatively short time scale of the diversification (all within ~120,000 years)
341 limiting the number of genes within each lineage with a high dN/dS ratio. It could also be due to the high
342 level of overall connectivity (introgression) detected between the lineages limiting the significant genes
343 resulting from the codeml approach which scans for genes with a strong signature of positive selection
344 when compared to the rest of the phylogeny (i.e. not detecting genes that are important for all or many
345 caribou lineages). There was overlap in the genes that pulled out as significant for each lineage (table 1),
346 which may be indicative of the diversification occurring from a large pool of standing genetic variation,
347 which is known to be a driver of diversification during adaptive radiations [84].

348 We used the codeml approach due to our questions relating to the diversification among
349 caribou lineages, as well as our phylogenomic framework, and have thus characterized some genes
350 putatively involved in the differential adaptation of the caribou lineages (electronic supplementary
351 material for full lists of results). We performed gene ontology (GO) analyses to assign functional
352 categories to the genes under both 'biological process' and 'molecular function', and found a number of
353 processes represented in the significant genes such as immune processes, stress responses,
354 carbohydrate binding, amongst many others (electronic supplementary material for all GO terms). We
355 then performed enrichment analyses for each lineage, also under both 'biological process' and
356 'molecular function', in order find specific pathways containing multiple genes with signatures of rapid
357 evolution. We found that caribou lineages had different numbers of enriched pathways, with some
358 lineages showing a large number under molecular function compared to the others (over twice as many;
359 table 1). In particular, the CSM, NWB, and PMG lineages all show high numbers of enriched pathways.
360 These lineages also show some of the highest signatures of introgression, with the CSM lineage showing
361 high introgression from multiple lineages, as well as into the PMG lineage (figure 3), and with the
362 admixture graphs showing potential gene flow between the CSM, NWB, and PMG lineages (electronic
363 supplementary material, figure S3). Similarly, the NM2 lineage also shows a high number of enriched
364 pathways (table 1), and has signatures of introgression from many other lineages (figure 3) as well as
365 with the NM1 and GRA lineages (electronic supplementary material, figure S3). Gene flow can facilitate
366 adaptive diversification, as well as inflating standing genetic variation as a whole [84-86], and may have
367 been a driver of adaptive diversification in these the caribou lineages.

368 Our gene ontology results also point towards some other interesting patterns. For example, only
369 in the NAL lineage are there more enriched pathways under biological processes, and a number of these
370 pathways (6/15) are all involved in calcineurin signalling (electronic supplementary material). Calcineurin
371 is required for T cell activation and interleukin expression during immune response in mammals [87],

372 which might point towards a difference in the pathogens encountered by the NAL lineage, perhaps in
373 the more southern refugia. Calcineurin signalling is also important in cardiovascular and skeletal muscle
374 development in vertebrates [88] and so alternatively could be involved in morphological differences.

375 Another intriguing pattern from the gene ontology enrichment analysis is that the enriched
376 pathways under biological processes for both the CSM and NM2 lineages are mostly involved in blood
377 and circulatory systems (5/9 and 3/3 of their enriched pathways, respectively), including five of the
378 same genes within these pathways for both lineages (electronic supplementary material). The
379 importance of circulatory pathways could be due to the elevational migrations undertaken by the
380 Northern, Central, and Southern Mountain caribou within these two lineages [24]. That both lineages
381 show signatures of rapidly evolving genes in these pathways, with some of the same genes, could be
382 due to selection on shared standing variation, or potentially due to introgression between these
383 lineages given the relatively strong signature of introgression we found from the CSM into the NM2
384 lineage (figure 3).

385 Under molecular function, the NM2 lineage shows the largest number of enriched pathways
386 compared to all other lineages (table 1). Many of the pathways are involved in the activity of
387 carbohydrate, lipid, glucose or other sugars (electronic supplementary material) which may also speak
388 to the energetic requirement of elevational migrations, or potentially some difference in diet. Fully
389 understanding the many genes and pathways we have found to have signatures of selection will likely
390 take further study. Although such bottom-up approaches to find adaptive genetic variation, as we have
391 used here, can be advantageous in discovering adaptive genes relating to many and/or unknown
392 phenotypes, it can be difficult to fully elucidate the effect these genes have on fitness in non-model
393 organisms [89,90]. However, gleaning overall patterns such as we have found and discussed allows
394 insight into the process of intra-specific adaptive diversification, such as the potential importance of
395 introgression of glacial lineages in increasing both neutral and adaptive standing variation. This is in

396 contrast to our hypothesis that gene flow would act as a homogenising force reducing unique adaptive
397 variation uncovered for introgressed lineages, and shows how introgression can, in fact, be a driving
398 force behind diversification. Further, the characterization of putatively rapidly evolving genes amongst
399 intra-specific lineages has rarely been done before and thus provides a benchmark for investigations of
400 ecotypic adaptive diversification for other species. This is particularly timely given the current
401 recognition of the importance of understanding, monitoring, and preserving intra-specific diversity
402 [12,20].

403

404 **(d) Conservation of caribou intra-specific diversity**

405 With the increasing availability of genomic data, for caribou and other non-model species, our
406 understanding of adaptation and adaptive genetic variation is increasing [91,92]. Our improvement of
407 the caribou reference genome into a chromosome-scale assembly with a high-quality annotation will be
408 an invaluable resource to enable progress and more in-depth analyses of adaptation, as well as other
409 important analyses for conservation such as demography and inbreeding/runs of homozygosity [93].
410 Our whole genome phylogenomic reconstruction of caribou across the continent showed nine
411 major phylogenomic lineages not concordant with current DUs (figure 2) [24], potentially indicating the
412 need to re-assess the designations. Although we do not include representatives from all of the 11 extant
413 DUs in the analyses, we have good geographic sampling from eight DUs (figure 1). Further, two of the
414 three DUs we have not included, Torngats and Atlantic-Gaspesie, have extremely restricted ranges
415 (figure 1). Based on previous genetic work all three likely fall within the NAL lineage [22,24], although
416 future work to include these to investigate possible phylogenomic substructure of the NAL lineage
417 would be beneficial.

418 Preserving the high genetic diversity of caribou, indicated here by the divergent intra-specific
419 lineages (figure 2) and the evidence of differential adaptive variation between lineages (table 1), may be
420 important for the persistence of caribou and their ability to adapt to environmental changes [18,19]. In
421 particular, maintaining the connectivity between lineages may be important as introgression appears to
422 have been a driver of genetic variation, and would continue to allow movement of adaptive genes [94].
423

424 ## 4. Conclusion

425 We have assembled and annotated a chromosome-scale caribou reference genome as an invaluable
426 resource for conservation genomic research. Continent-wide phylogenomic reconstruction using 66 re-
427 sequenced high-coverage genomes recovered nine major lineages, not concordant with DUs but
428 representing a high level of intra-specific diversity. Quaternary glacial cycles, coupled with caribou
429 vagility and life history, played a strong role in increasing both neutral and adaptive genetic diversity.
430 Divergence due to isolation in refugia followed by repeated secondary contact likely drove
431 diversification with our finding of widespread signals of introgression, particularly in those lineages with
432 the highest numbers of enriched functional pathways with signatures of rapidly evolving genes.
433 Introgression thus appears to have been a crucial process in caribou genetic diversification, and
434 potentially many other species that underwent repeated range shifts during glacial cycles, and therefore
435 connectivity should be maintained between lineages for future resilience of the species.

436

437 **Data accessibility.** The raw reads from 46 of the re-sequenced genomes are available on the National
438 Centre for Biotechnology (NCBI) under BioProject Accession numbers PRJNA634908, PRJNA694662, and
439 PRJNA754521. New genomes for this manuscript will be uploaded upon acceptance. All code used to
440 filter and map re-sequenced genomes, as well as for downstream analyses, can be found on GitHub

441 (https://github.com/BeckySTaylor/Phylogenomic_Analyses).The new reference genome and annotation
442 will also be made available through the NCBI upon acceptance.

443 **Author contributions.** R.S.T. helped to conceive the study, did the bioinformatics, and wrote the
444 manuscript. M.M. helped to conceive the study, provided funding, and edited the manuscript. S.K.
445 helped with the bioinformatics and edited the manuscript, and G.M. did the laboratory work to produce
446 the fibroblast cells for the reference genome and annotation. K.S. provided funding and support with
447 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
448 and edited the manuscript. P.J.W. helped to conceive the study, provided funding, and edited the
449 manuscript.

450 **Conflict of interest declaration.** We declare no competing interests.

451 **Funding.** Funding was provided through an NSERC Collaborative Research & Development (CRD) grant,
452 NSERC grant RGPIN-2015–04477, The Government of Canada's Genomics Research and Development
453 Initiative (GRDI), Genome Canada, the Species-at-Risk Stewardship Program (SARSP), WCS Garfield
454 Weston Fellowship, Manitoba Hydro, Saskatchewan Power and Weyerhaeuser Inc.

455 **Acknowledgments.** We would like to thank the many field collectors and hunters who supplied samples
456 for this work, as well as Bridget Redquest and Austin Thompson for technical support in the laboratory.
457 We are also thankful to the Shared Hierarchical Academic Research Computing Network (SHARCNET:
458 www.sharcnet.ca), Compute Canada, and Amazon Cloud Computing for high-performance computing
459 services.

460

461 ## References

462

463 1. Tingley R, Dubey S. 2012. Disparity in the timing of vertebrate diversification events between
464 the northern and southern hemispheres. *BMC Evol Biol* **12**, 244. (doi:10.1186/1471-2148-12-
465 244)
466 2. Roberts DR, Hamann A. 2015. Glacial refugia and modern genetic diversity of 22 western North
467 American tree species. *Proc R Soc B* **282**, 20142903. (doi:10.1098/rspb.2014.2903)
468 3. Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S,
469 Grivet D, Lascoux M, et al. 2003. Glacial refugia: hotspots but not melting pots of genetic
470 diversity. *Science* **300**, 1563–1565. (doi:10.1126/science.1083264)
471 4. Provan J, Bennett K. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends Ecol*
472 *Evol* **23**, 564–571. (doi:10.1016/j.tree.2008.06.010)
473 5. Alcala N, Vuilleumier S. 2014. Turnover and accumulation of genetic diversity across large time-
474 scale cycles of isolation and connection of populations. *Proc R Soc B* **281**, 20141369.
475 (doi:10.1098/rspb.2014.1369)
476 6. Hansen CCR, Hvilsom C, Schmidt NM, Aastrup P, Van Coeverden de Groot PJ, Siegismund HR,
477 Heller R. 2018. The Muskox Lost a Substantial Part of Its Genetic Diversity on Its Long Road to
478 Greenland. *Curr Biol* **28**, 4022-4028.e5. (doi:10.1016/j.cub.2018.10.054)
479 7. Ony M, Klingeman WE, Zobel J, Trigiano RN, Ginzel M, Nowicki M, Boggess SL, Everhart S,
480 Hadziabdic D. 2021. Genetic diversity in North American *Cercis Canadensis* reveals an ancient
481 population bottleneck that originated after the last glacial maximum. *Sci Rep* **11**, 21803. (doi:
482 10.1038/s41598-021-01020-z)
483 8. Salinas P, Morinha F, Literak I, García J, Milá B, Blanco G. 2021. Genetic diversity, differentiation
484 and historical origin of the isolated population of rooks *Corvus frugilegus* in Iberia. *J Avian*
485 *Biol* **52**, jav.02689. (doi:10.1111/jav.02689)

9. Lafontaine G, Napier JD, Petit RJ, Hu FS. 2018. Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology* **99**, 1530–1546. (doi:10.1002/ecy.2382)
10. Mimura M, Yahara T, Faith DP, Vázquez-Domínguez E, Colautti RI, Araki H, Javadi F, Núñez-Farfán J, Mori AS, Zhou S, et al. 2017. Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evol Appl* **10**, 121–139. (doi:10.1111/eva.12436)
11. Des Roches S, Post DM, Turley NE, Bailey JK, Hendry AP, Kinnison MT, Schweitzer JA, Palkovacs EP. 2018. The ecological importance of intraspecific variation. *Nat Ecol Evol* **2**, 57–64. (doi:10.1038/s41559-017-0402-5)
12. Hoban S, Archer FI, Bertola LD, Bragg JG, Breed MF, Bruford MW, Coleman MA, Ekblom R, Funk WC, Grueber CE, Hand BK. 2022. Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biological Reviews* **97**, 1511-1538. (doi:10.1111/brv.12852)
13. Yiming L, Siqi W, Chaoyuan C, Jiaqi Z, Supen W, Xianglei H, Xuan L, Xuejiao Y, Xianping L. 2021. Latitudinal gradients in genetic diversity and natural selection at a highly adaptive gene in terrestrial mammals. *Ecography* **44**, 206–218. (doi:10.1111/ecog.05082)
14. Des Roches S, Pendleton LH, Shapiro B, Palkovacs EP. 2021. Conserving intraspecific variation for nature's contributions to people. *Nat Ecol Evol* **5**, 574-82. (doi:10.1038/s41559-021-01403-5)
15. Leigh DM, van Rees CB, Millette KL, Breed MF, Schmidt C, Bertola LD, Hand BK, Hunter ME, Jensen EL, Kershaw F, et al. 2021. Opportunities and challenges of macrogenetic studies. *Nat Rev Genet* **22**, 791–807. (doi:10.1038/s41576-021-00394-0)
16. Lexer C, Mangili S, Bossolini E, Forest F, Stölting KN, Pearman PB, Zimmermann NE, Salamin N. 2013. 'Next generation' biogeography: towards understanding the drivers of species diversification and persistence. *J. Biogeogr* **40**, 1013–1022. (doi:10.1111/jbi.12076)

17. Paz-Vinas I, Loot G, Stevens VM, Blanchet S. 2015. Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Mol Ecol* **24**, 4586-604. (doi:10.1111/mec.13345)
18. Carvalho SB, Velo-Antón G, Tarroso P, Portela AP, Barata M, Carranza S, Moritz C, Possingham HP. 2017. Spatial conservation prioritization of biodiversity spanning the evolutionary continuum. *Nat Ecol Evol* **1**, 0151. (doi:10.1038/s41559-017-0151)
19. Andrello M, D'Aloia C, Dalongeville A, Escalante MA, Guerrero J, Perrier C, Torres-Florez JP, Xuereb A, Manel S. 2022. Evolving spatial conservation prioritization with intraspecific genetic data. *Trends Ecol Evol* **37**, 553–564. (doi:10.1016/j.tree.2022.03.003)
20. O'Brien D, Laikre L, Hoban S, Bruford MW, Ekblom R, Fischer MC, Hall J, Hvilsom C, Hollingsworth PM, Kershaw F, Mittan CS. 2022. Bringing together approaches to reporting on within species genetic diversity. *J Applied Ecol* **59,** 2227-33. (doi:10.1111/1365-2664.14225)
21. Banfield AWF. 1961. *A Revision of the Reindeer and Caribou, Genus Rangifer*. National Museum of Canada, Bulletin No. 177, Queen's Printer: Ottawa, ON, Canada.
22. Yannic G, Pellissier L., Ortego J. Lecomte N, Couturier S, Cuyler C, Dussault C, Hundertmark KJ, Irvine RJ, Jenkins DA, et al. 2014. Genetic diversity in caribou linked to past and future climate change. *Nat Clim Change* **4**, 132–137. (doi:10.1038/nclimate2074)
23. Taylor RS, Manseau M, Klütsch CFC, Polfus JL, Steedman A, Hervieux D, Kelly A, Larter NC, Gamberg M, Schwantje H, Wilson PJ. 2021. Population dynamics of caribou shaped by glacial cycles before the last glacial maximum. *Mol Ecol* **30**, 6121–6143. (doi:10.1111/mec.16166)
24. COSEWIC. 2011. Designatable units for caribou (*Rangifer tarandus*) in Canada. Committee on the Status of Endangered Wildlife in Canada.
25. Vors LS, Boyce MS. 2009. Global declines of caribou and reindeer. *Global Change Biol* **15**, 2626–2633. (doi:10.1111/j.1365-2486.2009.01974.x)

26. Festa-Bianchet M, Ray JC, Boutin S, Côté SD, Gunn A. 2011. Conservation of caribou (*Rangifer tarandus*) in Canada: An uncertain future. *Can J Zool* **89**, 419–434. (doi:10.1139/z11-025)
27. Weckworth BV, Hebblewhite M, Mariani S, Musiani M. 2018. Lines on a map: Conservation units, meta-population dynamics, and recovery of woodland caribou in Canada. *Ecosphere* **9**, e02323. (doi:10.1002/ecs2.2323)
28. COSEWIC. 2014a. COSEWIC assessment and status report on the Caribou *Rangifer tarandus*, Newfoundland population, Atlantic-Gaspésie population and Boreal population, in Canada. Committee on the Status of Endangered Wildlife in Canada.
29. COSEWIC. 2014b. COSEWIC assessment and status report on the Caribou *Rangifer tarandus*, Northern Mountain population, Central Mountain population and Southern Mountain population in Canada. Committee on the Status of Endangered Wildlife in Canada.
30. COSEWIC. 2015a. COSEWIC assessment and status report on the Peary Caribou *Rangifer tarandus pearyi* in Canada. Committee on the Status of Endangered Wildlife in Canada.
31. COSEWIC. 2015b. COSEWIC assessment process, categories and guidelines. Committee on the Status of Endangered Wildlife in Canada.
32. COSEWIC. 2016. COSEWIC assessment and status report on the Caribou *Rangifer tarandus*, Barren-ground population, in Canada. Committee on the Status of Endangered Wildlife in Canada.
33. COSEWIC. 2017a. COSEWIC assessment and status report on the Caribou Dolphin and Union population (*Rangifer tarandus*), in Canada. Committee on the Status of Endangered Wildlife in Canada.
34. COSEWIC. 2017b. COSEWIC assessment and status report on the Caribou *Rangifer tarandus*, Eastern Migratory population and Torngat Mountains population, in Canada. Committee on the Status of Endangered Wildlife in Canada.

557 35. Flagstad Ø, Røed KH. 2003. Refugial origins of reindeer (*Rangifer tarandus* L.) inferred from
558 mitochondrial DNA sequences. *Evolution* **57**, 658–670. (doi:10.1111/j.0014-
559 3820.2003.tb01557.x.)
560 36. Cronin MA, MacNeil MD, Patton JC. 2005. Variation in mitochondrial DNA and microsatellite
561 DNA in caribou (*Rangifer tarandus*) in North America. *J Mammal* **86**, 495–505.
562 (doi:10.1644/1545-1542(2005)86[495:VIMDAM]2.0.CO;2)
563 37. McDevitt AD, Mariani S, Hebblewhite M, Decesare NJ, Morgantini L, Seip D, Weckworth BV,
564 Musiani M. 2009. Survival in the Rockies of an endangered hybrid swarm from diverged caribou
565 (*Rangifer tarandus*) lineages. *Mol Ecol* **18**, 665–679. (doi:10.1111/j.1365-294X.2008.04050.x)
566 38. Klütsch CFC, Manseau M, Wilson PJ. 2012. Phylogeographical analysis of mtDNA data indicates
567 postglacial expansion from multiple glacial refugia in woodland caribou (*Rangifer tarandus*
568 *caribou*). *PLoS ONE* **7**, e52661. (doi:10.1371/journal.pone.0052661)
569 39. Weckworth BV, McDevitt A, Musiani M, Hebblewhite M, Mariani S. 2012. Reconstruction of
570 caribou evolutionary history in Western North America and its implications for conservation.
571 *Mol Ecol* **21**, 3610–3624. (doi:10.1111/j.1365-294X.2012.05621.x)
572 40. Latch EK, Heffelfinger JR, Fike JA, Rhodes Jr OE. 2009. Species-wide phylogeography of North
573 American mule deer (*Odocoileus hemionus*): cryptic glacial refugia and postglacial
574 recolonization. *Mol Ecol* **18**, 1730-45. (doi:10.1111/j.1365-294X.2009.04153.x)
575 41. Klütsch CFC, Manseau M, Anderson M, Sinkins P, Wilson PJ. 2017. Evolutionary reconstruction
576 supports the presence of a Pleistocene Arctic refugium for a large mammal species. *J*
577 *Biogeogr* **44**, 2729–2739. (doi:doi.org/10.1111/jbi.13090)
578 42. Klütsch CFC, Manseau M, Trim V, Polfus J, Wilson PJ. 2016. The eastern migratory caribou : the
579 role of genetic introgression in ecotype evolution. Roy Soc Open Sci 3, 150469.
580 (doi:0.1098/rsos.150469)

581 43. Polfus JL, Manseau M, Klütsch CFC, Simmons D, Wilson PJ. 2017. Ancient diversification in glacial
582 refugia leads to intraspecific diversity in a Holarctic mammal. *J. Biogeogr.* **44**, 386–396.
583 (doi:10.1111/jbi.12918)
584 44. Taylor RS, Manseau M, Horn RL, Keobouasone S, Golding GB, Wilson PJ. 2020. The role of
585 introgression and ecotypic parallelism in delineating intraspecific conservation units. *Mol Ecol*
586 **29**, 2793–2809. (doi:10.1111/mec.15522)
587 45. Taylor RS, Horn RL, Zhang X, Golding GB, Manseau M, Wilson PJ. 2019. The caribou (*Rangifer*
588 *tarandus*) genome. *Genes* **10**, 540. (doi:10.3390/genes10070540)
589 46. Yamaguchi K, Kadota M, Nishimura O, Ohishi Y, Naito Y, Kuraku S. 2021. Technical
590 considerations in Hi-C scaffolding and evaluation of chromosome-scale genome assemblies. *Mol*
591 *Ecol* **30**, 5923-34. (doi:10.1111/mec.16146)
592 47. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD,
593 Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-
594 range linkage. *Genome Res* **26**, 342–350. (doi:10.1101/gr.193474.115)
595 48. Bushnell B, Rood J, Singer E. 2017. BBMerge—Accurate paired shotgun read merging via
596 overlap. *PLoS ONE* **12**, e0185056. (doi:10.1371/journal.pone.0185056)
597 49. Li Z, Lin Z, Ba H, Chen L, Yang Y, Wang K, Qiu Q, Wang W, Li G. 2017. Draft genome of the
598 reindeer (*Rangifer tarandus*). *GigaScience* **6**, 1–5. (doi:10.1093/gigascience/gix102)
599 50. Solmundson K, Bowman J, Wilson P, Taylor RS, Horn RL, Keobouasone S, Manseau M. Genomic
600 islands of heterozygosity maintained across caribou populations despite inbreeding. Available
601 from: https://www.biorxiv.org/content/10.1101/2020.12.29.424772v1
602 51. Taylor RS, Manseau M, Redquest B, S. Keobouasone, P. Gagné, C. Martineau, Wilson PJ. 2022.
603 Whole genome sequences from non-invasively collected caribou faecal samples. *Cons Genet*
604 *Resour* **14**, 53-68. (doi:10.1007/s12686-021-01235-2)

605 52. Bolger AM, Lohse M, Usadel B. 2014. trimmomatic: A flexible trimmer for Illumina sequence
606 data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
607 53. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**,
608 357–359. (doi:10.1038/nmeth.1923)
609 54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000
610 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and
611 samtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
612 55. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,
613 Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: A MapReduce framework for
614 analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303.
615 (doi:10.1101/gr.107524.110)
616 56. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Hansaker RE, Lunter G, Marth
617 GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.
618 (doi:10.1093/bioinformatics/btr330)
619 57. Weldenegodguad M, Pokharel K, Ming Y, Honkatukia M, Peippo J, Reilas T, Røed KH, Kantanen J.
620 2020. Genome sequence and comparative analysis of reindeer (*Rangifer tarandus*) in northern
621 Eurasia. *Sci Rep* **10**, 1-4. (doi:10.1038/s41598-020-65487-y)
622 58. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: A fast and effective stochastic
623 algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* **32**, 268-
624 274. (doi:10.1093/molbev/msu300)
625 59. Galloway SM, Hanrahan V, Dodds KG, Potts MD, Crawford AM, Hill DF. 1996. A linkage map of
626 the ovine X chromosome. *Genome Res* **6**, 667–677. (doi:10.1101/gr.6.8.667)
627 60. Liu R, Low WY, Tearle R, Koren S, Ghurye J, Rhie A, Phillippy AM, Rosen BD, Bickhart DM, Smith
628 TPL, et al. 2019. New insights into mammalian sex chromosome structure and evolution using

629 high-quality sequences from bovine X and Y chromosomes. *BMC Genomics* **20**, 1000.
630 (doi:10.1186/s12864-019-6364-z)
631 61. Hoang DT, Chernomor O, von Haeseler A., Minh BQ, Vinh LS. 2018 UFBoot2: Improving the
632 ultrafast bootstrap approximation. *Mol Biol Evol* **35**, 518–522. (doi:10.1093/molbev/msx281)
633 62. Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated
634 individuals. *Genome Res* **19**, 1655–1664, 2009. (doi:10.1101/gr.094052.109)
635 63. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker
636 PIW, Daly MJ, Sham PC. 2007. plink: A tool set for whole-genome association and population-
637 based linkage analyses. *Am J Human Genet* **81**, 559–575. (doi:10.1086/519795)
638 64. RStudio Team. 2015. RStudio: Integrated development for R. RStudio Inc. Retrieved from
639 http://www.rstudio.com/
640 65. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for
641 Statistical Computing. Retrieved from https:// www.R-project.org/
642 66. Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag: New York, USA.
643 ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.
644 67. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
645 McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**,
646 giab008. (doi:10.1093/gigascience/giab008)
647 68. Malinsky M, Matschiner M, Svardal H. 2021. dsuite – Fast D statistics and related admixture
648 evidence from VCF files. *Mol Ecol Resour* **21**, 584–595. (doi:10.1111/1755-0998.13265)
649 69. Hudson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol
650 Biol Evol* **23**, 254-67. (doi:10.1093/molbev/msj030)

651 70. Macías LG, Barrio E, Toft C. 2020. GWideCodeML: a python package for testing evolutionary
652 hypotheses at the genome-wide level. *G3: Genes, Genomes, Genetics* **10**, 4369-72. (doi:
653 10.1534/g3.120.401874)
654 71. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-91.
655 (doi:10.1093/molbev/msm088)
656 72. Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Research* **9**.
657 (doi:10.12688/f1000research.23297.2)
658 73. Ge SX, Jung D, Yao R. 2020. ShinyGO: a graphical gene-set enrichment tool for animals and
659 plants. *Bioinformatics* **36**, 2628-9. (doi:10.1093/bioinformatics/btz931)
660 74. Gripenberg U, Nieminen M. 1986. The chromosomes of reindeer (*Rangifer tarandus*). *Rangifer* **6**,
661 109 (doi:10.7557/2.6.1-App.625)
662 75. Environment Canada. 2014. Recovery Strategy for the Woodland Caribou, Southern Mountain
663 population (*Rangifer tarandus caribou*) in Canada. *Species at Risk Act* Recovery Strategy Series.
664 Environment Canada, Ottawa. viii + 103 pp.
665 76. Jenkins DA, Yannic G, Schaefer JA, Conolly J, Lecomte N. 2018. Population structure of caribou in
666 an ice-bound archipelago. *Divers Distrib* **24**, 1092-108. (doi:10.1111/ddi.12748)
667 77. Cavedon M, Poissant J, VonHoldt B, Michalak A, Hegel T, Heppenheimer E, Hervieux D, Neufeld
668 L, Polfus JL, Schwantje H, Steenweg R. 2022. Population structure of threatened caribou in
669 western Canada inferred from genome-wide SNP data. *Cons Genet* **16**, 1-5.
670 (doi:10.1007/s10592-022-01475-1)
671 78. Boulet M, Couturier S, Cote SD, Otto RD, Bernatchez L. 2007. Integrative use of spatial, genetic,
672 and demographic analyses for investigating genetic connectivity between migratory, montane,
673 and sedentary caribou herds. *Mol Ecol* **20**, 4223-40. (doi:10.1111/j.1365-294X.2007.03476.x)

674 79. McLoughlin PD, Paetkau D, Duda M, Boutin S. 2004. Genetic diversity and relatedness of boreal
675 caribou populations in western Canada. *Biol Cons* **118**, 593–598.
676 (doi:10.1016/j.biocon.2003.10.008)
677 80. Zittlau K, Coffin J, Farnell R, Kuzyk G, Strobeck C. 1998. Genetic relationships of three Yukon
678 caribou herds determined by DNA typing. *Rangifer* **12**, 59–62. (doi:10.7557/2.20.5.1625)
679 81. Galbreath KE, Cook JA, Eddingsaas AA, DeChaine EG. 2011. Diversity and demography in
680 Beringia: Multilocus tests of paleodistribution models reveal the complex history of Arctic
681 ground squirrels. *Evolution* **65**, 1879–1896. (doi:10.1111/j.1558-5646.2011.01287.x)
682 82. Dussex N, Alberti F, Heino MT, Olsen R-A, van der Valk T, Ryman N, Laikre L, Ahlgren H, Askeyev
683 IV, Askeyev OV, et al. 2020. Moose genomes reveal past glacial demography and the origin of
684 modern lineages. *BMC Genomics* **21**, 854. (doi:10.1186/s12864-020-07208-3)
685 83. Maier PA, Vandergast AG, Ostoja SM, Aguilar A, Bohonak AJ. 2019. Pleistocene glacial cycles
686 drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). *Evolution* **73**,
687 2476-96. (doi:10.1111/evo.13868)
688 84. Berner D, Salzburger W. 2015. The genomics of organismal diversification illuminated by
689 adaptive radiations. *Trends Genet* **31,** 491–499. (doi:10.1016/j.tig.2015.07.002)
690 85. Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita MK. 2014. Diversification
691 and asymmetrical gene flow across time and space: lineage sorting and hybridization in
692 polytypic barking frogs. *Mol Ecol* **23**, 3273–3291. (doi:10.1111/mec.12814)
693 86. Lexer C, Marthaler F, Humbert S, Barbará T, de la Harpe M, Bossolini E, Paris M, Martinelli G,
694 Versieux LM. 2016. Gene flow and diversification in a species complex of *Alcantarea* inselberg
695 bromeliads. *Bot J Linn Soc* **181**, 505–520. (doi:10.1111/boj.12372)

696 87. Park H-S, Lee SC, Cardenas ME, Heitman J. 2019. Calcium-calmodulin-calcineurin signaling: a
697 globally conserved virulence cascade in eukaryotic microbial pathogens. *Cell Host & Microbe* **26**,
698 453–462. (doi:10.1016/j.chom.2019.08.004)
699 88. Schulz RA, Yutzey KE. 2004. Calcineurin signaling and NFAT activation in cardiovascular and
700 skeletal muscle development. *Dev Biol* **266**, 1–16. (doi:10.1016/j.ydbio.2003.10.008)
701 89. Ross-Ibarra J, Morrell PL, Gaut BS. 2007. Plant domestication, a unique opportunity to identify
702 the genetic basis of adaptation. *Proc Natl Acad Sci USA* **104**, 8641–8648.
703 (doi:10.1073/pnas.0700643104)
704 90. Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level.
705 *Nat. Reviews Genet* **12**, 767-780. (doi:10.1038/nrg3015)
706 91. Fuentes-Pardo AP, Ruzzante DE. 2017. Whole-genome sequencing approaches for conservation
707 biology: Advantages, limitations and practical recommendations. *Mol Ecol* 26, 5369-406. (doi:
708 10.1111/mec.14264)
709 92. Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalén L, De
710 Meester L, Ekblom R, et al. 2015. Genomics and the challenging translation into conservation
711 practice. *Trends Ecol Evol* **30**, 78–87. (doi:10.1016/j.tree.2014.11.009)
712 93. Taylor RS, Jensen EL, Coltman DW, Foote AD, Lamichhaney S. 2021. Seeing the whole picture:
713 what molecular ecology is gaining from whole genomes. *Mol Ecol* 30, 5917-5922.
714 (doi:10.1111/mec.16282)
715 94. Hanson JO, Fuller RA, Rhodes JR. 2019. Conventional methods for enhancing connectivity in
716 conservation planning do not always maintain gene flow. *J Appl Ecol* **56**, 913–922.
717 (doi:10.1111/1365-2664.13315)
718
719
720
721

722
723
724
725
726
727
728
729
730
731

732 **Table 1.** The overall number of significant rapidly evolving genes as well as the number of unique genes

733 for each lineage.

Lineage	Number of genes with a significant dN/dS ratio	Number of unique genes with a significant dN/dS ratio	Number of enriched pathways – biological processes	Number of enriched pathways – molecular function
NAL	764	245	15	5
NWB	758	250	1	39
PMG	779	287	10	38
CSM	787	270	9	30
ARC	791	282	0	14
BRG	777	276	0	9
GRA	749	271	1	4
NM1	777	265	3	3
NM2	742	278	3	51

734

735

736
737
738

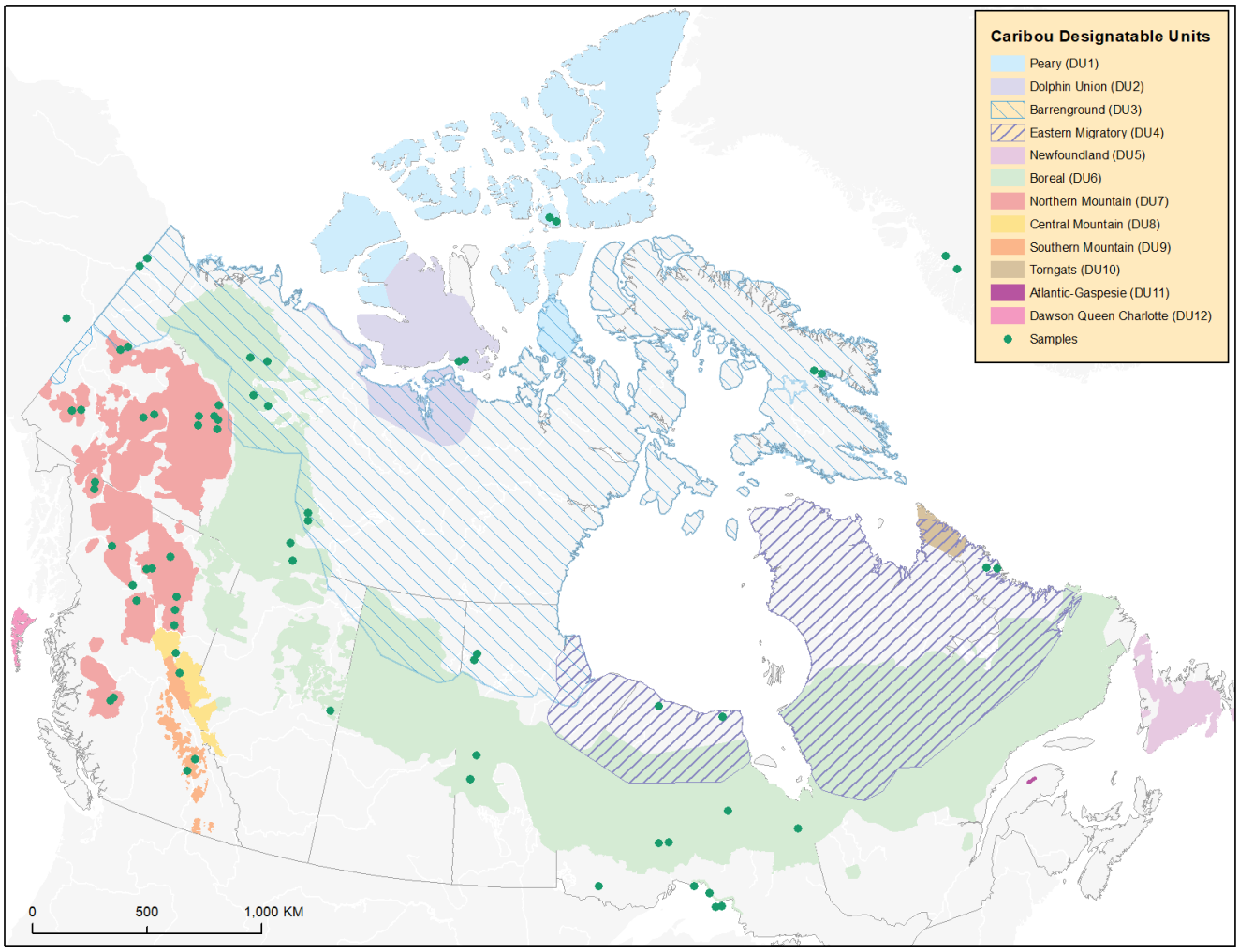

Figure 1. Map of sampling locations used for the re-sequenced whole genome sequences. Background colours show the Designatable Unit (DU) ranges.

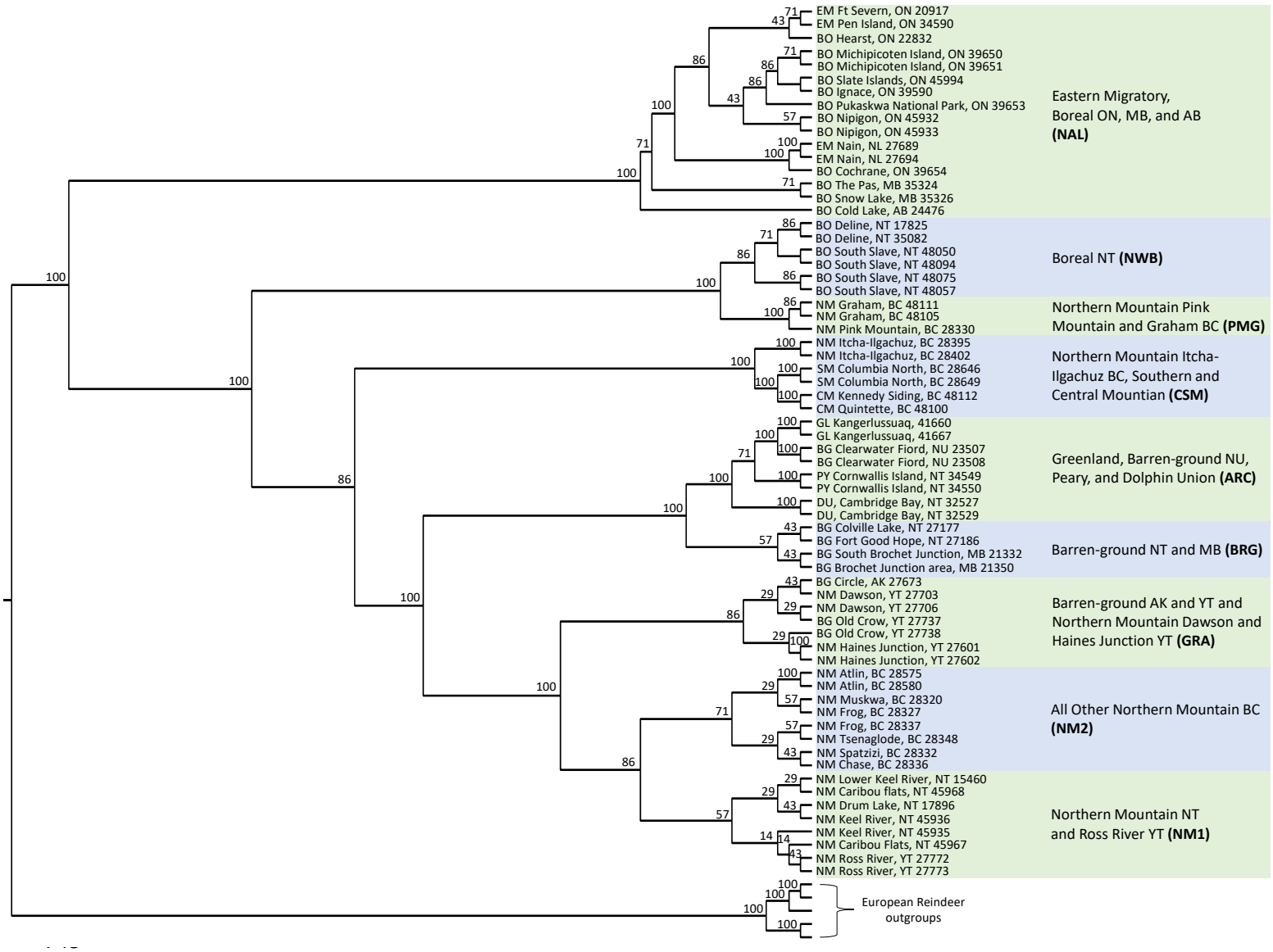

Figure 2. Maximum likelihood phylogenomic reconstruction using the whole genome sequences of the 66 caribou and five reindeer genomes as outgroups. Bootstrap supports are shown on the nodes. Each individual is identified on the tree with their PCID number (electronic supplementary material, table S1) 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.

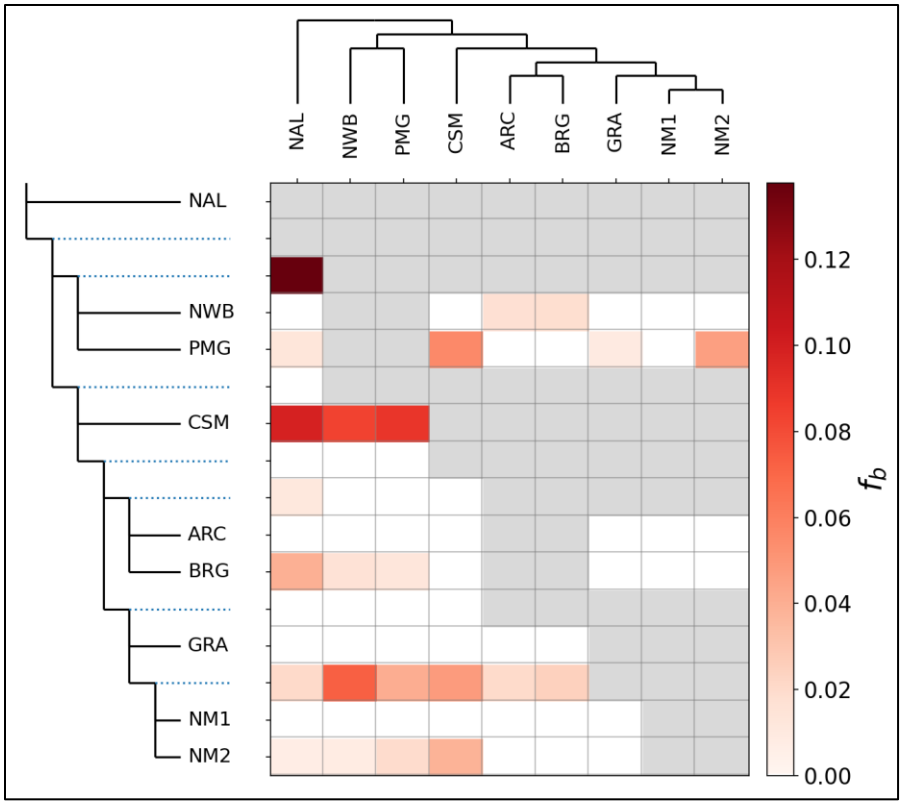

Figure 3. Heatmap showing the f-Branch statistics alongside the phylogeny; NAL – Eastern Migratory and Boreal from Ontario and Manitoba, NWB – Boreal from the Northwest Territories, PMG – Northern Mountain Pink Mountain and Graham, CSM – Central and Southern Mountain and Northern Mountain Itcha-Ilgachuz, ARC – Arctic caribou, BRG – mainland Barren-ground from the Northwest Territories and Manitoba, GRA – Barren-ground caribou from the Yukon and Alaska and Northern Mountain caribou from Dawson and Haines Junction in the Yukon, NM1 – Northern Mountain from the Northwest Territories and Ross River from the Yukon, NM2 – all other Northern Mountain from British Columbia. The gene flow is from the group 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. Greyed out squares indicate tests that could not be made with this statistic, and white squares indicate where no gene flow was detected.