

Article

Draft Genome Assembly of an Iconic Arctic Species: Muskox (*Ovibos moschatus*)

Erin Prewer^{1,*} , Susan Kutz², Lisa-Marie Leclerc³ and Christopher J. Kyle^{1,4,5}

¹ Environmental and Life Sciences Graduate Program, Trent University, 1600 West Bank Drive, Environmental Science Building, Peterborough, ON K9J 7B8, Canada; christopherkyle@trentu.ca

² Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, 3280 Hospital Drive NW, Calgary, AB T2N 4Z6, Canada; skutz@ucalgary.ca

³ Department of Environment, Government of Nunavut, P.O. Box 377, Kugluktuk, NU X0B 0E0, Canada; lleclerc@gov.nu.ca

⁴ Forensic Science Department, Trent University, 1600 West Bank Drive, DNA Building Block B, Peterborough, ON K9L 0G2, Canada

⁵ Natural Resources DNA Profiling and Forensic Centre, 2140 East Bank Drive, Peterborough, ON K9L 1Z8, Canada

* Correspondence: erinprewer@trentu.ca; Tel.: +1-(416)-659-7324

Abstract: Muskoxen (*Ovibos moschatus*) are Arctic species within the Caprinae subfamily that are economically and culturally significant to northern Indigenous communities. Low genetic diversity from repeated genetic bottlenecks, coupled with the effects of Arctic warming (e.g., heat stress, changing forage, pathogen range expansions), present conservation concerns for this species. Reference genome assemblies enhance our ecological and evolutionary understanding of species (which in turn aid conservation efforts). Herein, we provide a full draft reference genome of muskox using Illumina Hiseq data and cross-species scaffolding. The final reference assembly yielded a genome of 2,621,890,883 bp in length, a scaffold N50 of ~13.2 million, and an annotation identifying ~19.3 k genes. The muskox genome assembly and annotation were then used to reconstruct a phylogenetic tree which estimated muskoxen diverged from other ungulate species ~12 Mya. To gain insight into the demographic history of muskoxen we also performed pairwise sequentially Markovian coalescent (PSMC) that identified two population bottlenecks coinciding with major glaciation events contributing to the notoriously low genetic variation observed in muskoxen. Overall, this genome assembly provides a foundation for future population genomic studies, such as latitudinal analyses, to explore the capacity of muskoxen to adapt to rapidly changing environments.

Keywords: de novo genome assembly; genetically depauperate; muskoxen; Arctic; *Ovibos moschatus*; PSMC; phylogenetic tree



Citation: Prewer, E.; Kutz, S.; Leclerc, L.-M.; Kyle, C.J. Draft Genome Assembly of an Iconic Arctic Species: Muskox (*Ovibos moschatus*). *Genes* **2022**, *13*, 809. <https://doi.org/10.3390/genes13050809>

Academic Editor: Manuel Vera

Rodriguez

Received: 15 April 2022

Accepted: 29 April 2022

Published: 1 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Muskoxen (*Ovibos moschatus*) are an iconic Arctic species closely related to sheep and goats within the *Caprinae* subfamily and the only living member of the *Ovibos* genus. Endemic muskoxen are found on the mainland and Arctic Archipelago of the Northwest Territories (NWT) and Nunavut (NU), Canada and east Greenland (Denmark). Reintroduced or translocated populations of muskoxen are currently found in west Greenland, Russia, Alaska (USA), and portions of the Yukon and Quebec (Canada) [1]. Muskoxen are keystone Arctic species, facilitating nitrogen and soil nutrient turnover while also having significant cultural, nutritional, and economic roles for Indigenous People of the Arctic [2,3]. Muskoxen contribute to community identity in the creation of art, tools, and clothing, but also local food security, employment, and revenue through sport hunting and sales of muskox by-products such as its specialized wool, qiviut [3]. Harvests were suspended in 2012 due drastic declines in muskox populations on Banks Island (NWT) and

Victoria Island (NU) [4,5]. These population declines occurred concurrently with changing climatic conditions including severe icing events, vegetation shifts, and northward range expansions of pathogens [1,6–8]. As Banks and Victoria Island populations were two of the largest endemic populations, these declines incited concern for muskox health and sustainability [1,4].

Muskox populations have notoriously low levels of genetic diversity resulting from a combination of population bottlenecks, founder effects, and population fragmentation across their range [9–13]. Confounding the premise that low genetic diversity may be associated with population declines on Victoria and Banks Island is the fact that other populations across Canada remained stable or grew with unknown reasons for this dichotomy [1]. Past diversity estimates from microsatellite studies [9,12,13], as well as genotyping by sequencing by Hansen et al. [14], focused on neutral regions of the genome. While informative, these studies did not elucidate information related to the adaptive capacity of muskoxen that might be associated with demographic trends in context of changing selective factors on the landscape (such as those from climate change and a warming Arctic). Elucidating the functional genomic variation across muskox populations may provide insight into muskox population health and vulnerabilities but requires a better genomic foundation. Access to a reference genome would offer the ability to assess the genetic diversity of genes linked to evolutionary and local muskox adaptations such as increased digestion, cold resistance, and immune response [15–17]. As such, genome assemblies have the potential to provide insight into how evolutionary and demographic histories have influenced patterns of genomic diversity in muskoxen as well as their capacity to adapt to a rapidly changing environment.

Herein, we report the first assembly and annotation of the muskox genome using high throughput Illumina sequencing and cross-species *in silico* mate pair library construction. Our final genome assembly and annotation was used to reconstruct a phylogenetic tree and estimate divergence times of this unique genus from other ungulate species. The genome was also used to perform pairwise sequentially Markovian coalescent (PSMC) analyses to identify historical trends in an effective population size. Our main goal was to provide a reference genome from which further population genomic studies could be performed, such as assessing the distribution of genetic diversity in muskoxen, identifying genes involved in their unique Arctic adaptations, and to better understand their capacity to respond to rapid environmental change.

2. Materials and Methods

2.1. Genome Sequencing

Genomic DNA for the reference genome assembly was extracted from the hide of a male muskox from Holman, Victoria Island, Canada using the Qiagen DNeasy blood and tissue kit (Qiagen, Mississauga, ON, Canada). Quality of extracted DNA was assessed on a 1.5% agarose gel and quantified using PicoGreen. Extracted DNA was shipped on dry ice to the Centre for Applied Genomics at Sick Kids hospital, Toronto, ON, CA for library preparations and sequencing. Four paired-end libraries were prepared using Illumina TruSeq Nano DNA kit (Illumina, San Diego, CA, USA) with insert sizes of 200 bp, 350 bp, 550 bp and 700 bp and two mate-pair libraries were created using the NxSeq Long Mate Pair Library Kit with insert sizes of 5 kbp and 8 kbp. Two paired-end libraries, with insert sizes 200 bp and 350 bp, were sequenced on the HiSeqX which produced approximately 940 million (2×151 bp) paired reads. The two additional, paired-end libraries with insert sizes of 550 bp and 700 bp, as well as two mate-pair libraries, were sequenced on the HiSeq2500 producing approximately 237 million (2×126 bp) paired reads.

2.2. De Novo Genome Assembly

We performed a series of data filtering steps to remove read contamination, low quality reads, and duplicate reads. We used FastQC v0.11.5 [18] to check overall quality of the libraries after each step to assess how much data was removed and effectiveness

of each program. First, we removed adapter and primer sequence contamination using scythe v0.994 [19] and EA utils v. 805 [20], with mate-pair libraries undergoing additional processing to split and remove chimera code, junction code and linker sequences as per the Lucigen NxSeq long mate-pair library kit protocol. Next, we performed low quality base trimming using sickle v. 1.33 [21] with a quality score cut off set at $Q = 25$ and a minimum length set at 70% of the original read length. We then removed duplicate read pairs using FastUniq v. 1.1 [22] in order to remove any technical duplicates which may affect downstream scaffolding. Finally, we used bbsplit from the bbmap v37.22 [23] package to remove sequences belonging to known lab contaminants using viral, bacterial, fungal and human database created by DeconSeq [24]. After data preprocessing, approximately 1.23×10^{11} bp remained from paired-end libraries and approximately 4.93×10^6 bp remained from mate-pair libraries for a theoretical coverage of $\sim 85\times$ based on the genome size of the domestic goat (2.9 Gb).

We performed an initial de novo assembly using SOAPdenovo2 v.240 software [25], that is, a De Bruijn graph-based de novo assembler and then used Gapcloser v.1.12-r6 [25] and uses short read library data to fill gaps that occur during scaffolding. To further improve the de novo genome assembly, we used cross-species-scaffolding v 1.0.1 [26]. This reference-based scaffolding method uses genomes of closely related species to create scaffolding libraries in silico that can then be used to re-perform scaffolding on assembled genomes. Cross-species-scaffolding requires a closely related genome, so we tested both domestic goat (GCA_001704415.1) and domestic sheep genomes (GCA_002742125.1) to create two separate sets of 20 mate-pair libraries with inserts ranging from 500 bp to 50 kbp. To determine which in silico library (goat or sheep based) produced the best overall assembly we used both datasets to scaffold and assemble multiple genomes with SOAPdenovo2, with kmers ranging from 51 to 127 bp. These genomes were then ranked using metrics estimated by Quast v4.4 [27] and BUSCOv3 [28], including NG50, number of Ns per 100 kbp and completed BUSCOs (Supplementary Table S2) to identify which kmer value, and reference-based mate pair library, was optimal for downstream analyses.

Large sections of the muskox's mitochondrial genome were found within several long scaffolds of the top assembly. To address this issue, we further filtered cleaned paired-end and mate-pair reads to remove reads that corresponded to the mitochondrial genome. Mitochondrial reads were extracted from cleaned sequence libraries using the bbsplit tool from bbmap by binning reads that mapped to the reference mitochondrial genome (GenBank FJ207536.1). Then, to create the mitochondrial genome, these reads were aligned to the reference (FJ207536.1) and the vcfutils.pl tool from the samtools package was used to call consensus sequences using default parameters. Finally, we assembled a new whole nuclear genome using SOAPdenovo2 with the newly filtered clean read dataset. This genome assembly was then used to re-perform genome scaffolding with the previously optimized cross-species-scaffolding library and kmer value. The final genome assembly was completed by a final round of Gapcloser.

2.3. Genome Annotation

We used Repeat Modeler v.1.0.10 [29] software to create repeat libraries for the muskox genome assembly. We then used these libraries to perform hard masking of repetitive regions in the muskox genome using Repeat Masker v. 4.0.7 software [30]. We retrained Augustus v. 3.2.3 software [31] to perform gene prediction on the masked genome. In order to create a set of training genes for Augustus, we used Gmap v2017-08-15 [32], Genemark v4.35 [33], Exonerate v. 2.2.0 [34] and EVidenceModeler v.1.1.1 software [35]. Genemark-ES was used to perform ab initio gene prediction on both masked genomes. We downloaded all sheep and goat EST sequences from GenBank and used them to perform evidence-based gene prediction on the masked genomes using Gmap software. We also downloaded all sheep and goat SWISS-PROT reviewed protein sequences from the UniProt-Kb database and performed evidence-based gene prediction of masked genomes using exonerate software. We used EVidenceModeler v.1.1.1 software to create a set of consensus

genes from outputs produced by Genemark v4.35, Exonerate v. 2.2.0 and Gmap v2017-08-15 software [32–35] with more weight given to the evidence-based predictions. This set of consensus genes was used to retrain Augustus software, creating a new set of parameters for muskoxen. These parameters were used to perform gene prediction on the masked genomes with the consensus gene set used as hints.

2.4. Evolutionary and Demographic History

Pairwise Sequentially Markovian Coalescent (PSMC) model was used to estimate effective population history following the pipeline outlined by Li et al. [36]. Sequences from cleaned paired-end libraries were aligned to the muskox genome using Burrows-Wheeler alignments (BWA) v. 0.7.17 [37]. BAM alignments from each library were merged and duplicates were removed using MergeBamAlignment and MarkDuplicates from the Picard Toolkit (Broad Institute). The combined alignment was used to create consensus sequences using samtools v.0.1.15, vcfutils and bcftools [38]. Option -d was set to 20, as recommended to be set to a third of the coverage, while -D was set at 170, as recommended to be twice the average coverage. We then performed PSMC v.0.6.5 analyses using -p parameter suitable for modern humans ($4 + 25 \times 2 + 4 + 6$) and bootstrapped using 100 iterations. The mutation rate 2×10^{-9} was chosen as the average mutation rate in mammals with a generation time of 10 years [39].

To construct the phylogenetic tree, 1:1 orthologous genes were identified among 8 species (muskox-*Ovibos moschatus*, sheep-*Ovis aries*; GCA_000298735.1, goat-*Capra hircus*; GCA_001704415.1, cow-*Bos taurus*; GCA_002263795.2, horse -*Equus caballus*; GCA_002863925.1, pig-*Sus scrofa*; GCA_000003025.6, bison-*Bison bison bison*; GCA_000754665.1, yak-*Bos gruniens*; GCA_000298355.1). Using PorthoMCL [40], 1:1 orthologous genes were then identified and corresponding coding sequences were aligned using Clustal-Omega. All alignments were concatenated using FasConCat v.1.04 software [41], and Jmodeltest v.2.1.10 was used to predict the best-fit model of nucleotide substitution [42]. A maximum likelihood tree was then generated in RAXMLv8 [43] using the GTR+I+G model with 1000 bootstrap replicates and horse as the outgroup. MCMCtree from the PAMLv4 package [44] was used to convert the tree to an ultrametric format and to calibrate the nodes with known split times. Calibration minimums and maximums were given for sheep-goat divergence (3.9–8.1 MYA) from Chen et al. [45] and for sheep-cow divergence (18.3–28.5 MYA) from Benton and Donoghue [46].

3. Results

3.1. Genome Sequencing and Assembly

We sequenced the draft genome of a male muskox from Victoria Island using 4 paired-end libraries and 2 mate-pair libraries. Across all 6 libraries, a total of 1,178,984,124 paired reads were generated. After filtering for low quality, duplicate and contaminant reads, 78.9% (843,572,921) paired reads remained for paired-end libraries, however only 0.04% (39,179) paired reads remained for mate-pair libraries (Supplementary Table S1). Loss of mate pair library data was the result of duplicate read removal steps. Mate pair sequencing failure likely relates to a lack of sufficient quality DNA for mate pair library preparations, and thus PCR bias when sequencing. The overall result was a sequencing coverage of $\sim 85 \times$ for the muskox genome based on a goat genome size of 2.9 Gb.

An initial draft assembly was performed by SOAPdenovo2 software v.2.04-r240 [25] using all paired-end libraries and low coverage of mate pair reads that remained after filtering. The resulting genome was fragmented with over ~ 822 K contigs and an N50 of 26,107 bp, meaning that half of the genome was made up of contigs of this length or larger, and a complete BUSCO score of 65.6% which is below that of other ungulate genomes [45]. We assumed the lower genome quality as reflected by the metrics of # of contigs, N50 and BUSCO scores was likely associated with the lack of mate pair libraries, preventing SOAPdenovo2 v.240 [25] from combining contigs, especially those separated by large gaps or repetitive regions, into scaffolds. In order to improve genome quality,

we used in silico mate-pair libraries created by Cross-Species-Scaffolding (v.2.2) [26] by aligning the cleaned muskox reads to those of closely related species. Both sheep and goat genomes were used as a reference, after which assemblies were compared based on resulting N50, scaffold number and BUSCO scores amongst other common quality metrics (e.g., Ns per 100 kbp) to determine which genome would produce the best in silico mate-pair libraries, (Supplementary Table S2). Cross-species scaffolding libraries created by both reference genomes greatly improved genome quality, but the assembly with the goat genome as reference, using a kmer value of 121, ranked highest. The best ranked genome had 6495 contigs, a complete BUSCO score of 87.5% and an N50 of 1,705,149 bp, which is 65× higher than the assembly without cross-species scaffolding. The use of cross-species scaffolding allowed us to create large scaffolds without further sequencing of long read data, mate pair libraries, and/or sequencing by ligation, but there are potential drawbacks when using related species scaffolding methods. First, cross-species scaffolding is limited by the quality of the initial genome assembly as well as the existence of a high-quality genome from a related species, though the increased availability of genomic resources may allow this method to be more widely used [26,47]. For example, in silico mate pair libraries have been used to improve genomes of fin whales, narwhals, gray's beaked whales as well as addax (that similarly used the goat genome as a reference) [48–51]. Additionally, as the arrangement of the contigs within the scaffolds are based on the genome of another species, this can limit analyses of genomic architecture, such as gene copy number and gene rearrangements [26,47]. Previous karyotype mapping has found muskoxen to be highly homologous to ancestral Pecora chromosomal arrangements, with five fusions of different chromosome arms forming submetacentric chromosomes [52–55]. However, G–, C– and R– banding found many muskox chromosomes to be either identical or strikingly similar to those of goats. These data further strengthen the validity of using the goat genome as a reference for cross-species scaffolding of muskoxen, and likely explain why we found goats produced a better muskox assembly than the sheep genome [52–55]. The removal of mitochondrial reads did not greatly improve the quality of the initial genome assembly with a complete BUSCO score of 64.2% and an N50 of 26,274 however using this genome as the base to reperform cross-species scaffolding had a large impact on the final genome assembly quality.

The final draft genome produced 8659 scaffolds for a cumulative length of 2,621,890,883 bp, with a contig and scaffold N50 of 38,369 bp and 13,200,690 bp, respectively. The scaffold N50 was 7× better than our previous assembly that used cross-species scaffolding prior to mitochondrial read filtering. The longest scaffold was 50,595,910 bp long, where 2,601,612,364 bp of the assembly was made up of contigs over 25,000 bp in length. Based on cumulative length and Nx plots (Supplementary Figure S1) these quality metrics show that the draft genome was mainly composed of large scaffolds with few short sequences. Completeness of the draft genome was further assessed with Benchmarking Universal Single-Copy Orthologs (BUSCO v.3) [28] vertebrata gene set. Of the 3023 genes in the BUSCO gene set, 2658 (87.9%) complete single copy orthologs were identified in the muskox draft genome, indicating that gene prediction would identify a large percentage of completed genes in muskoxen. Additional BUSCO scores and basic statistics of the assemblies, pre- and post-cross-species scaffolding, as well as pre- and post- mitochondrial read removal, are shown in Table 1. The draft assembly was deposited in GenBank (JACAUE000000000).

Table 1. Comparison of quality metrics for genome assemblies pre- and post-cross-species-scaffolding and pre- and post-mitochondrial read filtering. Mx1 is the initial scaffold level assembly prior to both cross-species scaffolding and mitochondrial read filtering; Mx2 is the scaffold level genome assembly after cross-species scaffolding, but prior to mitochondrial DNA filtering; Mx3 is the scaffold level genome after mitochondrial read filtering, but prior to cross-species scaffolding; Muskox Final is the final genome assembly with columns for both contig and scaffold levels of assemblies.

Program	Quality Metric	Mx1	Mx2	Mx3	Muskox Final (Contig)	Muskox Final (Scaffold)
QUAST	Assembly length	2,528,575,139	2,704,466,437	2,527,322,807	2,473,404,411	2,621,890,883
	Assembly length over 25,000 bp	1,310,898,307	2,687,729,513	1,275,654,848	1,683,252,947	2,601,612,364
	Number of contigs	822,935	6495	868,473	114,189	8659
	Longest contig	328,372	9,453,317	301,459	388,735	50,595,910
	L50	27,014	494	26,422	18,651	61
	N50	26,107	1,705,149	26,274	38,369	13,200,690
	N per 100 kbp	5056.29	10,272	5035	0.93	5664
	GC content	41.56	41.81	41.55	44.53	44.53
BUSCO	Completed BUSCO (out of 3023)	1984 (65.6%)	2646 (87.6%)	1953 (64.6%)		2673 (88.4%)
	Single Copy				N/A	
	Completed BUSCO (out of 3023)	1973 (65.2%)	2623 (86.7%)	1941 (64.2%)		2658 (87.9%)
	Duplicated BUSCO (out of 3023)	14 (0.4%)	14 (0.4%)	12(0.4%)		15 (0.5%)
	Fragmented (out of 3023)	448 (14.8%)	223 (7.1%)	458(15.2%)		212 (7%)
	Missing (out of 3023)	591 (19.5%)	163 (5.3%)	612(20.2%)		138 (4.6%)

3.2. Genome Annotation

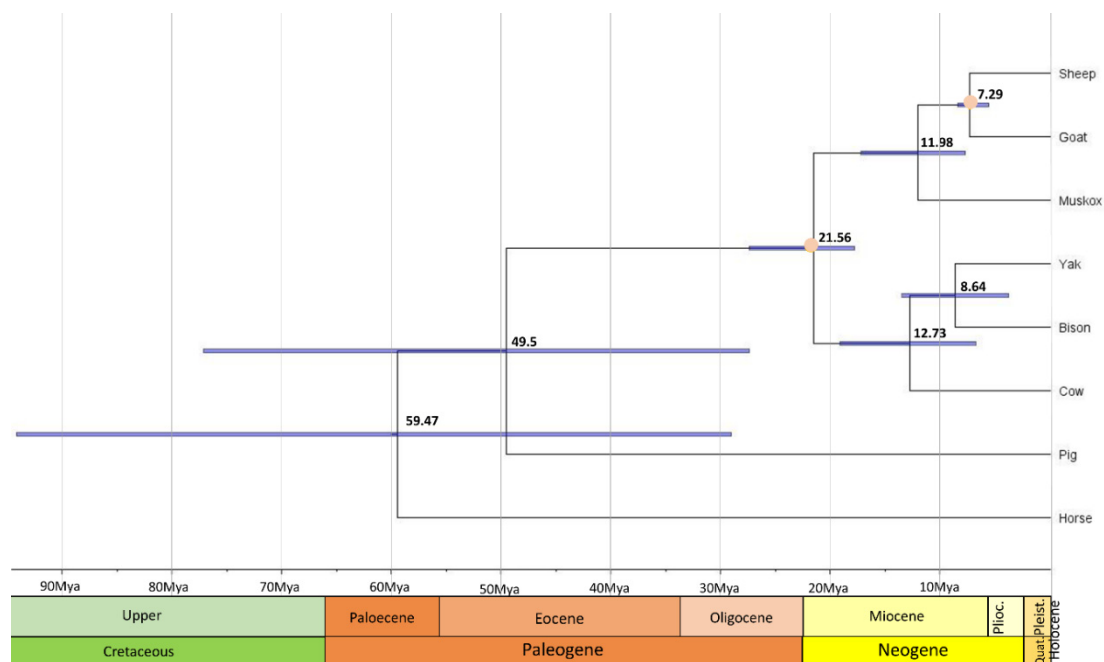
Repeat masking was performed prior to gene predictions using RepeatMasker (v.4.0.7) [30] that identified 42.39% of sequences were made up of interspersed repeats. Once repeats were masked, both homology and ab initio predictions were used to identify protein coding genes. Genes predicted by both methods were combined using Evidence Modeler (EVM). These genes were then used as a training set for Augustus (v.3.2.3) [31] which performed the final gene prediction with the combined consensus gene set as hints. A total of 19,132 genes were predicted with an average of 24.9 kb per gene, 1461 bp per coding DNA sequence, and 154 bp per exon. Of the predicted genes, 12,848 (67%) were annotated by Interpro for Gene Ontology and 11,349 genes (59%) aligned to the National Centre for Biotechnology Information (NCBI) non-redundant protein database (Supplementary Figure S2). In comparison to the genome annotations of other ungulate species, the number of genes and average coding sequence length of the muskox genome assembly are within the norm [45]. We also compared genomes using quality metrics of contig and scaffold N50s, and complete BUSCO scores. The quality of the draft muskox genome assembled herein falls within those presented by Chen et al. [45] for other ruminant species. Genome assembly and annotation metrics of the final muskox genome are shown in Table 2 in comparison with the highest quality ruminant genomes assembled by Chen et al. [45] as well as their assembly of another Arctic ruminant (reindeer).

Table 2. Comparison of the final muskox genome assembly to both another Arctic ruminant (reindeer) and the highest quality wild ruminant genomes assembled by Chen et al. [45].

	<i>Ovibos moschatus</i>	<i>Okapia johnstoni</i>	<i>Rangifer tarandus</i>	<i>Gervus albirostris</i>	<i>Procapra przewalskii</i>	<i>Capra ibex</i>	<i>Ovis Amon</i>
Common name	Muskox	Okapi	Reindeer	White-lipped deer	Przewalski's gazelle	Ibex	Argali
Scaffold N50	13,200,690	3,620,116	1,059,113	3,567,448	5,152,914	15,190,720	5,734,776
Contig N50	38,369	58,892	91,805	22,599	20,018	24,835	45,638
BUSCO %	88.4%	90.10%	90.40%	89.10%	89.10%	92.50%	93.10%
Number of genes	19,132	19,568	21,555	23,319	23,562	21,204	20,335
Average cds	1461	1518	1440	1440	1150	1544	1571

3.3. Evolutionary and Demographic History

In order to compare coding sequences of muskoxen to orthologous genes among Caprinae and other mammals, PorthoMCL was used to identify orthologs amongst 8 ungulate species. The ortholog set was then filtered to contain only 1:1 orthologs and resulting in 893 genes used for the phylogenetic tree. As expected, the ultrametric and time calibrated phylogenetic tree from mcmctree shows that muskoxen are more closely related to sheep and goats than cows (Figure 1) [11,56,57]. From this tree, muskoxen diverged from sheep and goats approximately 12 million years ago (Mya) and this sister clade diverged from cows approximately 21.6 Mya. The divergence time of the sheep and goats' clade is comparable with times previously estimated in the literature, as are the branch dates of other species included in the tree [45]. Previous studies using the mitochondrial genome estimate muskoxen diverged from sheep and goats ~8–15 Mya [58,59], values that fall within our confidence intervals of 8–17 Mya. Mitochondrial gene sequences have been found to have lower resolving power over nuclear exons when investigating phylogeny reconstructions, therefore the analyses performed based on the nDNA genome assembly should be more accurate [60] in describing the phylogeny and divergence times of this unique genus.

**Figure 1.** Evolution of gene families among muskoxen and non-Arctic relatives. The phylogenetic tree was constructed using 893 single copy orthologs across all 8 ungulate species. Divergence times are in black, orange dots represent the calibration points, and 95% credible intervals are represented by blue node bars.

We used the Pairwise Sequentially Markovian Coalescent (PSMC) model to assess historical patterns of effective population size (N_e) (Figure 2) [36]. At 3 Mya, there is a relatively low N_e at ~35,000, and over the next 500 thousand years, N_e increases until it reaches 100 k. At 1 Mya, N_e begins to steadily decline until it reaches an N_e of ~2 k. This low N_e remains from ~20–40 thousand years ago (Kya) before showing a slight recovery ~15 Kya, reaching an N_e ~5 k. Low points in muskox N_e starting at 3 Mya and 40 Kya coincide with major glaciation events. The start of the PSMC occurred during a major glaciation event occurring ~3.15–2.75 Mya, referred to as the climate crash [61] and we see N_e recovery follows immediately after this period. The second low point represents a bottleneck occurring during the last glacial maximum (LGM) ~26.5–19 Kya where once again a recovery is observed only after the LGM ended [62], though the decline of N_e was continuous from 1 Mya onward.

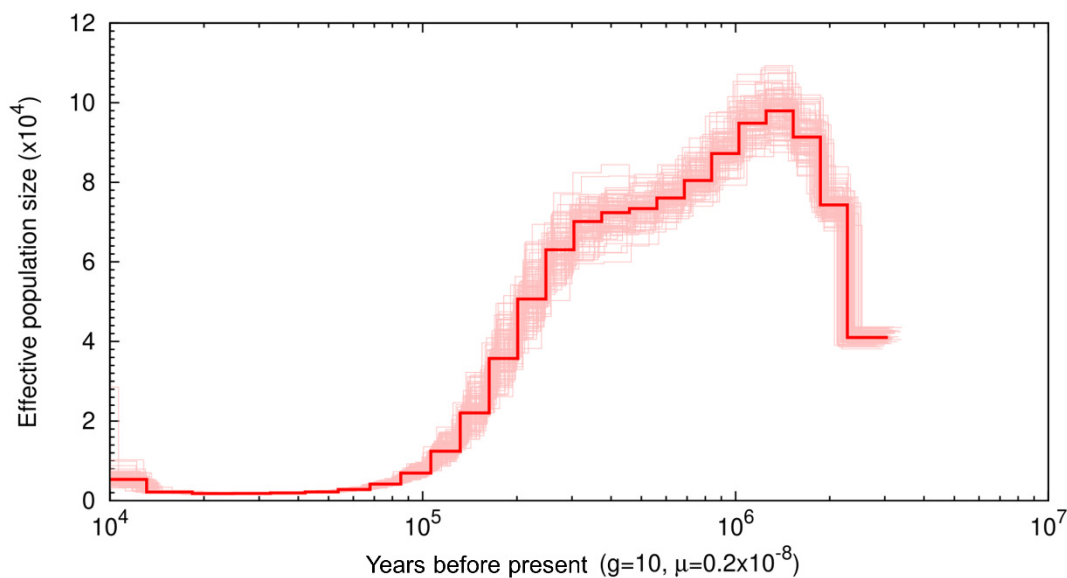


Figure 2. Pairwise sequentially Markovian coalescent (PSMC) plot of the muskox genome with 100 bootstrap repetitions assuming a generation time of 10 and substitution rate of 0.2×10^{-8} . The x-axis represents time before present on a log scale and y axis represents effective population size.

Previous muskox diversity analyses performed by Hansen et al., [14] used ddRAD sequencing to assess neutral genome wide variation in muskox populations, where stairway plot analyses were used to assess N_e through time. The analyses by Hansen et al. [14] of the mainland west population, the closest geographical population to the sample used in this study, showed that N_e was at a stable high of ~13 k at 50 Kya until ~27 Kya when N_e began to steadily decline. Hansen et al. [14] showed the population decline started to plateau at ~15 Kya with N_e estimates remaining at ~9 k at 10 Kya. In comparison to the PSMC analyses performed herein, N_e estimates at 10 Kya are nearly doubled in analyses by Hansen et al. [14]. Additionally, where Hansen et al. [14] found population declines occurred between ~27 Kya and ~15 Kya, the PSMC analyses performed in this study found N_e stable or increasing in size between this time frame. Finally, the high N_e reported in Hansen et al. [14] between ~27 Kya and 50 Kya coincides with an N_e low according to PSMC analyses with N_e estimates $6\times$ higher in Hansen et al. [14]. Discrepancies between these two results likely relate to the difference in reference genomes used to call genotypes, where Hansen et al. [14] used the sheep genome, and the PSMC analyses performed herein used the cross-species scaffolded muskox genome. Prasad et al. [47] performed variant calling using the reference genomes of related species of varying phylogenetic distance and compared them to variants called using cross-species scaffolded genomes. When used for downstream analyses such as PSMCs, analyses using cross-species scaffolded genomes were more reliable than mapping directly to the reference genome of related species, even if

the cross-species genome was highly fragmented [47]. Overall, Prasad et al. [47] found that if a suitable reference genome did not exist for a species of interest, cross-species scaffolding provides a good reference genome alternative. As such, the use of the muskox genome assembled herein may provide a better reference for demographic analyses such as the PSMCs than published sheep and goat genomes.

4. Conclusions

In this study, we present the first draft genome for muskoxen, the only living member of the *Ovibos* genus. A combination of paired-end and in silico mate-pair libraries resulted in an assembly with an N50 of 13 Mb and a BUSCO score of 88.7%. The application of in silico mate pair libraries greatly improved genome scaffolding and genome quality but does not fully replace the need for further sequencing of long read data, mate pair libraries, and/or sequencing by ligation to further enhance the description of the muskox genome. Beyond the development of a draft genome assembly and annotation, these data were used to reconstruct a phylogenetic tree, estimate the divergence time of this unique genus, and assess trends in historical effective population sizes via pairwise sequentially Markovian coalescent (PSMC) analyses. Divergence estimates were consistent with previous studies using mitochondrial DNA, while PSMC analyses found effective population lows coinciding with major glaciation events. With the addition of genomes from diverse muskox populations, future research should include the identification of positively selected genes to gain insight into the muskox's key Arctic adaptations whose genetic underpinnings remain unknown. Overall, these data provide a solid foundation for further genome sequencing to elucidate patterns of gene flow, drift, and selection to better understand the muskox's varying demographic and evolutionary histories and their capacity to adapt to rapid environmental change.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13050809/s1>, Figure S1: Nx and cumulative length plot of the muskox reference genome; Figure S2: Barplot of number of genes with Blast2Go output; Table S1: Library sequencing statistics for reference genome assembly; Table S2: Comparison of Genome assemblies.

Author Contributions: Conceptualization, C.J.K., E.P. and S.K.; methodology, E.P. and C.J.K.; software, E.P.; formal analysis, E.P.; investigation, E.P.; resources, C.J.K., S.K. and L.-M.L.; data curation, E.P.; writing—original draft preparation, E.P.; writing—review and editing, C.J.K. and S.K.; supervision, C.J.K.; funding acquisition, S.K. and C.J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC), ArcticNet, Canada North Outfitting, Polar Knowledge Canada, Canada North Outfitting, Government of Nunavut and the Government of the Northwest Territories.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data sets supporting the results of this article are available on GenBank, under the BioProject (Accession: PRJNA640637).

Acknowledgments: We would like to thank Fabien Mavrot, Matilde Tomaselli, Juliette DiFrancesco, Angie Schneider, and James Wang for their considerable work in the collection of samples, subsampling, archiving and data provisions needed for this work. Thank you to the Hunters and Trappers Organisation (HTO) and hunters from Ulukhaktok, Kugluktuk and Cambridge Bay who participated in the sampling program, the conservation officers, and Canada North Outfitting. We would like to thank M Dedato, A. Shafer, M. Donaldson and S. Vilaca for their insight and assistance. This research was enabled in part by support provided by Compute Canada (www.computecanada.ca; RRG gme-665-ab; accessed 2017–2020).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cuyler, C.; Rowell, J.; Adamczewski, J.; Anderson, M.; Blake, J.; Bretten, T.; Brodeur, V.; Campbell, M.; Checkley, S.L.; Cluff, H.D.; et al. Muskox Status, Recent Variation, and Uncertain Future. *Ambio* **2019**, *49*, 805–819. [[CrossRef](#)] [[PubMed](#)]
2. Mosbacher, J.B.; Michelsen, A.; Stelvig, M.; Hendrichsen, D.K.; Schmidt, N.M. Show Me Your Rump Hair and I Will Tell You What You Ate—The Dietary History of Muskoxen (*Ovibos Moschatus*) Revealed by Sequential Stable Isotope Analysis of Guard Hairs. *PLoS ONE* **2016**, *11*, e0152874. [[CrossRef](#)] [[PubMed](#)]
3. Tomaselli, M.; Gerlach, S.C.; Kutz, S.J.; Checkley, S.L.; Iqaluktutiaq, T.C. Iqaluktutiaq Voices: Local Perspectives about the Importance of Muskoxen, Contemporary and Traditional Use and Practices + Supplementary Appendices S1–S5 (See Article Tools). *ARCTIC* **2018**, *71*, 1–113. [[CrossRef](#)]
4. Kutz, S.; Bollinger, T.; Branigan, M.; Checkley, S.; Davison, T.; Dumond, M.; Elkin, B.; Forde, T.; Hutchins, W.; Niptanatiak, A.; et al. Erysipelothrix Rhusiopathiae Associated with Recent Widespread Muskox Mortalities in the Canadian Arctic. *Can. Vet. J.* **2015**, *56*, 560–563. [[PubMed](#)]
5. Tomaselli, M.; Kutz, S.J.; Gerlach, C.; Checkley, S.L. Local Knowledge to Enhance Wildlife Population Health Surveillance: Conserving Muskoxen and Caribou in the Canadian Arctic. Available online: <https://www.semanticscholar.org/paper/Local-knowledge-to-enhance-wildlife-population-and-Tomaselli-Kutz/b5af5af8200134d2c58a4369cf47b0bca284a2fb> (accessed on 16 April 2020).
6. Tomaselli, M.; Elkin, B.; Kutz, S.; Harms, N.J.; Nymo, H.I.; Davison, T.; Leclerc, L.-M.; Branigan, M.; Dumond, M.; Tryland, M.; et al. A Transdisciplinary Approach to Brucella in Muskoxen of the Western Canadian Arctic 1989–2016. *EcoHealth* **2019**, *16*, 488–501. [[CrossRef](#)] [[PubMed](#)]
7. Tomaselli, M.; Dalton, C.; Duignan, P.J.; Kutz, S.; van der Meer, F.; Kafle, P.; Surujballi, O.; Turcotte, C.; Checkley, S. Contagious Ecthyma, Rangeliferine Brucellosis, and Lungworm Infection in a Muskox (*Ovibos moschatus*) from the Canadian Arctic, 2014. *J. Wildl. Dis.* **2016**, *52*, 719–724. [[CrossRef](#)] [[PubMed](#)]
8. Mavrot, F.; Orsel, K.; Hutchins, W.; Adams, L.G.; Beckmen, K.; Blake, J.E.; Checkley, S.L.; Davison, T.; Francesco, J.D.; Elkin, B.; et al. Novel Insights into Serodiagnosis and Epidemiology of Erysipelothrix Rhusiopathiae, a Newly Recognized Pathogen in Muskoxen (*Ovibos moschatus*). *PLoS ONE* **2020**, *15*, e0231724. [[CrossRef](#)]
9. Prewer, E.; Kutz, S.; Leclerc, L.M.; Kyle, C.J. 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.* **2019**, *129*, 459–469. [[CrossRef](#)]
10. MacPhee, R.D.; Tikhonov, A.N.; Mol, D.; Greenwood, A.D. Late Quaternary Loss of Genetic Diversity in Muskox (*Ovibos*). *BMC Evol. Biol.* **2005**, *5*, 49. [[CrossRef](#)]
11. Groves, P.; Shields, G.F. CytochromeB Sequences Suggest Convergent Evolution of the Asian Takin and Arctic Muskox. *Mol. Phylogenet. Evol.* **1997**, *8*, 363–374. [[CrossRef](#)]
12. Holm, L.E.; Forchhammer, M.C.; Boomsma, J.J. Low Genetic Variation in Muskoxen (*Ovibos moschatus*) from Western Greenland Using Microsatellites. *Mol. Ecol.* **1999**, *8*, 675–679. [[CrossRef](#)] [[PubMed](#)]
13. Van Coeverden, P.J.; Groot, D.; Boag, P. Optimization of Novel Polymorphic Microsatellites in Muskox (*Ovibos moschatus*) Leads to an Increased Estimate of Muskox Microsatellite Diversity. *Mol. Ecol. Notes* **2004**, *4*, 713–715. [[CrossRef](#)]
14. Hansen, C.C.R.; Hvilsom, C.; Schmidt, N.M.; Aastrup, P.; Van Coeverden de Groot, P.J.; Siegismund, H.R.; Heller, R. The Muskox Lost a Substantial Part of Its Genetic Diversity on Its Long Road to Greenland. *Curr. Biol. CB* **2018**, *28*, 4022–4028.e5. [[CrossRef](#)] [[PubMed](#)]
15. Brix, O.; Bårdgard, A.; Mathisen, S.; Tyler, N.; Nuutinen, M.; Condo, S.G.; Giardina, B. Oxygen Transport in the Blood of Arctic Mammals: Adaptation to Local Heterothermia. *J. Comp. Physiol. B* **1990**, *159*, 655–660. [[CrossRef](#)]
16. Adamczewski, J.Z.; Kerr, W.M.; Lammerding, E.F.; Flood, P.F. Digestion of Low-Protein Grass Hay by Muskoxen and Cattle. *J. Wildl. Manag.* **1994**, *58*, 679–685. [[CrossRef](#)]
17. Adamczewski, J.Z.; Flood, P.F.; Chaplin, R.K.; Schaefer, J.A. Seasonal Variation in Intake and Digestion of a High-Roughage Diet by Muskoxen. *Can. J. Anim. Sci.* **1994**, *74*, 305–313. [[CrossRef](#)]
18. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 8 June 2017).
19. Scythe—A Bayesian Adapter Trimmer (Version 0.994 BETA). Available online: <https://github.com/vsbuffalo/scythe> (accessed on 23 December 2019).
20. Aronesty, E. Ea-Utils: Command-Line Tools for Processing Biological Sequencing Data. 2011. Available online: <https://github.com/ExpressionAnalysis/ea-utils> (accessed on 8 June 2017).
21. Joshi, N.; Fass, J. Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33). 2011. Available online: <https://github.com/najoshi/sickle> (accessed on 8 June 2017).
22. Xu, H.; Luo, X.; Qian, J.; Pang, X.; Song, J.; Qian, G.; Chen, J.; Chen, S. FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. *PLoS ONE* **2012**, *7*, e52249. [[CrossRef](#)]
23. Bushnell, B. BMAP: A Fast, Accurate, Splice-Aware Aligner. 2014. Available online: <https://www.osti.gov/servlets/purl/1241166> (accessed on 11 May 2017).
24. Schmieder, R.; Edwards, R. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. *PLoS ONE* **2011**, *6*, e17288. [[CrossRef](#)]

25. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler. *GigaScience* **2012**, *1*, 18. [CrossRef]
26. Grau, J.H.; Hackl, T.; Koepfli, K.-P.; Hofreiter, M. Improving Draft Genome Contiguity with Reference-Derived in Silico Mate-Pair Libraries. *GigaScience* **2018**, *7*, giy029. [CrossRef]
27. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinforma. Oxf. Engl.* **2013**, *29*, 1072–1075. [CrossRef] [PubMed]
28. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinforma. Oxf. Engl.* **2015**, *31*, 3210–3212. [CrossRef] [PubMed]
29. Smit, A.; Hubley, R. RepeatModeler Open-1.0. 2008. Available online: <http://www.repeatmasker.org> (accessed on 14 August 2017).
30. Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2013. Available online: <http://www.repeatmasker.org> (accessed on 14 August 2017).
31. Stanke, M.; Morgenstern, B. AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints. *Nucleic Acids Res.* **2005**, *33*, W465–W467. [CrossRef] [PubMed]
32. Wu, T.D.; Watanabe, C.K. GMAP: A Genomic Mapping and Alignment Program for MRNA and EST Sequences. *Bioinformatics* **2005**, *21*, 1859–1875. [CrossRef]
33. Besemer, J.; Borodovsky, M. GeneMark: Web Software for Gene Finding in Prokaryotes, Eukaryotes and Viruses. *Nucleic Acids Res.* **2005**, *33*, W451–W454. [CrossRef]
34. Slater, G.S.C.; Birney, E. Automated Generation of Heuristics for Biological Sequence Comparison. *BMC Bioinform.* **2005**, *6*, 31. [CrossRef]
35. Haas, B.J.; Salzberg, S.L.; Zhu, W.; Pertea, M.; Allen, J.E.; Orvis, J.; White, O.; Buell, C.R.; Wortman, J.R. Automated Eukaryotic Gene Structure Annotation Using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **2008**, *9*, R7. [CrossRef]
36. Li, H.; Durbin, R. Inference of Human Population History from Individual Whole-Genome Sequences. *Nature* **2011**, *475*, 493–496. [CrossRef]
37. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
38. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinforma. Oxf. Engl.* **2009**, *25*, 2078–2079. [CrossRef]
39. Kumar, S.; Subramanian, S. Mutation Rates in Mammalian Genomes. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 803–808. [CrossRef] [PubMed]
40. Tabari, E.; Su, Z. PorthomCL: Parallel Orthology Prediction Using MCL for the Realm of Massive Genome Availability. *Big Data Anal.* **2017**, *2*, 4. [CrossRef] [PubMed]
41. Kück, P.; Meusemann, K. FASconCAT: Convenient Handling of Data Matrices. *Mol. Phylogenet. Evol.* **2010**, *56*, 1115–1118. [CrossRef] [PubMed]
42. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. JModelTest 2: More Models, New Heuristics and High-Performance Computing. *Nat. Methods* **2012**, *9*, 772. [CrossRef] [PubMed]
43. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]
44. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]
45. Chen, L.; Qiu, Q.; Jiang, Y.; Wang, K.; Lin, Z.; Li, Z.; Bibi, F.; Yang, Y.; Wang, J.; Nie, W.; et al. Large-Scale Ruminant Genome Sequencing Provides Insights into Their Evolution and Distinct Traits. *Science* **2019**, *364*, eaav6202. [CrossRef]
46. Benton, M.J.; Donoghue, P.C.J. Paleontological Evidence to Date the Tree of Life. *Mol. Biol. Evol.* **2007**, *24*, 26–53. [CrossRef]
47. Prasad, A.; Lorenzen, E.D.; Westbury, M.V. Evaluating the Role of Reference-Genome Phylogenetic Distance on Evolutionary Inference. *Mol. Ecol. Resour.* **2022**, *22*, 45–55. [CrossRef]
48. Westbury, M.V.; Petersen, B.; Garde, E.; Heide-Jørgensen, M.P.; Lorenzen, E.D. Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size. *iScience* **2019**, *15*, 592–599. [CrossRef]
49. Westbury, M.V.; Petersen, B.; Lorenzen, E.D. Genomic Analyses Reveal an Absence of Contemporary Introgressive Admixture between Fin Whales and Blue Whales, despite Known Hybrids. *PLoS ONE* **2019**, *14*, e0222004. [CrossRef] [PubMed]
50. Westbury, M.V.; Thompson, K.F.; Louis, M.; Cabrera, A.A.; Skovrind, M.; Castruita, J.a.S.; Constantine, R.; Stevens, J.R.; Lorenzen, E.D. Ocean-Wide Genomic Variation in Gray’s Beaked Whales, *Mesoplodon Grayi*. *R. Soc. Open Sci.* **2021**, *8*, 201788. [CrossRef] [PubMed]
51. Hempel, E.; Westbury, M.V.; Grau, J.H.; Trinks, A.; Paijmans, J.L.A.; Kliver, S.; Barlow, A.; Mayer, F.; Müller, J.; Chen, L.; et al. Diversity and Paleodemography of the Addax (*Addax nasomaculatus*), a Saharan Antelope on the Verge of Extinction. *Genes* **2021**, *12*, 1236. [CrossRef] [PubMed]
52. Biltueva, L.S.; Sharshov, A.A.; Graphodatsky, A.S. G-Banding Homologies in Musk Ox, *Ovibos Moschatus*, and Other Bovids. *Hereditas* **1995**, *122*, 185–187. [CrossRef]
53. Desaulniers, D.M.; King, W.A.; Rowell, J.E.; Flood, P.F. The Banded Chromosomes of the Muskox (*Ovibos moschatus*). *Can. J. Zool.* **1989**, *67*, 1155–1158. [CrossRef]

54. Pasitschniak-Arts, M.; Flood, P.F.; Schmutz, S.M.; Seidel, B. A Comparison of G-Band Patterns of the Muskox and Takin and Their Evolutionary Relationship to Sheep. *J. Hered.* **1994**, *85*, 143–147. [[CrossRef](#)]
55. Proskuryakova, A.A.; Kulemzina, A.I.; Perelman, P.L.; Yudkin, D.V.; Lemskaya, N.A.; Okhlopkov, I.M.; Kirillin, E.V.; Farré, M.; Larkin, D.M.; Roelke-Parker, M.E.; et al. Comparative Chromosome Mapping of Musk Ox and the X Chromosome among Some Bovidae Species. *Genes* **2019**, *10*, 857. [[CrossRef](#)]
56. Shafer, A.B.A.; Hall, J.C. Placing the Mountain Goat: A Total Evidence Approach to Testing Alternative Hypotheses. *Mol. Phylogenet. Evol.* **2010**, *55*, 18–25. [[CrossRef](#)]
57. Pasitschniak-Arts, M.; Flood, P.F.; Schmutz, S.M.; Tedesco, S.; Seidel, B. The Phylogenetic Relationship of the Muskox and Takin Based on High Resolution, G-Banded, Chromosome Analysis. *Rangifer* **1992**, *12*, 203–205. [[CrossRef](#)]
58. Bibi, F. A Multi-Calibrated Mitochondrial Phylogeny of Extant Bovidae (Artiodactyla, Ruminantia) and the Importance of the Fossil Record to Systematics. *BMC Evol. Biol.* **2013**, *13*, 166. [[CrossRef](#)]
59. Toljagić, O.; Voje, K.L.; Matschiner, M.; Liow, L.H.; Hansen, T.F. Millions of Years Behind: Slow Adaptation of Ruminants to Grasslands. *Syst. Biol.* **2018**, *67*, 145–157. [[CrossRef](#)] [[PubMed](#)]
60. Springer, M.S.; DeBry, R.W.; Douady, C.; Amrine, H.M.; Madsen, O.; de Jong, W.W.; Stanhope, M.J. Mitochondrial Versus Nuclear Gene Sequences in Deep-Level Mammalian Phylogeny Reconstruction. *Mol. Biol. Evol.* **2001**, *18*, 132–143. [[CrossRef](#)] [[PubMed](#)]
61. Bartoli, G.; Sarnthein, M.; Weinelt, M.; Erlenkeuser, H.; Garbe-Schönberg, D.; Lea, D.W. Final Closure of Panama and the Onset of Northern Hemisphere Glaciation. *Earth Planet. Sci. Lett.* **2005**, *237*, 33–44. [[CrossRef](#)]
62. Clark, P.U.; Dyke, A.S.; Shakun, J.D.; Carlson, A.E.; Clark, J.; Wohlfarth, B.; Mitrovica, J.X.; Hostetler, S.W.; McCabe, A.M. The Last Glacial Maximum. *Science* **2009**, *325*, 710–714. [[CrossRef](#)]