G3: Genes|Genomes|Genetics Early Online, published on December 5, 2019 as doi:10.1534/g3.119.400747

Genome assembly and analysis of the North American mountain goat (*Oreamnos americanus*) reveals species-level responses to extreme environments

Daria Martchenko*, Rayan Chikhi†, Aaron B.A. Shafer*‡

*Environmental and Life Sciences Graduate Program, Trent University, K9J 7B8 Peterborough, Canada

†Institut Pasteur and CNRS, C3BI USR 3756, Paris, France

‡Forensics Program, Trent University, K9J 7B8 Peterborough, Canada

Genome assembly of *Oreamnos americanus*

Keywords: *de novo* genome assembly, HiRise scaffolding, demography, PSMC, mountain goat, Caprinae

Corresponding author: Daria Martchenko, Trent University, DNA Building, 2140 East Bank Drive, Peterborough, Ontario, Canada K9J 7B8, dariamartchenko@trentu.ca

ABSTRACT

The North American mountain goat (*Oreamnos americanus*) is an iconic alpine species that faces stressors from climate change, industrial development, and recreational activities. This species' phylogenetic position within the Caprinae lineage has not been resolved and their phylogeographic history is dynamic and controversial. Genomic data will be used to address these questions and provide valuable insights to conservation and management initiatives. We sequenced short-read genomic libraries constructed from a DNA sample of a 2.5-year-old female mountain goat at 80X coverage. We improved the short-read assembly by generating Chicago library data and scaffolding using the HiRise approach. The final assembly was 2,506 Mbp in length with an N50 of 66.6 Mbp, which is within the length range and in the upper quartile for N50 published ungulate genome assemblies. Comparative analysis identified 84 gene families unique to the mountain goat. The species demographic history in terms of effective population size generally mirrored climatic trends over the past one hundred thousand years and showed a sharp decline during the last glacial maximum. This genome assembly will provide a reference basis for future population and comparative genomic analyses.

INTRODUCTION

Mountain goats (*Oreamnos americanus*) are a symbol of alpine wilderness and belong to the Caprinae subfamily of ungulates (hoofed mammals) that are known for their exceptional climbing ability (Fig. 1). As northern alpine specialists, mountain goats are vulnerable to climate change (Pettorelli *et al.* 2007) and face pressures from industrial development, recreational activities, and hunting (Côté and Festa-Bianchet 2003). There

70 is a need to provide novel tools and information to support conservation and management
71 initiatives as it pertains to this enigmatic species.

72 Mountain goats are not true goats (i.e. genus *Capra*), but diverged early within the
73 Caprinae lineage and are typically associated with the chamois (*Rupicapra rupicapra*)
74 and goral (*Naemorhedus* spp.); their phylogenetic position within the Caprinae is not
75 resolved, but sampling of nuclear and mitochondrial genes suggested an early and
76 independent divergence within the lineage (Shafer and Hall 2010). Similarly, the mountain
77 goats phylogeographic history appears dynamic and is contentious (Shafer *et al.* 2011a,
78 2011b). Throughout the last glacial period most of the present-day range was covered
79 with ice (Shafer *et al.* 2010), indicative of multiple refugia and contraction and expansion
80 events of subpopulations (Shafer *et al.* 2011a). Within their present-day range, mountain
81 goat populations are not continuous and dispersal is limited (Côté and Festa-Bianchet
82 2003), that combined with regional variability in habitat and climate, support the
83 hypothesis of locally adapted populations across the range.

84 With the rapid development and lowering costs of sequencing technologies, the
85 use of genomics in conservation and wildlife management is becoming more widespread
86 [8]. Over 20 ungulate genomes have been assembled (Martchenko *et al.* 2018), with
87 chromosome-level assemblies for several commercially important ungulate species
88 (ARS-UCD1.2, Oar_v3.1, (Bickhart *et al.* 2017), and an increasing number of draft
89 assemblies for wild ungulates (Li *et al.* 2017; Zhang *et al.* 2018; Bana *et al.* 2018; Wang
90 *et al.* 2019). This draft genome of the North American mountain goat adds to this list of
91 ungulate genomes, and will be used for demographic analyses (e.g. Miller *et al.* 2015),

comparative genomic studies (e.g. Orlando *et al.* 2013; Agaba *et al.* 2016), and testing for putative local adaptation (e.g. Savolainen *et al.* 2013; Fitzpatrick and Keller 2015).

MATERIALS AND METHODS

Library construction

A tissue sample was obtained from a spleen of a 2.5-year-old female harvested on Revillagigedo Island, Alaska, USA and frozen promptly after harvest. High quality molecular DNA was extracted using the phenol chlorophorm method (Sambrook and Russell 2006). Two TruSeq Nano genomic library preparations with an insert size of 500 bp were constructed and sequenced on an Illumina HiSeqX platform with 150 bp paired-end reads. A total of 214 Gbp of raw short-read data were generated.

We trimmed the adapters and the low quality bases from the reads with BBDuk (part of BBTools) (Bushnell 2018) to a minimum base quality 20 and a minimum read length 50 bp after a quality assessment with FastQC (Andrews 2010). To avoid any potential contamination of the genome with bacterial or viral sequences, we screened the trimmed reads with Kraken (Wood and Salzberg 2014) using the full standard database (Oct. 2017 release). A total of 0.95% of the reads were classified as belonging to a known bacterial or viral taxon and were removed. The final short-read dataset comprised 176 Gbp.

Three Chicago libraries were prepared (Dovetail Genomics) following the approach in Putnam et al. (Putnam *et al.* 2016). Briefly, Chicago library preparation involves in vitro chromatin reconstitution to generate libraries with separations between reads up to maximum fragment size of the input DNA. These libraries were sequenced

on one lane of the Illumina HiSeqX instrument generating a total of 104 Gbp of sequence data.

Genome assembly

A k-mer size of 101 was selected for the short-read dataset with KmerGenie (Chikhi and Medvedev 2014). The draft *de novo* assembly was completed with the Meraculous assembler v. 2.0.4 (Chapman *et al.* 2011) using the following options: 'diploid_mode 1', 'min_depth_cutoff 0' to allow for auto-detection of k-mer frequency cutoff by the assembler, 'no_read_validation 0' to decrease the runtime as the reads were trimmed and screened prior to the assembly, and 'gap_close_aggressive 1' to remove the uniqueness requirement and use the most common sequence obtained from potential gap-closing reads. For comparison purposes, the short-read data was also assembled with MEGAHIT v. 1.1.1 (Li *et al.* 2015) using 'k-list 101' option.

We improved the short-read assembly by scaffolding with the Chicago library data using the HiRise pipeline software (Putnam *et al.* 2016). The HiRise software uses a likelihood model of the Chicago data to infer the genomic distance between the read pairs, which is then used to scaffold and check the orientation of draft scaffolds.

Quality assessment

To assess the assembly quality, we mapped the raw short-insert library reads to the assembled genome with bowtie2 (Langmead and Salzberg 2012). We also aligned the genome against itself with minimap2 (Li 2018) with options '-ax asm5 -X' to check for putatively artefactual duplications, reflected by sequence overlaps between scaffolds

(excluding self-hits). We ordered the SAM file by decreasing alignment scores in order to identify large regions that are similar between different scaffolds.

To validate the genome assembly and assess potential contamination and sequencing bias we analyzed the contig and scaffold assemblies with KAT (Mapleson *et al.* 2017). KAT plots the k-mer spectra – the number of distinct k-mers seen once, twice, three times etcetera and compares it between the raw reads and the genome assembly. We calculated an assembly quality value (QV) following the method of Bickhart *et al.* (2017) and ran FRC_align (Vezzi *et al.* 2012) to identify any problematic regions in the genome. We also evaluated the quality of the assembled genome using BUSCO V3 and the mammalia odb9 database.

Genome Annotation

We identified and classified the repeat regions of the assembled genome using RepeatMasker v. 4.0.8 (Smit *et al.* 2013). We configured RepeatMasker with RMBlast v. 2.6.0 sequence search engine, Tandem Repeat Finder v. 4.0.9 (Benson 1999), Dfam_Consensus database (20181026 release) and RepBase RepeatMasker edition (20181026 release) (Bao *et al.* 2015) and used '-species artiodactyl' parameter for the analysis.

We then annotated scaffolds that were greater than 2500 bp (n = 482) using both homology-based and *de novo* predictions. Proteins from *Homo sapiens, Equus caballus, Bos taurus,* and *Ovis aries* (all Ensembl 89 release (Hunt *et al.* 2018)) were aligned to genome using blastx v. 2.7.1. We also used NCBI mammalian RefSeq transcipt database v. 1.0 and BLAT v. 1.0 to help identify exon structure. For *de novo* prediction we first

mapped RNAseq data (SRX1947618) from *Capra hircus*, the most closely related species available on NCBI Sequence Read Archive, to the genome using HISAT2 v. 2.10 (Kim *et al.* 2015) and used these data for prediction with Augustus v. 3.1.1 (Stanke *et al.* 2006). EVidenceModeler v. 1.1.1 (Haas *et al.* 2008) was used to generate a consensus gene set with the following default weighting scheme (Humann *et al.* 2018): gene prediction via Augustus (1x weight); protein alignments via blastx (5x weight); transcript alignment via blat (10x weight). Lastly, we used PASA v. 2.3.3 (Haas *et al.* 2008) to refine our gene identifications.

To assign function to the newly annotated genes we aligned them to the NCBI mammalian RefSeq protein database using blastp v. 2.7.1 with a maximum HSP distance of 30,000 and e-value of $1e^{-8}$. The Interpro v 5.29-68 and KEGG databases were screened to annotate domains and identify pathways based on the top blast hit. We used Infernal v. 1.1.1 (Nawrocki and Eddy 2013) and the Rfam database release 14.1 to annotate miRNA, rRNA, and snRNA genes; tRNAs were detected using tRNAscan-SE version 2.0.1 (Lowe and Eddy 1997).

Species-specific genes, phylogeny and demographic history

Using the predicted protein sequences of *Oreamnos americanus,* we analyzed the orthologs shared between *Capra hircus, Equus caballus, Bos taurus, Sus scrofa*, *Ovis aries* and *Oreamnos americanus* with OrthoVenn (Wang *et al.* 2015). We then used PHYLUCE v. 1.5.0 (Faircloth 2016) to identify orthologous UCE sequences between *Sus scrofa, Equus caballus, Bos taurus, Capra hircus,* and *Ovis aries* and *Oreamnos americanus* using the 5k amniote probe set (Faircloth *et al.* 2012). We extracted UCEs

with 250 bp flanks totaling 11.9 aligned Mbp. We conducted a maximum likelihood phylogenetic analysis using RAxML v. 8 (Stamatakis 2014) under a GTRGAMMAI substitution model selected with jModelTest 2 (Darriba *et al.* 2012). The phylogenetic tree was constructed with PAML MCMCtree v. 4.9 (Yang 2007) and calibrated with the divergence time of goat and cow (18.3–28.5 Mya) (Benton and Donoghue 2007).

Lastly, using a random 30X subset of the sequencing data, we modeled the historical effective population size (N_e) for the North American mountain goat using PSMC [18]. We used the default parameters of 64 atomic time intervals (-p "4+25*2+4+6"), a generation time of 6 years (Li and Durbin 2011) and mutation rate $\mu = 1.33*10^{-8}$ mutations/site/generation, calculated as the average mammalian mutation rate of $2.22*10^{-9}$ mutations/site/year (Kumar and Subramanian 2002) multiplied by the generation time of 6 years.

Data availability

The raw sequence data have been deposited in the Short Read Archive (SRA) under accession number PRJNA510081. The genome assembly has been deposited in GenBank under accession number PRJNA510081.

RESULTS AND DISCUSSION

Genome assembly

The short-read assembly with Meraculous (Chapman *et al.* 2011) produced a genome consisting of 172,540 scaffolds with an N50 of 29.0 Kb and total genome size of 2.5 Gbp. For comparison purposes, the short-read data was also assembled with MEGAHIT v.

1.1.1 (Li *et al.* 2015), which produced a less continuous contig assembly of 2.6 Gbp compared to the Meraculous contig assembly (Tables S1 and S2).

The final assembly including Chicago libraries was 2.5 Gbp in length and consisted of 3,217 scaffolds that have an N50 of 66.6 Mbp (Table 1). The assembled genome is 93% of the predicted size estimated by KmerGenie (Chikhi and Medvedev 2014). Compared to the other ungulate species, the N50 of the mountain goat assembly is in the top quartile and the assembly length is mid-range (Table S3) (Miller *et al.* 2015; Bickhart *et al.* 2017; Williams *et al.* 2017; Bana *et al.* 2018). The genomic GC content is 41.67%, compared to 41.97% for non-primate mammalian animals (Li and Du 2014).

Quality assessment

Mapping the sequencing reads back to the assembled genome can be used to identify misassemblies and check assembly quality (Muller *et al.* 2013). The overall alignment rate of mapped the raw short-insert library reads to the final scaffolded assembly was 97%; 80% of the reads aligned concordantly exactly 1 time and 14% of the reads aligned concordantly more than once (Table S4), indicating the high quality of the assembly. For reference, when the same analysis was completed for the high quality assembly of the domestic goat, 89% of the raw sequencing reads mapped back to the assembled genome (Dong *et al.* 2013). To detect potential long duplicated regions in the final scaffolded assembly, we aligned the genome assembly against itself. Sixteen out of 30,472 entries in the SAM file had scores over 1000, ranging from 1011 to 1488. By examining the CIGAR strings of each of those entries, we concluded that the assembly did not contain duplicated segments across scaffolds longer than 2 Kbp. The QV score of the genome

was 41.8 and the FRC output was consistent with short-read domestic goat genome assemblies (Bickhart *et al.* 2017; Table S6).

For both contig and scaffold assemblies for the mountain goat (Fig. S1 and S2) the number of distinct k-mers (k=27) with over 1X coverage, i.e. k-mers that are not single-copy in the assembly, is low, which suggests low levels of sequencing bias and contamination. The 1X k-mer distribution has a much larger first peak, suggestive of an inbred individual, which is ideal for assembly purposes and why our sample was selected. This mountain goat is descended from a founding population of less than 17 individuals (Smith and Nichols 1984); further the mating system of mountain goats is such that only a small number of males sire most offspring (Mainguy *et al.* 2009), suggesting this individual should have low levels of heterozygosity.

A total of 4104 single-copy orthologues were screened against the assembled mountain goat genome with 3817 (93%) being complete, and 287 missing or fragmented BUSCOS (7%); this is in the same range as the BUSCO scores reported for other ungulate species (Li *et al.* 2017; Zhang *et al.* 2018). Similarly, we detected 471 of 481 ultraconserved elements (UCEs) (Bejerano *et al.* 2004) using scripts available in PHYLUCE v. 1.5.0 (Faircloth 2016) which is comparable to high quality mammalian genomes (Seemann *et al.* 2016).

Genome annotation

Long interspersed nuclear elements (LINEs) comprised 27.2% of the assembled genome, short interspersed nuclear elements (SINEs) comprised 11.4%, and in total repeats

represented 46.7% of the assembly (Table 2). The repeat content is consistent with other ungulate genomes (Dong *et al.* 2013; Li *et al.* 2017; Zhang *et al.* 2018).

Our annotation pipeline of the longest scaffolds resulted in 22,292 total genes with 16,012 being protein coding. Of these, the Interpro v 5.29-68 databases identified 13,874 (74%) genes with information, and 9,460 (51%) having a gene ontology assignment. Similarly, RefSeq protein database detected a a total of 14,470 genes (78%) with NCBI RefSeq match. Lastly, the Interpro v 5.29-68 databases annotated 13,874 (74%) genes with information, with 9,460 (51%) having a gene ontology assignment. A total of 34,819 putative noncoding RNA sequences were identified encompassing 0.1% of the genome and is comparable to other ungulates (Zhang *et al.* 2018).

Species-specific genes, phylogeny and demographic history

A total of 84 gene families were found to be specific to the mountain goat (Fig. 2a). Seventy-three orthologous clusters were uniquely shared between *Ovis aries* and *O. americanus,* 33 between *Capra hircus* and *O. americanus,* and 41 between *Bos taurus* and *O. americanus*. The gene families unique to mountain goat were enriched for ferroxidase activity, transcription regulation and protein folding and stability; we hypothesize that enrichment for the ferroxidase activity could potentially have allowed mountain goats to adapt to their alpine environment due to its link to erythropoiesis and altitude (Cherukuri *et al.* 2004).

The phylogenetic analysis (Fig. 2b) suggested that mountain goats are ancestral to sheep and goats as expected (Shafer and Hall 2010). There was no relationship between the assembly strategy, number of indels in UCEs, and phylogenetic placement

(Table S7). The mitochondrial genome phylogeny confirmed a similar relationship (Fig. S3). As the ice sheets progressed during the last glacial period, the Earth experienced a decrease in average surface temperature; the mountain goat N_e reflects these temperature patterns, with a general downward trend in N_e from 200 kya to 35 kya (Fig. 3). PSMC analysis suggested a near 10-fold decrease in the N_e of mountain goats at the end of the last glacial period and the start of the Holocene likely due to warming and loss of suitable habitat.

Conclusion

This is the first genome assembly and annotation for the North American mountain goat, and the first *de novo* assembly of a wild caprid. The biological sample came from an island with a small founding population (Smith and Nichols 1984) making it ideal for genome assembly. Using the newly assembled genome we identified species-specific genes and modeled the historical population demography of the species that showed a dramatic decline at the height of the last glaciation. Relative to other ungulate genomes, including economically valuable domesticated species that incorporated long-read sequencing strategies, this is among the highest quality wild ungulate genomes to date (Martchenko *et al.* 2018).

Figures

Figure 1. North American mountain goat female stomping to assert her dominance at Eulachon Creek, Alaska, USA. Mountain goats display a strong social structure and one of the highest levels of aggressiveness among ungulates. Photo by K.S. White.

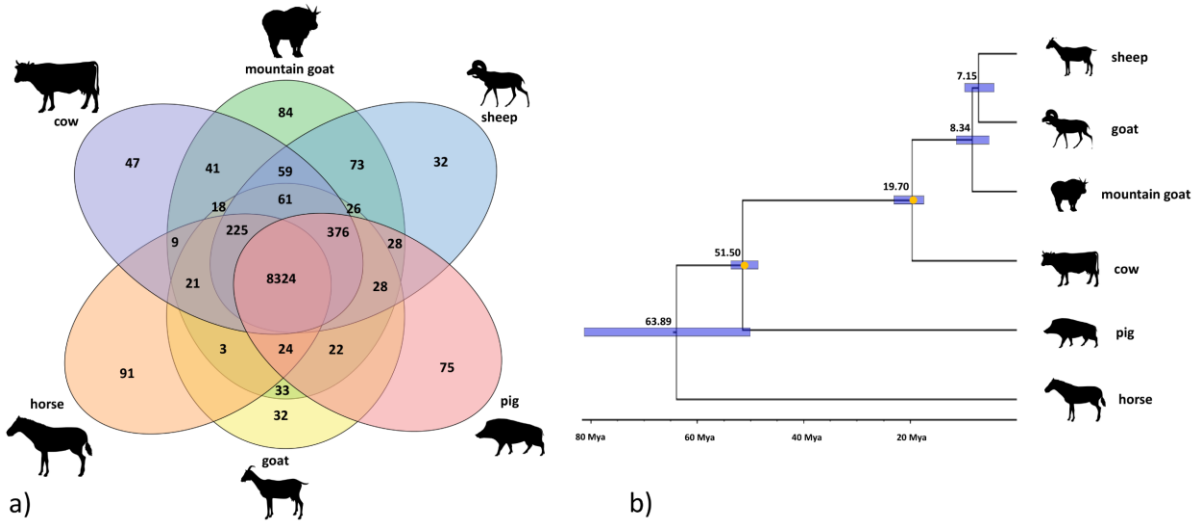

Figure 2. a) A Venn diagram of orthologous clusters shared among mountain goat (*Oreamnos americanus)*, pig (*Sus scrofa),* horse *(Equus caballus), cow (Bos taurus),* goat *(Capra hircus)* and sheep (*Ovis aries)*. The numbers represent the number of orthologous clusters and only the clusters unique to each species and shared with *O. americanus* are labelled for clarity. b) Phylogenetic relationships and divergence times between the six species estimated from ultraconserved elements (UCEs) and MCMCtree. The calibration points are indicated by orange dots; node bars encompass the 95% credible intervals.

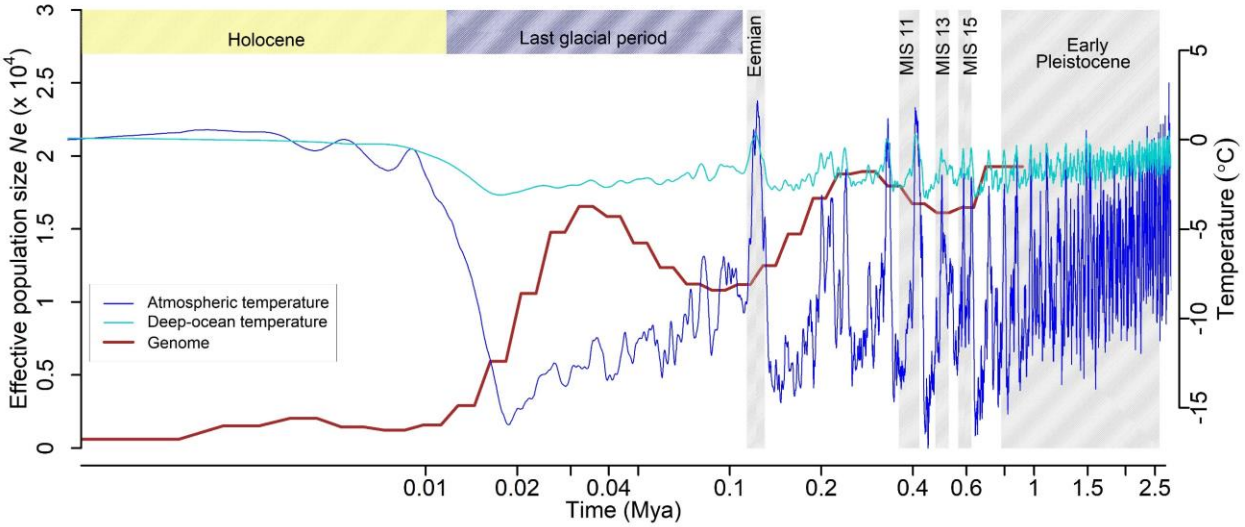

Figure 3. Historical effective population size (N_e) of the North American mountain goat estimated with PSMC analysis assuming a mutation rate μ of $1.33*10^{-8}$ mutations/site/generation and a generation time of 6 years. Atmospheric and ocean temperatures in °C are given on the vertical right axis, N_e in units of 10,000 individuals on the left vertical axis, and time in Mya on the logarithmic horizontal axis.

336 Table 1. Genome assembly statistics for the North American mountain goat for the short-

337 read data and subsequent scaffolding.

	Short-read assembly	**Scaffolded + Chicago assembly**
Total length	2,489 Mbp	2,506 Mbp
N50 / L50	29.0 Kbp / 24,137 scaffolds	66,617 Kbp / 13 scaffolds
N90 / L90	6.5 Kbp / 92,814 scaffolds	18,734 Kbp / 37 scaffolds
Number of scaffolds	172,540	3,217
N count (% of genome)	530,049 (0.02%)	17,462,549 (0.70%)
Gaps	23,758	193,083

338

339

340 Table 2. Summary of repeats in the assembled genome of *Oreamnos americanus*.

	Length (bp)	**Percentage of sequence**
Total interspersed repeats:	1,146,349,974	45.74%
SINEs	286,943,974	11.45%
LINEs	681,473,093	27.19%
LTR elements	122,655,871	4.89%
DNA elements	54,579,895	2.18%
Unclassified	697,141	0.03%
Satellites	3,429,492	0.14%
Simple repeats	17,168,694	0.69%
Low complexity	3,546,216	0.14%
Total:	1,170,891,890	46.72 %

341

342

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This research was supported by NSERC Discovery Grant, Alaska Department of Fish and Game Federal Aid, Compute Canada award to ABAS and INCEPTION grant (PIA/ANR-16-CONV-0005) to RC. We thank Maël Kerbiriou and Dr. Antoine Limasset from Université de Lille who provided valuable insights for this study. DM is supported by an Ontario Graduate Scholarship; the Globalink Travel Award that made this collaboration possible was provided by Mitacs and Inria. We thank two anonymous reviewers for comments that improved the manuscript.

Literature Cited

Agaba, M., E. Ishengoma, W. C. Miller, B. C. McGrath, C. N. Hudson *et al.*, 2016 Giraffe genome sequence reveals clues to its unique morphology and physiology. Nat. Commun. 7: 11519.

Andrews, S., 2010 Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bana, N. Á., A. Nyiri, J. Nagy, K. Frank, T. Nagy *et al.*, 2018 The red deer Cervus elaphus genome CerEla1.0: sequencing, annotating, genes, and chromosomes. Mol. Genet. Genomics 293: 665–684.

Bao, W., K. K. Kojima, and O. Kohany, 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6: 11.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent *et al.*, 2004 Ultraconserved elements in the human genome. Science 304: 1321–1325.

Benson, G., 1999 Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27: 573–580.

Benton, M. J., and P. C. J. Donoghue, 2007 Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24: 26–53.

Bickhart, D. M., B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie *et al.*, 2017 Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. Nat. Genet. 49: 643–650.

Bushnell, B., 2018 *BBTools. Available online at: https://jgi.doe.gov/data-and-tools/bbtools/*.

Chapman, J. A., I. Ho, S. Sunkara, S. Luo, G. P. Schroth *et al.*, 2011 Meraculous: de novo genome assembly with short paired-end reads. PloS One 6: e23501.

Cherukuri, S., N. A. Tripoulas, S. Nurko, and P. L. Fox, 2004 Anemia and impaired stress-induced erythropoiesis in aceruloplasminemic mice. Blood Cells. Mol. Dis. 33: 346–355.

Chikhi, R., and P. Medvedev, 2014 Informed and automated k-mer size selection for genome assembly. Bioinformatics 30: 31–37.

Côté, S. D., and M. Festa-Bianchet, 2003 Mountain goat, in *Wild Mammals of North America: Biology, Management, Conservation*, edited by G. A. Feldhamer, B. Thompson, and J. Chapman. The John Hopkins University Press, Baltimore, Maryland.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada, 2012 jModelTest 2: more models, new heuristics and high-performance computing. Nat. Methods 9: 772.

Dong, Y., M. Xie, Y. Jiang, N. Xiao, X. Du *et al.*, 2013 Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). Nat. Biotechnol. 31: 135–141.

Faircloth, B. C., 2016 PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32: 786–788.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield *et al.*, 2012 Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61: 717–726.

Fitzpatrick, M. C., and S. R. Keller, 2015 Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol. Lett. 18: 1–16.

Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen *et al.*, 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9: R7.

Humann, J., T. Lee, S. Ficklin, C. Cheng, H. Hough *et al.*, 2018 Next Generation Genome Annotation and Analysis Workshop., in San Diego, CA, USA.

Hunt, S. E., W. McLaren, L. Gil, A. Thormann, H. Schuilenburg *et al.*, 2018 Ensembl variation resources. Database 2018:.

Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360.

Kumar, S., and S. Subramanian, 2002 Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. 99: 803–808.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359.

Li, H., 2018 Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100.

Li, X.-Q., and D. Du, 2014 Variation, Evolution, and Correlation Analysis of C+G Content and Genome or Chromosome Size in Different Kingdoms and Phyla. PLOS ONE 9: e88339.

Li, H., and R. Durbin, 2011 Inference of human population history from individual whole-genome sequences. Nature 475: 493–496.

Li, Z., Z. Lin, H. Ba, L. Chen, Y. Yang *et al.*, 2017 Draft genome of the reindeer (Rangifer tarandus). GigaScience 6:.
Li, D., C.-M. Liu, R. Luo, K. Sadakane, and T.-W. Lam, 2015 MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinforma. Oxf. Engl. 31: 1674–1676.
Lowe, T. M., and S. R. Eddy, 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.
Mainguy, J., S. D. Côté, and D. W. Coltman, 2009 Multilocus heterozygosity, parental relatedness and individual fitness components in a wild mountain goat, Oreamnos americanus population. Mol. Ecol. 18: 2297–2306.
Mapleson, D., G. Garcia Accinelli, G. Kettleborough, J. Wright, and B. J. Clavijo, 2017 KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33: 574–576.
Martchenko, D., E. Prewer, E. K. Latch, C. J. Kyle, and A. B. A. Shafer, 2018 Population genomics of ungulates., in *Population Genomics*, Springer, Cham.
Miller, J. M., S. S. Moore, P. Stothard, X. Liao, and D. W. Coltman, 2015 Harnessing cross-species alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (Ovis canadensis). BMC Genomics 16:.
Muller, E. E. L., E. Glaab, P. May, N. Vlassis, and P. Wilmes, 2013 Condensing the omics fog of microbial communities. Trends Microbiol. 21: 325–333.
Nawrocki, E. P., and S. R. Eddy, 2013 Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29: 2933–2935.
Orlando, L., A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen *et al.*, 2013 Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499: 74–78.
Pettorelli, N., F. Pelletier, A. von Hardenberg, M. Festa-Bianchet, and S. D. Côté, 2007 Early Onset of Vegetation Growth Vs. Rapid Green-up: Impacts on Juvenile Mountain Ungulates. Ecology 88: 381–390.
Putnam, N. H., B. L. O'Connell, J. C. Stites, B. J. Rice, M. Blanchette *et al.*, 2016 Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26: 342–350.
Sambrook, J., and D. W. Russell, 2006 Purification of Nucleic Acids by Extraction with Phenol:Chloroform. Cold Spring Harb. Protoc. 2006: pdb.prot4455.
Savolainen, O., M. Lascoux, and J. Merilä, 2013 Ecological genomics of local adaptation. Nat. Rev. Genet. 14: 807–820.
Seemann, S. E., C. Anthon, O. Palasca, and J. Gorodkin, 2016 Quality Assessment of Domesticated Animal Genome Assemblies. Bioinforma. Biol. Insights 9: 49–58.
Shafer, A. B. A., S. D. Côté, and D. W. Coltman, 2011a Hot spots of genetic diversity descended from multiple Pleistocene refugia in an alpine ungulate. Evol. Int. J. Org. Evol. 65: 125–138.
Shafer, A. B. A., C. I. Cullingham, S. D. Côté, and D. W. Coltman, 2010 Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Mol. Ecol. 19: 4589–4621.
Shafer, A. B. A., and J. C. Hall, 2010 Placing the mountain goat: A total evidence approach to testing alternative hypotheses. Mol. Phylogenet. Evol. 55: 18–25.

Shafer, A. B. A., K. S. White, S. D. Côté, and D. W. Coltman, 2011b Deciphering translocations from relicts in Baranof Island mountain goats: is an endemic genetic lineage at risk? Conserv. Genet. 12: 1261–1268.
Smit, A., R. Hubley, and P. Green, 2013 *RepeatMasker Open-4.0.*
Smith, C. A., and L. Nichols, 1984 Mountain goat transplants in Alaska: restocking depleted herds and mitigating mining impacts, pp. 467–480 in *Proceedings of Biennial Symposium, Northern Wild Sheep and Goat Council*, Whitehorse YK, Canada.
Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.
Stanke, M., A. Tzvetkova, and B. Morgenstern, 2006 AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol. 7 Suppl 1: S11.1-8.
Vezzi, F., G. Narzisi, and B. Mishra, 2012 Reevaluating assembly evaluations with feature response curves: GAGE and assemblathons. PloS One 7: e52210.
Wang, Y., D. Coleman-Derr, G. Chen, and Y. Q. Gu, 2015 OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 43: W78-84.
Wang, Y., C. Zhang, N. Wang, Z. Li, R. Heller *et al.*, 2019 Genetic basis of ruminant headgear and rapid antler regeneration. Science 364: eaav6335.
Williams, J. L., D. Iamartino, K. D. Pruitt, T. Sonstegard, T. P. L. Smith *et al.*, 2017 Genome assembly and transcriptome resource for river buffalo, Bubalus bubalis (2n = 50). GigaScience 6: 1–6.
Wood, D. E., and S. L. Salzberg, 2014 Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15: R46.
Yang, Z., 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24: 1586–1591.
Zhang, C., L. Chen, Y. Zhou, K. Wang, L. G. Chemnick *et al.*, 2018 Draft genome of the milu (Elaphurus davidianus). GigaScience 7: gix130.