Received: 17 May 2021 | Revised: 1 September 2021 | Accepted: 7 September 2021

DOI: 10.1111/mec.16184

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep

Maulik Upadhyay[1] | Elisabeth Kunz[1] | Edson Sandoval-Castellanos[1] | Andreas Hauser[2] | Stefan Krebs[2] | Alexander Graf[2] | Helmut Blum[2] | Arsen Dotsev[3] | Innokentiy Okhlopkov[4] | Alexey Shakhin[3] | Vugar Bagirov[3] | Gottfried Brem[5] | Ruedi Fries[6] | Natalia Zinovieva[3] | Ivica Medugorac[1]

[1]Population Genomics Group, Department of Veterinary Sciences, LMU Munich, Munich, Germany

[2]Laboratory for Functional Genome Analysis, Gene Center, LMU Munich, Munich, Germany

[3]L.K. Ernst Federal Science Center for Animal Husbandry, Podolsk, Russia

[4]Institute for Biological Problems of Cryolithozone, Yakutsk, Russia

[5]Institute of Animal Breeding and Genetics, VMU, Vienna, Austria

[6]Lehrstuhl für Tierzucht, Technische Universität München, Freising, Germany

Correspondence
Natalia Zinovieva, L.K. Ernst Federal Science Center for Animal Husbandry, Podolsk, Russia.
Email: n_zinovieva@mail.ru

Ivica Medugorac, Population Genomics Group, Department of Veterinary Sciences, LMU Munich, Martinsried, Germany.
Email: Ivica.Medjugorac@gen.vetmed.uni-muenchen.de

Funding information
This work was partly supported by the Russian Science Foundation within Project No. 21-66-00007 (molecular genetic studies) and by the Ministry of Science and Higher Education of the Russian Federation within Theme No. 0445-2019-0024 (expedition studies). This work was also carried out by the support of LMUexcellent funding of LMU Munich.

Abstract

To predict species responses to anthropogenic disturbances and climate change, it is reasonable to use species with high sensitivity to such factors. Snow sheep (*Ovis nivicola*) could represent a good candidate for this; as the only large herbivore species adapted to the cold and alpine habitats of northeastern Siberia, it plays a crucial role in its ecosystem. Despite having an extensive geographical distribution among all ovine species, it is one of the least studied. In this study, we sequenced and analysed six genomes of snow sheep in combination with all other wild sheep species to infer key aspects of their evolutionary history and unveil the genetic basis of their adaptation to subarctic environments. Despite their large census population size, snow sheep genomes showed remarkably low heterozygosity, which could reflect the effect of isolation and historical bottlenecks that we inferred using the pairwise sequential Markovian coalescent and runs of homozygosity. F_4-statistics indicated instances of introgression involving snow sheep with argali (*Ovis ammon*) and Dall (*Ovis dalli*) sheep, suggesting that these species might have been more widespread during the Pleistocene. Furthermore, the introgressed segments, which were identified using mainly minimum relative node depth, covered genes associated with immunity, adipogenesis and morphology-related traits, representing potential targets of adaptive introgression. Genes related to mitochondrial functions and thermogenesis associated with adipose tissue were identified to be under selection. Overall, our data suggest introgression as a mechanism facilitating adaptation in wild sheep species and provide insights into the genetic mechanisms underlying cold adaptation in snow sheep.

KEYWORDS
Ovis nivicola, whole genome sequencing, hybridization, cold adaptation

1 | INTRODUCTION

Anthropogenic disturbances, including climate change, constitute the main causes of habitat loss for arctic and subarctic species. Large cold-adapted mammals are particularly vulnerable because of their reduced population sizes and relatively limited capacity to migrate (in comparison to, for example, birds). Studying the evolutionary history and genetic diversity of representative species can help with evaluating the species' survival potential when under anthropogenic pressures and to establish conservation policies or interventions for other Holarctic species.

The snow sheep is the largest herbivorous mammal in the mountains of northeast Asia. It is adapted to upper mountain habitats above the forest belt, where, owing to the excellent insulating properties of fur and fat reserves combined with an optimal choice of shelters, they can endure temperatures as low as −60 to −50°C with strong winds (Zheleznov-Chukotsky, 1994). However, slow reproduction rates, disjointed habitat zones and low population densities make this species vulnerable to environmental change. Moreover, snow sheep cannot withstand even small anthropogenic loads, and one of the reasons for its mountainous distribution is centuries-old persecution by humans (Zheleznov-Chukotsky, 1994). Another reason for the current snow sheep distribution is its dependence on crooked vegetation (e.g., cedar elfin, stone birch, willow), which are being pushed to higher altitudes by ongoing global warming, causing a significant decrease of available habitat (Ermolin & Medvedev, 2020). Snow sheep are so reactive to habitat change that some populations that had lived in hills and low mountain systems have disappeared within the last 35 years (Zheleznov-Chukotsky, 2007). Due to all these factors, snow sheep can serve as a representative species for studying the effects of climate change and anthropogenic disturbances on cold-adapted mammals. Furthermore, despite having the largest distribution range and population size of all wild sheep, snow sheep are the least studied species of the genus, *Ovis*.

The genus *Ovis* comprises species adapted to a wide range of habitats. Seven species have been identified, based on body size, horn morphology, coat colour and patterns, the number of chromosomes, and geographical distribution (Rezaei et al., 2010). The European mouflon (*Ovis musimon*) and the Asiatic mouflon (*Ovis orientalis*), both with 54 (2*n*) chromosomes, are dispersed across Europe and western Asia. The Asiatic mouflon is the wild ancestor of domestic sheep (*Ovis aries*, 2*n* = 54). The argali sheep (*Ovis ammon*, 2*n* = 56) is found in the mountainous areas of central Asia, while the urial sheep (*Ovis vignei*, 2*n* = 58) is found throughout the areas southwest of the Caspian Sea. The Dall sheep, or thinhorn sheep (*Ovis dalli*, 2*n* = 54), is native to the mountainous regions of northwestern America, and the bighorn sheep (*Ovis canadensis*, 2*n* = 54) lives in the rocky mountainous regions of the United States, Canada and Mexico. Snow sheep (*Ovis nivicola*, 2*n* = 52) mainly inhabit the mountainous areas of northeastern Siberia.

Snow sheep, Dall sheep and bighorn sheep have been proposed to form the subgenus *Pachyceros*, whose origin is explained by two main hypotheses: Cowan (1940) proposed that American wild sheep (Dall and bighorn sheep) evolved in Beringia (a continuous land bridging present-day eastern Siberia and Alaska) from snow-sheep-like species that expanded into North America. The other hypothesis suggests that snow sheep and North American wild sheep evolved from argali-like species. Analyses of mitochondrial DNA (mtDNA; Bunch et al., 2006) have supported Cowan's hypothesis, but this phylogenetic inference could be inaccurate due to incomplete lineage sorting (ILS) and introgression.

Introgression, including adaptive introgression, is increasingly being recognized as an important source of diversity and adaptation across taxa (Arnold & Kunte, 2017). In sheep, introgression from wild species into domestic sheep has been recorded by numerous studies that have identified regions containing genes associated with climatic adaptation and immunity (Barbato et al., 2017; Cao et al., 2020). For instance, alleles in the *PKN2* gene that were introgressed from argali sheep into Chinese domestic breeds correlated with resistance to bacterial pneumonia (Cao et al., 2020). This suggests that introgression between wild sheep species is probable, and whole genome sequencing (WGS) data would answer persistent questions about the phylogeny of *Ovis* and allow for creation of the first map of introgression among wild sheep.

Mammalian species adapted to arctic and antarctic conditions survive these extremely cold environments by using the three basic strategies of shivering and nonshivering thermogenesis (NST), increasing subcutaneous adipose tissue, and hibernation (Yudin et al., 2017). While the molecular mechanisms contributing to these physiological changes are still poorly understood, it stands to reason that these changes must have resulted from the accumulation of genetic changes occurring in parallel in diverse gene families. Indeed, several gene sets associated with cold adaptation have been identified in a variety of species (Yudin et al., 2017), including polar bears and Yakutian horses (Librado et al., 2015; Welch et al., 2014). Characterizing regions under selection in snow sheep will contribute to understanding the mechanisms behind adaptations to cold environments.

In our study, we analysed new WGS data from six snow sheep along with published data from 32 sheep specimens to (i) characterize their genetic diversity and demographic history, (ii) resolve the phylogenetic relationships and introgression patterns to shed light on specific questions such as the origin *of Pachyceros*, and (iii) identify regions and genes under selection in snow sheep and discuss their potential role in mechanisms of adaptation to cold.

2 | MATERIALS AND METHODS

2.1 | Sample collection

For WGS, six individuals from the Yakut snow sheep subspecies (Figure 1a) were sampled from different habitats throughout the Verkhoyansk Mountain chain (Figure 1b): Kharaulakh Ridge, Tiksi (TIK), Orulgan Ridge (ORU), Central Verkhoyansk Ridge (VER), Chersky Range (CHE; also Tas-Kystabyt), Sarychev Ridge (SAR) and

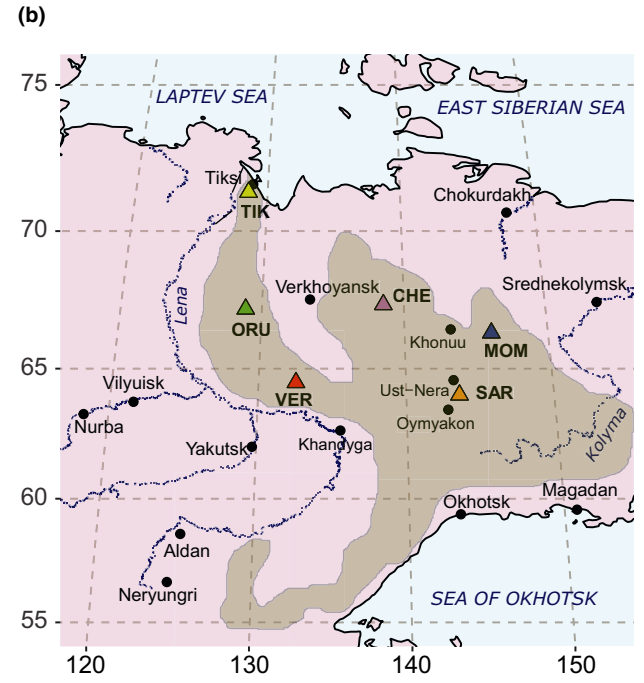

FIGURE 1 (a) Photo showing a group of snow sheep. (b) Sampling locations of snow sheep samples used in the present study. Different colour-filled triangles refer to the approximate location of the sampling sites: yellow colour-filled triangle marks the location of Tiksi (TIK), green Orulgan Ridge (ORU), red Central Verkhoyansk Ridge (VER), pink Chersky Range (CHE), blue Momsky Ridge (MOM), and yellow colour-filled tringle marks the location of Sarychev Ridge (SAR)

Momsky Ridge (MOM). The sampling locations were selected after examination of aerial observations made in 2008–2010 in an area between 63°N and 71°N and between 128°E and 146°E.

2.2 | Illumina whole genome sequencing

The library construction for Illumina sequencing was carried out with 200 ng of genomic DNA following the Illumina library preparation protocols (NexteraFlex DNA library kit; Illumina). The resulting amplified library was quantified and controlled on an Agilent Bioanalyzer 2100 (Agilent) and sequenced in 2 × 100-bp paired-end mode on an Illumina HiSeq1500, yielding ~200 million read pairs.

2.3 | WGS alignment, SNP calling, filtering and annotation

We collated the six snow sheep WGS data with additional 32 samples representing all ovine species and one outgroup sample of domestic goat. These 33 samples were downloaded from the NCBI SRA. All the WGS data were trimmed with SICKLE using default parameters except that the length threshold was set to 50 bp (Joshi & Fass, 2011). The trimmed reads were aligned onto the autosomes of the domestic sheep assembly Oar_v4.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2) using the default parameters of the BWA-MEM algorithm in BWA version 0.7.17 (Li, 2013). Following the alignment, PICARD's MarkDuplicates module was used to remove the duplicated reads. Local realignment was carried out using the RealignerTargetCreator and IndelRealigner modules of GATK version 3.8.0. In the first step, all variants were called using the command *bcftools mpileup* on all the samples simultaneously and using the parameters of -q 20 -Q 20 -C 50 -I. In the next step, we filtered the single nucleotide polymorphisms (SNPs) using these criteria: (i) remove multi-allelic SNPs and (ii) remove SNPs with mapping and base quality less than 40. We set genotypes of samples as missing if they displayed a depth greater than 3*X* or less than *X*/3, where *X* was the average genome-wide depth of the respective sample. Annotations of the filtered SNPs were carried out using SNPEFF (Cingolani et al., 2012).

2.4 | Estimation of heterozygosity

Genetic heterozygosity was estimated directly from the bam file of each individual using MLRHO version 2.9 (Haubold et al., 2010). The method implemented in MLRHO co-estimates various parameters, including the population mutation rate and sequencing error rate. For this analysis, we used only the sites with a mapping quality and a base quality greater than 20. Additionally, we also estimated the

genome-wide heterozygosity by estimating the Θ-Watterson estimator in the bin of 100-kb nonoverlapping blocks using the Python module SCIKIT-ALLEL version 1.3.2 (Miles et al., 2020). Genetic diversity was also assessed as a count of heterozygous sites per sample per 100 kbp. The heterozygous sites were counted for the entire autosomal genome in sliding windows of 100 kbp with a step size of 10 kbp. We only considered those windows where more than 80% of the sites were called with high confidence. Further runs of homozygosity (ROH) were identified using the command *bcftools roh* with default settings (Narasimhan et al., 2016).

2.5 | Demographic history

We carried out a pairwise sequentially Markovian coalescent (PSMC) analysis (Li & Durbin, 2011) on individuals with the depth ≥8× to estimate the demographic history back to 10,000 years ago. For this purpose, a consensus sequence of each bam file was built as described previously (Li & Durbin, 2011). In brief, first, SAMTOOLS mpileup command was run with the -C50 option to minimize the effect of reads having excessive mismatches; lastly, a vcfutils.pl script was used to convert a vcf file (generated in the first step) to fastq format after removing sites with less than a third or more than a third of the average depth of the sample. The analysis was run using these parameters: -N30 -t15 -r5 -p '4 + 25*2 + 4 + 6. To convert coalescent time to years, we set the average mutation rate (μ) as 2.5×10^{-8} per base per generation (Li et al., 2020) and the generation interval to 4 years.

2.6 | Phylogenetic relationships within the genus Ovis

Genetic clustering was carried out using pairwise identity by state (IBS) distances between every possible pair of samples. IBS distances were calculated using PLINK (Purcell et al., 2007). Subsequently, this distance matrix was imported using the PAUP tool (Swofford, 2003) to construct neighbour joining (NJ)-based clustering with goat as an outgroup. Next, we used a maximum likelihood-based approach as implemented in TREEMIX (Pickrell & Pritchard, 2012) to construct a phylogenetic tree after grouping the samples into the respective populations. TREEMIX was run 100 times (100 bootstrapping with blocks of 5000 SNPs) using these parameters: -k 5000, -noss, -bootstrap. Subsequently, the Consense module of the PHYLIP package (Felsenstein, 1989) was used to generate a consensus phylogenetic tree.

2.7 | Divergence time estimation using *MCMCtree*

Divergence times within the genus *Ovis* were estimated using the MCMCtree module of PAML (dos Reis & Yang, 2019; Yang, 2007). Because MCMCtree works only with one sample per species, we randomly selected one sample per species, including the outgroup. First, a consensus sequence for each of these eight samples was generated; for this, the command *bcftools mpileup* was ran in multiple-sample mode within each species separately to call genotype at each position, and a consensus sequence was then created for a single individual per species by considering these parameters: (i) sites with mapping quality less than or equal to 40, (ii) sites with depth less than a third or greater than a third of the average coverage of a sample was set to missing, and (iii) heterozygous SNPs were replaced by their respective IUPAC nucleotide codes. Subsequently, an alignment file was prepared by concatenating the coding sequences of one-to-one orthologues of goat and sheep. Next, a phylogenetic tree was inferred using the default settings in RAXML-NG (Kozlov et al., 2019; Kozlov & Stamatakis, 2019). We chose the GTR + G substitution model because of its generality and based on published suggestions (Abadi et al., 2019). The concatenated coding sequences and the phylogenetic tree were used as inputs for the MCMCtree analysis. Divergence times were estimated using the approximate likelihood function as previously described (dos Reis & Yang, 2019). In the first step, branch lengths, their gradients, and the Hessian of the log likelihood (usedata = 3) were estimated using the HKY + Γ model of nucleotide sequence evolution. We employed the options of gradient estimation and Hessian matrix (usedata = 2) for Markov chain Monte Carlo (MCMC) sampling of the posterior distribution. The independent log-normal clock model (clock = 2) with priors was used for the node ages to be uniformly distributed (BDparas = 1 1 0). The prior for the substitution rates was gamma distributed with a mean of 0.05 substitutions per 100 million years (rgene_gamma = 2 40 1). The rate variance parameter was set as sigma2_gamma = 1 10 1. The MCMC chain was run for 50,000,000 iterations and sampled every 100 iterations, with the first 500,000 iterations discarded as burn-in. One internal fossil calibration was used, a recently identified fossil (Wang et al., 2016) from an extinct mountain sheep that lived ~3.2–5.4 million years ago (Ma). Therefore, the most recent common ancestor (MRCA) of the genus *Ovis* was set with a uniform prior at 3.2–5.4 Ma with a probability of violating the upper bound of 0.20.

2.8 | Investigation of introgression hypotheses using F_4 statistics, and testing for incomplete lineage sorting

To test the hypotheses of introgression in the genus *Ovis* involving the snow sheep lineage, we estimated the F_4 statistic. This statistic was first introduced by Reich et al. (2009) and is a statistically powerful measure for differentiating between ILS and introgression. It assumes the unrooted tree topology ((A, B), (C, D)), where A–B and C–D are pairs of sister taxa. A significant deviation from zero is an indication of introgression. Guided by the inferred phylogeny, we tested introgression between snow sheep and Dall or bighorn sheep (A and B), with the remaining sheep species in C and D. Furthermore, we replaced D with the outgroup (O), or, in other words, ((A, B), (C, O)), in the quadruplet in which the statistic indicated deviation from

a tree-like model of the tree topology. This was done to identify the exact lineage (i.e., C or D) responsible for deviation from a tree-like pattern or to identify whether C or D shared gene flow with either A or B. These hypotheses are represented in the form of phylogenetic trees. The analysis was carried out on the 241,026 SNPs that remained after filtering out SNPs that were a distance from each other of <10 kb (Liu et al., 2019).

To rule out the hypothesis of ILS, the significance of the statistic was assessed by a *Z*-value obtained from block jackknifing, as well as with a simulation-based approach as implemented in the F_4 tool (Meyer et al., 2016). The F_4 tool uses coalescent simulation to produce the sequence data sets of the quadruplet A, B, C and D with similar sizes and amounts of missing data as the empirical data. The demographic model implemented by the F_4 tool does not assume migration, and parameters (divergence times and population size) are optimized until the simulated sequences are similar to the observed data. We ran 1000 such simulations and reported the probability (p_{sim}) as the proportion of simulated data sets producing an F_4 statistic at least as extreme as that observed.

2.9 | Identification of introgressed segments

To identify the introgressed genomic segments, variations of the relative node depth (RND) were estimated along the genome (Feder et al., 2005; Rosenzweig et al., 2016). BEAGLE 5.1 was used to phase the filtered data (Browning et al., 2018). Using a custom Python script, we first calculated nucleotide divergence (D_{xy}) between each haplotype of snow sheep and argali in the nonoverlapping window size of 50 kbp. Subsequently, the minimum nucleotide divergence value (D_{min}) for each window was normalized by the average distance (D_{out}) between each of the wild sheep (argali, snow sheep and the outgroup) to calculate the minimum RND (RND_{min}). The introgressed segments in the genotypes are expected to show exceptionally low RND_{min} values. Only those segments that displayed RND_{min} values <2× the standard deviation from the mean were used as putative introgressed segments (I_{ps}). To compare the observed RND_{min} values against that of the values that could be obtained under the null hypothesis of isolation with no migration, we simulated the data following this study (Liu et al., 2019). In brief, considering the divergence time estimates of MCMCtree, constant effective population size of 100,000 individuals for all eight species (seven sheep species and an outgroup), and a generation interval of 4 years, we simulated 1000 loci of 100 kbp using MOMI2 (Kamm et al., 2020) with the demographic model that does not consider migration. The RND_{min} values of these simulated data were used to assess significance values of the empirical RND_{min} data. However, we also note that such simulations normally require knowledge of detailed and precise demographic parameters. Therefore, inaccurate assumptions about the effective population size coupled with unknown bottleneck/expansion events may lead to under-/over-representation of ILS. Furthermore, for each I_{ps}, an NJ algorithm based on F_{ST} distances was applied to assess the clustering pattern of the species involved in introgression. Additionally, a five-taxon-based phylogenic test as implemented in DFOIL (Pease & Hahn, 2015) was also carried out for each I_{ps} to assess the direction of introgression. DFOIL analysis was carried out on a consensus sequence generated for each species separately considering the multiple samples and these parameters: (i) sites with mapping quality less than or equal to 40, (ii) sites with depth less than a third or greater than a third of the average pooled samples were set to missing, and (iii) heterozygous SNPs were replaced by their respective IUPAC nucleotide codes. A segment was only considered introgressed if that conclusion was supported by all three of the above-mentioned approaches (RND_{min}, NJ-based genetic clustering and DFOIL statistics).

2.10 | Selection scans

To identify the genomics segments under selection, we first applied the branch-site model as implemented in CODEML (Yang, 2007). For this purpose, protein sequences from a slightly improved annotation of the snow sheep genome were used (Note S1; Table S1); the transcriptome sequences that are used for the purpose of annotation have been described previously (Upadhyay et al., 2020). Protein-coding sequences from nine additional species (Table S2; opossum, human, horse, dog, pig, cattle, goat, sheep and Marco Polo sheep) were downloaded from Ensembl and NCBI, and the domestic sheep sequences (reference, Oar_v4.0) were downloaded from NCBI. Subsequently, for each of these species, the protein sequences of the longest isoform per gene were extracted. Next, the ORTHOFINDER tool (Emms & Kelly, 2019) was used to identify one-to-one orthologues across these species. Later, coding sequences (CDS) of these orthologous proteins were aligned using the codon mode in PRANK (Löytynoja, 2014). The branch-site model requires an a priori phylogenetic tree; therefore, fourfold degenerate sites were extracted from the aligned CDS, and RAXML-NG was used to reconstruct the phylogenetic tree. To identify a positive selection in the snow sheep lineage, a model was run with snow sheep as the foreground branch and allowing some sites to be under selection. Likelihood ratio test (LRT) was calculated by comparing this Model A vs. the Model A_{Null}, in which neutral evolution/negative selection were assumed on the foreground branch. To reduce the false positives, these steps were followed: (i) hard-trimming of the alignment was carried out using GBLOCKS (Castresana, 2000), (ii) next, orthologous sequences in the domestic sheep genome were searched for the CDS identified under selection in snow sheep using BLAT, and (iii) finally, it was verified whether the Illumina WGS of snow sheep aligned with the domestic sheep genome and if the domestic sheep genome had the same variants that were present in the snow sheep genome assembly. Furthermore, we only considered those genes for which CODEML also output the sites under selection with a Bayes empirical Bayes (BEB) value >0.85.

Additionally, Fay and Wu's *H* test (FWH) was also carried out to identify genomic segments under selection in snow sheep. FWH statistics were estimated directly from the alignment files of snow

sheep samples using the ANGSD tool (Korneliussen et al., 2014). FWH was calculated in windows of 50 kbp with 10 kbp as the step size, and the consensus sequence of goat was used as an outgroup. The top 0.1% segments (245 of ~245,000) showing the most negative values of FWH were selected as putative segments under selection. Furthermore, for these segments, the *p*-values for FWH were also estimated using coalescent simulations as implemented in Zeng's DH program (Zeng et al., 2006).

3 | RESULTS

3.1 | Characterizing genetic diversity and demographic history

We carried out WGS on six snow sheep samples and collated these data with an additional 32 whole genome sequences (Table S3) belonging to all wild species of the genus *Ovis* and one goat sample to be used as an outgroup. After aligning these data against the domestic sheep reference genome, Oar_v4.0, and applying stringent filtering criteria, we obtained genome-wide coverage in the range of ~4.3 to ~16.6 (Table S3), with an average of ~11.4 for all 38 samples (excluding the outgroup sample). The average genome-wide coverage for snow sheep was about 13×. By using *bcftools mpileup* for variant calling, applying stringent filtering criteria and removing goat-specific single nucleotide variants (SNVs), we identified ~56.4 million SNVs across the genus *Ovis*. Within snow sheep, a total of ~18.6 million SNVs were identified. Similar to what is shown in Figure 2a, the greatest number of shared SNVs was with Dall sheep (shared SNPs, ~13.3 million) and bighorn sheep (shared SNPs, ~13.5 million). A total of ~2.4 million SNVs were specific to snow sheep, ~1.6 million of which were heterozygous, while the remaining ~0.8 million SNVs were found only in homozygous form. Of these ~0.8 million SNVs, ~0.3 million were completely fixed in the snow sheep lineage. It is noteworthy that this number of snow-sheep-specific SNVs may be slightly inflated because of the stringent filtering criteria, possibly because it is likely that a small number of SNVs designated as missing in other *Ovis* lineages and identified in snow sheep samples might also have been designated as snow-sheep-specific. Nevertheless, these species-specific SNVs represent valuable resources for wildlife management of snow sheep.

Nucleotide diversity was assessed using genome-wide heterozygosity values inferred from the Θ-Watterson estimator in 100-kb blocks. The results (Figure 2b) revealed that genome-wide diversity varies considerably across sheep species. The estimators suggested overall low diversity in snow sheep (0.0005 ± 0.0002), Dall sheep (0.0003 ± 0.0002) and bighorn sheep (0.0004 ± 0.0002), while the Asiatic mouflon (0.0032 ± 0.0008), domestic sheep (0.0023 ± 0.0007) and urial sheep (0.0026 ± 0.0007) harboured significantly high diversity. Because the sample sizes varied across the different sheep species, mean heterozygosity was also estimated across the individual genomes in 100-kbp windows (Figure 2c). On average, snow sheep samples had ~91 ± 5 heterozygous sites per 100 kbp, which was the lowest of all species, while urial sheep (422 ± 16) and the Asiatic mouflon (367 ± 46) had the highest number of heterozygous sites. It is noteworthy that argali displayed intermediate values for both the Θ-Watterson estimator and heterozygous sites. Additionally, we also calculated the population mutation rate for each individual sample using MLRHO. The results were in agreement (Table S4) with the Watterson estimator (Figure 2b) and mean heterozygosity (Figure 2c).

Runs of homozygosity were identified for all samples (Figure 2d) using *bcftools roh*. On average, it was observed that snow sheep had about 68.38 ± 5.08% (translating to 1.67 ± 0.12 Gbp) of the total genome covered by ROH. This average was the highest of all the *Ovis* species. In fact, on average, Dall sheep and bighorn sheep also had a high fraction of their genomes under ROH (56.05 ± 2.74% and 54.26 ± 5.62%, respectively); however, these are lower than the fraction observed in snow sheep. Among all *Ovis* species, on average, urial sheep had the lowest fraction (14.10 ± 1.19%) of its genome under ROH. Interestingly, a large fraction of the genome identified in snow sheep, Dall sheep and bighorn sheep were under short ROHs (≤1 Mbp). Argali samples displayed relatively higher fractions of the genome under long (>2.5 Mbp) and medium (>1 and <2.5 Mbp) ROH.

In the demographic histories inferred by PSMC (Figure 2e), the five snow sheep samples with lower coverage (<15×) and one snow sheep sample with higher coverage (~17×) displayed concordant demographic trends. They showed a population expansion of between 0.3 and 0.2 Ma and a population reduction thereafter until about 80–60 thousand years ago (ka), after which the population stabilized. The demographic trajectories of Dall sheep and bighorn sheep mirrored the patterns of snow sheep except for the population increase after ~40–30 ka in Dall and, especially, bighorn sheep. On the other hand, the demographic trajectories of domestic sheep (Figure S1) were similar to those previously reported (Li et al., 2020), with two apparent population expansions and two severe bottlenecks. It is noteworthy that Asiatic mouflon and urial sheep displayed demographic trajectories (Figure 2e) similar to the domestic sheep, with the exception of the most recent period for which we estimate a population expansion of the Asiatic mouflon.

3.2 | Phylogenetic relationships and introgression patterns

Two phylogenetic approaches and one clustering-based approach (Figures 3 and S2) resulted in a consistent tree topology that was similar to those reported using mtDNA. With goat as an outgroup, the NJ-based clustering (Figure 3a) based on pairwise IBS distance separated the samples into two major groups. One group included the clades of snow sheep, Dall sheep and bighorn sheep, while the other group included the clades of argali, Asiatic mouflon, urial sheep and domestic sheep. Molecular clock dating (Figure 3b) indicated that ovine species shared their MRCA ~3.8 Ma. Subsequently, argali sheep diverged from the MRCA of

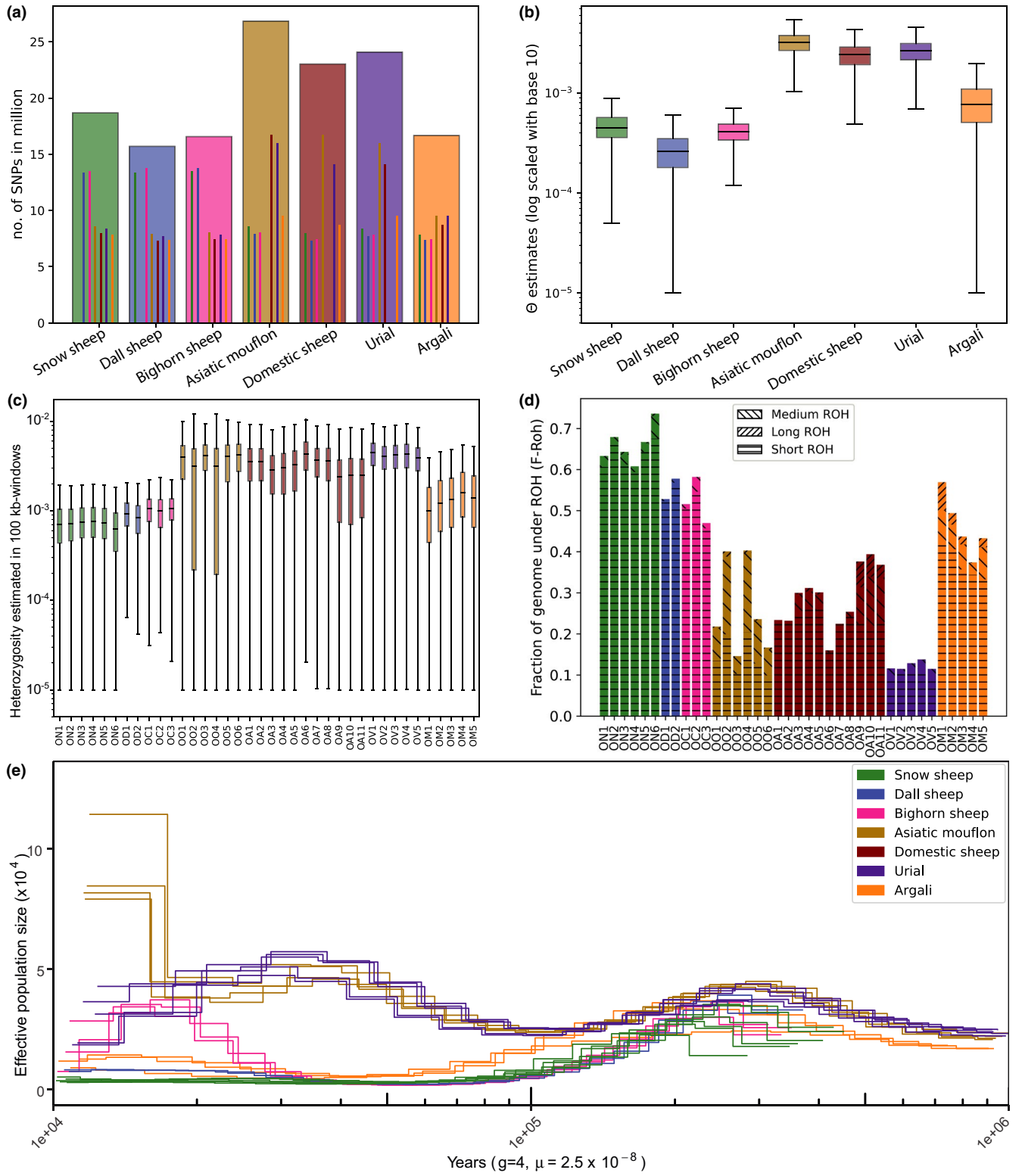

FIGURE 2 **SNP overlap, genome-wide heterozygosity, individual-sample based heterozygosity and ROH. (a) Number of SNPs (in millions) identified with respect to the domestic sheep reference genome (Oar_4.0). Within each large bar, smaller bars indicate the overlap with other species, note that the number of samples representing each species ranged from two (dall sheep) to 11 (domestic sheep). (b) Genome-wide distribution of heterozygosity values based on the Θ-Watterson estimator. (c) Heterozygosity values calculated for each sample. (d) Estimation of the fraction of genome under ROH for each sample. (e) Demographic trends for all wild Ovine species between 10,000 and 1 Million years ago. Two Asiatic mouflon showing abnormal expansion in the recent times are not shown in the panel. The colour schemes shown beside (e) apply to all the subpanels of this figure. Abbreviations on the *x*-axis of (c) and (d): snow sheep (ON), bighorn sheep (OC), dall sheep (OD), Asiatic mouflon (OO), domestic sheep (OA), urial (OV), argali (OM)**

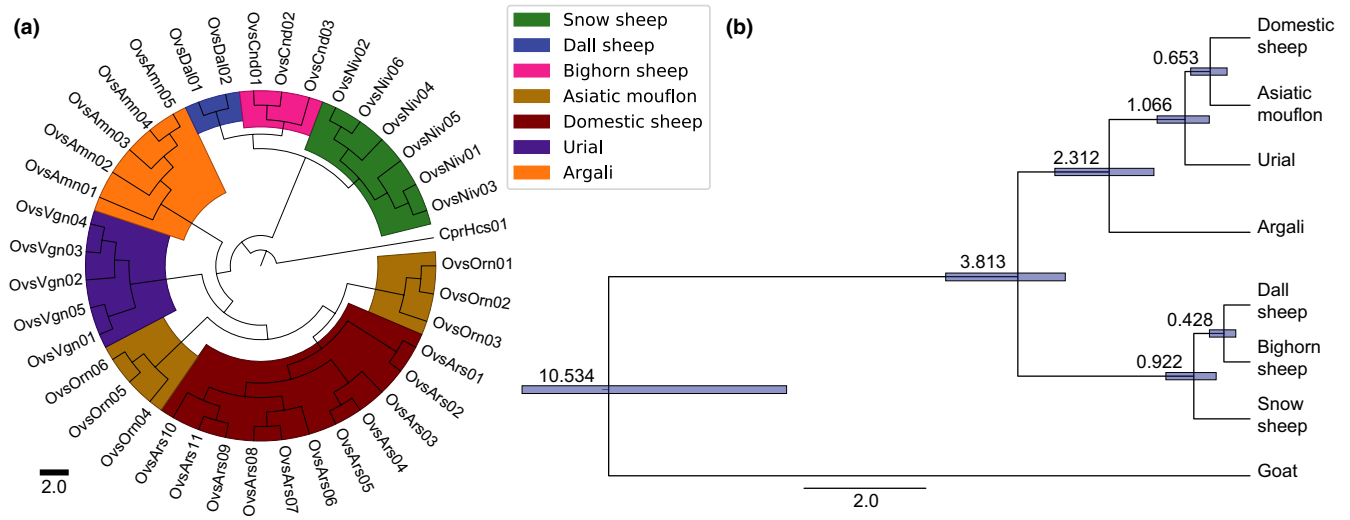

FIGURE 3 (a) IBS distance-based phylogenetic tree, and (b) RAXML-based phylogenetic tree along with divergence time estimates (node values) in units of 1 Ma. Abbreviations in (a): OvisNiv: snow sheep, OvsCnd: bighorn sheep, OvsDal: Dall sheep, OvsAmn: argali, OvsVgn: urial, OvsOrn: Asiatic mouflon, OvsArs: domestic sheep, CprHcs: goat

domestic sheep, urial sheep and Asiatic mouflon ~2.3 Ma. The urial lineage diverged from the MRCA of domestic sheep and mouflon ~1.0 Ma. Snow sheep diverged from its MRCA with North American sheep ~0.9 Ma, which was followed by the divergence of Dall sheep and bighorn sheep ~0.43 Ma.

Using snow sheep and bighorn sheep, or snow sheep and Dall sheep, as the sisters and the remaining wild sheep as the other sister group in the tree topology ((A, B), (C, D)), we tested whether the snow sheep lineage was involved in hybridization with urial sheep, argali, Asiatic mouflon and/or domestic sheep (Figure 4). Interestingly, we observed deviations ($p < .05$) from a tree-like model of population branching in all the cases that involved argali with urial sheep (Figure 4a,b) or Asiatic mouflon or domestic sheep as the sisters. Furthermore, using goat lineage as an outgroup, with argali as the donor/recipient lineage and snow sheep and bighorn sheep as the sisters, we observed significant *z*-values, indicating deviation from the expected tree topology (Figure 4e). These results indicate that argali could have hybridized with snow sheep after the latter diverged from the lineage of North American wild sheep. To validate this hypothesis and rule out the possibility of ILS, we used simulated data as implemented in the F_4 tool. In all quadruplets showing a significant *z*-value, we observed that the probability (p_{sim}) of producing an F_4 statistic at least as extreme as the observed value was <.05, indicating that the most probable cause of the deviation from the expected tree topology is gene flow between argali and snow sheep.

With Dall and snow sheep as focal sister species (Figure 4d), the argali lineage did not produce significant *z*-values, possibly due to an excess of shared derived alleles between snow sheep and Dall sheep, in turn caused by gene flow between these lineages. Indeed, using goat lineage as an outgroup with snow sheep as the donor/recipient lineage (Figure 4c) and Dall and bighorn sheep as the focal sister species, we observed significant *z*-values as well as $p_{sim} < .05$, suggesting introgression between snow sheep and Dall sheep after the latter diverged from bighorn sheep.

3.3 | Genes underlying introgression between snow sheep and argali

Introgressed segments between argali and snow sheep were identified based on the consistency of our results from three approaches—RND_{min}, F_{ST}-based phylogenetic tree, and DFOIL statistics (Figure 5a,b). First, 223 putative introgressed segments of 50 kbp were identified based on an RND_{min} threshold of 0.12 (Figure 5b). This threshold was significantly lower than the lowest value of the RND_{min} statistics (~0.28) that were observed in simulated data under ILS (Figure S3). Next, the F_{ST}-based phylogenetic approach identified 151 (151/223) segments for which argali clustered within the nodes of snow sheep, Dall sheep or bighorn sheep. A subsequent DFOIL analysis supported introgression in 42 of these 151 segments. Furthermore, for almost all of the introgressed segments (41/42), the DFOIL statistics implied gene flow between the MRCAs of the *Pachyceros* and argali lineage or from the snow sheep to argali lineage. These segments encompassed 45 genes (Figure 5a; Table S5 excluding uncharacterized loci) that might be potential targets of adaptive introgression in the argali and snow sheep lineage. These segments involved genes associated with immunity, such as *LRRC19* and *PKN2* (Figure 5c; Ng & Xavier, 2011; Peckham et al., 2017); genes associated with cold adaptation, such as *STAU2*, *RPTOR* (Figure 5d) and *RALGAPA1* (Ding et al., 2018; Sun et al., 2010); and genes associated with spermatogenesis, such as *NEK1* and *EZH2* (Cai et al., 2020; Holloway et al., 2011). Interestingly, at least eight genes (Kim et al., 2012; Morandi et al., 2016; Nakagami, 2013; Walden et al., 2012; Table S5) are associated with adipogenesis or obesity-related traits, and five genes (*HOXC8*,

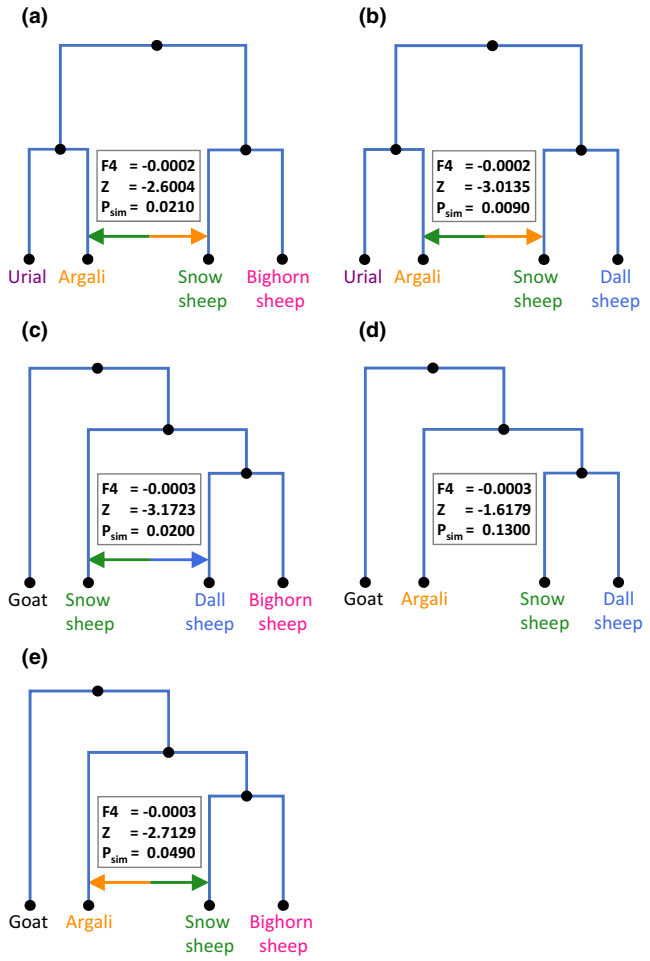

FIGURE 4 Different tree topologies to investigate the hypotheses of introgression across the genus *Ovis* that involved the snow sheep lineage. F4 refers to the F_4 statistic. Z refers to the deviation of the F_4 statistic from zero in units of the standard error. P_{sim} refers to proportion of simulation having the F_4 value as extreme as that observed

HOXC9, *HOXC10*, *HOXC11* and *HOXC12*) were homeobox genes that are also involved in body morphology.

3.4 | Genes under positive selection

To identify positive selection in the snow sheep lineage, the branch-site likelihood ratio test was estimated for 7084 protein-coding sequences that had been identified as one-to-one orthologues across the 10 species (opossum, human, horse, dog, pig, cattle, goat, domestic sheep, Marco Polo sheep and snow sheep). After the first round of analysis, a total of 606 genes were identified under positive selection. However, visual inspection revealed that many of these genes could be the result of errors in sequencing, annotation (i.e., incorrect gene structure in gene prediction) and/or alignment. Therefore, a pipeline was developed to minimize false positives (see Section 2). This pipeline led to the identification of eight amino acid substitutions (Table 1) in seven genes under positive selection in the snow sheep lineage. One of these was *transmembrane protein 131* (*TMEM131*), which is the gene that encodes a protein that regulates endoplasmic reticulum (ER) stress response and collagen production. Interestingly, a previous study (Vasan et al., 2007) associated SNPs in this gene with heart rate and exercise performance in humans. The predicted structure (Figure 6a) for this gene in the snow sheep assembly is very similar to that of cattle, goats and domestic sheep. Genes related to hypoxia and environmental adaptations, such as *ZFYVE16* and *MGARP*, were also identified under selection in the snow sheep lineage (Pan et al., 2017). The *THBD* gene, associated with blood clotting along with immunity and defence, was also identified under selection. Next, a detailed investigation of the Illumina WGS data revealed that six amino acid substitutions identified under selection were fixed in all six of the snow sheep samples, with corresponding variants also observed in RNA-sequencing data for four substitutions (Table 1; Figure 6b). Using the FWH and DH-based simulations, a total of 81 genes (Table S6) were identified under selection in the snow sheep lineage. Of these, ~30 genes were associated with functions related to either olfactory receptors (OR) or immune response. It is well known that OR and immunity-related genes evolve rapidly through mechanisms associated with copy number variations. However, a previous study has shown that OR genes are strong candidates for positive selection in the Homininae branch because of their role in taste and smell (Daub et al., 2017). FWH analysis also identified (Figure 6c,d) genes such as *Dynamin-like 120 kDa protein, mitochondrial* (*OPA1*), a gene associated with lung morphology/development (Smits et al., 2010), and Growth/differentiation factor 3 (*GDF3*), a gene involved in adipose-tissue homeostasis under selection (Figure 6c,d; Andersson et al., 2008).

4 | DISCUSSION

In this study, we used newly generated WGS data from snow sheep and combined those data with publicly available data for all remaining sheep species. Our analyses showed that snow sheep consistently displayed the least genetic diversity among wild sheep, which could be related to its demographic history and life history traits. Interestingly, *Ovis* species adapted to colder environments (subgenus *Pachyceros*) show significantly lower diversity than species adapted to temperate or semi-arid climates (domesticated sheep, Asiatic mouflon and urial). Argali is in an intermediate position in terms of both diversity and environmental conditions. The demographic trends of snow sheep were similar to those of other sheep, especially the other members of the subgenus *Pachyceros*, the Dall and bighorn sheep. This agrees with the phylogenetic and divergence time inference. Our results also revealed introgression within *Pachyceros* and between *Pachyceros* and argali lineages, indicating that these lineages must have had larger and overlapping

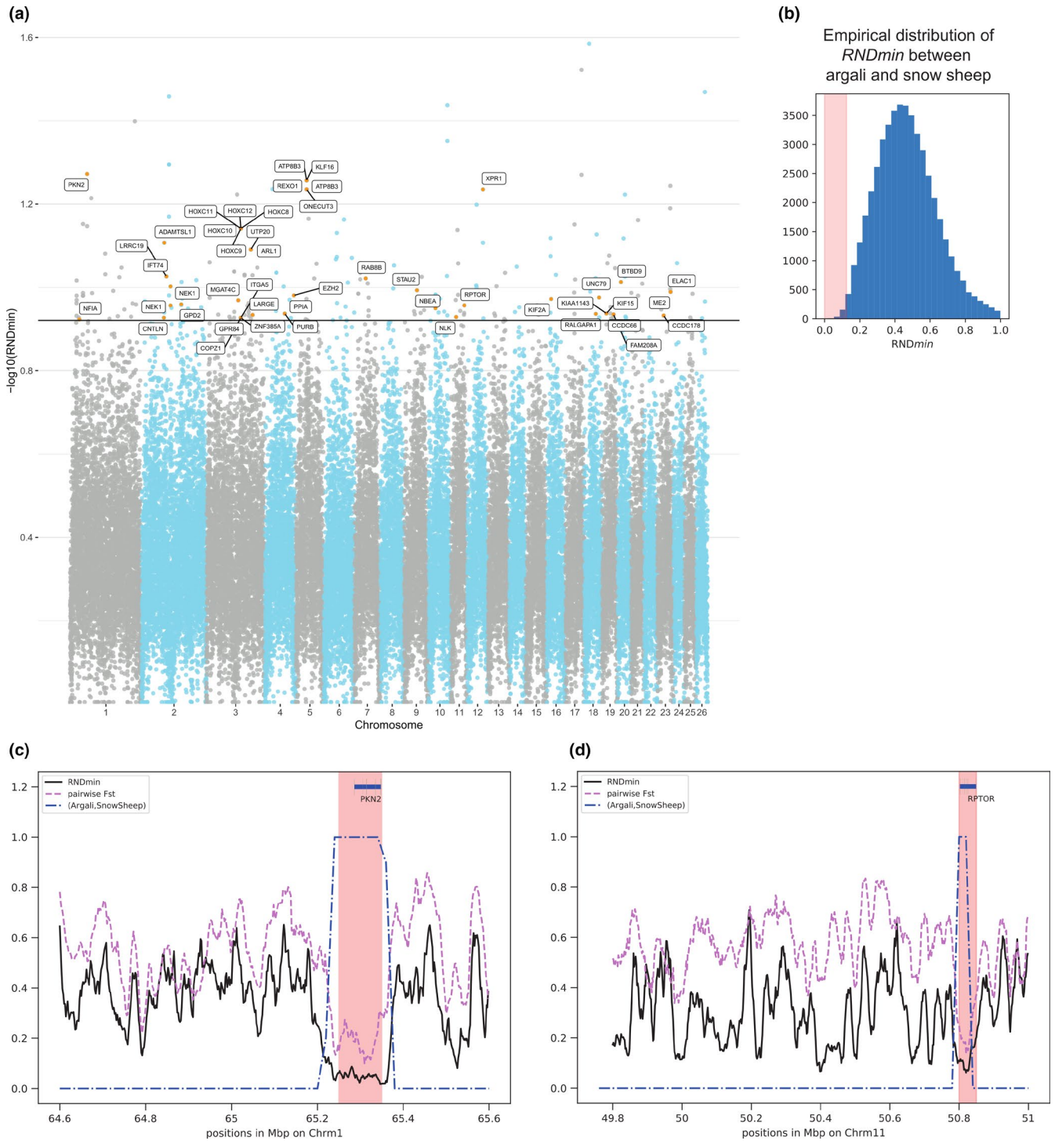

FIGURE 5 Introgression segments between snow sheep and argali. (a) Genome-wide distribution of RND_{min} between argali and snow sheep; black horizontal line indicates the threshold of 0.12 (<2× the standard deviation from the mean), and yellow circles highlight the genes (annotated in thick black boxes) that covered segments with significant RND_{min} values. (b) Histogram of observed RND_{min} values; the portion highlighted in light pink indicates the number of segments with RND_{min} values less than 0.12 and also supported by NJ-based clustering and DFOIL statistics approaches. (c) Distribution of RND_{min} values (black solid line), pairwise F_{ST} values (violet dashed line) and proportion of phylogenetic trees having argali and snow sheep as sisters (blue dash-dotted line) for the introgressed segment encompassing the *PKN2* gene. (d) Distribution of RND_{min} values (black solid line), pairwise F_{ST} values (violet dashed line) and proportion of phylogenetic trees having argali and snow sheep as sisters (blue dash-dotted line) for the introgressed segment encompassing the *RPTOR* gene

TABLE 1 Genes and amino acid substitutions identified to be under selection using the branch-site model in CODEML. d_n stands for nonsynonymous substitutions and d_s stands for synonymous substitutions.

Gene	Notation	Coordinates on Oar4.0	BEB value	LogL_Ho	LogL_Ha	d_n/d_s value	Fixed in all the six Snow sheep samples?	Variant also seen in RNA-seq?
TMEM131	T1459M	chrm3: 103138639–103138640	0.975	−12,452.90	−12,448.86	417.12	Yes	Yes
MAGI3	Y1125G	chrm1: 89999818–89999819	0.983	−13,104.20	−13,099.29	381.52	Yes	No read covered the positions
MGARP	V161D	chrm17: 18068322–18068323	0.888	−2735.68	−2731.70	999.00	Yes	Yes
ZFYVE16	S1033K	chr5: 77843168–77843169	0.886	−14,189.00	−14,180.37	999.00	Yes	No read covered the positions
PALM2	S540S	chr2: 13192566–13192567	0.920	−7187.12	−7182.70	250.60	Yes	Yes
THBD	A240V	chr13: 40896497	0.978	−5853.10	−5848.98	999.00	Yes	Yes
THBD	S367L	chr13: 40896116–40896117	0.948	−5853.10	−5848.97	999.00	No (fixed in 5/6 samples)	Yes
EXOC3L1	R77Q	chr14: 34078400–34078401	0.982	−7514.78	−7509.38	999.00	No (segregating in all samples)	No read covered the positions

geographical distributions during the Pleistocene. Our data also indicated that introgression and selection involving prominent genes contributed to the adaptive fitness of snow sheep.

4.1 | Genetic diversity and demography in *Pachyceros*

Snow sheep, Dall sheep and bighorn sheep all displayed similar demographic trends (Figure 2e) after their lineages split. These populations experienced demographic declines starting around 150 ka and lasting until 60–80 ka, after which their populations stabilized or increased. The demographic minimum of those patterns apparently fell in the youngest part of Marine Isotope Stage 5 (MIS 5) and MIS 4, which were preceded by the colder MIS 2–4 last glacial periods. Considering that MIS 5 has several interglacial substages, the demographic pattern of snow sheep and North American wild sheep supports refugium dynamics of cold-adapted species. In such dynamics, the populations of cold-adapted species shrink and retreat to northern and alpine refugia during interglacial periods and then expand during glacial periods (Stewart et al., 2010). It is possible that during interglacial periods, snow sheep and North American wild sheep became confined to high-altitude refugia, increasing their genetic isolation. Such isolation could be the cause of the low heterozygosity and relatively high proportion of short genomic ROH observed in snow sheep (Figure 2 d). Furthermore, the latest census estimated around 78,000 individuals in the overall snow sheep population. However, the low degree of heterozygosity in the genome suggests that snow sheep have had low effective population sizes since their origin.

4.2 | Origin and divergence of *Pachyceros*

The two hypotheses concerning the origin of snow sheep and North American wild sheep (Dall and bighorn sheep) address their origin as being an argali-like ancestor or a snow-sheep-like ancestor distributed throughout Beringia (Bunch et al., 2006; Cowan, 1940). In agreement with a previous mtDNA-based study (Bunch et al., 2006), our results (Figure 3) support a phylogenetical closer relationship of snow sheep with North American wild sheep than with argali sheep. We estimated their MRCA at around 0.9 Ma and the Dall and bighorn sheep divergence around 0.43 Ma. These estimates are much younger than the estimations reported from mtDNA (Bunch et al., 2006). This could be due to the different time constraints used for calibration in the different studies, the difference in the substitution rates between mtDNA and nuclear DNA, or most importantly, to an increased chance of lineage sorting in mtDNA, given that it constitutes a single locus (Arbogast et al., 2002). Because of this, it is even possible that mtDNA has been introgressed from an old lineage within the *Pachyceros genus*. This possibility is not remote given that we discovered that introgression has occurred within the subgenus *Pachyceros*, and moreover, that the distantly related argali sheep and snow sheep each had some degree of introgression as well.

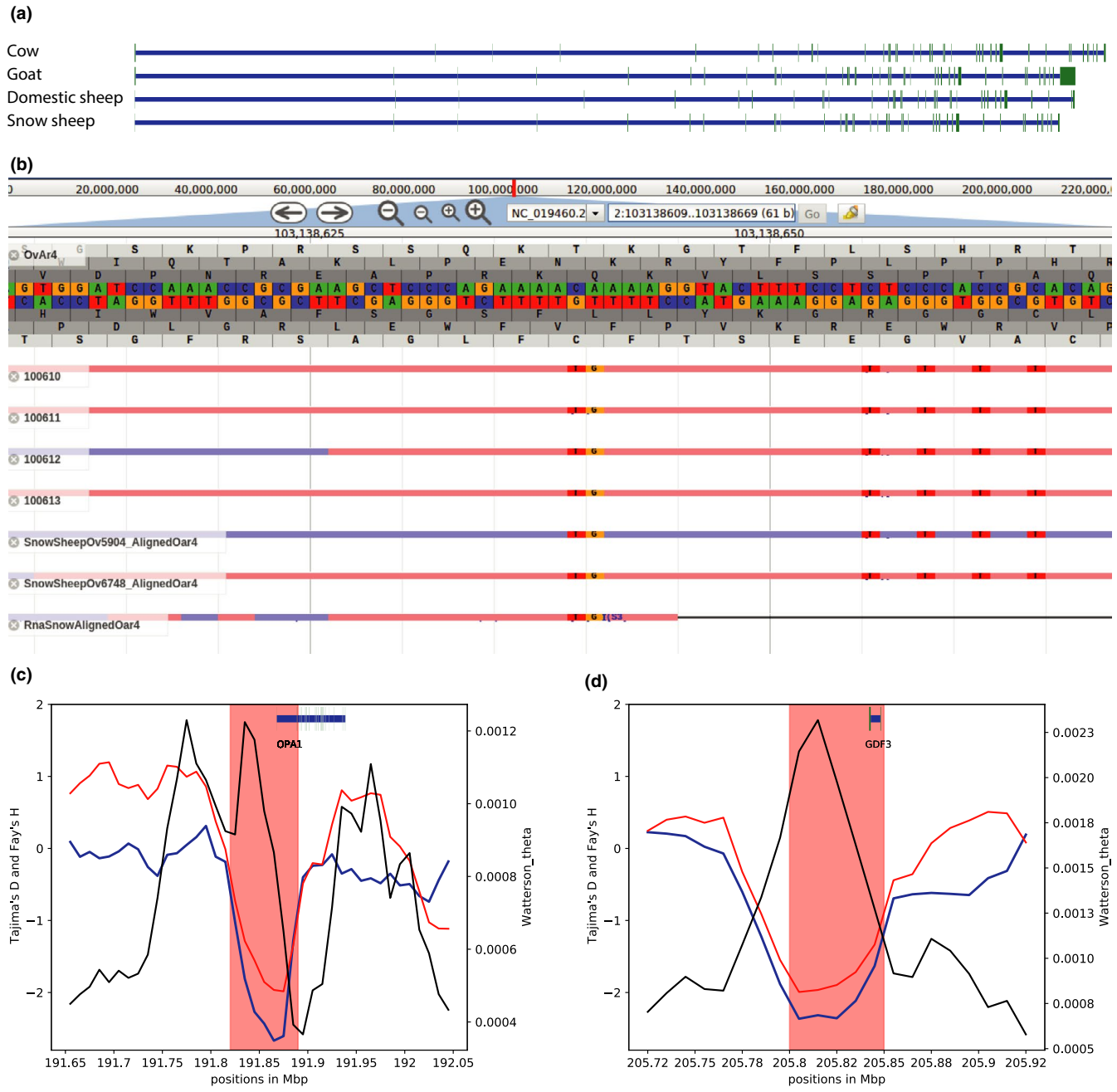

FIGURE 6 Examples of genes and sites identified to be under selection in the snow sheep lineage. (a) Comparative gene structure of the *TMEM131* gene of cow, goat, domestic sheep and snow sheep. (b) JBROWSE screen-shot of the codon under selection (two consecutive SNPs fixed in all the snow sheep samples) in the *TMEM131* gene; the RNA-seq samples as well as whole genome sequences were aligned onto the domestic sheep reference genome. (c) The 50-kb genome segment identified under selection using the FWH statistic; the segment covered a part of the *OPA1* gene. (d) The 50-kb genome segment identified under selection using the FWH statistic; the segment covered the entire *GDF3* gene. In (c) and (d), the red, blue and black lines represent Tajima's *D*, the FWH statistic and Watterson's theta as calculated using the tool ANGSD. In (a), (c) and (d), the green vertical lines indicate the position of the exons

4.3 | Introgression involving the *Pachyceros* and argali lineages

The F_4 statistics of real and simulated data (Figure 4) indicated that, after speciation, snow sheep had gene flow with Dall sheep (Figure 4c). This was possible through the Bering land bridge that connected Siberia and Alaska until about 12 ka and acted as a migration route during glacial periods when sea level was lower. Indeed, a biogeographical role of the Bering bridge in the dispersal of other megafaunal species is well known. For instance, a recent analysis of ancient DNA of brown bears and lions found that these mammalian species underwent multiple dispersals between North America and

Eurasia, and that these events coincided with glacial periods of low sea levels (Salis et al., 2020).

The F_4 statistics as well as DFOIL statistics revealed a complex introgression scenario involving ancestral lineages of snow sheep and argali sheep (Figure 4a–b). Indeed, as a part of the middle-to-large sized group of megafaunas that expanded during the Pleistocene, snow sheep were much more widespread during the Pleistocene than during the Holocene. They inhabited the south of western Siberia, a significant part of eastern Siberia, and the Far East, including both mountainous and lowland territories. According to Boeskorov (2014), the remains of snow sheep belonging to the Pleistocene were found in a number of places outside the modern area, including the southern part of western Siberia. There is abundant evidence (Danilkin, 2005), including fossil-based evidence (Note S2), suggesting that the snow sheep and argali habitats overlapped in many areas of southern Siberia during the Middle and Late Pleistocene. The distribution overlap was so large that Ovodov (2003) proposed that the main reasons for the extinction of the snow sheep in the southern part of central Siberia and the Baikal region at the end of the Pleistocene were the expansion of the Siberian Forest belt and competition with argali sheep. Therefore, the evidence seems consistent that, during the Pleistocene, snow sheep had gene flow with argali.

4.4 | Genes underlying the introgressed segments

The introgressed segments between the snow sheep lineage and argali contain genes mainly involved in immune response, adipogenesis and adaptation (Figure 5a; Table S5). For instance, the *PKN2* gene has been associated with responses to viral respiratory disorders in the domestic pig and *Mycoplasma ovipneumoniae* infection in sheep (Dekkers et al., 2017; Mousel et al., 2021). Indeed, the adaptive introgression of immune-related genes has been hypothesized in other mammalian species, including humans, Iberian hares and alpine ibex (Deschamps et al., 2016; Grossen et al., 2014; Holloway et al., 2011). Furthermore, we observed a strikingly high number of adipogenesis-related genes. For instance, genes in the homeobox cluster, such as *HOXC9* and *HOXC8*, are highly expressed in brite adipose tissue (Walden et al., 2012). Brite adipose tissues could play an important role in thermogenesis under prolonged cold conditions (Rosenwald et al., 2013). Another introgressed gene, *NFIA*, was identified as a positive transcriptional regulator of adipogenesis that is associated with brown adipose tissue (BAT) (Hiraike et al., 2017). Interestingly, a previous study (Cao et al., 2020) found that *NFIA* was introgressed from argali into Chinese domestic sheep that are adapted to high altitude. In fact, of the 45 genes that we identified as introgressed between the argali and snow sheep lineages, six were also inferred as introgressed from argali or a related species into domestic sheep (mainly of China) in the previous study (Cao et al., 2020). Therefore, our data suggest that adaptive segments can cross the boundary of multiple species because of multiple introgression events.

4.5 | Genes under positive selection in snow sheep

We identified several genes under selection in the snow sheep lineage (Tables 1 and S6) that might be contributing to adaptation to subarctic climates. Mammals adapted to arctic and subarctic climates employ several behavioural, physical and physiological adaptations to survive harsh cold, food scarcity and seasonal sunlight variations (Blix, 2016). One of the most important physiological adaptations is the accumulation of a large amount of body fat in autumn to survive food scarcity in winter. Furthermore, accumulated adipose tissues are used to increase metabolic heat production through the process termed NST. The FWH analyses (Figure 6c) identified the *GDF3* gene as one of the candidates under selection in snow sheep. *GDF3* is highly expressed in adipose tissue and plays an important role in adipose tissue homeostasis. One study (Andersson et al., 2008) showed that *Gdf3*$^{-/-}$ mutant mice accumulated less adipose tissue than wild mice. Furthermore, this gene is regulated by the activin receptor *ALK7*, which plays an important role in NST (Marmol et al., 2020). *CADM2*, a gene expressed in adipocytes and associated with obesity and energy homeostasis (Yan et al., 2018), was also identified as under recent positive selection in a human Mongolian population (Nakayama et al., 2017). The *HS6ST1* gene encodes heparan sulphate proteoglycans that contribute to lung morphology and function (Smits et al., 2010). The syndecan-2 (*SDC2*) gene has also been identified to be under positive selection in Ethiopian populations adapted to altitude-associated hypoxia (Huerta-Sanchez et al., 2013). The *USH2A* gene is associated with hearing and vision and has been identified as under selection in grey wolves (Schweizer et al., 2016). Moreover, the variants in this gene have been identified to be under convergent and divergent selection across several mammalian species (Marcovitz et al., 2019).

The ratio of nonsynonymous (d_n) to synonymous substitutions (d_s), d_n/d_s, as estimated using CODEML identified the transmembrane gene *TMEM131* as under selection (Figure 6a,b). This gene was also identified as under selection in racing pigeons (Gazda, 2019) and also under long-term selection in Scandinavian Mesolithic and modern Finnish human populations (Gunther et al., 2018) adapted to cold environments. A recent study (Zhang et al., 2020) revealed that TMEM131 proteins play an important role in regulating the ER stress response and collagen production. It is noteworthy that it has been reported that ER stress is associated with cardiac function under hypoxic conditions (Yuan et al., 2017). Therefore, it can be hypothesized that codon/sites (SNPs) identified as under selection might be contributing to improvements in cardiac performance, providing long-term adaptation to a cold climate. One more gene, the *Zinc finger protein* (*ZFPM2*), plays a central role in cardiac morphogenesis and the development of coronary vessels in the epicardium. This gene was also identified as under positive selection in Tibetan pigs (Dong et al., 2014).

It is well known that mitochondria play a vital role in acclimatization to hypoxic conditions (Murray & Horscroft, 2016). Our analyses (Figure 6c; Tables 1 and S6) identified two genes under positive selection that have been associated with mitochondrial function and morphology in previous studies. The *OPA1* gene encodes a protein

that helps regulate mitochondrial stability and energy output by maintaining a mitochondrial cristae morphology (Frezza et al., 2006). Because many proteins and enzymes associated with oxidative phosphorylation are located in the mitochondrial cristae membrane, its structure is vital for mitochondrial function. A recent study (Wang et al., 2019) has shown that *OPA1* was over-expressed in the tissues of Tibetan sheep but not in small-tail Han sheep. Furthermore, the numbers of mitochondrial cristae in all four of the tissues they collected were much higher than those in small-tailed Han sheep. The study proposed that high numbers of the cristae allow Tibetan sheep to optimize oxygen use in energy consumption. Interestingly, our transcriptome data (Figure S4) also showed that the *OPA1* gene is highly expressed in snow sheep tissues. Another gene, *MGARP*, is also located within the inner/cristae membrane of mitochondria and has been proposed to play an important role in mitochondrial remodelling and cristae shaping (Zhang, 2019).

Snow sheep share a biogeography, evolutionary history, morphology, metabolism and behaviour with other cold-adapted species. The factor that decouples snow sheep from most Quaternary mammalian megafauna that remarkably disappeared in the last 50 ka is probably its capacity to adapt to alpine habitats, becoming less dependent on a steppe–tundra habitat, which also allowed the snow sheep to stay in the regions historically avoided by humans.

5 | CONCLUSION

In this paper, we presented the snow sheep, *Ovis nivicola*, as a reference species for monitoring anthropogenic disturbances and climate change. Our data show that despite large census population sizes, the snow sheep genome has low heterozygosity, which could be attributed to historical climate change. Therefore, the genomic data generated in the present study have great implications for establishing conservation policies and interventions in the management of snow sheep populations. Hybridization events, as is implied between Dall sheep and snow sheep, also confirm the role of the Bering land bridge in shaping the genetic diversity of Arctic and subarctic species. Complex introgression events, such as that implied between argali and the snow sheep lineage, also revealed that wild sheep species had a vast geographical distribution during the middle and early Pleistocene. Characterizing introgressed genomic regions revealed genes associated with immunity and adipogenesis, highlighting the role of adaptive introgression in wild sheep species. Additionally, the genes identified as under selection in the snow sheep lineage provide insight into the genetic basis of adaptation to cold environments. These results hint at new points of interest for future research regarding environmental adaptation of species threatened by anthropogenic disturbance.

ACKNOWLEDGEMENTS

This work was partly supported by the Russian Science Foundation within Project No. 21-66-00007 (molecular genetic studies) and by the Ministry of Science and Higher Education of the Russian Federation within Theme No. 0445-2019-0024 (expedition studies). This work was also carried out by the support of LMU excellent funding of LMU Munich. We would also like to thank the International Sheep Genomics Consortium (ISGC) as well as the sample donors of the project PRJNA160933 for making the WGS data available to us. We would also like to thank Dr Alvarez-Carretero Sandra for her advice in running the MCMCtree analysis. We also acknowledge Jurgen Klawatsch for his inputs concerning functionality of a small number of genes discussed in the paper. Upadhyay, M., Kunz, E., Sandoval-Castellanos, E., Hauser, A., Krebs, S., Graf, A., Helmut, B., Dotsev, A., Okhlopkov, I., Shakhin A., Bagirov, V., Brem, G., Fries, R., Zinovieva, N., Medugorac, I.; 2021; Whole genome sequences reveal adaption of Snow sheep to subarctic climates and complex introgression across wild sheep species; EMBL-ENA, PRJEB38329.

AUTHOR CONTRIBUTIONS

I.M., N.Z and G.B. conceived the study. A.D., I.O., A.S. and V.B. collected the data. S.K., H.B. and R.F. contributed reagents, and carried out sequencing. M.U. performed entire data analyses with inputs from E.K., E.C., A.H., S.K. and A.G., and M.U., I.M., E.C. and N.Z. contributed to the interpretation of the results. M.U. wrote the first original draft and carried out visualization of the results. A.D., I.O., A.S., V.B., G.B., E.K. and N.Z. reviewed and edited the manuscript. I.M. and N.Z supervised the study.

DATA AVAILABILITY STATEMENT

The snow sheep genome assembly (including raw promethION sequences) used in this paper and one Illumina whole genome sequence of snow sheep are already available on EMBL-ENA with project accession PRJEB38329. The remaining five snow sheep sequences are available in EMBL-ENA with project accession PRJEB44668. The pdf file containing the commands can be downloaded from https://doi.org/10.6084/m9.figshare.16611829

ORCID

Ivica Medugorac https://orcid.org/0000-0002-6891-3039

REFERENCES

Abadi, S., Azouri, D., Pupko, T., & Mayrose, I. (2019). Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications*, *10*(1), 934. https://doi.org/10.1038/s41467-019-08822-w

Andersson, O., Korach-Andre, M., Reissmann, E., Ibanez, C. F., & Bertolino, P. (2008). Growth/differentiation factor 3 signals through ALK7 and regulates accumulation of adipose tissue and diet-induced obesity. *Proceedings of the National Academy of Sciences USA*, *105*(20), 7252–7256. https://doi.org/10.1073/pnas.0800272105

Arbogast, B. S., Edwards, S. V., Wakeley, J., Beerli, P., & Slowinski, J. B. (2002). Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, *33*(1), 707–740. https://doi.org/10.1146/annurev.ecolsys.33.010802.150500

Arnold, M. L., & Kunte, K. (2017). Adaptive genetic exchange: A tangled history of admixture and evolutionary innovation. *Trends in Ecology & Evolution*, *32*(8), 601–611. https://doi.org/10.1016/j.tree.2017.05.007

Barbato, M., Hailer, F., Orozco-terWengel, P., Kijas, J., Mereu, P., Cabras, P., Mazza, R., Pirastru, M., & Bruford, M. W. (2017). Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Scientific Reports*, *7*(1), 7623. https://doi.org/10.1038/s41598-017-07382-7

Blix, A. S. (2016). Adaptations to polar life in mammals and birds. *Journal of Experimental Biology*, *219*(Pt 8), 1093–1105. https://doi.org/10.1242/jeb.120477

Boeskorov, G. G. (2014). A new subspecies of fossil Siberian snow sheep (Ovis, Artiodactyla, Bovidae) from Yakutia. *Zoologicheskii Zhurnal*, *93*(7), 868–876. https://doi.org/10.7868/S004451341407006X

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Bunch, T. D., Wu, C., Zhang, Y. P., & Wang, S. (2006). Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. *Journal of Heredity*, *97*(1), 21–30. https://doi.org/10.1093/jhered/esi127

Cai, Y. U., Deng, M., Liu, Z., Zhang, G., Pang, J., An, S., Wang, Z., Zhang, Y., & Wang, F. (2020). EZH2 expression and its role in spermatogonial stem cell self-renewal in goats. *Theriogenology*, *155*, 222–231. https://doi.org/10.1016/j.theriogenology.2020.06.013

Cao, Y.-H., Xu, S.-S., Shen, M., Chen, Z.-H., Gao, L., Lv, F.-H., Xie, X.-L., Wang, X.-H., Yang, H., Liu, C.-B., Zhou, P., Wan, P.-C., Zhang, Y.-S., Yang, J.-Q., Pi, W.-H., Hehua, E. E., Berry, D. P., Barbato, M., Esmailizadeh, A., … Li, M.-H. (2020). Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Molecular Biology and Evolution*, *38*, 838–855. https://doi.org/10.1093/molbev/msaa236

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Cowan, I. M. (1940). Distribution and variation in the native sheep of North America. *The American Midland Naturalist*, *24*(3), 505–580. https://doi.org/10.2307/2420858

Danilkin, A. A. (2005). *Hollow-horned ruminants (Bovidae)*. KMK Scientific Press.

Daub, J. T., Moretti, S., Davydov, I. I., Excoffier, L., & Robinson-Rechavi, M. (2017). Detection of pathways affected by positive selection in primate lineages ancestral to humans. *Molecular Biology and Evolution*, *34*(6), 1391–1402. https://doi.org/10.1093/molbev/msx083

Dekkers, J., Rowland, R. R. R., Lunney, J. K., & Plastow, G. (2017). Host genetics of response to porcine reproductive and respiratory syndrome in nursery pigs. *Veterinary Microbiology*, *209*, 107–113. https://doi.org/10.1016/j.vetmic.2017.03.026

Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., Patin, E., & Quintana-Murci, L. (2016). Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. *American Journal of Human Genetics*, *98*(1), 5–21. https://doi.org/10.1016/j.ajhg.2015.11.014

Ding, D., Liu, G., Hou, L., Gui, W., Chen, B., & Kang, L. (2018). Genetic variation in PTPN1 contributes to metabolic adaptation to high-altitude hypoxia in Tibetan migratory locusts. *Nature Communications*, *9*(1), 4991. https://doi.org/10.1038/s41467-018-07529-8

Dong, K., Yao, N. A., Pu, Y., He, X., Zhao, Q., Luan, Y., Guan, W., Rao, S., & Ma, Y. (2014). Genomic scan reveals loci under altitude adaptation in Tibetan and Dahe pigs. *PLoS One*, *9*(10), e110520. https://doi.org/10.1371/journal.pone.0110520

dos Reis, M., & Yang, Z. (2019). Bayesian molecular clock dating using genome-scale datasets. In M. Anisimova (Ed.), *Evolutionary genomics: Statistical and computational methods* (pp. 309–330). Springer New York.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Ermolin, A. B., & Medvedev, D. G. (2020). About the southern border of the snow sheep areal in northern Asia. *Bulletin of Hunting Science*, *17*(3), 189–195.

Feder, J. L., Xie, X., Rull, J., Velez, S., Forbes, A., Leung, B., & Aluja, M. (2005). Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in Rhagoletis. *Proceedings of the National Academy of Sciences*, *102*(Suppl. 1), 6573. https://doi.org/10.1073/pnas.0502099102

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, *5*, 164–166.

Frezza, C., Cipolat, S., Martins de Brito, O., Micaroni, M., Beznoussenko, G. V., Rudka, T., Bartoli, D., Polishuck, R. S., Danial, N. N., De Strooper, B., & Scorrano, L. (2006). OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. *Cell*, *126*(1), 177–189. https://doi.org/10.1016/j.cell.2006.06.025

Gazda, M. A., Andrade, P., Afonso, S., Dilytė, J., Archer, J. P., Lopes, R. J., Faria, R., & Carneiro, M. (2018). Signatures of Selection on Standing Genetic Variation Underlie Athletic and Navigational Performance in Racing Pigeons. *Molecular Biology and Evolution*, *35*(5), 1176–1189. http://dx.doi.org/10.1093/molbev/msy030

Grossen, C., Keller, L., Biebach, I., International Goat Genome & Croll, D. (2014). Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. *PLoS Genetics*, *10*(6), e1004438. https://doi.org/10.1371/journal.pgen.1004438

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., Munters, A. R., Coutinho, A., Simões, L. G., Vicente, M., Sjölander, A., Jansen Sellevold, B., Jørgensen, R., Claes, P., Shriver, M. D., … Jakobsson, M. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biology*, *16*(1), e2003703. https://doi.org/10.1371/journal.pbio.2003703

Haubold, B., Pfaffelhuber, P., & Lynch, M. (2010). mlRho—A program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Molecular Ecology*, *19*(s1), 277–284. https://doi.org/10.1111/j.1365-294X.2009.04482.x

Hiraike, Y., Waki, H., Yu, J., Nakamura, M., Miyake, K., Nagano, G., & Kadowaki, T. (2017). NFIA co-localizes with PPARgamma and transcriptionally controls the brown fat gene program. *Nature Cell Biology*, *19*(9), 1081–1092. https://doi.org/10.1038/ncb3590

Holloway, K., Roberson, E. C., Corbett, K. L., Kolas, N. K., Nieves, E., & Cohen, P. E. (2011). NEK1 facilitates cohesin removal during mammalian spermatogenesis. *Genes*, *2*(1), 260–279. https://doi.org/10.3390/genes2010260

Huerta-Sánchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., Cardona, A., Montgomery, H. E., Cavalleri, G. L., Robbins, P. A., Weale, M. E., Bradman, N., Bekele, E., Kivisild, T., Tyler-Smith, C., & Nielsen, R. (2013). Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Molecular Biology and Evolution*, *30*(8), 1877–1888. https://doi.org/10.1093/molbev/mst089

Joshi, N. A., & Fass, J. N. (2011). *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.21)* [Software]. Available at https://github.com/najoshi/sickle

Kamm, J., Terhorst, J., Durbin, R., & Song, Y. S. (2020). Efficiently inferring the demographic history of many populations with allele count data. *Journal of American Statistical Association*, *115*(531), 1472–1487. https://doi.org/10.1080/01621459.2019.1635482

Kim, Y. H., Park, T. C., Lee, G., & Shin, J. C. (2012). Gene expression pattern of human chorion-derived mesenchymal stem cells during adipogenic differentiation. *Yonsei Medical Journal*, *53*(5), 1036–1044. https://doi.org/10.3349/ymj.2012.53.5.1036

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kozlov, A. M., & Stamatakis, A. (2019). *Using RAxML-NG in Practice*. p. 25, doi:https://doi.org/10.20944/preprints201905.0056.v1

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. ArXiv, 1303.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, X., Yang, J. I., Shen, M., Xie, X.-L., Liu, G.-J., Xu, Y.-X., Lv, F.-H., Yang, H., Yang, Y.-L., Liu, C.-B., Zhou, P., Wan, P.-C., Zhang, Y.-S., Gao, L., Yang, J.-Q., Pi, W.-H., Ren, Y.-L., Shen, Z.-Q., Wang, F., … Li, M.-H. (2020). Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nature Communications*, *11*(1), 2815. https://doi.org/10.1038/s41467-020-16485-1

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M. A., Gamba, C., Seguin-Orlando, A., Mortensen, C. D., Petersen, B., Hoover, C. A., Lorente-Galdos, B., Nedoluzhko, A., Boulygina, E., Tsygankova, S., Neuditschko, M., Jagannathan, V., … Orlando, L. (2015). Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences USA*, *112*(50), E6889–6897. https://doi.org/10.1073/pnas.1513696112

Liu, L., Bosse, M., Megens, H.-J., Frantz, L. A. F., Lee, Y.-L., Irving-Pease, E. K., Narayan, G., Groenen, M. A. M., & Madsen, O. (2019). Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nature Communications*, *10*(1), 1992. https://doi.org/10.1038/s41467-019-10017-2

Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. In D. J. Russell (Ed.), *Multiple sequence alignment methods* (pp. 155–170). Humana Press.

Marcovitz, A., Turakhia, Y., Chen, H. I., Gloudemans, M., Braun, B. A., Wang, H., & Bejerano, G. (2019). A functional enrichment test for molecular convergent evolution finds a clear protein-coding signal in echolocating bats and whales. *Proceedings of the National Academy of Sciences*, *116*(42), 21094–21103. https://doi.org/10.1073/pnas.1818532116.

Marmol, P., Krapacher, F., & Ibanez, C. F. (2020). Control of brown adipose tissue adaptation to nutrient stress by the activin receptor ALK7. *Elife*, *9*, https://doi.org/10.7554/eLife.54721

Meyer, B. S., Matschiner, M., & Salzburger, W. (2016). Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for lake Tanganyika cichlid fishes. *Systematic Biology*, *66*(4), 531–550. https://doi.org/10.1093/sysbio/syw069

Miles, A., Murillo, R., Ralph, P., Harding, N., Pisupati, R., Summer, R., & Millar, T. (2020). *scikit-allel-A Python package for exploring and analysing genetic variation data (Version 1.3.2)*. Zenodo. http://doi.org/10.5281/zenodo.822784

Morandi, E. M., Verstappen, R., Zwierzina, M. E., Geley, S., Pierer, G., & Ploner, C. (2016). ITGAV and ITGA5 diversely regulate proliferation and adipogenic differentiation of human adipose derived stem cells. *Scientific Reports*, *6*, 28889. https://doi.org/10.1038/srep28889

Mousel, M. R., White, S. N., Herndon, M. K., Herndon, D. R., Taylor, J. B., Becker, G. M., & Murdoch, B. M. (2021). Genomic regions associated with Mycoplasma ovipneumoniae presence in nasal secretions of domestic sheep. *bioRxiv* preprint. https://doi.org/10.1101/2021.02.04.429710

Murray, A. J., & Horscroft, J. A. (2016). Mitochondrial function at extreme high altitude. *Journal of Physiology*, *594*(5), 1137–1149. https://doi.org/10.1113/JP270079

Nakagami, H. (2013). The mechanism of white and brown adipocyte differentiation. *Diabetes & Metabolism Journal*, *37*(2), 85–90. https://doi.org/10.4093/dmj.2013.37.2.85

Nakayama, K., Ohashi, J., Watanabe, K., Munkhtulga, L., & Iwamoto, S. (2017). Evidence for very recent positive selection in mongolians. *Molecular Biology and Evolution*, *34*(8), 1936–1946. https://doi.org/10.1093/molbev/msx138

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, *32*(11), 1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Ng, A., & Xavier, R. J. (2011). Leucine-rich repeat (LRR) proteins: integrators of pattern recognition and signaling in immunity. *Autophagy*, *7*(9), 1082–1084. https://doi.org/10.4161/auto.7.9.16464

Ovodov, N. D. (2003). Pleistocene snow sheep of Siberia. In E. E. Antipina, & E.-N.-M. Chernykh (Eds.), *The latest archeozoological investigations in Russia* (pp. 198–214). Moscow: Languages of Slavic culture.

Pan, S., Zhang, T., Rong, Z., Hu, L. I., Gu, Z., Wu, Q. I., Dong, S., Liu, Q., Lin, Z., Deutschova, L., Li, X., Dixon, A., Bruford, M. W., & Zhan, X. (2017). Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai-Tibetan Plateau in a predatory bird. *Molecular Ecology*, *26*(11), 2993–3010. https://doi.org/10.1111/mec.14090

Pease, J. B., & Hahn, M. W. (2015). Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, *64*(4), 651–662. https://doi.org/10.1093/sysbio/syv023

Peckham, D., Scambler, T., Savic, S., & McDermott, M. F. (2017). The burgeoning field of innate immune-mediated disease and autoinflammation. *The Journal of Pathology*, *241*(2), 123–139. https://doi.org/10.1002/path.4812

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489–494. https://doi.org/10.1038/nature08365

Rezaei, H. R., Naderi, S., Chintauan-Marquier, I. C., Taberlet, P., Virk, A. T., Naghash, H. R., Rioux, D., Kaboli, M., & Pompanon, F. (2010). Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution*, *54*(2), 315–326. https://doi.org/10.1016/j.ympev.2009.10.037

Rosenwald, M., Perdikari, A., Rulicke, T., & Wolfrum, C. (2013). Bi-directional interconversion of brite and white adipocytes. *Nature Cell Biology*, *15*(6), 659–667. https://doi.org/10.1038/ncb2740

Rosenzweig, B. K., Pease, J. B., Besansky, N. J., & Hahn, M. W. (2016). Powerful methods for detecting introgressed regions from population genomic data. *Molecular Ecology*, *25*(11), 2387–2397. https://doi.org/10.1111/mec.13610

Salis, A. T., Bray, S. C. E., Lee, M. S. Y., Heiniger, H., Barnett, R., Burns, J. A., Doronichev, V., Fedje, D., Golovanova, L., Harington, C. R., Hockett, B., Kosintsev, P., Lai, X., Mackie, Q., Vasiliev, S., Weinstock, J., Yamaguchi, N., Meachen, J., Cooper, A., & Mitchell, K. J. (2020). Lions and brown bears colonized North America in multiple

synchronous waves of dispersal across the Bering Land Bridge. *bioRxiv*. doi:https://doi.org/10.1101/2020.09.03.279117

Schweizer, R. M., vonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., Novembre, J., & Wayne, R. K. (2016). Genetic subdivision and candidate genes under selection in North American grey wolves. *Molecular Ecology*, *25*(1), 380–402. https://doi.org/10.1111/mec.13364

Smits, N. C., Shworak, N. W., Dekhuijzen, P. N., & van Kuppevelt, T. H. (2010). Heparan sulfates in the lung: Structure, diversity, and role in pulmonary emphysema. *Anatomical Record*, *293*(6), 955–967. https://doi.org/10.1002/ar.20895

Stewart, J. R., Lister, A. M., Barnes, I., & Dalen, L. (2010). Refugia revisited: Individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1682), 661–671. https://doi.org/10.1098/rspb.2009.1272

Sun, C., Southard, C., Witonsky, D. B., Kittler, R., & Di Rienzo, A. (2010). Allele-specific down-regulation of RPTOR expression induced by retinoids contributes to climate adaptations. *PLoS Genetics*, *6*(10), e1001178. https://doi.org/10.1371/journal.pgen.1001178

Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4*. Sinauer Associates.

Upadhyay, M., Hauser, A., Kunz, E., Krebs, S., Blum, H., Dotsev, A., Okhlopkov, I., Bagirov, V., Brem, G., Zinovieva, N., & Medugorac, I. (2020). The first draft genome assembly of snow sheep (*Ovis nivicola*). *Genome Biology and Evolution*, *12*(8), 1330–1336. https://doi.org/10.1093/gbe/evaa124

Vasan, R. S., Larson, M. G., Aragam, J., Wang, T. J., Mitchell, G. F., Kathiresan, S., Newton-Cheh, C., Vita, J. A., Keyes, M. J., O'Donnell, C. J., Levy, D., & Benjamin, E. J. (2007). Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study. *BMC Medical Genetics*, *8*(Suppl. 1), S2. https://doi.org/10.1186/1471-2350-8-S1-S2

Walden, T. B., Hansen, I. R., Timmons, J. A., Cannon, B., & Nedergaard, J. (2012). Recruited vs. nonrecruited molecular signatures of brown, "brite," and white adipose tissues. *American Journal of Physiology. Endocrinology and Metabolism*, *302*(1), E19–E31. https://doi.org/10.1152/ajpendo.00249.2011

Wang, G., He, Y., & Luo, Y. (2019). Expression of OPA1 and Mic60 genes and their association with mitochondrial cristae morphology in Tibetan sheep. *Cell and Tissue Research*, *376*(2), 273–279. https://doi.org/10.1007/s00441-018-2975-y

Wang, X., Li, Q., & Takeuchi, G. T. (2016). Out of Tibet: An early sheep from the Pliocene of Tibet, Protovis himalayensis, genus and species nov. (Bovidae, Caprini), and origin of Ice Age mountain sheep. *Journal of Vertebrate Paleontology*, *36*(5), e1169190. https://doi.org/10.1080/02724634.2016.1169190

Welch, A. J., Bedoya-Reina, O. C., Carretero-Paulet, L., Miller, W., Rode, K. D., & Lindqvist, C. (2014). Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. *Genome Biology and Evolution*, *6*(2), 433–450. https://doi.org/10.1093/gbe/evu025

Yan, X., Wang, Z., Schmidt, V., Gauert, A., Willnow, T. E., Heinig, M., & Poy, M. N. (2018). Cadm2 regulates body weight and energy homeostasis in mice. *Molecular Metabolism*, *8*, 180–188. https://doi.org/10.1016/j.molmet.2017.11.010

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yuan, F., Zhang, L., Li, Y. Q., Teng, X., Tian, S. Y., Wang, X. R., & Zhang, Y. (2017). Chronic intermittent hypobaric hypoxia improves cardiac function through inhibition of endoplasmic reticulum stress. *Scientific Reports*, *7*(1), 7922. https://doi.org/10.1038/s41598-017-08388-x

Yudin, N. S., Larkin, D. M., & Ignatieva, E. V. (2017). A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC Genetics*, *18*(Suppl. 1), 111. https://doi.org/10.1186/s12863-017-0580-9

Zeng, K., Fu, Y. X., Shi, S., & Wu, C. I. (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*, *174*(3), 1431–1439. https://doi.org/10.1534/genetics.106.061432

Zhang, S. (2019). MGARP is ultrastructurally located in the inner faces of mitochondrial membranes. *Biochemical and Biophysical Research Communications*, *516*(1), 138–143. https://doi.org/10.1016/j.bbrc.2019.06.028

Zhang, Z., Bai, M., Barbosa, G. O., Chen, A., Wei, Y., Luo, S., Wang, X., Wang, B., Tsukui, T., Li, H., Sheppard, D., Kornberg, T. B., & Ma, D. K. (2020). Broadly conserved roles of TMEM131 family proteins in intracellular collagen assembly and secretory cargo trafficking. *Science Advances*, *6*(7), eaay7667. https://doi.org/10.1126/sciadv.aay7667

Zheleznov-Chukotsky, N. K. (1994). *Ecology of snow sheep of northern*. Nauka.

Zheleznov-Chukotsky, N. K. (2007). OVIS-ZH-CH1 project "Conservation and multiplication of snow sheep (*Ovis nivicola* Eschscholtz, 1829) in Russia" 2007-2017. *Bulletin of Hunting Science*, *4*(3), 285–315.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Upadhyay, M., Kunz, E., Sandoval-Castellanos, E., Hauser, A., Krebs, S., Graf, A., Blum, H., Dotsev, A., Okhlopkov, I., Shakhin, A., Bagirov, V., Brem, G., Fries, R., Zinovieva, N., & Medugorac, I. (2021). Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Molecular Ecology*, 30, 6701–6717. https://doi.org/10.1111/mec.16184