



OPEN

Chromosomal-level reference genome assembly of muskox (*Ovibos moschatus*) from Banks Island in the Canadian Arctic, a resource for conservation genomics

Si Lok^{1,2}, Timothy N. H. Lau^{1,2}, Brett Trost^{1,2,3}, Amy H. Y. Tong⁴, Tara Paton^{1,2}, Richard F. Wintle^{1,2}, Mark D. Engstrom⁵, Anne Gunn⁶ & Stephen W. Scherer^{1,2,7,8}

The muskox (*Ovibos moschatus*), an integral component and iconic symbol of arctic biocultural diversity, is under threat by rapid environmental disruptions from climate change. We report a chromosomal-level haploid genome assembly of a muskox from Banks Island in the Canadian Arctic Archipelago. The assembly has a contig N50 of 44.7 Mbp, a scaffold N50 of 112.3 Mbp, a complete representation (100%) of the BUSCO v5.2.2 set of 9225 mammalian marker genes and is anchored to the 24 chromosomes of the muskox. Tabulation of heterozygous single nucleotide variants in our specimen revealed a very low level of genetic diversity, which is consistent with recent reports of the muskox having the lowest genome-wide heterozygosity among the ungulates. While muskox populations are currently showing no overt signs of inbreeding depression, environmental disruptions are expected to strain the genomic resilience of the species. One notable impact of rapid climate change in the Arctic is the spread of emerging infectious and parasitic diseases in the muskox, as exemplified by the range expansion of muskox lungworms, and the recent fatal outbreaks of *Erysipelothrix rhusiopathiae*, a pathogen normally associated with domestic swine and poultry. As a genomics resource for conservation management of the muskox against existing and emerging disease modalities, we annotated the genes of the major histocompatibility complex on chromosome 2 and performed an initial assessment of the genetic diversity of this complex. This resource is further supported by the annotation of the principal genes of the innate immunity system, genes that are rapidly evolving and under positive selection in the muskox, genes associated with environmental adaptations, and the genes associated with socioeconomic benefits for Arctic communities such as wool (qiviut) attributes. These annotations will benefit muskox management and conservation.

Keywords *Ovibos moschatus*, Muskox, Umingmak, Continuous long read, Chromosomal-level assembly, Climate adaption, Climate change, Conservation genomics

The muskox (*Ovibos moschatus*), a large iconic and charismatic animal closely related to sheep and goats within the *Caprinae* subfamily, is one of the few surviving relics of the Pleistocene megafauna^{1,2}. The muskox has an

¹The Centre for Applied Genomics, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, 686 Bay Street, Rm 13.9713, Suite 03-6577, Toronto, ON M5G 0A4, Canada. ²Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada. ³Program in Molecular Medicine, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada. ⁴Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada. ⁵Department of Natural History, Royal Ontario Museum, Toronto, ON M5S 2C6, Canada. ⁶Salt Spring Island, Canada. ⁷McLaughlin Centre, University of Toronto, Toronto, ON M5G 0A4, Canada. ⁸Department of Molecular Genetics, Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada. ✉email: si.lok@sickkids.ca; Stephen.scherer@sickkids.ca

intriguing recent history. It was largely unknown by the early Europeans until the first published account in 1720. For the next several decades, classification attempts were debated and confounded naturalists by the muskox's apparent sheep, cow, and buffalo-like features^{3,4}. The muskox was formally introduced into systematic zoology by Zimmermann in 1780 under the name *Bos moschatus*⁵. After further debate, it was removed from the Linnean genus *Bos* and was reassigned to a new genus, *Ovibos*, by Blainville in 1816, where the name stands today with the muskox being the only living member of that genus⁶. However, the classification debate continues to this day over subspecies designation, their number and distribution^{7–9}. In the present study, we follow the last formal taxonomic revision of the species, which regarded *O. moschatus* as a monotypic species⁷.

Muskoxen are physiologically and behaviourally adapted to live year-around in the cold conditions of the Arctic tundra. Once widely distributed across the Holarctic, the present endemic population is restricted to Northern Canada and Greenland. In the last century, muskoxen have been reintroduced to Northern Quebec, Norway, Russia, Southwest Greenland, and Alaska (see Fig. 1). The reintroduction effort in Alaska was to replace populations extirpated by overhunting. Muskox populations today inhabit a large and relatively diverse range that extends from the sub-arctic northern edge of the boreal forest (56° N) to the high-arctic tundra (83° N).

Known by the Inuit as *Umingmak* (the Bearded One), the muskox is an integral and evolving part of Inuit cultural and social-economic heritage^{10,11}. In addition to being a meat source, and providing direct economic benefits such as tourism and the commercialisation of hide and luxury qivut wool^{12–15}, the muskox is a key species in the Arctic tundra ecosystem. Muskox grazing critically impacts plant communities and nutrient flow^{16,17} as well as carbon dioxide and methane fluxes in the Arctic biosphere¹⁸, which are expected to be disrupted by climate change.

The world-wide population of muskox in 2020 is estimated to be 170,000¹⁹. As such, the muskox holds the International Union for Conservation of Nature (IUCN) Conservation status of Least Concern²⁰. However, of the 55 muskox populations reported in 2020¹⁹, six are in decline, including the population on Banks Island, which is well-known for its historic fluctuations in number²¹. The population on Banks Island was the largest in Canada in the 1990s but has suffered a 70% decline since 2000. Among the factors contributing to this decline include infection by a new bacterial pathogen *Erysipelothrix rhusiopathiae*, which normally infects farm animals and poultry^{22,23}. Together with the expanding range of muskox lungworms^{24–26}, and other diseases including brucellosis²⁷, these could be early harbingers of emerging disease modalities from climate disruption.

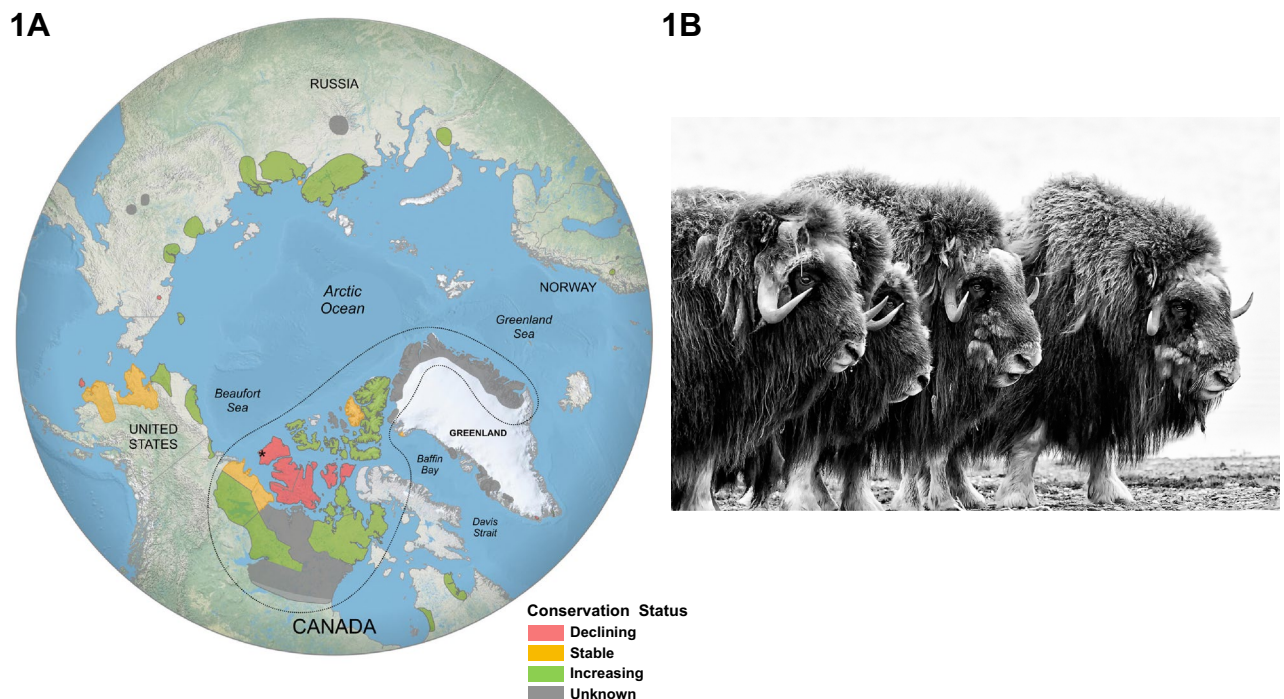


Figure 1. Muskox distributions. **(A)** Population distributions and regional conservation status of the muskox in 2023 (Tomaselli M, Cuyler C, Kutz S, Adamczewski J, Brodeur V, Campbell M, Cluff D, Gorn T, Hughes LJ, Leclerc L-M, Nelson M, Parr B, Sipko T, Suiter M, Taillon J and Gunn A. Muskox health, status and trends—crossing the boundaries. Unpubl. Poster, Arctic Ungulate Conference, Anchorage, Alaska, May 2023). Endemic populations in Canada and Greenland are separated from the translocated (introduced or re-introduced) populations by the enclosed dotted boundary line. The Sachs Harbour area of Banks Island, Northwest Territories, Canada, where our specimen (Royal Ontario Museum archive: ROMMRAN27; NCBI BioSample: SAMN26661894) originated is denoted by an asterisk (*). **(B)** Muskoxen in a defensive posture. Despite their well-known group defence, muskoxen are prey for wolves (*Canis lupus*) and grizzly bears (*Ursus arctos*). The muskoxen in the photograph are descendants of individuals translocated in 1935 from eastern Greenland to the Arctic Wildlife Refuge, Alaska, USA. (Photograph: ©Peter Mather, 2010).

Climate records supported by mitochondrial and microsatellite markers and by whole genome genotyping have provided evidence that the muskox underwent multiple population bottlenecks, which have resulted in contemporary populations having very low genetic diversity^{9,28–30}. Although the muskox has been shown to have the lowest genome-wide heterozygosity reported in an ungulate, there are no overt signs of inbreeding depression in native muskox populations³¹. Additionally, Prewer and colleagues commented that the extremely low variability is a variance with many increases in muskox numbers and expanding distribution raising the possibility that the adaptive capacity has been under-estimated³². A likely mechanism is that over time, deleterious genes can be purged as have been suggested for the endangered kākāpō³³ and the Svalbard reindeer³⁴. Bayesian modelling suggests the muskox might have reached a genetic diversity minimum³², and it is not clear whether their remaining genetic capacity is sufficient for them to adapt to a rapidly changing Arctic.

The past decade has experienced an unprecedented increase in temperature in the Arctic. The Intergovernmental Panel on Climate Change 2023 climate models predict a mean global increase between 1.4 and 4.4 °C by 2100³⁵ with a disproportionate effect in the Arctic. Large herbivores are particularly at risk³⁶. Thus, whether the muskox could physiologically and behaviourally adapt to a warmer Arctic would be of fundamental consequence for their conservation and management. An initial step to understand the muskox's adaptive capacity has been to move toward the use of whole-genome re-sequencing³¹ based on the alignments to the draft short-read assemblies of the muskox^{2,37}. While these assemblies provided an important first glimpse of the muskox genome, they are relatively incomplete and fragmented, typical of assemblies derived from short reads. As a resource for a conservation genomics-based approach for the management of the muskox, we report a chromosomal-level assembly derived from long-reads of a male muskox from Banks Island in the Canadian Arctic Archipelago. The ability of long reads to span repetitive sequences and other difficult regions of the genome³⁸ has enabled the present assembly to have a more complete representation of the coding genes, isoforms of these genes, and the elucidation of their genomic organization. Moreover, the anchoring of our assembly to the 24 muskox chromosomes allows a direct cytogenetic comparison to other members of Bovidae, notably to the useful quantitative trait loci (QTLs) identified from economically important livestock. These QTLs could potentially be mapped to the muskox. From this assembly, we also provide exon-level annotation of genes that could be useful in muskox management, including genes of the immune system, those associated with environmental adaptations, particularly to heat stress, and genes associated with relevant socioeconomic benefits as direct drivers for species conservation.

Materials and methods

Specimen selection for genome assembly

Kidney tissue (NCBI: BioSample SAMN26661894) from a male muskox (*Ovibos moschatus*) was obtained from the Royal Ontario Museum Tissue Archive (catalog number: ROMMRAN27). The tissue was collected on 25 October 1991 near Sachs Harbor, Banks Island, NWT, Canada.

Species identity of the specimen was verified by matching a segment of the cytochrome oxidase 1 (COX1) gene from the specimen's assembled mitochondrial genome against the Barcode of Life Database (<http://www.boldsystems.org>). The top nine hits in the database were *Ovibos moschatus* at a 100% match. Male specimens are typically selected for genome assembly to have representation of the X and Y chromosomes. The male sex of our muskox specimen was denoted by the presence in our assembly of the X and Y chromosome-specific genes, ZFX (Acc KAL1286804) and ZFY (Acc KAL1286379)³⁹, and the Y chromosome-specific gene, SRY (Acc KAL1286397).

DNA extraction for sequencing and genome assembly

Genomic DNA was extracted from frozen kidney tissue using a Puregene Reagent Kit (Qiagen, Hilden, Germany), with the resulting purified kidney DNA having 260 nm/280 nm and 260 nm/230 nm absorbance ratios of 1.90. DNA had a peak length of 22.6 kb on an Agilent TapeStation (Agilent, Santa Clara, CA). Although DNA from flash-frozen fresh whole blood is often preferred for long-read sequencing, when blood samples are not available due to logistical or regulatory issues, DNA purified from well-maintained archival tissue banks (particularly kidney tissues) is acceptable for PacBio sequencing and genome assembly. DNA for library construction was quantified by fluorometry using the Qubit DNA HS Assay (ThermoFisher, Waltham, MA).

Genome sequencing and pre-assembly filtering

Long-read libraries for genome sequencing were prepared from five µg of un-sheared genomic DNA using an Express template prep kit (version 2.0) (Pacific Biosciences, Menlo Park, CA) followed by a post-library-construction sizing step of > 15 kb on the BluePippin system (Sage Science, Beverly, MA). Size-selected library was sequenced on a Sequel II sequencer (Pacific Biosciences; PacBio) in the Continuous Long-Read (CLR) mode with a 15-h movie acquisition time. Raw reads were processed on PacBio's P filter to remove low-quality reads and adapter sequences. The longest sub-read was selected from each productive zero-mode waveguide on the flow-cell. Sub-reads of less than 2 kb length were discarded. 16,756,911 reads remained with a read length N50 value of 15,868 bp, similar to the peak length of the starting DNA (22.6 kb). Using an estimated genome size of 3 Gb inferred by tabulating k-mer frequencies⁴⁰, the resulting CLRs used for the assembly comprised 207.6 Gb of sequences, representing 71.6× genome coverage. Read statistics, including genomic coverage at different ranges of read-length N50 values are provided in Supplementary Figure S1.

Short reads used to polish the assembled genome were produced from 700 ng of the same genomic DNA used for long-read sequencing. DNA was randomly fragmented to 500–600 bp using a Covaris LE220 Focused Ultrasonicator (Covaris, Woburn, MA), followed by library construction carried out in accordance with the Illumina TruSeq DNA PCR-Free protocol (Document 1000000039279v00) (Illumina, San Diego, CA). Sequencing was performed on the HiSeq X sequencer (Illumina), producing 2 × 150 nt paired-end reads. Only sequences with quality scores Q35 or better were used. To achieve this level of accuracy, the first 10 bases were discarded

in read 1 and read 2, the last 20 bases were discarded in read 1, and the last 30 bases were discarded in read 2. After trimming the least accurate portions of the reads and removal of adapter sequences, 777 million paired-end reads remained, comprising 178.6 Gb of sequences (61.6× coverage).

De novo genome assembly and chromosome assignment

A two-step assembly workflow was used to assemble the muskox genome. In the first step, de novo genome assembly was carried out directly from uncorrected PacBio CLR (71.6× coverage) using Flye 2.8.2, which was designed to assemble error-prone reads⁴¹. To resolve mis-assemblies and other errors, the primary assembly was polished three times with Flye-polish⁴¹ using the same CLR that were used to construct the assembly. To correct residual PacBio sequencing errors remaining in the primary assembly, typically single base or small insertions and deletions (indels) inherent to CLR, the assembly was subjected to eight rounds of additional polishing with high-quality trimmed Illumina short-reads (61.6× coverage) using FreeBayes 1.3.1 (<https://github.com/freebayes/freebayes>).

In the second step, multiple cross-species scaffolding against reference genomes of related Cervidae was used to generate super-scaffolds to produce a chromosomal-level assembly for the muskox. To mitigate potential bias during this process, all input de novo assembled contigs and scaffolds of the muskox were to be left intact and unaltered. This cross-species super-scaffolding step drew heavily on the previous comparative studies of karyotypes in the Cervidae family⁴² and on the comparative chromosome painting-based map of the muskox, dromedary (*Camelus dromedaries*) and human⁴³, and similar information from 43 other bovid genomes⁴⁴. The chromosome number of the muskox was established at $2n = 48$ ^{43,45,46}. Making use of the finding that autosomal arms are highly syntenic in typical Bovidae family members^{42,47}, we used the chromosomal-level reference genomes of the closely related takin (*Budorcas taxicolor*) (GCF_023091745.1), goat (*Capra hircus*) (GCF_001704415.2), and sheep (*Ovis aries*) (GCA_016772045.2) to scaffold our muskox contigs into provisional chromosome arms. Muskox contigs were aligned against the aforementioned reference genomes using minimap2 (v2.24)⁴⁸ with parameter $-x$ asm20, allowing alignments with up to 20% divergence in the nonrepetitive sequences⁴⁹. As a quality control step, the resulting alignments were then reviewed and filtered to ensure there were no contradictions with the Repeat Graph generated by the Flye assembler. The Repeat Graph represented all assembled contigs and their junctions, thus providing a guide as to whether viable paths between contigs in question were plausible during scaffolding. The resulting scaffolded chromosome arms were then localised and assigned to 24 provisional chromosomes using the cattle genome (GCF_002263795.3) as an anchor in accordance with the cross-species relationships previously established⁴³. Chromosomal scaffolding, orientation and assignments were further refined and adjusted with a final alignment against the comparative chromosome map established by chromosome painting for the muskox, human (GCF_000001405.40) and the dromedary genomes (GCF000803125.2)⁴³. In another quality control step, Minimap2 was used once again to return the best pairwise unique matching segments against the human and dromedary genomes, enabling our assembled scaffolds to be anchored to the 24 chromosomes of the muskox with high confidence. In a final quality control step, long reads were tiled along the length of the assembly to identify regions of discordance for curation. The resulting assembly, O.moschatus_RAN27-v1.0, has a contig N50 of 44.7 Mbp, a scaffold N50 of 112.3 Mbp, and a completeness of 93.49% as determined by the tabulation of k-mers derived from highly accurate Illumina short-reads using Merqury⁵⁰.

Sequence contamination in muskox assembly

As PacBio CLR could have up to 15–20% pseudorandom errors in the forms of base substitutions and small insertions and deletions⁵¹, a search for DNA contamination from environmental or laboratory sources might not be sensitive at the read level. Likewise, Illumina short reads might be too short to make reliable and unambiguous calls of potential contaminating DNA. As a consequence, contamination assessment was carried out in the final assembly after the ensuring correction steps for sequencing errors.

Potential microbial and other environmental DNA contaminants in the DNA sample that made it into the final assembly were assessed using the workflow depicted in Supplementary Fig. 2. BLAST analysis of non-overlapping 5 kb windows in the assembly revealed five windows showing a match (95–99%) to *Delftia acidovorana* (Acc CP058970.1), a common aerobic environmental organism found in soil. The five positive windows were adjacent to each other and were all confined to a single contig of 26 kb length, which was removed from the assembly. No microbial or fungal sequences were found amongst the top 50 ranked hits in the remaining windows, indicating there were no other overt environmental contaminants in the final muskox assembly.

Human DNA contamination has been reported in many non-primate genome databases, presumably from laboratory sources⁵². Using a primate-specific SINE, *AluY*, we found no discernible primate sequences in the muskox assembly.

BUSCO and genome annotation

The completeness of our muskox assembly was qualitatively assessed using BUSCO v5.2.2 (Benchmarking Universal Single Copy Orthologs; mammalia_odb10)⁵³. BUSCO designates the status of a set of 9226 mammalian lineage marker genes as complete, fragmented, or missing in an assembly. Muskox genes that were failed by BUSCO were subjected to a verification step at exon-level resolution to determine their true status in our assembly. In this verification procedure, we first identified a seed exon for that the gene query in the assembly using a set of orthologous human, sheep, and cow exon probes (NCBI RefSeq). Typically, the seed exon has the highest BLAST score amongst the exon queries for the gene under investigation. From the seed exon, we then scanned up-stream and down-stream along the assembly or scaffold using BLAST in incremental 10 kb size windows in the search for the adjacent exon. Specific search criteria include nucleotide and amino acid sequence identity to the query exon, and the conservation of the reading frame relative to the consensus splice signals at the exon boundaries.

Once an adjacent exon was identified, the process was repeated until all exons were identified. For a gene to be designated as complete in the muskox assembly, the full-length polypeptide encoded by the predicted exons must begin with an initiation codon and end with a termination codon, and to be at least 70% identical to the full-length human, sheep, takin, or cow protein encoded by the query as reported in RefSeq.

Correction of sequencing errors

Sequencing errors in the PacBio platform are typically between 10 and 15%, comprising single base substitutions or small indels. These errors were corrected in the primary assembly using Flye-polish⁴¹ followed by eight rounds of “polishing” with high-quality trimmed Illumina short-reads of Q35 or better using Freebayes 1.3.1 (<https://github.com/freebayes/freebayes>). From monitoring the kinetics of base changes in the primary assembly at different rounds of polishing, we found base changes from polishing with Freebayes do not approach a plateau value until after the fifth or sixth rounds, after which further base changes are confined to a small number of positions where they alternate between what appear to be allelic bases in successive rounds (data not shown).

Assessment of residual sequencing errors in the assembly

In the absence of a reference DNA of a known sequence to make a direct comparison, it was not possible to tabulate the base accuracy of our assembly following the polishing steps. However, we can tabulate base substitution errors indirectly when errors generate premature termination codons disrupting the Open Reading Frames (ORFs) we have annotated.

The following formula depicts the probability of creating a spurious termination codon at different levels of sequencing accuracies:

$$T = (1 - A^3) \times 3/63$$

where A is the probability of a correct base call (i.e. the sequencing accuracy for that nucleotide position); A^3 is the joint probability of three consecutive correct calls forming a correct codon; $1 - A^3$ is the probability of a codon being incorrect with at least one incorrect base call; $3/63$ is the probability of an incorrect codon being a termination codon*; T is the probability of an in-frame termination codon created from an in-frame non-termination codon due to sequencing error; * Of the 64 possible codons, only one is correct and 63 are incorrect at a given position. Since there are three possible termination codons, therefore, the probability of an incorrect codon being a termination codon is $3/63$.

The muskox gene with the longest coding capacity, *Titin* (*TTN*, Acc KAL1287845), in conjunction with the annotated genes in Supplementary Tables S2, S3, S8, and S9, yield ORFs comprising 782,521 codons. Since all of these genes were found to encode uninterrupted ORFs where none of the 782,521 codons are termination codons, a value for T can be set at $< 1/782,521$ or $< 0.00013\%$. Substituting this value for T and solving for A in the aforementioned formula yields an inferred sequence accuracy in the assembly of not less than 99.99910%, corresponding to a Phred score⁵⁴ of no less than 50.4839 in these coding regions.

Identification of repetitive DNA in the genome

Repetitive elements described in RepBase were identified and tabulated in the muskox genome using RepeatMasker⁴⁹. Supplementary Table S1 summarizes the distributions of repetitive sequences in the muskox.

Assessment of genetic diversity in muskox

Genetic diversity was assessed for the Banks Island muskox specimen assembled in this study, and was compared to the genetic diversity of selected muskox specimens collected by Van Coeverden de Groot and colleagues⁵⁵ and Pečnerová and colleagues³¹. Illumina short reads from selected specimens with the highest sequence read depth were downloaded from NCBI SRA (see Supplementary Table S11 for accession numbers) from each geographic region depicted in Fig. 1 of Pečnerová’s paper³¹. Reads from the selected specimens comprised an average 14× coverage of the muskox genome (median coverage of 12×).

Sequence reads were aligned to the muskox reference genome in the present study using BWA 0.7.17r1198⁵⁶. Duplicate reads were marked using the “MarkDuplicate” function of GATK 4.1.9.0⁵⁷. The average depth of coverage of each sample was calculated using the “depth” function of SAMtools 1.9⁵⁶. To ensure that heterozygosity estimates were not confounded by differences in read depths, reads from samples with high read depths were subsampled to 10–15× average depth. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were detected using Genome Analysis Toolkit (GATK) best practices⁵⁸.

Using the alignment file from BWA as input, SNVs and indels were identified using the “HaplotypeCaller” function of GATK with parameter—minimum-mapping quality 20 followed by the GATK function “GenotypeGVCFs”. The FILTER column of the VCFs was populated using the hard-filtering criteria suggested by the authors of GATK (<https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-varia>nts). For SNVs the criteria were: QD < 2.0, MQ < 40.0, SQR > 3.0, SOR > 3.0, FS > 60.0, MQRankSum < −12.5, and ReadPosRankSum < −8.0. For indels the criteria were: QD < 2.0, SQR > 10.0, FS > 200.0, and ReadPosRankSum < −20.0. Only variants passing these filters were used for assessing heterozygosity or to make variant calls against the reference muskox genome. The degree of heterozygosity for each genome was tabulated based on the number of heterozygous variants per megabase of reference sequence. To assess genetic diversity in the muskox immune genes including genes in the MHC, we determined the exon coordinates for the genes of interest and tabulated the calls within those coordinates from the alignment files for each muskox specimen, as well as the variant counts and profiles aggregated across all of the specimens.

Mitochondrial genome assembly

The muskox mitochondrial genome was assembled with Abyss V2.1.5 (parameters $k = 103$, $kc = 6$)⁵⁹ from 40 million Illumina paired-end reads. The NCBI muskox mitochondrion sequence NC_020631.1 was used as a BLAST query to select the contig containing the presumptive mitochondrion sequence. Circularization of the selected contig yielded a complete mitochondrial genome of 16,431 bp. The accuracy of the mitochondrial assembly was supported by the uninterrupted tiling of mapped Illumina short reads across the length of the assembly in 50 bp moving windows of single base increments. Tiling depths of greater than 7500 reads were tabulated across 100% of the assembled circular mitochondrion genome. The mitochondrial assembly was further supported by the alignment of PacBio long reads, yielding a read depth of 90×. We also performed a de novo assembly of the aligned long reads and produced a circular contig of 16,341 bp in length that is identical to the one assembled from short reads.

Results and discussion

Assembly strategy and assessment of assembly quality

Pacific Biosciences (PacBio) Continuous Long-Reads (CLR) were assembled directly using the Flye assembler⁴¹ into a high-quality haploid assembly for the muskox. The resulting assembly, *O. moschatus*_RAN27-v1.0, has a contig N50 of 44.7 Mbp, a scaffold N50 of 112.3 Mbp, and a completeness of 93.49% as determined by the tabulation of k-mers derived from highly accurate Illumina short-reads using Merquy⁵⁰. The incomplete portion of the assembly can be attributed to the incomplete Chromosome Y, centromeres, telomeres, and other unresolved repeats that are missing in the assembly.

As shown here and in our previous study of the North American wolverine genome⁶⁰, the 15–20% pseudorandom sequencing errors typically associated with CLR⁵¹ can be effectively mitigated by the Flye assembler and the ensuing polishing regimens to yield a high-quality assembly that is essentially free of residual sequencing errors. Using the assessment method for the estimation of residual sequencing errors described in Materials and Methods, we estimate the sequence accuracy in our muskox assembly after the polishing steps to be not less than 99.99910% corresponding to a Phred score of not less than 50.4839 (corresponding to < 1 substitution error in 111,787 bases in the coding regions). This level of accuracy exceeds the recommend value for “finished genomes”⁶¹.

We favour the error mitigation strategy used in this study over the computationally intensive approach of correcting individual PacBio reads prior to assembly⁶² or the use of PacBio's recent Circular Consensus Sequencing also known as HiFi Sequencing (HiFi/CCS)^{63,64} where reads are partially corrected during sequencing. HiFi/CCS holds promise as the technology continues to improve in accuracy and achieves a more favourable cost profile; however, at the time of this study, HiFi/CCS technology could achieve a uniform Q30 or better correction for only a small portion of the flow cell, with the majority of the other reads having very heterogeneous and lower degrees of correction. This heterogeneity in read accuracy complicates the models used by many assemblers to build contigs. Moreover, the HiFi/CCS process itself often has a detrimental effect on the overall lengths of the reads compared to CLR, with concomitant impact on the efficiency of scaffolding and assembly. By contrast, assemblers such as Flye could easily model the uniform, but albeit higher overall error rate across all CLR, to build accurate contigs. At least at the time of this study, CLR, arguably offer a better functional trade-off for cost effective de novo assembly of large, complex genomes.

Table 1 compares the assembly metrics of our muskox assembly, *O. moschatus*_RAN27-v1.0, against the two current short-read assemblies for the muskox (ASM214823v1)² and (ASM2253363v2)³⁷. All three muskox assemblies have similar ungapped lengths of between 2.5 and 2.6 Gb. The improved continuity of our long-read assembly is evident by having a much higher contig and scaffold N50 values and lower contig and scaffold counts over the previously reported short-read assemblies for the muskox and other ruminant genomes assembled from short reads⁶⁵. Of the two short-read assemblies for the muskox, the assembly reported by Prewer and colleagues³⁷ is more fragmented, which could be due to the lower quality of the source DNA extracted from hide as opposed to the DNA extracted from fresh blood used by Li and colleagues². The low contig and scaffold L50 values (i.e., the number of contigs or scaffolds that represent half the genome) further reflect the improved contiguity offered by long-read technologies. In terms of the usual continuity metrics, our muskox assembly is at par with, or is better than, the other chromosomal-level reference assemblies currently available for Bovidae, the takin, goat, sheep, and cow.

Assembly completeness and residual sequence errors

The BUSCO (Benchmarking Universal Single-Copy Orthologs) program has long been used to assess the completeness of genome assemblies⁵³. When BUSCO v5.2.2 with its set of 9226 mammalian lineage marker genes was applied to our muskox assembly, 8852 genes (95.9%) were scored as complete, and 374 genes (4.1%) were scored as fragmented or missing in the assembly. At face value, these BUSCO scores placed our muskox assembly in the top echelon for reported genome assemblies derived from long reads. Figure 2 compares the BUSCO scores of our muskox assembly with current NCBI designated reference genome assemblies for sheep (ARS-UI_Ramb_v3.0; GCF_016772045.2) and human (GRCh39.14; GCF_000001405.40).

An important and useful feature of BUSCO is its ability to assess the completeness of genome assemblies across a broad range of species. This was accomplished by balancing sensitivity and specificity to create a single universal set of consensus gene profiles that can be used generically across all mammalian species. To enable this feature, gene profiles are not optimized for any particular species. Due to this compromise, it is possible that subsets of BUSCO profiles could have difficulties detecting their cognate genes in species whose orthologs deviate significantly from the consensus. Accordingly, BUSCO could under-count those genes, resulting in an underestimation of the quality of an assembly being evaluated. Undercounting of genes by BUSCO was exemplified by

	<i>Ovibos moschatus</i> (Muskox) male 2024/04/26 O.moschatus _RAN27_v1.0	<i>O. moschatus</i> * (Muskox) male 2021/01/13 ASM2146233v1	<i>O. moschatus</i> (Muskox) male 2022/05/16 ASM2253363v2	<i>Budorcas taxicolor</i> * (Takin) male 2022/04/22 Takin1.1 ARS-UI- Ramb_v3.0	<i>Ovis aries</i> * (Sheep Breed: Rambouillet) female 2023/07/20 ARS1.2	<i>Capra hircus</i> * (Goat) male 2016/08/24 ARS-UCD2.0	<i>Bos taurus</i> * (Cattle Breed: Hereford) female 2023/07/01
Total sequence length (bp)	2,602,202,933	2,615,835,778	2,621,888,837	2,851,954,952	2,654,063,983	2,922,617,086	277,068,120
Total ungapped length (bp)	2,602,159,733	2,601,841,192	2,473,402,365	2,851,579,369	2,654,021,983	2,922,578,899	2,770,657,958
Number of contigs	654	14,255	114,188	1515	227	30,378	2344
Contig N50 (bp)	44,730,146	843,055	38,369	68,053,581	43,178,051	26,244,591	26,402,946
Contig L50	20	832	18,652	15	24	32	32
Number of scaffolds	222	6709	8658	1363	143	29,886	1958
Scaffold N50 (bp)	112,334,376	46,819,416	13,200,690	109,747,698	101,274,418	87,277,232	103,308,737
Scaffold L50	9	20	61	10	8	13	12
	72× PacBio Sequell II CLR Flye 2.8.2 Freebayes 1.3.1	60× Illumina Hi-Seq SOAPdenovo 1.12 (Li et al. ²)	85× Illumina Hi-Seq SOAPdenovo 240 (Prewer et al. ³⁷)	24× PacBio Sequel I CLR Illumina NovaSeq Bionano Hifiasm v0.13	55× PacBio RSII Nanop- ore Illumina HiSeq Canu 1.8 Nanopolish 0.12.5 Salsa 2.2 PB Jelly 15.8.24 Freebayes 1.3.1	50× PacBio Bionano Hi-C Celera 8.2 Bionano Irys April 2015 Lachesis HiC June 2015	Various Falcon v FEB-2016

Table 1. Assembly metrics. Comparison of the muskox assembly of this study with other muskox and Bovidae assemblies. Asterisk (*) denotes current NCBI reference genomes.

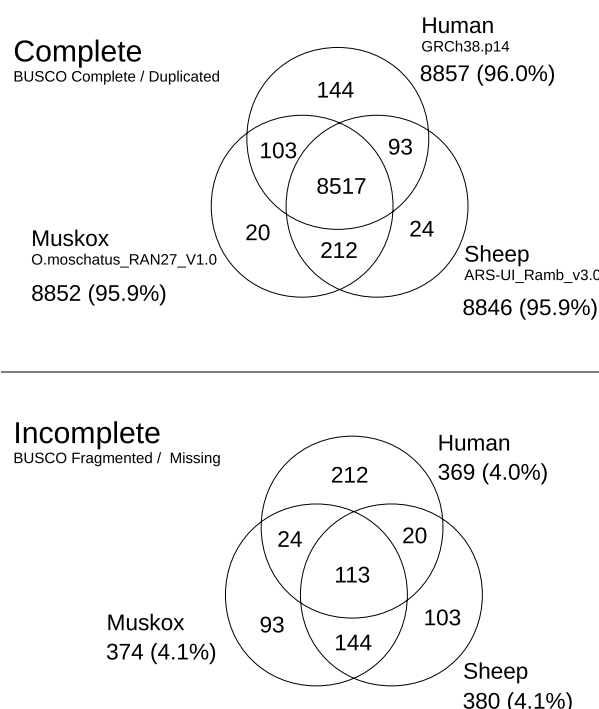


Figure 2. Venn diagram of BUSCO results. Analysis by Benchmarking Universal Single Copy (BUSCO) v5.2.2 of the muskox *O.moschatus*_RAN27_v1.0, human GRCh38.p14 (GCF_000001405.40) and sheep ARS-UI_RAMB_v3.0 (GCF_016772045.2) assemblies. The mammalian lineage (mammalia_odb10) comprised of 9226 ortholog groups was used. “Complete genes” refers to those scored as “Complete” or “Duplicated” by BUSCO, while incomplete genes refer to those scored as “Fragmented” or “Missing” by BUSCO. The official human reference GRCh38.p14 only had 96.0% BUSCO complete even though it is the most thoroughly annotated genome to date. All muskox incomplete genes were found and annotated in the muskox assembly (Supplementary Table S2).

our recent genome assembly for the wolverine⁶⁰ and by the current build of the gold standard human reference genome, GRCh38.p14, where 369 BUSCO genes were called as fragmented or missing (Fig. 2), despite those genes being annotated by the NCBI Genome Data Viewer as complete and experimentally verified.

To address the possibility of an undercounting in our muskox assembly, the 374 genes that BUSCO scored as fragmented or missing (Fig. 2) were annotated at exon-level resolution. The results revealed that all 374 genes are present and complete (Supplementary Table S2), revising the BUSCO score to 100%. Interestingly, amongst the genes BUSCO deemed to be missing in the muskox, *LRP1B* (LDL receptor related protein 1B), might have evaded BUSCO solely due to its size. In many species, *LRP1B* is among a small number of genes that span more than a million base pairs (2.20 Mbp in sheep and 1.90 Mbp in human). In the muskox, *LRP1B* (Acc KAL1287620) comprises 91 highly conserved exons spanning 2.17 Mbp, encoding 4596 amino acids, and is present in its entirety on contig MUS01_AAA20220901_F8-ctg00025 (Acc JBFTXH010000025).

Assessment of genome-wide genetic diversity

Genetic heterozygosity was tabulated for our specimen and compared to the muskox previously sampled across Greenland and Canada^{31,55}. Figure 3 depicts plots of the number of heterozygous variants per megabase for the muskox, two threatened mammalian species, and a ~21,000-year-old ancient muskox sample from Siberia. Consistent with previous findings measured by the tabulation of heterozygous positions, contemporary muskoxen have very low overall genetic diversity across the genome, although mainland muskoxen have higher diversity than those from the Arctic Islands, possibly because the muskox survived the 1900s over-harvesting in six geographically separated refugia. The ancient sample from Siberia showed that genetic diversity was higher in the past. Muskox genetic diversity is lower than two mammalian species under threat that have undergone recent genetic bottlenecks, the Tasmanian devil (*Sarcophilus harrisii*)⁶⁶ and the cheetah (*Acinonyx jubatus*)⁶⁷. Interestingly, despite the muskox having the lowest genome-wide heterozygosity reported in an ungulate, there are no overt signs of inbreeding depression in native muskox populations³¹. The other large Arctic ruminant, the reindeer, is relatively diverse in comparison.

Muskox subspecies

Although muskoxen have not been viewed to be sufficiently variable to warrant sub-species recognition⁷, recent genome analysis showed a distinct clustering of barren-ground muskoxen separate from the Greenland and Arctic Island muskoxen^{9,31}. However, the interpretation of the taxonomic significance of this clustering pattern is not clear in a species with such low genetic diversity. Moreover, phenotypic differences due to diet and nutrition,

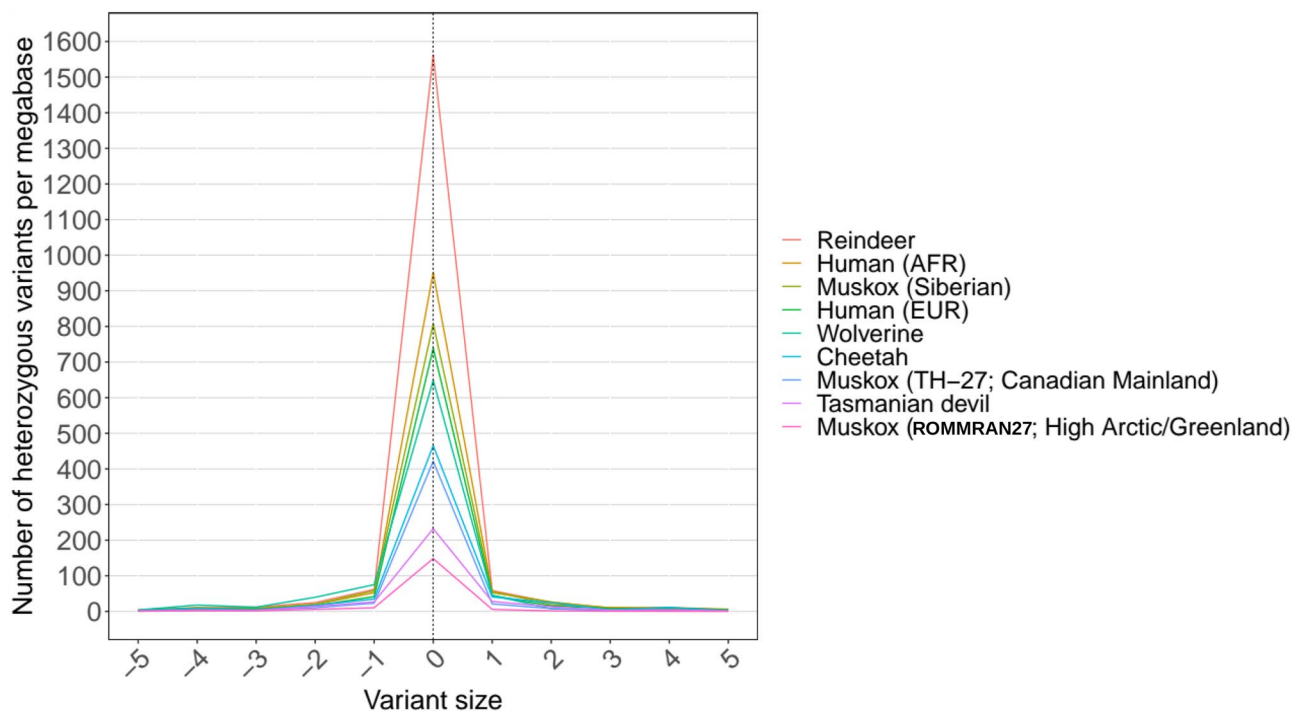


Figure 3. Number of heterozygous variants per megabase of reference sequence for variants of various sizes. Negative values on the x-axis indicate deletions, positive values indicate insertions, and a value of zero indicates single nucleotide variants. Heterozygous variant frequencies of different muskox populations are plotted against two diverse outbred species (human of African and European descent and wolverine), and two species that had undergone recent genetic bottleneck, the cheetah and the Tasmanian devil. The heterozygosity of the Canadian mainland muskox is on par with the cheetah, while that of the High Arctic/Greenland muskox is lower than that of the Tasmanian devil. The other large Arctic ruminant, the reindeer, is relatively diverse in comparison.

age, or other environmental factors have not been rigorously investigated. As such, the designation of muskox subspecies is still an open question awaiting more studies.

Muskox mitochondrial genome assembly and nuclear mitochondrial DNA

Identical, full-length mitochondrial genome sequences of 16,431 bp were assembled from long and short reads (Acc CM083088.1). Sequence comparisons with the other muskox mitochondrial genomes reported showed only minor differences, notably in the control-region, consistent with the muskox being monotypic, and do not allow definition of muskox subspecies³⁰.

Blast analysis of the muskox mitochondrial genome against the *O.moschatus*_RAN27_V1.0 assembly at a threshold of 1E4 revealed 597 potential mitochondrial pseudogene loci of varying lengths and degeneracies (data not shown). These “Nuclear Mitochondrial DNAs” (NUMTs), are believed to originate from invasion of the nuclear genome by mitochondrial DNA via nonhomologous recombination⁶⁸. It is generally viewed that the accumulation of NUMTs is a continuous evolutionary process⁶⁹. Using the same threshold, the mouse genome (*Mus musculus*) has 190 copies, the rat genome (*Rattus norvegicus*) has 61 copies, and the human genome has 1356 copies⁶⁸.

Chromosome assignment and chromosome evolution

As described in the Materials and Methods, we drew heavily on the previous comparative studies of karyotypes in the Cervidae family^{42,43} to anchor our assembly to the 24 chromosomes of the muskox. The muskox karyotype differed from the ancestral pecoran karyotype by six major fusions, one fission, and three inversions. The muskox karyotype includes six submetacentric and 17 acrocentric autosomes, and the sex chromosomes. Our assembly provides near complete representation of all the autosomes. The assembly of the X chromosome is essentially complete. However, the assembly of the Y chromosome is fragmented due to the extremely high composition of repetitive DNA, despite the use of the current generation of long reads. Further resolution of the Y chromosome would require yet longer and more accurate reads, perhaps combined with optical mapping. The comparative assemblies of the X and Y chromosomes in the muskox, takin, sheep and human are shown in Fig. 4. Table 2 presents the correspondence between conserved chromosome segments in the muskox, takin, sheep, goat, cow, pecoran ancestral karyotype⁷⁰, and human.

Conservation genomics

Global ecosystems are under threat from rapid climate change^{35,72}. In addition to direct threats to food supplies, and alterations to prey-to-predator or host-to-pathogen dynamics, there are climate induced anthropogenic disruptions affecting the Arctic ecology⁷³. The latter include increased human intrusions and industrial activities. Large herbivores, such as the muskox, are particularly at risk from such ecological disruptions^{19,36}.

Arctic animals are adapted to live year-round in extreme seasonal variations^{2,34,74–76}, characterized by a short season of plant growth and a period of less digestible forage¹⁹. Temperature (and duration) and photoperiod changes are major cues regulating host–pathogen interactions⁷⁷. The disruption of these seasonal cues by climate change could have profound consequences in host–pathogen dynamics in the Arctic^{78–81}.

Conservation genomics is application of advanced genomics analyses for the preservation of the viability of populations and the biodiversity of living organisms. This important arsenal of tools helps to mitigate the effects of climate change^{82–87}. High-quality reference genome assemblies are increasingly recognized as an important foundation of conservation genomics^{84–91}. This is particularly so as the new chromosomal-level assemblies derived from long reads are superseding the earlier, more fragmented assemblies with fewer gene representations.

One immediate use for the new assemblies is for the identification of comprehensive sets of relevant genes at exon level resolution. These gene sets are the starting points to identify specific genes or variants to elucidate adaptive pathways, and to develop genomic and epigenomic biomarkers to monitor gene flow, hybridization and introgression, disease susceptibility, and other responses to a changing environment. To provide this important resource for the muskox, we have annotated selected families or classes of genes that could participate in the conservation and management of Arctic wildlife. These include genes encoding components of the innate immune system, and are provided in Table 3 and in Supplementary Tables S3 through S9.

Muskox genes associated with Arctic adaptation

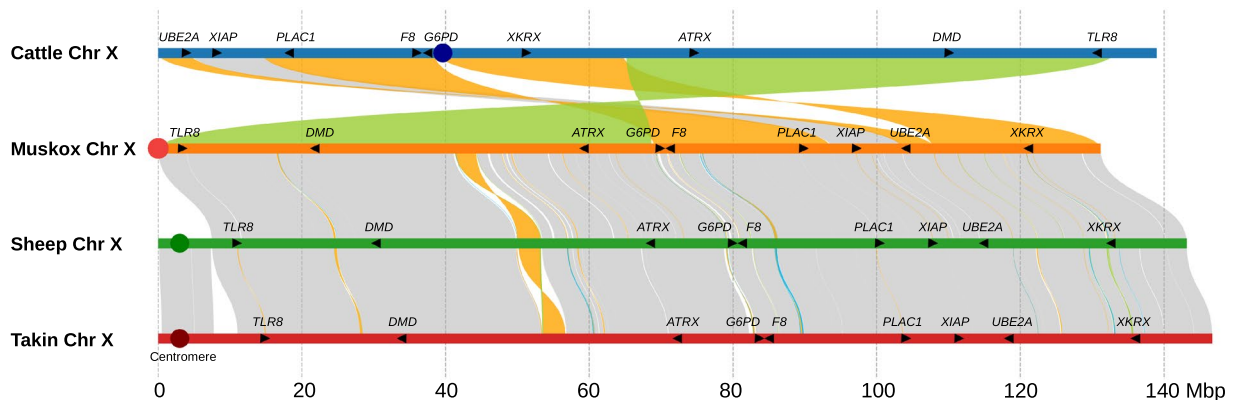
Supplementary Table S3 provides exon-level annotation, chromosome location, and citations of genes that are associated with Arctic adaptation⁷⁴. These genes are involved in metabolic processes, including various aspects of brown adipose tissue (BAT)⁹² and circadian rhythm^{93,94}. Supplementary Table S3 also includes candidate genes in the muskox previously identified² as having convergent amino acid substitutions, are rapidly evolving, or that are under positive selection.

Genes of the innate immunity pathways

There has been a marked increase in outbreaks of infectious disease among wildlife attributable to environmental disruptions from anthropogenic climate changes^{79,80}. Arctic species are particularly vulnerable as pathogens invade new northern niches. The innate immune system is the first to respond to pathogens^{95,96}. As a resource to support infection genomics and wildlife management, we provide gene annotations and accession numbers for the principal members of the innate immunity pathways in the muskox (Table 3 and Supplementary Tables S6 and S7).

Table 3 highlights the major classes of cell surface and intracellular innate receptors of the innate immunity system, and their principal signalling intermediaries in the muskox. These receptors include the Toll-like⁹⁷, Nod-like⁹⁸, C-type lectin⁹⁹, and the RIG-I¹⁰⁰ families of receptors that recognize pathogen-associated patterns

4A



4B

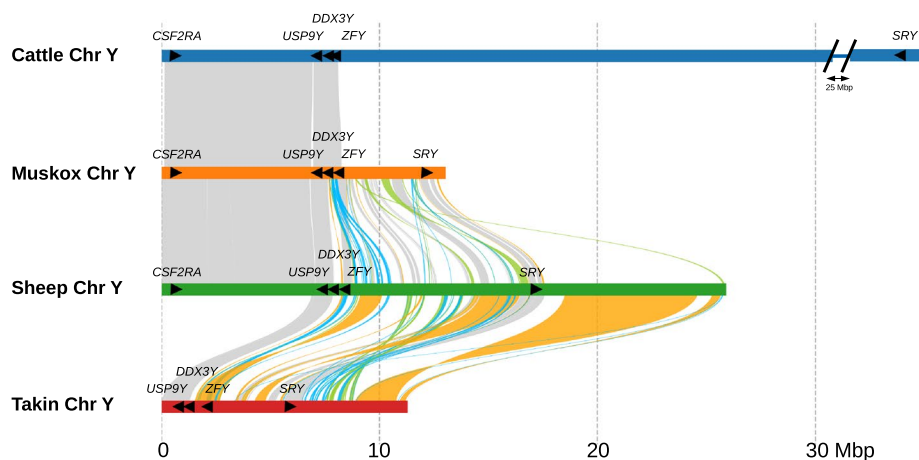


Figure 4. Synteny of the sex chromosomes between muskox, cattle, sheep and takin. Grey-shaded ribbons indicate regions of broad syntenicity, orange ribbons represent inversions, green ribbons represent translocations, and blue ribbons indicate duplications. *Plotstr*⁷¹ was used in the construction of the figure. **(A)** Comparison of chr X between the four animals. The cattle differs from the rest by three inversions and one translocation with inversion. The sheep mainly differs from muskox and takin by one inversion. The beginning of sheep and takin chromosome X represents a centromeric region not assembled in muskox. Selected genes associated with X-linked disorders or discussed in the paper are marked along with their orientation. **(B)** Comparison of chromosome Y between the four animals. Chromosome Y proved challenging to assemble and remained highly fragmented and incomplete for muskox, sheep and takin. Selected Y-specific genes were annotated and marked denoting their orientation.

or damage-associated molecular patterns. RIGI (DHX58) and IFIH1 (MDA5) are the principal sensors for nucleic acids from viral pathogens¹⁰¹. Once receptors are activated, signals are potentiated into the cell nucleus by specific adaptor molecules such as MYD88, TICAM1 (TRIF), MAVS, and STING¹⁰², and a series of signalling intermediaries, notably the TRAFs, STATs and members of the IRAK, JAK¹⁰³ and the mitogen-activated protein (MAPKs)¹⁰⁴ families of kinases. Signals potentiated through MAPKs could also provide cross-talk with the adaptive immunity pathways, and the support of apoptosis or autophagy leading to the removal of damaged or infected cells. The 55 MAPK genes in muskox are provided in Supplementary Table S6.

Host responses upon receptor activation include the expression of interferons^{105–107} (Table 3), specific inflammatory cytokines¹⁰⁸, chemokines^{109,110}, and anti-microbial peptides such as the defensins¹¹¹. Chemokines, notably CXCL7, CXCL9–11, CCL20, and CCL28 (Supplementary Table S6), appear to have direct antimicrobial activities similar to the defensins¹⁰⁹. We identified 16 beta defensin genes in muskox (Supplementary Table S6). All paralogs of the large extended human interleukin-17 (*IL17A* to *IL17F*) and interleukin-17 receptor (*IL17RA* to *IL17RE*; *IL17REL*) families are present in muskox, pointing to a broad evolutionary conservation of all individual family members in the rapid response to infectious agents¹¹². The chemokines, interleukins, and tumor necrosis factors and receptors involved in muskox innate immunity are presented in Supplementary Table S6.

MUSKOX	Sheep (GCA_016772045.2)	Goat (GCF_001704415.2)	Takin (GCF_023091745.1)	Pecoran Ancestral karyotype (PAK)	Cattle (GCF_002263795.3)	Human (GCF_000001405.40)
1p	17	17	17	N1	17	22q/12q/4pq
1q	1q	1	1q	A2	1	21/3/21
2p	20	23	11p	R	23	6p
2q	2q	2	2q	B2	2	2q"/1
3p	24	25	2p	T	25	16p/7
3q	3q	5	5q	C2	5	12pq/22q"/12pq/22q"
4p	26	27	24	V	27	4/8p"
4q	3p	11	11q	C1	11	2pq/9
5p	21	29	25	W	29	11
5q	23	24	22	s	24	18
6p	25	28	5p	u	28	10q
6q	19	22	1p	Q	22	3
7	6	6	6	F	6	4pq
8	5	7	7	E	7	19p/5
9	1p	3	3	A1	3	1
10	2p	8	8	B1	8	8p7 9/8p'/9
11	12	16	16	K	16	1
12	4	4	4	D	4	7
13	7	10	10	G	10	15/14/15/14
14	8	9	9	H1	9	6q
15	14	18	18	M	18	16q/19q
16	18	21	21	P	21	15/14
17	10	12	12	I	12	13
18	13	13	13	J	13	20/10p/20
19	15	15	15	L	15	11
20	11	19	19	N2	19	17
21	9	14	14	H2	14	8q
22	16	20	20	0	20	5
23	22	26	23	U	26	10q
X	X	X	X	X	X	X
Y	Y		Y		Y	Y

Table 2. Correspondence of conserved chromosomal segments between muskox, sheep, goat, takin, pecoran ancestral karyotype (PAK), cattle and human. The relationships between muskox, PAK, cattle, and human were established by Proskuryakova and colleagues through chromosome painting⁴³ and validated in this study by sequence alignment. The correspondence between muskox and sheep, goat and takin were established in this study by sequence alignment. Corresponding segments were ordered beginning from the centromere (most relevant for the human column). Chromosome 7–23 and X in muskox are acrocentric. The karyotype of muskox represents a departure from the PAK by six fusions (N1 + A2, R + B2, T + C2, V + C1, W + S, U + Q) and one fission (U), maintaining an autosomal fundamental number (i.e. autosomal arm count) of 58, a stable number for most karyotyped Bovidae species with varying diploid chromosome number (2n).⁴⁷

Genes of the major histocompatibility complex

The major histocompatibility complex (MHC) is the most important region in the genome with respect to adaptive and innate immunity^{95,96,113–116}. In human, the MHC is located on the short arm of chromosome 6 (6p21.3). The MHC is particularly gene rich, occupying more than 7.5 Mbp of the genome, comprising several hundred genes encoding ligands, receptors, and other mediators and regulators of the inflammatory and immune responses. As such, the MHC plays a vital role in fighting infection by pathogenic agents. Apart from regulating immunity, the MHC may also have roles in reproduction and social behaviors, such as mate selection and kin recognition^{117,118}.

The MHC is divided into three main regions loosely based on related structure or functions. Recently, the MHC has been extended to include the flanking genes¹¹³. The Class I region is located nearest to the telomere and includes the Class I leukocyte antigen presenting genes. These genes encode highly polymorphic cell surface receptor subunits, which when paired with a beta-2-microglobulin subunit (B2M), the functional receptor present endogenous antigens processed from intercellular pathogens to T-lymphocytes (CD8 + T-cells). The MHC Class II region includes the Class II leukocyte antigen presentation genes, which also encode highly polymorphic cell surface receptor subunits. However, they differ from the Class I receptors in that they form functional receptors through heterodimers comprised of alpha and beta Class II receptor subunits. By contrast to the widespread expression of the Class I leukocyte receptors, Class II receptor expression is restricted to the

Gene symbol	Alternative name/symbol	Accession	Muskox chromosome
Toll-like receptors			
<i>TLR1</i>	<i>CD281</i>	KAL1288282	Chr 7
<i>TLR2</i>	<i>CD282</i>	KAL1286936	Chr 1
<i>TLR3</i>	<i>CD283</i>	KAL1287412	Chr 4
<i>TLR4</i>	<i>CD294</i>	KAL1288035	Chr 10
<i>TLR5</i>	<i>CD295</i>	KAL1288066	Chr 11
<i>TLR6</i>	<i>CD296</i>	KAL1288284	Chr 7
<i>TLR7</i>	<i>CD287</i>	KAL1286564	Chr X
<i>TLR8</i>	<i>CD288</i>	KAL1286565	Chr X
<i>TLR9</i>	<i>CD289</i>	KAL1287817	Chr 6
<i>TLR10</i>	<i>CD290</i>	KAL1288283	Chr 7
Nod-like receptors			
	NLR family pyrin domain containing		
<i>NLRP1</i>	<i>NALP1; CARD7</i>	KAL1287737	Chr 20
<i>NLRP2</i>	<i>NALP2</i>	KAL1286518	Chr 15
<i>NLRP3</i>	<i>NALP3</i>	KAL1288227	Chr 8
<i>NLRP5</i>	<i>NALP5</i>	KAL1286519	Chr 15
<i>NLRP6</i>	<i>NALP6</i>	KAL1286925	Chr 5
<i>NLRP8</i>	<i>NOD16</i>	KAL1286520	Chr 15
<i>NLRP12</i>	<i>NALP12</i>	KAL1286516	Chr 20
<i>NLRP13</i>	<i>NOD14; NALP13</i>	KAL1286517	Chr 15
<i>NLRP14</i>	<i>NALP14</i>	KAL1287893	Chr 19
<i>NLRX1</i>	<i>NOD5</i>	KAL1287894	Chr 19
<i>NOD1</i>	<i>NLRC1; CARD4</i>	KAL1288359	Chr 12
<i>NOD2</i>	<i>NLRC2; CARD15</i>	KAL1287942	Chr 15
<i>NLRC3</i>	<i>NOD3</i>	KAL1287486	Chr 3
<i>NLRC5</i>	<i>NOD4</i>	KAL1287941	Chr 15
C-type receptors			
	C-type lectin domain family		
<i>CD209</i>	<i>CLEC4L; DC-SIGN</i>	KAL1286854	Chr 8
<i>OLR1</i>	<i>CLEC8A</i>	KAL1288126	Chr 3
<i>CLEC1A</i>	<i>CLEC1</i>	KAL1288098	Chr 3
<i>CLEC1B</i>	<i>CLEC2; CLEC2B</i>	KAL1288099	Chr 3
<i>CLEC4D</i>	<i>CD368; CLEC6; CLECSF8; DECTIN-3</i>	KAL1288101	Chr 3
<i>CLEC4E</i>	<i>CLECSF9; MINCLE</i>	KAL1288102	Chr 3
<i>CLEC4G</i>	<i>DTTR431</i>	KAL1286857	Chr 8
<i>CLEC5A</i>	<i>CLECSF5</i>	KAL1287060	Chr 12
<i>CLEC7A</i>	<i>CLECSF12; DECTIN-1</i>	KAL1288104	Chr 3
<i>CLEC9A</i>	<i>CD370; DNGR-1</i>	KAL1288105	Chr 3
<i>CLEC12A</i>	<i>CD371</i>	KAL1288097	Chr 3
Rig-like receptors			
	Retinoic acid-inducible gene-I-like receptors		
<i>RIGI</i>	<i>DDX58; DExD/H-box helicase 58; RLR1</i>	KAL1287283	Chr 10
<i>DHX58</i>	<i>RLR-3; LPG2; RLR3</i>	KAL1286474	Chr 20
<i>MAVS</i>	<i>CARDIF; IPS-1; IPS1; VISA</i>	KAL1287541	Chr 18
<i>IFIH1</i>	<i>MDA5; IDDM19; RLR; 2IDDM19</i>	KAL1287835	Chr 2
<i>ZBP1</i>	<i>Z-DNA binding protein 1</i>	KAL1287555	Chr 18
Adaptors			
<i>MYD88</i>	<i>IMD68; MYD88D</i>	KAL1287811	Chr 6
<i>STING1</i>	<i>ERIS</i>	KAL1288231	Chr 8
<i>TICAM1</i>	<i>TRIF; MyD88-3</i>	KAL1287077	Chr 8
<i>TIRAP</i>	<i>MYD88-2</i>	KAL1287636	Chr 5
<i>UNC93B1</i>	<i>UNC93; TLR signaling regulator</i>	KAL1286931	Chr 5
<i>TRAF1</i>	<i>TNF receptor associated factor 1</i>	KAL1288038	Chr 10
<i>TRAF2</i>	<i>TNF receptor associated factor 2</i>	KAL1288514	Chr 4
<i>TRAF3</i>	<i>TNF receptor associated factor 3</i>	KAL1286829	Chr 16
<i>TRAF4</i>	<i>TNF receptor associated factor 4</i>	KAL1287767	Chr 20
<i>TRAF5</i>	<i>TNF receptor associated factor 5</i>	KAL1287044	Chr 11
<i>TRAF6</i>	<i>TNF receptor associated factor 6</i>	KAL1287434	Chr 19
Continued			

Gene symbol	Alternative name/symbol	Accession	Muskox chromosome
TRAF7	TNF receptor associated factor 7	KAL1286545	Chr 3
Janus and other kinases			
JAK1	Janus kinase 1	KAL1286939	Chr 9
JAK2	Janus kinase 2	KAL1287446	Chr 10
JAK3	Janus kinase 3	KAL1286737	Chr 8
TYK2	Janus kinase JTK1	KAL1286878	Chr 8
IRAK1BP1	AIP70; SIMPL	KAL1286883	Chr 14
IRAK1	IRAK-1	KAL1286446	Chr X
IRAK2	IRAK-2	KAL1287806	Chr 6
IRAK3	IRAKM	KAL1288117	Chr 3
IRAK4	IRAK-4	KAL1286989	Chr 3
RIPK1	RIP-1	KAL1287230	Chr 2
RIPK2	RIP-2; CARD3	KAL1288301	Chr 21
STATs	Signal transducer/activation of transcription		
STAT1	CANDEF7; IMD31A; IMD31B; IMD31C; ISGF-3; STAT91	KAL1287623	Chr 2
STAT2	IMD44; ISGF-3; P113; PTORCH3; STAT113	KAL1288135	Chr 3
STAT3	ADMIO; ADMIO1; APRF; HIES	KAL1286482	Chr 20
STAT4	DPMC; SLEB11	KAL1287624	Chr 2
STAT5A	MGF	KAL1286483	Chr 20
STAT5B	GHISID2	KAL1286484	Chr 20
STAT6	D12S1644; HIES6; IL-4-STAT; STAT6B; STAT6C	KAL1288136	Chr 3
Interferon pathway			
IFNA1	IFN-alpha 1	KAL1286370	Chr 10
IFNA2	IFN-alpha 2	KAL1286361	Chr 10
IFNA3	IFN-alpha 3	KAL1286362	Chr 10
IFNA4	IFN-alpha 4	KAL1286377	Chr 10
IFNB1	IFN-beta 1	KAL1287444	Chr 10
IFNE	IFN-epsilon	KAL1287280	Chr 10
IFNG	IFN-gamma	KAL1288112	Chr 3
IFNK	IFN-kappa	KAL1287281	Chr 10
IFNT1	IFN-tau 1	KAL1286371	Chr 10
IFNT2	IFN-tau 2	KAL1286372	Chr 10
IFNT3	IFN-tau 3	KAL1286373	Chr 10
IFNT4	IFN-tau 4	KAL1286374	Chr 10
IFNT5	IFN-tau 5	KAL1286375	Chr 10
IFNT6	IFN-tau 6	KAL1286358	Chr 10
IFNT7	IFN-tau 7	KAL1286359	Chr 10
IFNT8	IFN-tau 8	KAL1286355	Chr 10
IFNT9	IFN-tau 9	KAL1286356	Chr 10
IFNT10	IFN-tau 10	KAL1286363	Chr 10
IFNT11	IFN-tau 11	KAL1286364	Chr 10
IFNT12	IFN-tau 12	KAL1286365	Chr 10
IFNT13	IFN-tau 13	KAL1286378	Chr 10
LTA	Interferon B	KAL1286623	Chr 2
IFNAR1	Interferon-alpha R1	KAL1286550	Chr 1
IFNAR2	Interferon-alpha R2	KAL1286551	Chr 1
IFNGR1	CD119; IFN-gamma R1	KAL1288382	Chr 14
IFNGR2	IFN-gamma R2	KAL1286552	Chr 1
IFNLR1	IFN-lambda R1; IL-28 R1; IL-28 RA	KAL1287147	Chr 2
IL10RB	CD210B; CRFB4; IL-10R2; IL-10 RB; IL-28R	KAL1286553	Chr 1
IRF1	Interferon regulatory factor 1	KAL1287074	Chr 8
IRF2BP1	Interferon regulatory factor 2 BP1	KAL1286769	Chr 15
IRF2BP2	Interferon regulatory factor 2 BP2	KAL1287784	Chr 6
IRF2BPL	Interferon regulatory factor 2 BPL	KAL1287051	Chr 13
IRF2	Interferon regulatory factor 2	KAL1287409	Chr 4
IRF3	Interferon regulatory factor 3	KAL1286770	Chr 15
Continued			

Gene symbol	Alternative name/symbol	Accession	Muskox chromosome
<i>IRF4</i>	<i>Interferon regulatory factor 4</i>	KAL1286770	Chr 2
<i>IRF5</i>	<i>Interferon regulatory factor 5</i>	KAL1288356	Chr 12
<i>IRF6</i>	<i>Interferon regulatory factor 6</i>	KAL1287038	Chr 11
<i>IRF7</i>	<i>Interferon regulatory factor 7</i>	KAL1286920	Chr 5
<i>IRF8</i>	<i>Interferon regulatory factor 8</i>	KAL1287935	Chr 15
<i>IRF9</i>	<i>Interferon regulatory factor 9</i>	KAL1287250	Chr 13
<i>MAVS</i>	<i>IPSI; VISA; CARDIF</i>	KAL1287541	Chr 18
<i>IFRD1</i>	<i>Interferon dev regulator 1</i>	KAL1288353	Chr 12

Table 3. Selected genes of the innate immunity pathways. Muskox orthologs of the pattern recognition receptors, interferons and intermediaries. See Supplementary Table S6 for the mitogen-activated protein kinases (MAPs), beta-defensins, interleukins, tumor necrosis factors (TNFs), and the chemokines.

professional antigen-presenting cells such as macrophages, dendritic and B cells, and present exogenous antigens to CD4+ helper T-cells. The Class III region contains genes encoding immune regulatory molecules, such as tumor necrosis factor (TNF), TNF family members, heat shock proteins, and complement factors.

Figure 5A compares the chromosomal organizations of the MHC in human, muskox, sheep, goat and cow. The chromosome arms in the five species share extensive synteny along their lengths, with the cow less so as indicated by the shorter lengths of the synteny blocks. Overall, muskox MHC contains a similar number of principal genes and are organized similarly as in the other species. Figure 5B is a schematic of the functional MHC genes we annotated in the muskox. However, when compared to human MHC, a striking finding is that in muskox, sheep, and cow, the MHC Class II region is interrupted, where a large segment comprising the distal portion of the MHC Class II region through to the Extended MHC Class II region is inverted in orientation, resulting in the distal portion of the MHC Class II region being translocated 17.4 Mb away from the rest of the complex in the direction of the centromere. Otherwise, we observed a high degree of synteny in the overall organization of muskox chromosome 2 with human chromosome 6 in GRCh38.p14.

In the muskox MHC Class I region we identified five Class I leukocyte antigen presenting genes, *HLA-L1*, *HLA-L2*, *HLA-L5*, *HLA-L6* and *HLA-L7*. Two other genes, *HLA-L3** and *HLA-L4**, were predicted to be pseudogenes based on the presence of in-frame termination codons, which were verified using Illumina short reads. *MICA* and *MICB* in human are adjacent stress-inducible genes related to the MHC Class I leukocyte antigen presenting genes¹¹⁹, but *MICA* and *MICB* use *KLRK1* (*NKG2D*) as the co-receptor subunit instead of *B2M*¹²⁰. *KLRK1* (*CD314*, *NKG2-D*) and *B2M* co-receptor genes are located on muskox chromosomes 3 and 13, respectively. Interesting, we observed a tandem duplication of *B2M* on muskox chromosome 13. The duplicated gene appears to be functional and was assigned the symbol *B2ML*. In the muskox, we only find *MICB*. Since the two genes might have overlapping functions, the absence of *MICA* in the muskox might be mitigated by redundancy; however, its absence could impair the muskox's response to certain stressors.

In the muskox MHC Class II region, we identified thirteen Class II leukocyte antigen presenting genes: *HLA-DRA-L1*, *HLA-DRB-L1*, *HLA-DRB-L2*, *HLA-DPA-L1*, *HLA-DQB-L1*, *HLA-DQA-L1*, *HLA-DQB-L2*, *HLA-DOA-L1*, *HLA-DMA-L1*, *HLA-DMB-L1*, *HLA-DOB-L1*, *HLA-DQA-L2*, and *HLA-DRB-L3*. As denoted by the A and B designations in their assigned gene symbols, six genes encode potential alpha-receptor subunits and seven genes encode potential beta-receptor subunits.

The MHC Class III genes are generally heterogeneous in structure. None present antigens to T-cells, but most are involved in some other aspects of immune regulation or inflammation^{121,122}. We identified and annotated 61 genes in the muskox MHC Class III region, including members of the TNF family^{123,124}, the Lymphocyte Antigen-6 family members¹²⁵, and the members of the complement cascade.¹²⁶ Compared to human, we observed a tandem duplication of the *C4* gene in the muskox. The second gene was assigned the symbol *C4L*. Including the 121 genes in the extended Class I and Class II regions, the final number of annotated genes in the muskox extended MHC is 245. Exon-level annotations for these genes are provided in Supplementary Table S4.

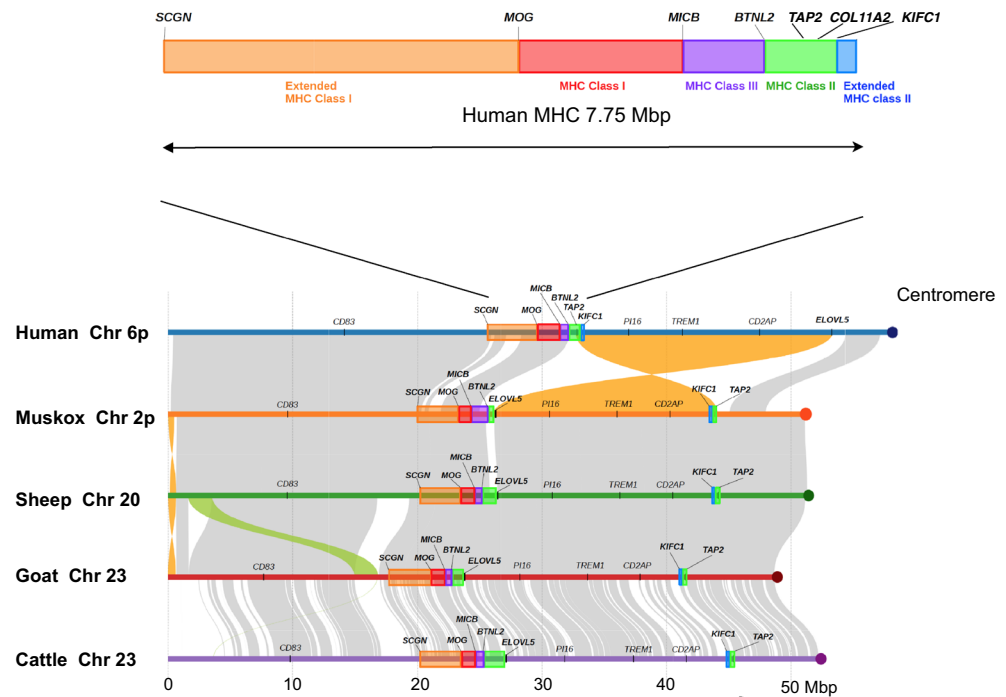
Two CD antigens are encoded by the MHC, *CD167* (*DDR1*) and *CD337* (*NCR3*). CD (Cluster of Differentiation) molecules or antigens are cell surface markers, many of which are receptors or ligands involved in activation or inhibitory pathways.¹²⁷ As such, many CD antigens are biomarkers of, or are participants in, infectious disease. A list of the several hundred annotated CD antigens in the muskox can be found in Supplementary Table S7.

Genetic diversity of muskox MHC

We have shown the muskox MHC region is similar in structure and overall genomic organization to the other Bovidae, indicating no potential deficiencies in muskox due to gross deletions or arrangements in the MHC region. The number of MHC Class I and Class II genes in the muskox is also similar to many other mammals, suggest that at face value, the muskox is immune-competent in interactions with T-cells. Genetic diversity within a population is viewed to be important in evolutionary ecology and conservation biology^{128–131}. We thus examined the genetic (polymorphic) diversity of the MHC genes in muskox.

In an early study of the muskox, no genetic diversity was found using an exon 2 probe that spanned the MHC *DRB* locus (corresponding to *HLA-DRB-L1* in the present study)¹³². This early result points to a lack of MHC diversity in muskox. However, the study was constrained by the small size of the study cohort, and by the limited sensitivity of amplicon sequencing that targeted only a single exon. With the complete characterization

5A



5B

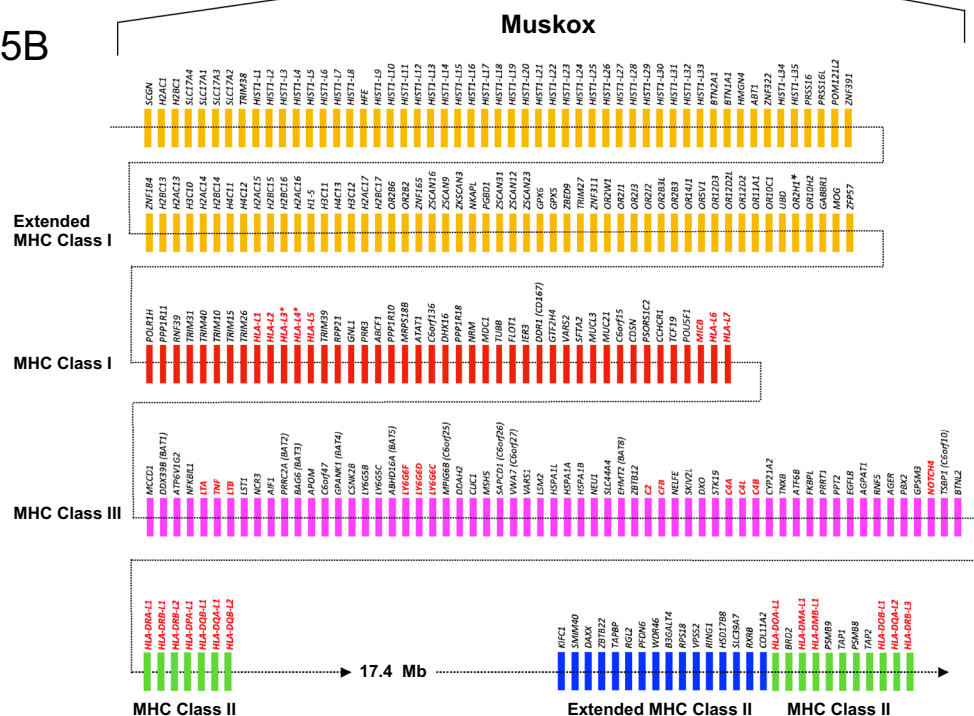


Figure 5. Major histocompatibility complex (MHC) on muskox chromosome 2. (A) Upper panel: Organization of the human MHC. Landmark genes serving as region boundaries are highlighted. Region lengths are drawn to scale. Lower panel: Comparison of the chromosome arm containing the MHC between human, muskox, sheep, goat and cattle was visualized using plotsr.⁷⁰ Grey-shaded ribbons indicate regions of broad synteny, orange ribbons represent inversions, and green ribbons indicate translocations. When compared to human, the MHC Class II region in muskox, sheep, and cow is interrupted, where a segment is inverted in orientation from the rest of the cluster as well as being translocated 17.4 Mb toward the direction of the centromere. There is broad synteny across the chromosome arm among the Bovidae species, with cattle being the outgroup of the four. (B) Gene map of the 245 genes in the extended MHC in muskox. Class I and Class II leukocyte antigen presenting genes are in red type. Asterisk (*) denotes a pseudogene.

of muskox MHC combined with the use of whole genome re-sequencing and mapping-based variant calling methods^{133–135} we could now overcome the sensitivity and throughput issues associated with previous amplicon sequencing attempts. To test the feasibility of this approach, we took advantage of whole-genome short-reads from a recent population survey of the muskox³¹ to provide an initial assessment of the genetic diversity of all the genes in the MHC of the muskox. These reads are available through NCBI SRA and are comprised of sets of Illumina reads with an average 14× genome coverage of individuals from 18 distinct populations of muskoxen across their native range. For our study, we downloaded reads from one representative specimen from each population. We included reads from a 21,100-year-old individual from Siberia from which inferences could be made of the muskox's demographic history.

Reads from individual muskox samples were mapped onto our newly constructed reference assembly, O.moschatus_RAN27-v1.0, and variant calls were made from the mapped reads against this reference genome. Figure 6A presents the variant calls made for each muskox specimen across the coding regions of genes in the MHC regions (see Fig. 5B for a schematic of the gene organization in the muskox). All variant calls were found to be single base substitutions. We did not observe any insertions or deletions among the calls. The specimen from Banks Island, ROMMRAN27, from which DNA was used to assemble the reference genome is denoted by an asterisk (*). Variant calls made for ROMMRAN27 were heterozygous Single Nucleotide Polymorphisms (SNPs) present in our sample. The two-letter prefix for each sample is a geographic identifier denoting the origin of the 18 muskox specimen shown in Fig. 6B. The prehistoric sample from Siberia is designated SIB.

Variant calls spanning the coding regions of the 245 genes in the MHC across all 20 muskox samples are tabulated in Supplementary Table S5. It is likely the calls presented in Supplementary Table S5 are an under-representation, since we adopt stringent mapping criteria in an effort to minimize false positive variant calls. The sensitivity of a short-read mapping approach to call variants is highest for the regions of the genome that are free

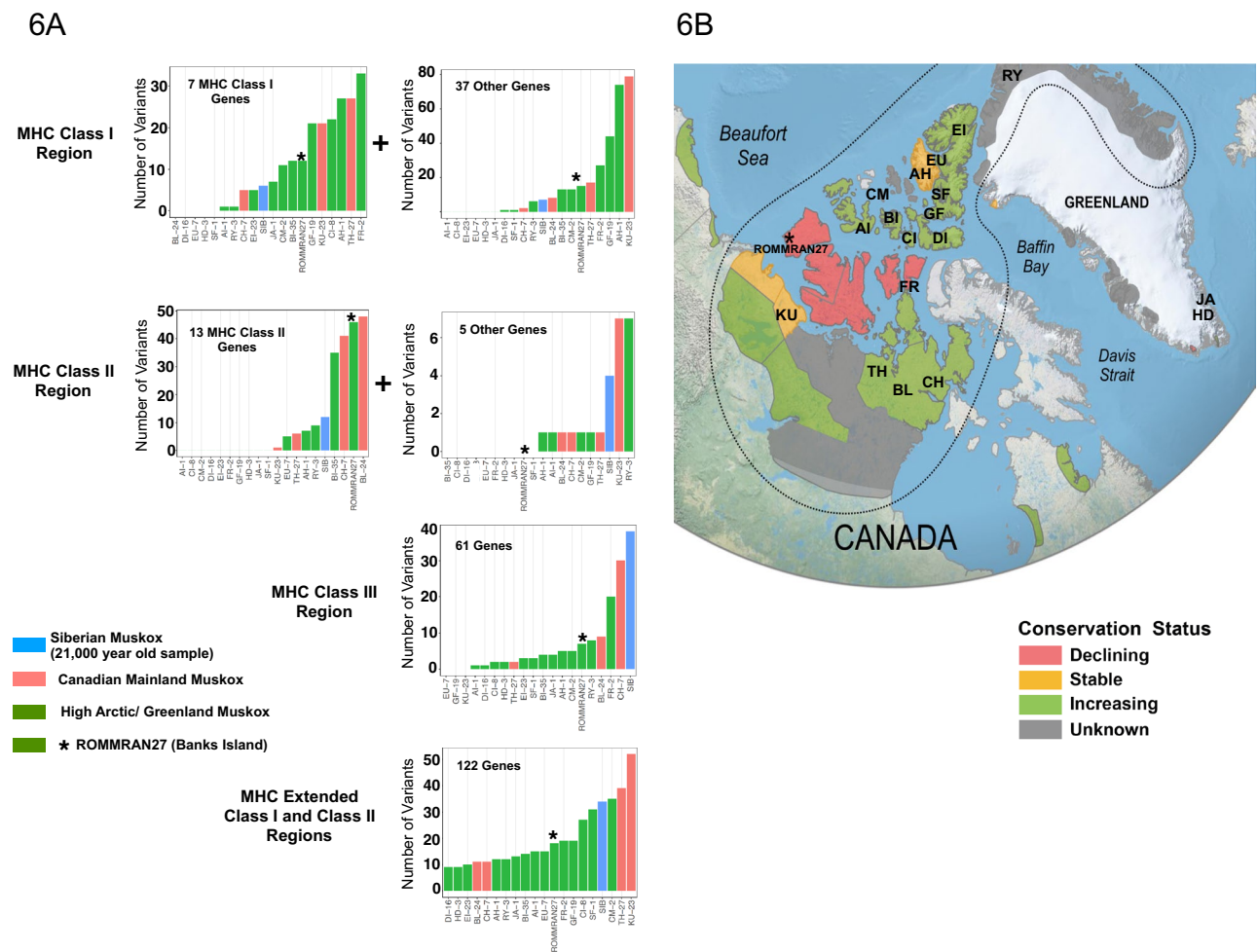


Figure 6. Gene diversity in the MHC region of muskoxen across various geographic locations. (A) Illumina reads of 20 muskoxen, including a 21,000-year-old individual from Siberia, were used to tabulate variants in the coding region of the 245 genes in the MHC (see Material and Methods). ROMMRAN27, from which we have assembled the genome is denoted by an asterisk (*). The two-letter prefix is a geographic abbreviation. The prehistoric sample from Siberia is designated SIB. (B) Geographic origins of the 19 muskox samples in northern Canada and Greenland. The Siberian sample is not depicted.

of major mapping impediments, such as the presence of highly repetitive sequences and other features associated with genomic regions of low complexity. Historically, the MHC region was renowned for the presence of such impediments, which could not be fully resolved until the recent advent of using long reads with better mapping specificity for variant calling^{136,137}. However, until the implementation of using long reads to make variant calls for the muskox, the present use of shortreads might result in significant false-negative calls, but it would still provide a useful first look at the genetic diversity in this region of the genome.

From the initial set of variants depicted in Fig. 6A, there appears to be no clear and consistent regional pattern, despite that some populations have been through human-caused bottlenecks, while others have been through climate-caused ones. Quite possibly, low population sampling has hampered a more complete assessment of the species. For example, there are no representations from Victoria Island, and Banks Island is represented only by our reference specimen, which seemed to be neither particularly invariant nor have a high number of variants compared to those from other Arctic islands. Interestingly, the prehistoric sample from Siberia, SIB, does not consistently have the highest number of variants across all six groups of genes.

The MHC Class I and Class II genes function to present processed antigens to CD8+ T-cells and CD4+ helper T-cells, respectively, and are critical for mounting an immune response. Table 4 summarizes variant calls for the seven MHC Class I genes, and the thirteen MHC Class II genes across the twenty muskox samples. Column 3 depicts the number of sets of unique combinations of single base variants (variant profile) for each gene across the 19 modern muskox specimens. However, in the absence of phasing information, such as long reads, RNA-seq reads, or other genetic information, to establish the two haplotypes, we cannot reconstruct the two alleles from each variant profile with certainty. Nevertheless, these profiles provide a direct indication, albeit a lower limit, of the coding diversity of the muskox MHC.

Even though the small sample size precludes an in-depth analysis for the muskox, some general observations can be made. For instance, *HLA-L1* and *HLA-L2* appear to be particularly polymorphic across many samples. Mainland muskoxen, with the exception of the Helena Island sample (sample BI), and our Banks Island sample (ROMMRAN27), are generally more diverse than samples from the other Arctic islands, especially those from Eastern Greenland (samples JA and HD). Interestingly, Class I pseudogenes, *HLA-L3** and *HLA-L4**, accumulated many variants principally across the Arctic Island samples, none of which however restore the coding sequence to create a functional gene. Finally, we observed that *HLA-L5*, *HLA-L6*, *HLA-DQA-L1*, *HLA-DQ-L1*, *HLA-DMB-L1*, *HLA-DQB-L1*, and *HLA-DRB-L3* are markedly free of variants across all samples, when compared to our reference specimen, ROMMRAN27.

Genetic diversity in the MHC is believed to impact wildlife health and conservation^{128–131}. In contrast to early work pointing to low or no genetic diversity in the muskox MHC¹³², our tabulation of variants across multiple specimens showed a significant diversity across the MHC. Although an extensive direct gene-to-gene comparison could not be made, at least qualitatively in selective regions that could be compared, the diversity of the muskox MHC appears to be similar to the level reported for the reindeer, another Arctic ruminant under threat¹³⁸. While the muskox has been shown to have the lowest genome-wide heterozygosity reported in an ungulate³¹ (Fig. 3), this lack of genetic diversity is not necessarily extended to the MHC region. This finding, along with the muskox having no overt signs of inbreeding depression in native populations³¹, is consistent with the view of Teixeira and Huber¹³⁹ that argues against the perceived importance of neutral genetic diversity over functional genetic diversity for the conservation of wild populations and species.

Conservation management

The present study provides genomics tools that will advance muskox conservation and management in at least three ways. First, the creation of a high-quality reference genome provides a resource to develop genetic markers that could be used estimate gene flow, effective population size, the timing of extirpation, and to monitor progress of future recolonization efforts. As shown by our initial survey of diversity of the MHC in the muskox, whole genome reference-assisted re-sequencing is a sensitive and powerful tool to identify markers. In the immediate future, the use of re-sequencing is likely to be used in a discovery phase, where selective cohorts are re-sequenced to generate a comprehensive set of informative sequence markers, after which these markers would be interrogated on lower cost and higher throughput genotyping panels that are constructed for specific applications.

Second, the genomic tools especially the annotated genes will contribute to management and conservation. Muskox population structure and connectivity can now be assessed from both a neutral and adaptive basis, as currently muskox populations are delimited based on hunting areas. The annotated genes can contribute to measure connectivity between populations and for understanding the dynamics of expanding muskox distribution on the mainland or options for translocations to rescue depleted populations. Regional populations can be sampled to assess for local adaptive loci identification from the annotated genes, and regional planning undertaken to conserve local adaptations. However, there are still considerable uncertainties in the present muskox population numbers and trends, thereby limiting the usefulness of the new genomics tools. In their recent review¹⁹, Cuyler and colleagues have acknowledged these shortcomings and have proposed increased frequency of muskox surveys, standardizing monitoring protocols, and the inclusion other informative metrics.

Finally, the remarkable super-fine under-wool of the muskox has garnered much attention for its economic potential, and has driven some early attempts to domesticate the muskox to improve yield and ease of qiviut harvest^{12–15}. While the role of domestication as a driver for conservation and management of endangered species is controversial^{140–142}, it could be one facet of an integrated adaptive solution that is locally implemented. However, domestication is a long and difficult process, but the genetics of domestication are relatively well understood for livestock^{143–148}. Although muskox domestication has been attempted, it has not yet succeeded^{149–151}. Efforts were too small-scale and brief to allow the selection for particular traits, although the muskox became tame. As a first step to support a possible future marker assisted selection scheme^{152–154} for the muskox, we provide

Category	Gene	Unique combinations of variants in 19 modern specimens	Siberia	Canada mainland				Canada Arctic Islands											Greenland			
			SIB	BL-24	CH-7	KU-23	TH-27	*ROM	AI-1	AH-1	BI-35	CI-8	CM-2	DI-16	EI-23	EU-7	FR-2	GF-19	SF-1	HD-3	JA-1	RY-3
MHC Class I	HLA-L1	9			5	21	8	3		17	3	12	3				12	11				
MHC Class I	HLA-L2	4	3				13			1		1					9					
MHC Class I	HLA-L3*	8					5	6			7		6				12	1			7	
MHC Class I	HLA-L4*	5	2				1	2			2		2		5							
MHC Class I	HLA-L5	1																				
MHC Class I	HLA-L6	1	1																			
MHC Class I	HLA-L7	2						1	1												1	
MHC Class II	HLA-DRA-L1	3			2	2		1		2											2	
MHC Class II	HLA-DRB-L1	7			7	6		5	7		5					5					6	
MHC Class II	HLA-DRB-L2	2	3			1		1														
MHC Class II	HLA-DPA-L1	2	1		2	2		2			2											
MHC Class II	HLA-DQB-L1	3	3		26	2		26			26											
MHC Class II	HLA-DQA-L1	1	3																			
MHC Class II	HLA-DQB-L2	3	1		9	9		9			7											
MHC Class II	HLA-DOA-L1	2																			1	
MHC Class II	HLA-DMA-L1	2			1	1		1														
MHC Class II	HLA-DMB-L1	1																				
MHC Class II	HLA-DOB-L1	1																				
MHC Class II	HLA-DQA-L2	2			1		1															
MHC Class II	HLA-DRB-L3	1	1																			

Table 4. Number of variants in the MHC Class I and Class II genes across 20 muskox samples of different geographic locations. Variant calls for the seven MHC Class I genes and thirteen MHC Class II genes were performed in as described in the Materials and Methods. The assembled sample ROMMRAN27 (*ROM) is in bold. Pseudogenes are designated by an asterisk after the gene name. Column 3 depicts the number of sets of unique combinations of single base variants (variant profile) for each gene across the 19 modern muskox specimens.

exon-level annotation of genes that have contributed to livestock domestication (Supplementary Table S8) and desirable commercial traits such as meat and dairy production, wool quality, prolificacy, and heat tolerance^{155–159} (Supplementary Table S9).

Given the high market price of qiviut, the genetic signatures for qiviut may become useful in the future, especially if the genetics of its production differ regionally with different muskox populations. To this end, Fig. 7 and Supplementary Table S10 show the organization and annotation of the principal keratin and keratin-associated protein gene clusters in muskox. Keratins and the keratin-associated proteins are conserved across multiple species with family members encoding the principal structural components of hair and wool^{160–163}. Importantly, variants of these genes have been shown to associate with wool quality, and are potential selection markers for livestock improvements^{164–167}.

Arctic conservation management can be complex with intersecting and at times conflicting environmental, cultural and socioeconomic issues^{90,168,169}. However, co-management (i.e., joint management between Indigenous people and government) and adaptive management^{169,170}, where multiple integrated solutions are differentially implemented locally, can help navigate many of those complexities. In particular, a co-management regime gives a clear voice to local communities in the balancing of cultural traditions, and costs incurred by environmental protection can be offset against the potential returns from developing economic values of muskoxen, including meat and leather production, sport hunting, tourism, and production of qiviut.

In conclusion, we report a chromosomal-level assembly of a muskox from Banks Island in the Canadian Arctic Archipelago with a view toward genomics-based approaches for the management and conservation of the muskox, with relevant benefits to Inuit communities and muskox conservation.

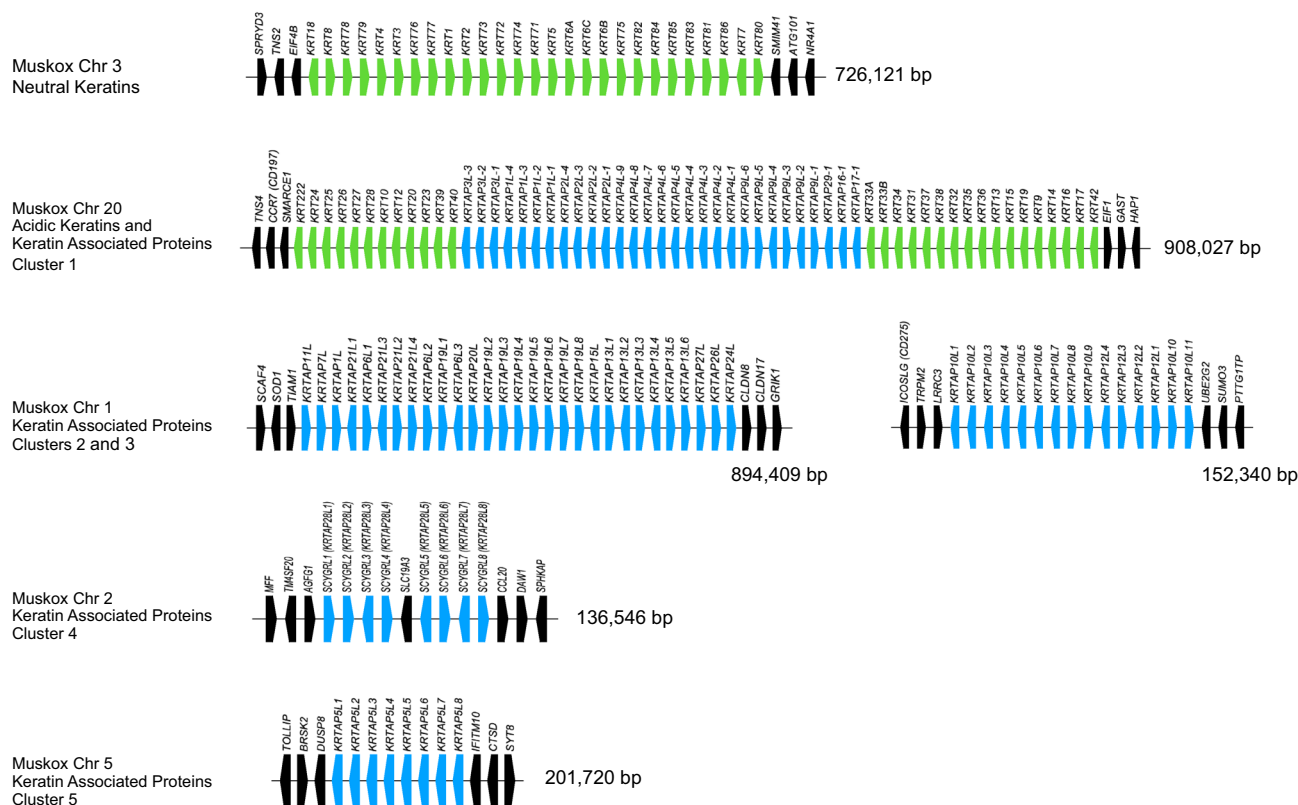


Figure 7. Schematic of Muskox keratin and keratin-associated protein gene clusters. Gene clusters are depicted in the same orientation as in the assembly. Genes flanking the clusters are depicted by black boxes. Keratin genes are depicted by green boxes. Keratin-associated protein genes are depicted by blue boxes. The direction of transcription is indicated by the box outdent. Gene annotations can be found in Supplementary Table S10.

Data availability

Sequences described for the muskox have been submitted to GenBank and other public archives and repositories. This study has been assigned BioProject number PRJNA1071053 to facilitate data dissemination. Accessions for muskox genes depicted in the manuscript are listed in the relevant Supplementary Tables.

Received: 27 February 2024; Accepted: 9 July 2024

Published online: 16 September 2024

References

- Lent, P. C. *Ovibos moschatus*. *Mamm. Species* **302**, 1–9 (1988).
- Li, M. *et al.* Convergent molecular evolution of thermogenesis and circadian rhythm in arctic ruminants. *Proc. R. Soc. B* **290**, 20230538 (2023).
- Allen, J. A. Ontogenetic and other variations in muskoxen, with a systematic review of the muskox group, recent and extinct. *Mem. Am. Mus. Nat. Hist.* **1**(6), 105–226 (1913).
- Glover, R. The muskox (*Ovibos moschatus*). *Oryx* **2**(2), 76–86 (1953).
- Zimmermann, F. A. W. *Geographische Geschichte des Menschen und der Vierfüßigen Thiere* Vol. 2, 86–88 (Weygand, 1780).
- de Blainville, H. H. D. Sur plusieurs espèces d'animaux mammifères de l'ordre de ruminants. *Bull. Sci. Soc. Philomath. Paris* **1816**(5), 73–82 (1816).
- Tener, J. S. *Muskoxen in Canada, a Biological and Taxonomic Review*. Canadian Wildlife Service Monograph No. 2 166 (Queen's Printer, 1965).
- Wilson, D. E. & Reeder, D. M. (eds) *Mammal Species of the World. A Taxonomic and Geographic Reference* 3rd edn. (Johns Hopkins Press, 2005).
- Hansen, C. C. R. *et al.* The muskox lost a substantial part of its genetic diversity on its long road to Greenland. *Curr. Biol.* **28**, 4022–4028 (2018).
- Kutz, S. *et al.* Muskox health ecology symposium 2016: gathering to share knowledge on Umingmak in a time of rapid change. *Arctic* **70**(2), 225–236 (2017).
- Tomaselli, M. *et al.* Iqalukutiaq voices: Local perspectives about the importance of muskoxen, contemporary and traditional use and practices. *Arctic* **71**(1), 1–4 (2018).
- Barr, W. The commercial trade in muskox hides in the Northwest Territories in 1860–1916. *Rangifer* **11**(2), 81–81 (1991).
- Von Bergen, W. Muskox wool and its possibilities as a new textile fiber. *Melliand Text. Mon.* **3**(10), 844–846 (1932).
- Rowell, J. E. *et al.* Fiber characteristics of qiviut and guard hair from wild muskoxen (*Ovibos moschatus*). *J. Anim. Sci.* **79**, 1670–1674 (2001).
- Helfferich, D. The muskox: Woolly and warm in a northern fiber industry. *Agroboreal* **39**(1), 29–37 (2007).

16. Mosbacher, J. B. *et al.* Quantifying muskox plant biomass removal and spatial relocation of nitrogen in a high arctic tundra ecosystem. *Arct. Antarct. Alp. Res.* **48**(2), 229–240 (2016).
17. Mosbacher, J. B. *et al.* Mushoxen modify plant abundance, phenology, and nitrogen dynamics in a high arctic fen. *Ecosystems* **22**(5), 1095–1107 (2019).
18. Falk, J. M. *et al.* Large herbivore grazing affects the vegetation structure and greenhouse gas balance in a high arctic mire. *Envir. Res. Lett.* **10**(4), 45001 (2015).
19. Cuyler, C. *et al.* Muskox status, recent variation, and uncertain future. *Ambio* **49**, 805–819 (2020).
20. Gunn, A. & Forchhammer, M. *Ovibos moschatus*. *The IUCN Red List of Threatened Species* (2022).
21. Gunn, A. *et al.* The history, status and management of the muskoxen on Banks Island. *Arctic* **44**(3), 188–195 (1991).
22. Kutz, S. *et al.* Cross-Canada disease report—*Erysipelothrix rhusiopathiae* associated with recent widespread muskox mortalities in the Canadian Arctic. *Can. Vet. J.* **56**, 560–563 (2015).
23. Forde, T. L. *et al.* Bacterial genomics reveal the complex epidemiology of an emerging pathogen in Arctic and Boreal Ungulates. *Front. Microbiol.* **7**(11), 1759 (2016).
24. Kutz, S. J. *et al.* Invasion, establishment, and range expansion of two parasitic nematodes in the Canadian Arctic. *Glob. Change Biol.* **19**, 3254–3262 (2013).
25. Okulewicz, A. The impact of global climatic change on the spread of nematodes. *Ann. Parasitol.* **63**(1), 15–20 (2017).
26. Kafle, P. *et al.* Range expansion of muskox lungworms track rapid arctic warming: Implications of geographic colonization under climate forcing. *Sci. Rep.* **10**, 17323 (2020).
27. Tomaselli, M. *et al.* A transdisciplinary approach to Brucella in muskoxen of the Western Canadian Arctic 1989–2016. *EcoHealth* **16**, 488–501 (2019).
28. Holm, L.-E. *et al.* Low genetic variation in muskoxen (*Ovibos moschatus*) from western greenland using microsatellites. *Mol. Ecol.* **8**, 675–679 (1999).
29. MacPhee, R. D. E. *et al.* Late quaternary loss of genetic diversity in muskox (*Ovibos*). *BMC Evol. Biol.* **5**, 49 (2005).
30. Groves, C. P. Interspecific variation in mitochondrial DNA of muskoxen, based on control-region sequences. *Can. J. Zool.* **75**(4), 568–575 (2011).
31. Pečnerová, P. *et al.* Population genomics of the muskox's resilience in the near absence of genetic variation. *Mol. Ecol.* **33**(2), e17205 (2024).
32. Prewer, E. *et al.* Already at the bottom? Demographic declines are unlikely further to undermine genetic diversity of a large Arctic ungulate: Muskox, *Ovibos moschatus* (Artiodactyla: Bovidae). *Biol. J. Linn. Soc.* **129**, 459–469 (2020).
33. Dussex, N. *et al.* Population genomics of the critically endangered kākāpō. *Cell Genom.* **1**(1), 100002 (2021).
34. Dussex, N. *et al.* Adaptation to the high-Arctic island environment despite long-term reduced genetic variation in Svalbard reindeer. *iScience* **26**, 107811 (2023).
35. IPCC Summary for Policymakers. In: *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Core Writing Team, H. Lee & J. Romero) 1–34 (IPCC). <https://doi.org/10.59327/IPCC/AR6-9789291691647.001>.
36. Atwood, T. B. Herbivores at the highest risk of extinction among, bird, and reptiles. *Sci. Adv.* **6**, eabb9458 (2020).
37. Prewer, E. *et al.* Draft genome assembly of an iconic species: Muskox (*Ovibos moschatus*). *Genes* **13**, 809 (2022).
38. Warburton, P. E. & Sebra, R. P. Long-read DNA sequencing: Recent advances and remaining challenges. *Annu. Rev. Genom. Hum. Genet.* **24**, 109–132 (2023).
39. Aasen, E. & Medrano, J. F. Amplification of the ZFY and ZFX genes for sex identification in humans, cattle, sheep and goats. *Bio/Technology* **8**, 1279–1281 (1990).
40. Simpson, J. T. Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* **30**(9), 1228–1235 (2014).
41. Kolmogorov, M. *et al.* Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**(5), 540–546 (2019).
42. Proskuryakova, A. A. *et al.* Comparative studies of karyotypes in Cervidae family. *Cytogenet. Genome Res.* **162**, 312–322 (2022).
43. Proskuryakova, A. A. *et al.* Comparative chromosome mapping of Musk ox and X chromosome among some Bovidae species. *Genes* **10**, 857 (2019).
44. Rubes, I. *et al.* Comparative molecular cytogenetics in Cetartiodactyla. *Cytogenet. Genome Res.* **137**(2–4), 194–207 (2012).
45. Tietz, W. J. & Teal, J. J. Chromosome number of the muskox. *Can. J. Zool.* **45**, 235–237 (1967).
46. Heck, H. D. *et al.* Chromosome study of members of the subfamily Caprinae and Bovinae, family Bovidae, the muskox, ibex, aoudad, Congo buffalo, and gaur. *Säugetierk. Mitt.* **33**, 172–179 (1968).
47. Gallagher, D. S. Jr. & Womack, J. E. Chromosome conservation in the Bovidae. *J. Hered.* **83**, 287–298 (1991).
48. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
49. Tarailo, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **4**, 10 (2009).
50. Rhie, A. *et al.* Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
51. Zhang, H. *et al.* A comprehensive evaluation of long read error correction methods. *BMC Genom.* **21**(Suppl 6), 889 (2020).
52. Longo, M. S. *et al.* Abundant human DNA contamination identified in non-primate genome databases. *PLoS ONE* **6**, e16410 (2011).
53. Simão, F. A. *et al.* BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**(19), 3210–3212 (2015).
54. Ewing, B. & Green, P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
55. Van Coeverden de Groot, P. J. V. C. *Conservation Genetic Implications of Microsatellite in the Muskox Ovibos moschatus: The Effect of Refugial Isolation and the Arctic Ocean on Genetic Structure*. PhD thesis, Queen's University, Canada (2001).
56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **15**, 1754–1760 (2009).
57. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**(9), 1297–1303 (2010).
58. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**(1110), 1 (2013).
59. Jackman, D. *et al.* ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* **27**(5), 768–777 (2017).
60. Lok, S. *et al.* Chromosomal-level reference genome assembly of the North American wolverine (*Gulo gulo luscus*): A resource for conservation genomics. *Genes Genomes Genet.* **12**(8), jkac138 (2022).
61. Chain, P. S. G. *et al.* Genome project standards in a new era of sequencing. *Science* **326**, 236–237 (2009).
62. Koren, S. *et al.* Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**(7), 693–700 (2012).
63. Lou, D. I. *et al.* High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl. Acad. Sci. (U.S.A.)* **110**(49), 19872–19877 (2013).

64. Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**(10), 1155–1162 (2019).
65. Chen, L. *et al.* Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* **364**, 1152 (2019).
66. Morris, K. *et al.* Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics. *Bio. Lett.* **9**, 20120900 (2012).
67. Dobrynin, P. S. *et al.* Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol.* **16**(1), 277 (2015).
68. Richly, E. & Leister, D. NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* **21**(6), 1081–1084 (2004).
69. Triant, D. A. & DeWoody, J. A. Extensive mitochondrial DNA transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications. *Gene* **401**(1–2), 61–70 (2007).
70. Kulemzina, A. I. *et al.* Comparative chromosome painting of pronghorn (*Antilocapra americana*) and saola (*Pseudoryx nghetinhensis*) karyotypes with human and dromedary probes. *BMC Genet.* **15**, 68 (2014).
71. Goel, M. & Schneeberger, K. Plotsr: Visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922–2926 (2022).
72. WMO Report. The global climate 2011–2020—A decade of accelerating climate change. WMO-No. 1338 (2023).
73. Crépín, A.-M. *et al.* Arctic climate change, economy and society (ACCESS): Integrated perspectives. *Ambio* **46**(Suppl. 3), S341–S354 (2017).
74. Blix, A. S. Adaptation to polar life in mammals and birds. *J. Exp. Biol.* **219**, 1093–1105 (2016).
75. Castruita, J. A. S. *et al.* Analysis of key genes involved in Arctic adaptation in polar bears suggest selection on both standing variation and de novo mutations played an importation role. *BMC Genom.* **21**, 543 (2020).
76. Tong, C. Convergent genomics and Arctic adaption of ruminants. *Proc. R. Soc. B* **291**, 20232448 (2024).
77. Hueffer, K. *et al.* Adaption of mammalian host–pathogen interactions in a changing arctic environment. *Acta Vet. Scand.* **53**, 17 (2011).
78. Parkinson, A. J. *et al.* Climate change and infectious diseases in the Arctic: Establishment of a circumpolar working group. *Int. J. Circumpolar Health* **73**, 25163 (2014).
79. Omazic, A. *et al.* Identifying climate-sensitive infectious diseases in animals and human in Northern regions. *Acta Vet. Scand.* **61**, 53 (2019).
80. Cohen, J. M. *et al.* Divergent impacts of warming weather on wildlife disease risk across climates. *Science* **370**, 933 (2020).
81. Rohr, J. R. & Cohen, J. M. Understanding how temperature shifts could impact infectious disease. *PLoS Biol.* **18**(11), e3000938 (2020).
82. Allendorf, F. W. *et al.* Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **11**(10), 697–709 (2010).
83. Wulschleger, S. D. *et al.* Genomics in a changing arctic: Critical questions awaiting the molecular ecologist. *Mol. Ecol.* **24**, 2301–2309 (2015).
84. Khan, S. *et al.* Overview on the role of advance genomics in the conservation biology of endangered species. *Int. J. Genom.* **2016**, 3460416 (2016).
85. Colella, J. P. *et al.* Conservation genomics in a changing Arctic. *Trends Ecol. Evol.* **35**(2), 149–162 (2020).
86. Wright, B. R. *et al.* A demonstration of conservation genomics for threatened species management. *Mol. Ecol. Resour.* **20**(6), 1526–1541 (2020).
87. Theissinger, K. *et al.* How genomics can help biodiversity conservation. *Trends Genet.* **39**(7), 545–559 (2023).
88. Brandies, P. E. *et al.* The value of reference genomes in the conservation of threatened species. *Genes* **10**(11), 846 (2019).
89. Formenti, G. *et al.* The era of reference genomes in conservation genomics. *Trends Ecol. Evol.* **37**(3), 197–202 (2022).
90. Hohenlohe, P. A. *et al.* Population genomics for wildlife conservation and management. *Mol. Ecol.* **30**(1), 62–82 (2021).
91. Paez, S. *et al.* Reference genomes for conservation. *Science* **37**(6604), 364–366 (2022).
92. Nam, M. & Cooper, M. P. Role of energy metabolism in the brown fat gene program. *Front. Endocrinol.* **6**, 104 (2015).
93. Van Oort, B. E. H. *et al.* Circadian organization on reindeer. *Nature* **438**, 1095–1096 (2005).
94. Van Beest, F. M. *et al.* Environmental conditions alter behavioral organization and rhythmicity of a large Arctic ruminant across the annual cycle. *R. Soc. Open Sci.* **7**, 201614 (2020).
95. Aristizábal, B. & González, A. Chapter 2 Innate immune system. In *Autoimmunity from Bench to Bedside* (eds Anaya, J.-M. *et al.*) (El Rosario University Press, 2013).
96. Paludan, S. R. *et al.* Constitutive immune mechanisms: Mediators of host defence and immune regulation. *Nat. Rev. Immunol.* **21**(3), 137–150 (2021).
97. Kawai, T. & Akira, S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **24**(5), 637–650 (2011).
98. Franchi, L. *et al.* Function of nod-like receptors in microbial recognition and host defence. *Immunol. Rev.* **227**(1), 106–128 (2009).
99. Bermejo-Jambrina, M. *et al.* C-type lectin receptors in antiviral immunity and viral escape. *Front. Immunol.* **9**, 590 (2018).
100. Rehwinkel, J. & Gack, M. U. RIG-I-like receptors: Their regulation and roles in RNA sensing. *Nat. Rev. Immunol.* **20**(9), 537–551 (2020).
101. Kawai, T. & Akira, S. Innate immune recognition of viral infection. *Nat. Rev. Immunol.* **7**(2), 131137 (2006).
102. Chen, H. & Jiang, Z. The essential adaptors of innate immune signalling. *Protein Cell* **4**(1), 27–39 (2013).
103. Bousoik, E. & Alibadi, H. M. Do we know jack about JAK? A closer look at JAK/STAT signalling. *Front. Oncol.* **8**, 287 (2018).
104. Arthur, J. S. C. & Ley, S. C. Mitogen-activated protein kinases in innate immunity. *Nat. Rev. Immunol.* **13**(9), 679–692 (2013).
105. McNab, F. *et al.* Type I interferons in infectious disease. *Nat. Rev. Immunol.* **15**(2), 87–103 (2015).
106. Kak, G. *et al.* Interferon-gamma (IFN-gamma): Exploring its implications in infectious disease. *Biomol. Concepts* **9**, 64–79 (2018).
107. Mesev, E. V. *et al.* Decoding type I and III interferon signalling during viral infection. *Nat. Microbiol.* **4**(6), 914–924 (2019).
108. Abrahá, R. Review on the role and biology of cytokines in adaptive and innate immune system. *Arch. Vet. Anim. Sci.* **2**(2), 2 (2020).
109. Esche, C. *et al.* Chemokines: Key players in innate and adaptive immunity. *J. Investig. Dermatol.* **125**, 615–628 (2005).
110. Sokol, C. L. & Luster, A. D. The chemokine systems in innate immunity. *Cold Spring Harb. Perspect. Biol.* **7**, a016303 (2015).
111. Xu, D. & Lu, W. Defensins: A double-edged sword in host immunity. *Front. Immunol.* **11**, 764 (2020).
112. Valeri, M. & Raffatellu, M. Cytokines IL-17 and IL-22 in host response to infection. *Pathog. Dis.* **74**, ftw111 (2016).
113. Horton, R. *et al.* Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**, 889–899 (2004).
114. Shiina, T. *et al.* The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* **54**, 15–39 (2009).
115. Cruz-Tapias, P. *et al.* Chapter 10 Major histocompatibility complex: antigen processing and presentation. In *Autoimmunity from Bench to Bedside* (eds Anaya, J.-M. *et al.*) (El Rosario University Press, 2013).
116. Kulski, J. K. *et al.* Genomic diversity of the major histocompatibility complex in health and disease. *Cells* **8**, 1270 (2019).
117. Ziegler, A. *et al.* Female choice and the MHC. *Trends Immunol.* **26**(9), 496–502 (2005).
118. Knapp, L. A. The ABC of the MHC. *Environ. Anthropol.* **14**, 28–37 (2005).
119. Groth, V. *et al.* Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl. Acad. Sci. (U.S.A.)* **93**, 12445–12450 (1996).
120. Bauer, S. *et al.* Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285**, 727–729 (1999).
121. Deakin, J. E. *et al.* Evolution and comparative analysis of the MHC class III inflammatory region. *BMC Genom.* **7**, 281 (2006).

122. Gruen, J. R. & Weissman, S. M. Human MHC class III and class IV genes and disease associations. *Front. BioSci.* **6**, d960–972 (2001).
123. Sedy, J. *et al.* Tumor necrosis factor superfamily in innate immunity and inflammation. *Cold Spring Harb. Perspect. Biol.* **7**, a016279 (2015).
124. Wallach, D. The tumor necrosis factor family: Family conversions and private idiosyncrasies. *Cold Spring Harb. Perspect. Biol.* **10**, a028431 (2018).
125. Upadhyay, G. Emerging role of lymphocyte antigen-6 family of genes in cancer and immune cells. *Front. Immunol.* **10**, 819 (2019).
126. Dunkelberger, J. R. & Song, W.-C. Complement and its role in innate and adaptive immune responses. *Cell Res.* **20**, 34–50 (2010).
127. Cruse, J. M. *et al.* (eds) *Cluster of Differentiation (CD) Antigens* (Immunology Guidbook, 2007). <https://doi.org/10.1016/B978-012198382-6/50027-3>.
128. Sommer, S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front. Zool.* **2**, 16 (2005).
129. Radwan, J. *et al.* Does reduced MHC diversity decrease viability of vertebrate population?. *Biol. Conserv.* **143**, 537–544 (2010).
130. Ujvari, B. & Belov, K. Major histocompatibility complex (MHC) markers in conservation biology. *Int. J. Mol. Sci.* **12**, 5168–5186 (2011).
131. Coker, O. M. *et al.* Major histocompatibility complex (MHC) diversity and its implications for human and wildlife health and conservation. *Genet. Biodivers. J.* **7**(2), 1–11 (2023).
132. Mikko, S. *et al.* Monomorphism and polymorphism at the Mhc *DRB* loci in domestic and wild ruminants. *Immunol. Rev.* **167**, 169–178 (1999).
133. Zverinova, S. & Guryev, V. Variant calling: Considerations, practices, and developments. *Hum. Mutation* **43**, 976–985 (2022).
134. Pan, B. *et al.* Assessing reproducibility of inherited variants detected with short-read whole genome sequencing. *Genome Biol.* **23**, 2 (2022).
135. Brlek, P. *et al.* Implementing whole genome sequencing (WGS) in clinical practice: Advantages, challenged, and future perspectives. *Cells* **13**, 504 (2024).
136. Mahmoud, M. *et al.* Structural variant calling: the long and the short of it. *Genome Biol.* **20**, 246 (2019).
137. Edge, P. & Bansal, V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat. Commun.* **10**, 4660 (2019).
138. Lukacs, M. *et al.* Functional immune diversity in reindeer reveals a high arctic population at risk. *Front. Ecol. Evol.* **10**, 1058674 (2023).
139. Teixeira, J. C., Huber, C. & D.,. The inflated significance of neutral genetic diversity in conservation genetics. *Proc. Natl. Acad. Sci. (U.S.A.)* **118**(10), e2015096118 (2021).
140. Abbott, B. & Van Kooten, G. C. Can domestication of wildlife lead to conservation? The economics of tiger farming in China. *Ecol. Econ.* **70**(4), 721–728 (2011).
141. Teletchea, F. Wildlife conservation: Is domestication a solution? In *Global Exposition of Wildlife Management, Chapter 1* (ed. Lameed, G. A.) (IntechOpen Ltd, 2017). <https://doi.org/10.5772/65660>.
142. Moyers, B. T. *et al.* Genetic costs of domestication and improvement. *J. Hered.* **109**(2), 103–116 (2018).
143. Larson, G. Genetics and domestication. *Curr. Anthropol.* **52**, S485–S495 (2011).
144. Hunter, P. The genetics of domestication. *EMBO Rep.* **19**(2), 201–205 (2018).
145. Alberto, F. J. *et al.* Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* **9**, 813 (2018).
146. Ahmad, H. I. *et al.* The domestication makeup: Evolution, survival, and challenges. *Front. Ecol. Evol.* **8**, 103 (2020).
147. Zheng, Z. *et al.* The origin of domestication genes in goats. *Sci. Adv.* **6**, eaaz5216 (2020).
148. Dou, M. *et al.* A missense mutation in *RRM1* contribute to animal tameness. *Sci. Adv.* **9**, eadf4068 (2023).
149. Wilkinson, P. F. The history of musk-ox domestication. *Polar Rec.* **17**(106), 13–22 (1974).
150. Helfferich, D. The muskox: A new northern farm animal. *Agroborealis* **38**(2), 18–21 (2006).
151. Helfferich, D. The muskox: Muskox husbandry. *Agroborealis* **39**(2), 10–19 (2007).
152. Georges, M. Mapping, fine mapping and molecular dissection of quantitative trait loci in domestic animals. *Annu. Rev. Genom. Hum. Genet.* **8**, 131–162 (2007).
153. Wakchaure, R. *et al.* Marker assisted selection (MAS) in animal breeding: A review. *J. Drug Metab. Toxicol.* **6**(5), 1000e127 (2015).
154. Sharma, P. *et al.* Overview of marker-assisted selection in animal breeding. *J. Adv. Biol. Biotechnol.* **27**(5), 303–318 (2024).
155. Georges, M. *et al.* Harnessing genomics information for livestock improvement. *Nat. Rev. Genet.* **20**(3), 135–156 (2019).
156. Gavran, M. *et al.* Candidate genes associated with economically important traits of sheep—A review. *Agric. Conspec. Sci.* **86**(3), 195–201 (2021).
157. Zeng, L. *et al.* Genes related to heat tolerance in cattle—A review. *Anim. Biotechnol.* **34**(5), 1840–1848 (2023).
158. Worku, D. *et al.* Candidate genes associated with heat stress and breeding strategies to relieve its effects in dairy cattle: A deeper insight into the genetic architecture and immune response to heat stress. *Front. Vet. Sci.* **10**, 1151241 (2023).
159. Arzik, Y. *et al.* Genome-wide scan of wool production traits in Akkaraman sheep. *Genes (Basel)* **14**(3), 713 (2023).
160. Moll, R. *et al.* The human keratins: Biology and pathology. *Histocom. Cell Biol.* **129**, 705–733 (2008).
161. Wu, D.-D. *et al.* Molecular evolution of the keratin associated protein gene family in mammals, role in the evolution of mammalian hair. *BMC Evol. Biol.* **8**, 241 (2008).
162. Gong, H. *et al.* Wool keratin-associated protein genes in sheep—A review. *Genes* **7**, 24 (2016).
163. Zhou, H. *et al.* The complexity of the ovine and caprine keratin-associated protein genes. *Int. J. Mol. Sci.* **22**, 12838 (2021).
164. Li, S. *et al.* Variation in the ovine KAP6-3 gene (*KRTAP6-3*) is associated with variation in mean fibre diameter-associated wool traits. *Genes* **8**, 204 (2017).
165. Sulayman, A. *et al.* Association analysis of polymorphisms in six keratin genes with wool traits in sheep. *Asian Australas. J. Anim. Sci.* **31**(6), 775–783 (2018).
166. Kalds, P. *et al.* Genetics of the phenotypic evolution in sheep: A molecular look at diversity-driving genes. *Genet. Sel. Evol.* **54**, 61 (2022).
167. Zhang, W. *et al.* Whole-genome resequencing reveal selection signal related to sheep wool fitness. *Animals* **13**, 2944 (2023).
168. Willi, Y. *et al.* Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *Proc. Natl. Acad. Sci. (U.S.A.)* **119**, e2105076119 (2022).
169. Buschman, V. Q. & Sudlovenick, E. Indigenous-led conservation in the Arctic supports global conservation practices. *Arctic Sci.* **9**, 714–719 (2023).
170. Madsen, J. *et al.* Implementation of the first adaptive management plan for a European migratory waterbird population: The case of the Svalbard pink-footed goose *Anser brachyrhynchus*. *Ambio* **46**(suppl. 2), S275–S289 (2017).

Acknowledgements

We are grateful to the staff of the Informatics and Biostatistics, and Sequencing Facilities of The Centre for Applied Genomics (TCAG) at The Hospital for Sick Children in Toronto. TCAG is part of CGEn, a Major Science Initiative of the Canada Foundation for Innovation. We acknowledge the contributions of Karen Ho and Sachin Desai for expert library construction and DNA sequencing, Joe Whitney and Omar Hamdan for informatics

support, and Peter J. Van Coeverden de Groot for sample collection. The sample of muskox analyzed in this study was originally obtained from the tissue collection of the University of Alberta under the direction of Dr. Curtis Strobeck in 1991. We thank the Inuvialuit harvesters at the time from Sachs Harbour, NWT and Dr. Strobeck for this critical contribution. We acknowledge the Inuvialuit's long-standing role and sovereign responsibilities in muskox management.

Author contributions

SL and SWS conceived the project. SL directed the project, analysed results, performed gene annotation, and wrote the manuscript. TNHL analysed results and performed genome assembly, informatics and gene annotation. BT performed genetic diversity analysis. AHYT contributed to gene annotation. TA contributed to the analysis of mitochondrial DNA. AG contributed to aspects of biology, phylogenetic analysis, ecology and conservation management. All authors contributed to writing and editing the manuscript.

Funding

This study was funded by the Lau Family Endowment for Genome Science Development and by Genome Canada's Disruptive Innovation in Genomics Program (Grant number 9425). The work was conducted at The Centre for Applied Genomics (TCAG) using infrastructure funded by the Canada Foundation for Innovation, and was operationally supported by the Canada Foundation for Innovation (CGEn Major Science Initiative), the Government of Ontario, Genome Canada through Ontario Genomics, and The Hospital for Sick Children (SickKids) Foundation. SWS holds the Northbridge Chair in Paediatric Research, a joint Hospital-University Chair between the University of Toronto, The Hospital for Sick Children, and the SickKids Foundation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-67270-9>.

Correspondence and requests for materials should be addressed to S.L. or S.W.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024