A highly contiguous reference genome for the Alpine ibex (*Capra ibex*)

F. Gözde Çilingir[1,2,*], Fabio Landuzzi[3,*], Alice Brambilla[2,4], Debora Charrance[3], Federica Furia[3], Sara Trova[5], Alberto Peracino[4], Glauco Camenisch[2], Dominique Waldvogel[2], Jo Howard-McCombe[6], Yeraldin Chiquinquira Castillo De Spelorzi[7], Edoardo Henzen[7], Andrea Bernagozzi[8], Alessandro Coppe[3], Jean Marc Christille[8], Manuela Vecchi[5], Diego Vozzi[7], Andrea Cavalli[3,9], Bruno Bassano[4], Stefano Gustincich[5,10], Daniel Croll[11], Luca Pandolfini[5,10,✉] and Christine Grossen[1,✉]

[1] Swiss Federal Research Institute WSL, Biodiversity and Conservation Biology Research Unit, Birmensdorf, Switzerland

[2] University of Zurich, Department of Evolutionary Biology and Environmental Studies, Zurich, Switzerland

[3] Computational and Chemical Biology, Italian Institute of Technology (IIT), CMP^{3}VdA, Aosta, Italy

[4] Alpine Wildlife Research Center, Gran Paradiso National Park, Torino, Italy

[5] Non-coding RNA and RNA-based therapeutics, Italian Institute of Technology (IIT), CMP^{3}VdA, Aosta, Italy

[6] WildGenes Laboratory, Royal Zoological Society of Scotland, Edinburgh, UK

[7] Genomics Facility, Istituto Italiano di Tecnologia (IIT), Via Enrico Melen, 83 16152 Genoa, Italy

[8] Fondazione Clément Fillietroz ONLUS, Astronomical Observatory of Autonomous Region of Aosta Valley, Fraz Lignan 39, 11020,Nus, AO, Italy

[9] CECAM-EPFL, 1015 Lousanne, Switzerland

[10] Non-coding RNA and RNA-based therapeutics, Center for Human Technology, Italian Institute of Technology (IIT), via Enrico Melen, 83 16152 Genova

[11] Institute of Biology, University of Neuchâtel, 2000 Neuchâtel, Switzerland

* These authors equally contributed to the work

✉ co-corresponding authors (luca.pandolfini@iit.it, christine.grossen@wsl.ch)

Abstract

Anthropogenic hybridization, the unintentional hybridization with a non-native or domestic species in human-dominated environments, is a major concern for species conservation and a challenge for conservation management decisions. Genetically depleted species are expected to be particularly vulnerable to hybridization and introgression since hybridization can restore or introduce new adaptive genetic variation and alleviate the effects of inbreeding through hybrid vigour. However, defining the precise sets of deleterious or beneficial mutations resulting from anthropogenic hybridization is complex and limited by the quality of genomic resources. The Alpine ibex (*Capra ibex*), a species native to the Alps, faced near-extinction two centuries ago, but conservation programs have successfully restored its populations. Alpine ibex are known to hybridize with the domestic goat (*Capra hircus*) occasionally leading to hybrid swarm formation. Past introgression has been observed at immune-related genes and was suggested to have had an adaptive effect. Alpine ibex also carry deleterious mutation load from the reintroduction bottlenecks, which could be alleviated through admixture. Here, we produced a chromosome-level reference genome for Alpine ibex based on Oxford Nanopore sequencing coupled with high-throughput chromosome conformation capture. The highly contiguous assembly of 2.66 Gbp reveals 30 chromosomes and is augmented by a 98.8% complete RNAseq-assisted gene model prediction. The Alpine ibex genome presents overall a high degree of synteny compared to the domestic goat, with a number of structural variants spanning 38 Mb of insertion and deletions, as well as 49 Mb of inversions. We also identified structural variants near but not within the major histocompatibility complex (MHC), an immune-relevant gene complex, where previous studies found signals of introgression. The high degree of synteny between the Alpine ibex and domestic goat chromosomes likely facilitates recombination between haplotypes of the two species and is, therefore, in accordance with the observation of hybrid swarms. To determine the precise impact of recent admixture, we resequenced eight hybrid individuals sampled from two hybrid swarms in Northern Italy. Swarm individuals carried between 18-80% goat genome representing up to 3rd generation hybrids, including one F1 hybrid. The reference genome will facilitate quantifying maladaptive variation introduced from domestic goats and guide management efforts.

Introduction

The Anthropocene, characterized by significant human impact on Earth's ecosystems, has profoundly affected species and populations through habitat degradation, over-exploitation, climate change, and pollution (Bongaarts, 2019). One critical consequence is anthropogenic hybridization, where human-induced habitat modifications facilitate interbreeding between previously isolated species, or human-mediated translocations cause unintended hybridization with non-native or domestic species (Grabenstein & Taylor, 2018). This can lead to introgression, the incorporation of genetic material from one taxon into another, posing complex challenges for conservation, especially given the increased frequency of such events (Allendorf et al., 2001). Natural hybridization is considered an important evolutionary force, contributing to genetic variation and playing a major role in the evolution of many taxonomic groups (Genovart, 2009; Mallet, 2005). However, it also risks compromising genetic integrity, potentially causing native populations to lose local adaptations through the introgression of non-native alleles (Allendorf et al., 2001; Barton, 2001; Burke & Arnold, 2001). Bottlenecked (and hence genetically depleted) species are expected to be particularly receptive to hybridization and introgression since hybridization can alleviate the effects of inbreeding via hybrid vigour. Introgression can also play an important role in introducing new or previously lost adaptive genetic variation (Moran et al., 2021). However, the interplay between putative adaptive and maladaptive variation resulting from anthropogenic hybridization is complex and remains poorly understood. To effectively characterize introgression patterns along individual genomes and understand the local effects of introgression, it is necessary to use whole-genome data mapped to a reference genome (Theissinger et al., 2023). Consequently, high-quality reference genomes are indispensable for investigating the genomic extent and impact of hybridization and introgression.

The mountain ungulate Alpine ibex (*Capra ibex*) is endemic to the European Alps and has experienced severe bottlenecks since the early 19th century (Stüwe & Nievergelt, 1991) (Figure 1A, B). Active conservation and reintroduction efforts have restored the species to its original distribution area, with a current population of about 55,000 individuals (Brambilla et al., 2020). However, these bottlenecks have resulted in Alpine ibex having one of the lowest levels of genetic variation among mammals (Biebach & Keller, 2010; Grossen et al., 2020). This reduced genetic diversity has led to increased inbreeding and inbreeding depression with negative effects

both at the individual (Brambilla et al., 2015) and population level (Bozzuto et al., 2019). Recent studies have reported the occurrence of hybrid swarms between Alpine ibex and domestic goats (*Capra hircus*) (Brambilla et al., 2023; Moroni et al., 2022). Such populations of hybrids have survived beyond the initial hybrid generation and continue interbreeding between hybrid individuals and parental species. Extensive hybridization raises concerns of maladaptation and loss of ibex identity. Genome-wide introgression analyses showed that successful hybridizations likely occurred during the severe historic bottleneck when ibex numbers were at their lowest (Münger et al. 2024; Grossen et al. 2014). Introgression notably occurred at immune-relevant genes where Alpine ibex variability is very low (hence introgression likely favoured) and potentially conferred an adaptive advantage (Münger et al. 2024). However, hybridization with domestic goats is likely to be detrimental for Alpine ibex, with potential loss of adaptations to the alpine environment. Quantifying the long-term effects of anthropogenic hybridization on wild species would be important for their conservation, yet, not many studies in this regard have been conducted on mammals. The availability of the appropriate genetic resources is the first and fundamental step to fill this gap.

In this study, we generated the first high-quality chromosome-level genome assembly of Alpine ibex. To assess the utility of this highly contiguous reference genome, we generated high-resolution domestic goat versus ibex ancestry maps of eight recent hybrids based on whole-genome Illumina sequencing data from eight recent hybrids, 20 domestic goats, and 29 non-hybrid Alpine ibex.

Results and Discussion

A high-quality chromosome-level assembly

A chromosome-level genome was assembled from samples of a wild female Alpine ibex (named Odissea) collected in Lauson Valley (Cogne, Gran Paradiso National Park, Italy). We obtained a total of 130 Gb, high-quality ONT data to produce a draft assembly of 1040 contigs and a total size of 2.66 Gb. The contig N50 was 58.7 Mb, and the assembly was polished with Illumina short-read sequencing data. There was no evidence of contamination except for two contigs (3.9 and 47.8 Kb, respectively) representing reverse transcribed viral RNA, which were therefore

removed. The BUSCO and k-mer-based completeness were estimated as 96.1% (Figure 1C) and 98.6%, respectively. In total, 99.5% of the Hi-C reads aligned to the draft genome assembly and ca. 88% of the reads were found to have valid Hi-C contacts. A total of 30 pseudo-chromosomes were anchored, corresponding to 98.9% of the estimated genome size, yielding a chromosome-level reference genome with an N50 of 101.2 Mb, a BUSCO completeness of 96.1%, k-mer-based completeness of 98.6% and a consensus quality value (QV) of 49.9 (Figure 1C; Supplementary Table S1). The Alpine ibex assembly also includes the full mitochondrial genome (16.7 Kb, Supplementary Figure S1).

The genome is constituted by 43.4% repetitive elements (1.15 Gb). Retroelements were identified as the predominant class of repetitive elements, comprising 38.22% of the total, within which Long Interspersed Nuclear Elements (LINEs) were notably prevalent, accounting for 32.56%. Subsequently, Long Terminal Repeats (LTRs) were observed to represent 4.87% (Supplementary Table S2).

Gene model prediction and functional annotation

Gene model prediction yielded a BUSCO completeness level of 89.7% (Supplementary Table S3). Integrating annotations lifted over from sheep (*Ovis aries*) coding sequences, the BUSCO completeness increased to 99.4%. Through the annotation curation, 98.8% completeness in protein mode and 99.7% completeness in transcriptome mode were estimated. The resulting gene model set consisted of 32,783 protein-coding genes with a mean gene length of 9,487 bp (including introns) and a median of 8 exons per coding sequence (Supplementary Table S4). Functional annotation of the gene models showed that 54,205 out of 61,420 (88.2%) transcripts have matches in protein databases based on eggNOG. This corresponds to 26,581 out of 32,783 genes (81.1%) with homology matches. The mitochondrial genome encodes 13 proteins, including seven complex I proteins (ND1, ND2, ND3, ND4L, ND4, ND5, ND6), four complex IV proteins (COXI, COXII, COXIII, and CytB), and 2 ATP synthase subunits (ATPase 6 and 8; Supplementary Figure S1).

Structural variants segregating between Alpine ibex and domestic goats

We conducted gene order conservation analyses between Alpine ibex and three related domesticated mammals with chromosome-scale assemblies, including cattle (*Bos taurus*), sheep, domestic goat, and the human genome as an outgroup (Figure 1D; Supplementary Tables S5 and S6). The Alpine ibex assembly links the subchromosome contigs of the X chromosome (Xa and Xb) contigs from domestic goat consistent with sheep and cattle X chromosomes (Figure 1D). In addition, several small unplaced goat scaffolds were assigned to Alpine ibex chromosomes 12 and 17, matching chr17 and chr10 in sheep and cattle. We also identified a translocation between domestic goat chromosomes 23 and 18. The same regions in ibex, sheep, and cattle are on syntenic chromosomes 23, 20, and 23, respectively. Finally, we identified two chromosome fusions in the sheep genome on chromosomes 1 and 2. Overall, domestic goat and Alpine ibex genomes share 2.4 Gb of syntenic regions and 37.3 Mb of rearrangements, corresponding to 90.9% and 1.4% of the assemblies, respectively. The rearrangements include 1.2 Mb of duplications (0.017%), 10.8 Mb of translocations (0.41%), and 26.1 Mb of inversions (0.97%).

The levels of detected synteny were consistent with structural variant analyses performed with SyRI. Major rearrangements between the domestic goat and Alpine ibex were found on chromosomes 7, 18, and 23 (Figure 2A; Supplementary Figure S2). The rearrangements include primarily indels (38 Mb, 1.4% of the genome) and inversions (49 Mb, 1.8%), with duplications and translocations being the rarest variant types (Figure 2B, C). A total of 43 Mb (1.6%) were highly divergent regions (Figure 2B, C). Nearly half (45%) of all insertions, deletions and highly divergent regions between domestic goat and Alpine ibex overlapped with genes, while the gene overlap proportion was even higher for inversions (52%), duplications (54%) and translocations (78%).

Conserved synteny also at the MHC and high diversity gain in hybrids

The MHC is a gene complex important for immunity segregating high levels of polymorphisms within species and is prone to duplications (Piertney & Oliver, 2006). As a consequence, assembling MHC regions is challenging [see *e.g.* challenges for human MHC assembly (Chin et al., 2020; Jain et al., 2018; Koren et al., 2018)]. The MHC is of particular interest for Alpine ibex because hybridization with domestic goats led to geographically widespread introgression at the MHC (Grossen et al., 2014; Münger et al., 2024). We were able to generate a contiguous

assembly of the Alpine ibex MHC, with high degrees of synteny with cattle, sheep, and domestic goat at both MHC regions (Figure 3A). Structural variation analyses using SyRi detected no major structural variation between Alpine ibex and domestic goat within the region of MHC class II (Figure 3B). We detected one small translocation (~110 Kb) very close to the MHC class II cluster between position 27.14-27.26 Mb in the Alpine ibex genome to 20.87-20.98 Mb in the domestic goat genome, not affecting any MHC genes (Figure 3B). There was one major translocation (~1.92 Mb) further away from the MHC class II on chromosome 23 from 48.15-50.07 Mb on Alpine ibex to 14.94-16.86 Mb on domestic goats (Figure 3B). As reported for the domestic goat, the MHC class II cluster of Alpine ibex encodes DR- and DQ- genes but no DP genes, in contrast to the human MHC (Beck & Trowsdale, 1999; Kelley et al., 2005). The DR- and DQ-genes are exceptionally polymorphic, with several hundred alleles at the DRB gene in humans (Korman et al., 1985; Kuhner & Peterson, 1992; Marsh & Bodmer, 1993). Alpine ibex populations carry introgressed alleles from the domestic goat at the DRB gene reestablishing variation after a complete loss (Grossen et al., 2014). We found that average observed individual heterozygosity was lowest in non-hybrid Alpine ibex (especially ~27.25-27.5 Mb, a region containing MHC class I genes, Figure 3C). Hybrids showed high heterozygosity, especially in the DR- and DQ-gene regions (from ~25.4 - 25.7 Mb; Figure 3C). Diversity at the MHC is associated with disease resistance in a number of species (Fu et al., 2023; Gowane et al., 2014; Hedrick et al., 1999, 2001; Langefors et al., 2001; Paterson et al., 1998; Westerdahl, 2007), including Alpine ibex (Brambilla et al., 2018). The high diversity at the MHC in hybrids may therefore confer an advantage to early hybrids. This advantage may increase the probability of introgression and, therefore, potentially increase the risk of disrupting local adaptation in wild populations.

Genome-wide patterns of hybridization: recombination between Alpine ibex and domestic goat haplotypes

The high level of synteny and lack of major rearrangements between the Alpine ibex and domestic goat genomes described in this study should facilitate recombination between genomes in hybrids. This is also in accordance with observations of hybrids beyond the first generation (Brambilla et al., 2023). To investigate the genome-wide effect of recent hybridization and recombination patterns, we performed whole genome resequencing of eight recent hybrids

(previously genotyped at 63 diagnostic SNPs, Brambilla et al., 2023) from two hybrid swarms in Northern Italy and four domestic goats from the same two regions (Figure 4, Supplementary Table S7). A principal component analysis (PCA) based on 29 non-hybrid Alpine ibex, eight hybrids, and 20 domestic goats (Figure 4C) shows that the most important genetic differentiation (PC1, 32% of the variance explained) is separating domestic goats from non-hybrid Alpine ibex. The eight hybrids were at intermediate distances between the two species, as expected.

To define the precise contributions of Alpine ibex and domestic goat haplotypes in hybrids, we analyzed genome-wide polymorphisms of 8 hybrids with different levels of hybridization (including one F1 and 7 backcrosses). The chromosome paintings showed extensive long domestic goat and Alpine ibex haplotype blocks interrupted by a number of recombination events in hybrids (Figure 4D). The F1 hybrid (individual 12, Figure 4D) clearly showed the expected inheritance pattern of chromosome-length haplotypes from the two parental species, while backcrosses (e.g. Figure 4A and B) revealed more complex haplotype contributions from the two species consistent with additional recombination events (Figure 4D). The patterns confirmed previous findings based on 63 diagnostic SNPs, suggesting the existence of hybrids up to fourth generations (here up to third generation), including backcrosses in both directions (Brambilla et al., 2023).

Concluding remarks

Chromosome-level reference genomes are crucial tools for the conservation of threatened taxa (Brandies et al., 2019; Formenti et al., 2022; Paez et al., 2022). In this study, we assembled the first high-quality, chromosome-level genome for the Alpine ibex, a species with a history of dramatic recovery from near extinction. The high degree of synteny between the Alpine ibex and domestic goat chromosomes, including in the MHC region, is in accordance with the observation of hybrid swarms (Brambilla et al., 2023). The similarity of the two genomes likely facilitates recombination between haplotypes of the two species. Furthermore, the detected high heterozygosity in hybrids may confer early-generation hybrids an advantage over non-hybrid individuals suggesting a high risk of introgression with potential loss of local adaptations. Yet, it remains unclear whether introgression of domestic goat genetic material could be maladaptive

for Alpine ibex. The new Alpine ibex reference genome enables detailed analysis of ongoing hybridization processes and historical introgression consequences, including accurate identification of introgressed genome segments to distinguish between adaptive and maladaptive variations in hybrids. Given the increased rate of human-induced hybridization (Ottenburghs, 2021), such insights will be highly beneficial for other threatened taxa experiencing hybridization with domestic species.

Figures:

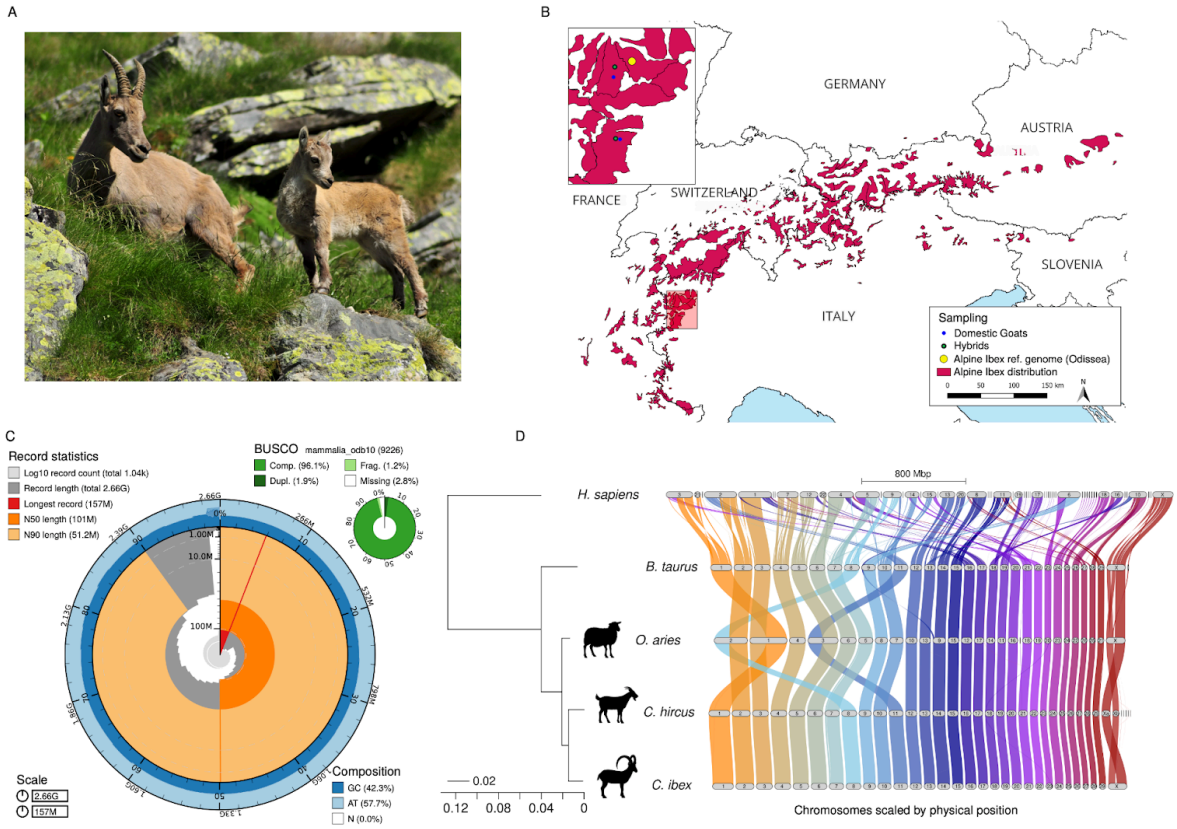


Figure 1: **Sampling and genome description**

A) Picture of a female Alpine ibex with her offspring (photo by Dario de Siena, Archivio PNGP). B) Map of species distribution and zoom in on the sampling location of Odissea (shown in yellow), the hybrids (green) and domestic goats (blue). C) Snail plot visualizing assembly statistics: scaffold statistics, sequence composition proportions, and BUSCO completeness scores. D) Riparian plot showing the results of gene synteny analysis conducted on the cattle, sheep, domestic goat and Alpine ibex genomes using GENESPACE. The human genome was used as an outgroup.

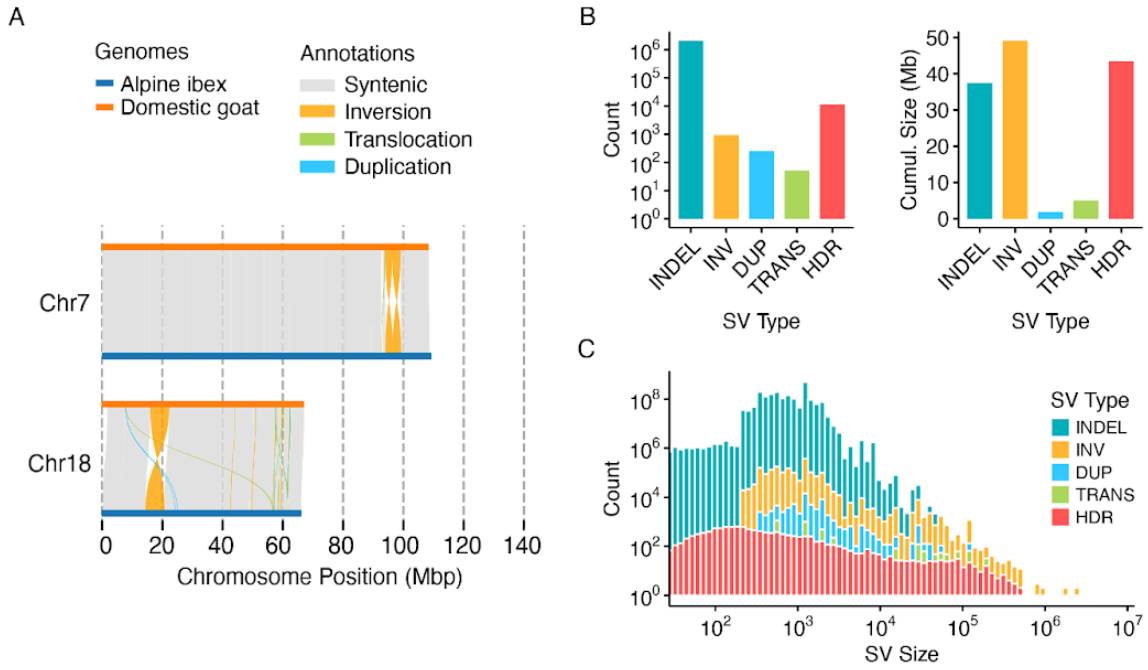
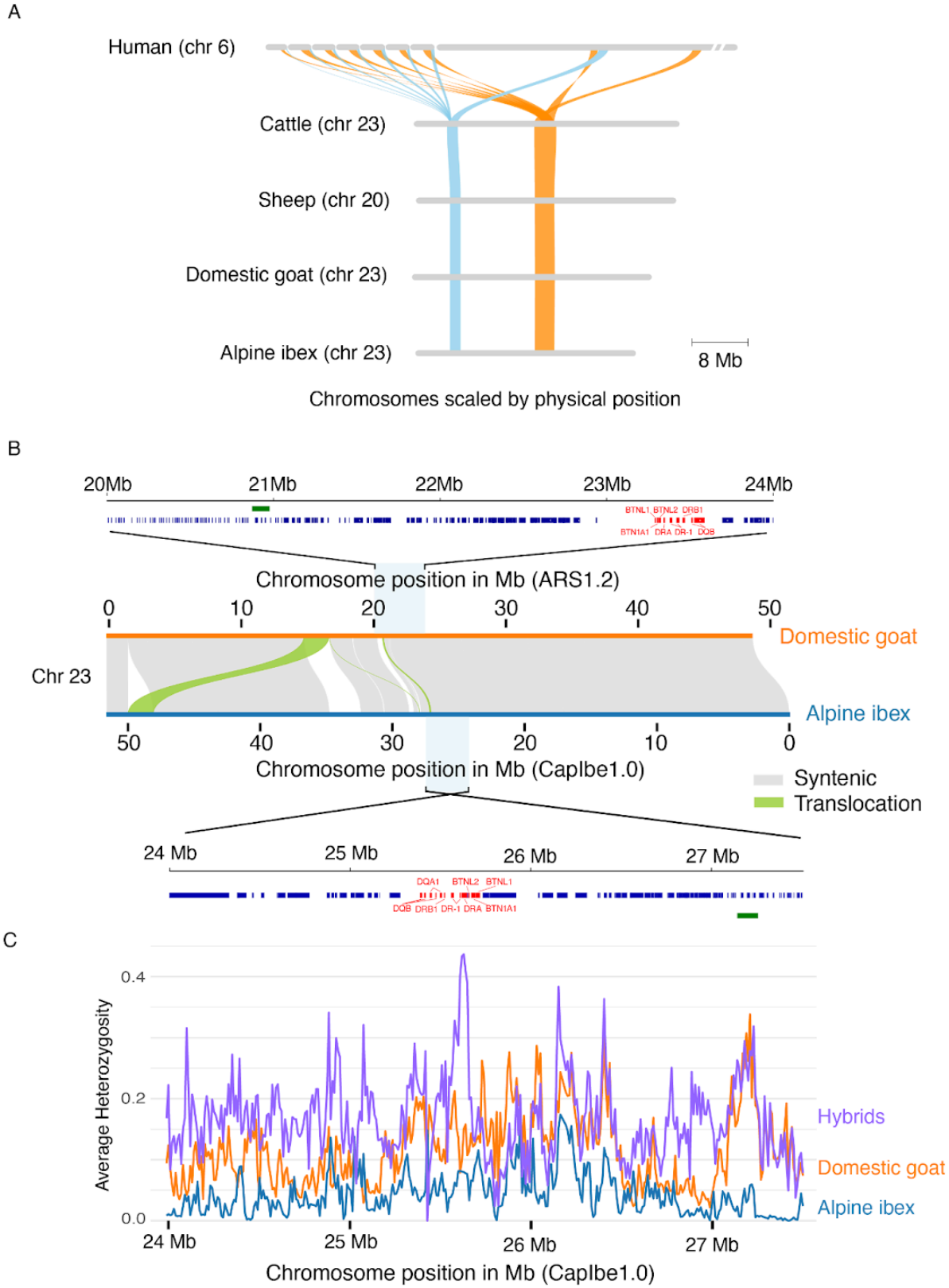


Figure 2: Structural variation

A) Major translocations between domestic goat and Alpine ibex observed on chromosomes 7 and 18 based on SyRI. B) Total counts and cumulative sizes of the structural variants identified by SyRI between the Alpine ibex and domestic goat genomes. C) Size distribution of translocations shown as a histogram from the same variants as in A (INDEL: insertion/deletion, INV: inversion, DUP: duplication, TRANS: translocation, HDR: highly divergent region).


A
Human (chr 6)
Cattle (chr 23)
Sheep (chr 20)
Domestic goat (chr 23)
Alpine ibex (chr 23)
8 Mb
Chromosomes scaled by physical position
B
20Mb
21Mb
22Mb
23Mb
24Mb
BTNL1 BTNL2 DRB1
BTN1A1 DRA DR-1 DQB
Chromosome position in Mb (ARS1.2)
0
10
20
30
40
50
Domestic goat
Chr 23
Alpine ibex
50
40
30
20
10
0
Chromosome position in Mb (CapIbe1.0)
Syntenic
Translocation
24 Mb
25 Mb
26 Mb
27 Mb
DQA1 BTNL2 BTNL1
DQB DRB1 DR-1 DRA BTN1A1
C
Average Heterozygosity
0.4
0.2
0.0
Hybrids
Domestic goat
Alpine ibex
24 Mb
25 Mb
26 Mb
27 Mb
Chromosome position in Mb (CapIbe1.0)

267

Figure 3: Conserved synteny at the MHC and high diversity in hybrids

A) Closeup of the synteny analysis, conducted with GENESPACE, highlighting two MHC regions within Alpine ibex chromosome 23 (light blue and orange, respectively).
B) Structural variation along chromosome 23 based on SyRI, including annotations of domestic goat (top) and Alpine ibex (bottom) in the range encompassing the major MHC II region. The green boxes highlight the genomic position of the small translocation shown in the synteny plot.
C) Average individual heterozygosity shown as observed heterozygosity in bins of 10 Kb per each group of Alpine ibex (orange), hybrids (purple), and domestic goat (blue).

Figure 4

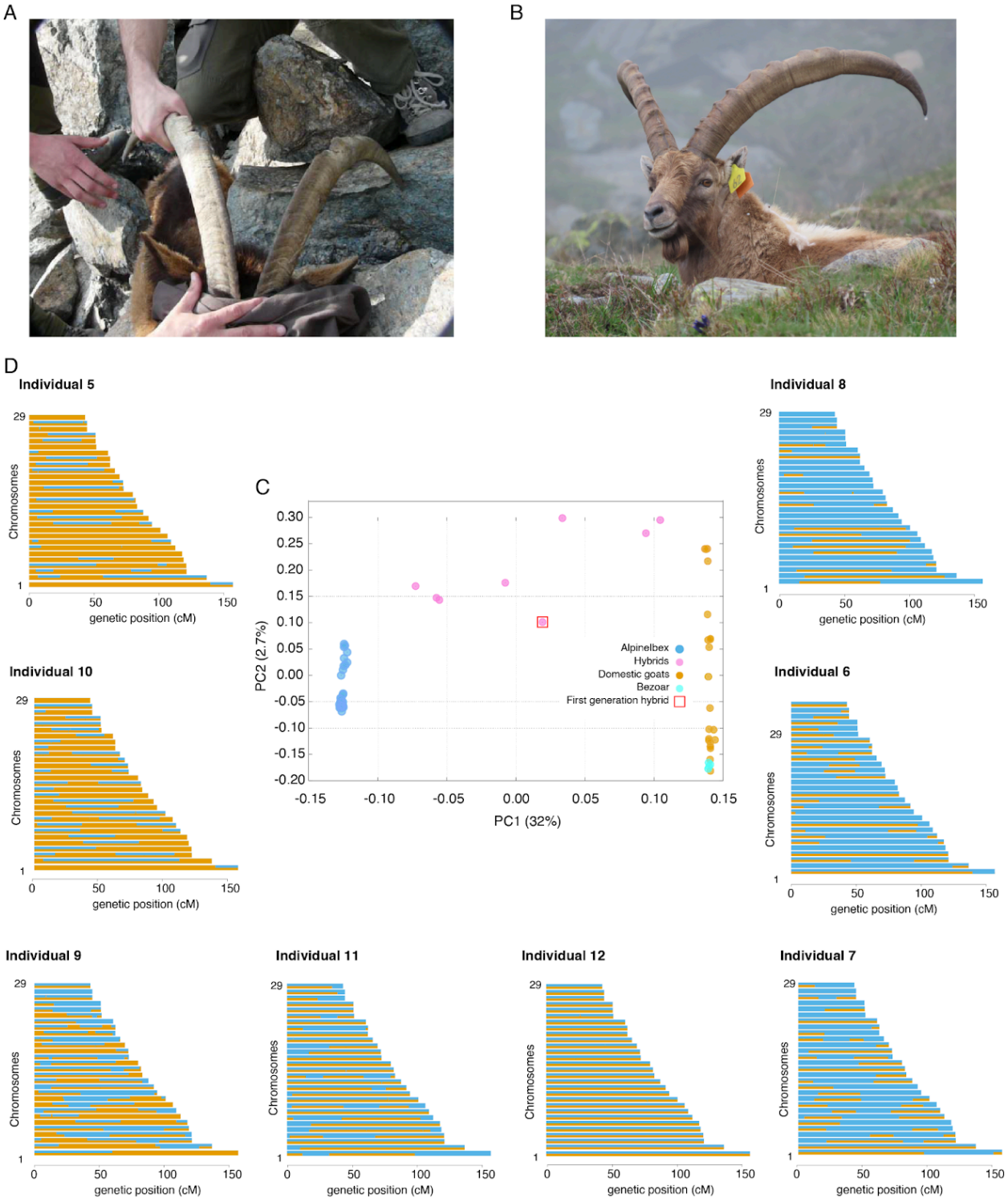


Figure 4: Genome-wide patterns of hybridisation

A) Picture of hybrid individual n 10.

B) Picture of hybrid individual n 8.

C) Principal Component Analysis showing the clustering of species and the distribution of hybrids along the conjunction between Alpine ibex and domestic goat/bezoar.
D) Chromosome paintings drawn with Mosaic, ordered according to the proportion of domestic goat ancestry (decreasing proportion of goat genome shown in orange from left to right).

Supplementary Material:

Supplementary Figure S1: Scheme of the Alpine ibex mitochondrial genome.

Supplementary Figure S2: Structural variant calling between domestic goat and Alpine ibex (based on SyRI) across all 29 autosomes.

Supplementary Figure S3: HiC contact map showing the data used for genome scaffolding.

Supplementary Table S1: Genome contiguity statistics of the draft and scaffolded Alpine ibex assemblies.

Supplementary Table S2: Summary of repeat annotations.

Supplementary Table S3: BUSCO statistics for the scaffolded assembly and the protein-coding gene annotation of Alpine ibex.

Supplementary Table S4: Genome annotation stats.

Supplementary Table S5: Syntenic blocks identified in the GENESPACE analysis.

Supplementary Table S6: Gene orthogroups identified in the GENESPACE analysis.

Supplementary Table S7: Goat and goat/ibex hybrid individuals sequenced in the present study.

Supplementary Table S8: Summary of the ONT runs used for genome assembly.

Supplementary Table S9: Species and IDs of the available datasets used in this study.

Materials and Methods

Sampling and DNA extraction for genome assembly

We collected fresh samples of different organs (spleen, muscles, liver, kidney) from a wild female Alpine ibex from the population of Gran Paradiso in the North-Western Italian Alps (Figure 1B). Immediately after the animal died of natural causes in Lauson Valley, Cogne (AO, Italy) on 18 June 2021, we brought it to the dissection laboratory of Gran Paradiso National Park and immediately dissected it. We collected the samples within four hours of the animal's death, placed the samples in cryo-stable plastic tubes (three aliquots from each organ), and immediately froze them at -80°C.

We performed High Molecular Weight DNA extraction to allow long-read sequencing for d*e novo* assembly of the reference genome of Alpine ibex. We carried out four separate extractions from spleen (N=2) and liver (N=2) tissue with MagAttract HMW DNA Kit (QIAGEN). We also performed a fifth extraction with spleen tissue using Monarch® HMW DNA Kit (New England BioLabs) for Ultra-Long library preparation.

Sampling and DNA extraction of hybrids

We collected blood and tissue samples from eight hybrid individuals between Alpine ibex and goats, and four domestic goats (Supplementary Table S7). The samples were collected within the framework of an extensive sampling carried out in two hotspots of suspected hybrid presence (Moroni et al., 2021, Brambilla et al., 2023): Gran Paradiso National Park (Valsavarenche, Aosta, Italy) and Lanzo Valley (Torino, Italy). Suspected hybrids were originally identified based on phenotypic characteristics and either captured with chemical immobilization or culled, and their hybrid status was subsequently confirmed based on 63 diagnostic SNPs (Brambilla et al., 2023). Domestic goat samples were collected in the same areas where the suspected hybrids were observed, during routine sanitary monitoring of the domestic herds.

Blood samples were collected in EDTA vacutainer tubes. Tissue samples were kept in absolute ethanol. Samples were stored at room temperature until DNA extraction, which was performed using QIAmp DNA mini and Biosprint 96 systems (QIAGEN), applying the appropriate protocol for each type of sample.

Library preparation and sequencing

Short-reads

We performed whole genome sequencing of gDNA (representing the reference individual, eight hybrids, and four domestic goats) on an Illumina NovaSeq 6000 platform (Illumina Inc., CA, USA) to generate paired 151 base-pair reads. We pooled individually barcoded hybrids during library preparation and performed sequencing whereas the library preparation and sequencing of the reference individual's gDNA were performed separately. For all samples, 350 ng of gDNA were used for library preparation with Illumina DNA PCR-Free Prep, Tagmentation library

preparation kit (Cat number 20041795), and IDT® for Illumina® DNA/RNA Unique Dual Indexes Set A or B, Tagmentation (96 Indexes, Cat number 20027214). Libraries were pooled by volume to reach a 90X coverage for the Alpine Ibex reference genome and a 25X coverage for hybrid samples. The Illumina bcl2fastq2 software was used to convert the original raw data format (bcl) to fastq.gz read files. Initial quality checks (including coverage and base qualities) were performed using FastQC (Andrews, 2010).

Long reads

For long-read sequencing, we utilized a PromethION 24 instrument (PRO-PRC024) following the instructions from Oxford Nanopore Technologies (Oxford, UK), using R9.4.1 chemistry flowcells (FLO-PRO002) and several library preparation kits, including Ligation Sequencing Kit (SQK-LSK110), Rapid Sequencing Kit (SQK-RAD004) and Ultra-Long Sequencing Kit (SQK-ULK001).

Reads were basecalled using Guppy version 6.1.1 in Super-High Accuracy mode (dna_r9.4.1_450bps_sup_prom.cfg). Only reads with a Q-score greater than 10 were retained for subsequent analysis. For the final yield of each run, see Supplementary Table S8.

Hi-C sequencing

The Hi-C sequencing library was prepared using 500 mg of liver tissue. Initially, the liver sample was treated with 1% formaldehyde for 15 minutes at room temperature for fixation. Next, solid glycine powder was added to achieve a final concentration of 125 mM, and the mixture was incubated for 15 minutes at room temperature with intermittent mixing. After centrifugation, the resulting pellet was resuspended in PBS + 1% Triton-X solution, followed by a 15-minute incubation at room temperature. The tissue was then collected by centrifugation of the mixture, transferred to a liquid nitrogen-cooled mortar, and ground to a fine powder. Finally, the cross-linked sample was shipped to Phase Genomics (Seattle, Washington, USA), while kept frozen on dry ice for sequencing. The Hi-C library construction was performed using the Phase Genomics Proximo Animal kit version 4.0. In summary, the DNA sample was digested with a fragmentation enzyme mix (including enzymes DpnII, DdeI, HinfI, and MseI), the 5’-overhangs were filled with a biotinylated nucleotide, the resulting fragments with blunt ends were ligated, and then sheared. The biotinylated ligation junctions were captured using streptavidin beads.

Finally, the resulting DNA fragments were sequenced using a NovaSeq 6000 instrument with a 150 bp paired-end run.

Genome assembly and polishing

The ONT sequencing data were assembled into the draft genome using the software Flye v2.9 (Kolmogorov et al., 2019). Then, the short-read data were aligned to this initial assembly using the GPU-accelerated fastq2bam tool of the NVIDIA Clara Parabricks v3.8 suite (Franke & Crowgey, 2020) with standard settings. The tool performs the following steps: alignment with BWA-mem (v0.7.15); sorting the coordinate with SortSam [GATK v4.1.0 (McKenna et al., 2010)]; and finally, applying MarkDuplicates (GATK v4.1.0). The resulting aligned data were then used in the Pilon v1.23 (Walker et al., 2014) to polish the assembly.

Contamination Scan

In order to identify and remove potential contaminant sequences in the assembled genome, we used the tool Foreign Contamination Screen developed by NCBI (https://github.com/ncbi/fcs). We scanned the genome with FCS-adaptor to identify short foreign sequences coming from laboratory processes and subsequently with FCS-GX (Astashyn et al., 2024) to detect contamination from foreign organisms coming from environmental sources.

Genome scaffolding using Hi-C sequencing data

Hi-C sequencing reads were quality filtered with Trimmomatic v0.39 (SLIDINGWINDOW:30:10 LEADING:19 TRAILING:19 AVGQUAL:20 MINLEN:30) (Bolger et al., 2014). Quality-filtered reads were mapped on the draft Alpine ibex genome by following the Arima Genomics mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline). Briefly, reads were mapped in single-end mode, chimeric joins across ligation junctions were filtered, and lastly, PCR duplicates were removed with Picard v2.6.0 (Broad Institute, 2019). The alignment files were then incorporated into the genome scaffolding using YaHS v1.1 (Zhou et al., 2023) with default settings. The scaffolded contigs were examined visually within JBAT v2.10.01 (Supplementary Figure S3) (Durand et al. 2016), and no manual refinement was found necessary.

Assembly quality evaluation

Main genome contiguity statistics were estimated with QUAST v5.0.2 (Mikheenko et al., 2018).

The k-mer completeness analysis and the reference-free QV estimation were performed with merqury v1.3 (Rhie et al., 2020) using the k-mer multiplicity of Illumina WGS reads utilized for the polishing of the Alpine ibex draft genome (see the section for Library Preparation and Sequencing of Short Reads) at k=21. The completeness of the assembly was also assessed based on evolutionarily informed expectations of gene content from near-universal single-copy orthologs via BUSCO v5.4.4 (Seppey et al., 2019) with default parameters, using the mammalian dataset (mammalia_odb10) in the genome mode.

Genome assembly statistics were visualized with a snail plot in BlobToolKit v4.3.2 (Challis et al., 2020) (Figure 1C).

Repeat annotation and masking

A species-specific *de novo* repeat library for the Alpine ibex genome was generated using RepeatModeler v2.0.3 (Flynn et al., 2020). For this, a species-specific repeat database was created with the BuildDatabase function of RepeatModeler using NCBI as the search engine. Then, *de novo* repeat family identification and modeling was performed with RepeatModeler utilizing RECON v1.08 (Bao & Eddy, 2002), RepeatScout v1.0.6 (Price et al., 2005), Tandem Repeats Finder (TRF) v4.09.1 (Benson, 1999), and RMBlast v2.11.0 (Johnson et al., 2008) on blast databases formatted with BuildDatabase. These consensus sequences were then used to softmask the genome with RepeatMasker v4.1.2 (https://www.repeatmasker.org/) using "-xsmall" flag. Additional masking by TRF (Benson, 1999) with maximum repeat period size of 500 was applied in order not to skip long tandem repeats since the default run of RepeatModeler does not identify tandem repeats with repeat pattern length of more than 10. Finally, merging of the repeat coordinates obtained from RepeatMasker outputs and additional masking with TRF was performed, and the softmasking of the genome with these merged coordinates was applied with bedtools v2.30.0 (Quinlan & Hall, 2010).

Gene model prediction, annotation, and curation

The gene model prediction was performed by using transcriptome and protein-homology-based approaches. Evidence for the transcriptome-based prediction was based on the already available RNA-seq dataset generated from retina/uvea, skin, heart, lung, lymph, bladder, ovary, kidney, liver, and spleen tissues of an Alpine ibex individual [NCBI BioProject PRJNA517635; (Grossen et al., 2020)]. The scaffolded Alpine ibex genome was indexed with STAR v2.7.2 (Dobin et al., 2013), and the RNA-seq data were aligned to the indexed genome using the *rna_fq2bam* tool of NVIDIA Clara Parabricks v3.8 (Franke & Crowgey, 2020).

The evidence for the protein-homology-based prediction was produced by combining all vertebrate proteins in the OrthoDB v10 (Kriventseva et al., 2019). This dataset was aligned against the softmasked, chromosome-level assembled reference genome via the ProtHint pipeline v2.6.0 (Brůna et al., 2020; Buchfink et al., 2015; Gotoh et al., 2014; Iwata & Gotoh, 2012; Lomsadze et al., 2005).

RNA-seq and homology-based evidence were incorporated for the braker2 v2.1.6 pipeline separately by running the software with --etmode and --epmode, respectively (Barnett et al., 2011; Brůna et al., 2021; Buchfink et al., 2015; Hoff et al., 2016, 2019; Li et al., 2009; Lomsadze et al., 2014; Stanke et al., 2006, 2008) with pre-trained parameter sets for human (*Homo sapiens)*, which is the evolutionarily closest taxon for Alpine ibex available within the software. Finally, all gene models derived from these runs were integrated into a high-confidence, non-redundant gene set using TSEBRA v1.1.1 (Gabriel et al., 2021) with the default configuration set.

Consequently, an annotation lift-over from the sheep reference genome [ARS-UI_Ramb_v2.0, NCBI Genome GCF_016772045.1, (International Sheep Genomics Consortium et al., 2010; R. Li et al., 2021)] was conducted with Liftoff v1.6.3 (Shumate & Salzberg, 2021) with “-copies” and “-polish” options. This way, extra copies of genes that were not annotated in the reference after the initial lift-over were searched, and the exons were realigned for polishing via restoring proper coding sequences in cases where the lift-over resulted in start/stop codon loss or introduced an in-frame stop codon. Finally, gene model prediction and lift-over results were

merged with TSEBRA (Gabriel et al., 2021), using "--keep_gtf" option and the default configuration set.

In order to reduce the redundancy of the transcript set obtained, we employed an isoform pruning pipeline as described in (Ivankovic et al., 2023). Briefly, transcripts were grouped according to the genomic locus position using gffread v0.12.7 (Pertea & Pertea, 2020). The longest Open Reading Frame (ORF) from each transcript was predicted using TransDecoder v5.5.0, and any shorter putative ORFs were excluded from further consideration. ORF clustering was performed using CD-HIT v4.8.1 (L. Fu et al., 2012) with a 90% identity threshold to group transcripts into clusters based on sequence similarity. Among the transcripts clustered within each gene, the top two transcripts with the longest ORFs were retained for each CD-HIT cluster. Genes lacking predicted ORFs in any of their transcripts were identified and separated into a distinct file. This step was implemented to safeguard very short monoexonic genes or non-coding RNAs (ncRNAs) from being overlooked during subsequent analyses. Finally, gene models were annotated using eggNOG mapper v2.1.11 (Cantalapiedra et al., 2021) in DIAMOND-search mode.

The completeness of the final annotation was assessed based on single-copy orthologs via BUSCO v5.4.4 (Seppey et al., 2019) with default parameters and the mammalian dataset (mammalia_odb10) in the protein mode.

Synteny analyses and structural rearrangement identification

We used GENESPACE v1.2.3 (Lovell et al., 2022) with default parameters to investigate synteny between human (GRCh38.p14, NCBI RefSeq GCF_000001405.40), cattle (ARS-UCD1.3, NCBI RefSeq GCF_002263795.2), sheep (ARS-UI_Ramb_v2.0, NCBI RefSeq GCF_016772045.2), domestic goat (ARS1.2, NCBI RefSeq GCF_001704415.2) and the Alpine ibex genomes. The transcript isoform corresponding to the longest ORF for each gene in a genome annotation was extracted. Only primary chromosome scaffolds were taken into account. Riparian plots were generated, setting the domestic goat genome as a reference. Further, syntenic regions and genome-wide rearrangements were identified only between closely related domestic goat (ARS1.2, NCBI RefSeq GCF_001704415.2) and the Alpine ibex genomes. For this, all chromosomes from each genome except for the sex chromosome were aligned to each other

using minimap2 v2.26 (Li, 2018) with default parameters (-ax asm5). The resulting alignments were processed with SyRI v1.6.3 (Goel et al., 2019) to identify syntenic regions and structural rearrangements. The syntenic regions and structural rearrangements were visualized with plotsr v1.1.1 (Goel & Schneeberger, 2022).

SNP genotyping

We compiled a dataset comprising the newly produced whole-genome re-sequencing data from eight hybrids and four domestic goats, available WGS data of Alpine ibex (N=29), domestic goat (N=18), bezoar, *Capra aegagrus* (N=4), Iberian ibex, *Capra pyrenaica* (N=4), Nubian ibex, *Capra nubiana* (N=2), Siberian ibex, *Capra sibirica* (N=2) and markhor, *Capra falconeri* (N=1). Additionally, four individuals representing outgroup species (chamois, *Rupicapra rupicapra*, cattle, sheep and the bighorn sheep, *Ovis canadensis*) were downloaded. BioProject and SRA accession IDs for each sample of this analysis are listed in Supplementary Table S9.

The downloaded WGS data were aligned to the newly generated chromosome-level assembly of Alpine ibex using the pipeline *germline* of the NVIDIA Clara Parabricks v3.8 (Franke & Crowgey, 2020), activating the flag *-gvcf* and without specifying a reference data set (no *--known-indels* option). The pipeline initially executed the *fastq2bam* tools (as described in the section "Genome assembly and polishing"). Then a GPU-accelerated version of *HaplotypeCaller* (GATK v4.1.0) for the variant calling was utilized. Later, a consolidation step was performed, where all the GVCFs, generated in the previous step were merged into a cohort file with *CombineGVCFs* (GATK v4.1.0). To speed-up the execution, the genome was divided into intervals of similar size (each interval could contain one or multiple contigs). Each cohort GVCF was then used for the joint genotyping with GenotypeGVCFs (GATK v4.1.0). Finally, the compressed VCF obtained from each region was merged into a single file containing all SNPs and indels from all samples.

The VCF containing all variants underwent a hard-filtering process: first, quality flags were set for each variant call using the function *VariantFiltation* (GATK v4.1.0) with thresholds set to QD>=2.0, FS<=40.0, SOR<=4.0, MQ>=30.0, MQRankSum>=-4.0, MQRankSumH<=4.0, ReadPosRankSum>=-2.0, ReadPosRankSumH<=2.0 and AN>=37. Subsequently, the variants

that did not match these quality standards were filtered out using the SelectVariants function (GATK v4.1.0). Of these, all SNPs polymorphic among the *Capra spp*. (removing all SNPs only polymorphic among the outgroup) were extracted, which were autosomal, bi-allelic, with a minimal genotype quality of 20 and a minimal genotyping rate of 80% using VCFtools v0.1.16 (Danecek et al., 2011).

Average observed heterozygosity per-site was estimated for each species/group using the "--hardy" option of VCFtools (Danecek et al., 2011).

Principal component analysis

The SNPs polymorphic within the *Capra spp.* were used to perform a Principal Component Analysis (PCA) with PLINK. All variants with minor allele frequency below 0.05 and genotype missingness rate of over 10% were discarded (--maf 0.05 --geno 0.1). Then, the variants in linkage-disequilibrium were filtered out (*--indep-pairwise 1000 200 0.2*). The variants that passed these filters were used to perform the PCA. The first two principal components explained 32.8% of the data variation [respectively PC_1 = 24.46 (32%) and PC_2 = 2.70 (2.7%), Figure 4C].

Genome-wide analyses of hybridization

To analyze Alpine ibex and domestic goat haplotypes (and perform chromosome painting), we used the R-toolbox Mosaic (Salter-Townshend & Myers, 2019). Mosaic searches for haplotypes that are shared between defined target individuals (here eight hybrids and four domestic goats) and individuals in several putative source panels (here domestic goat, bezoar -the wild ancestor of domestic goat-, Iberian ibex, Siberian ibex, and Nubian ibex). Then it builds an admixture model, which is consistent with mixing between unseen groups giving rise to the observed data. Before running Mosaic, vcftools was used to thin the data down to a minimal inter-SNP distance of 1 Kb. Next, Phasing was performed using beagle assuming a linear relationship between genetic and physical distance (1 Mb = 1 cM) because no recombination map is available for Alpine ibex. Custom scripts were then used to prepare the required input files (snpfiles, sample lists and genotype files) before running mosaic in R.

Chromosome painting visualization was performed using the function karyogram in Mosaic.

Data availability

The final Alpine ibex genome assembly and high-confidence coding gene annotations are publicly available in the NCBI repository XXXXXX (currently undergoing NCBI processing). Raw sequencing experiment data were submitted to the SRA archive under the project identifier PRJNA1104824 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1104824).

Acknowledgments:

For domestic goat sampling we thank Stefania Zanet, Alice Naudin and Emilio Gugliermetti. For hybrid sampling we thank the Surveillance Service of Gran Paradiso National Park, the Forestry Service of Ragione Valle d'Aosta, Claudio Scaini and the personnel of Città Metropolitana di Torino and of CATO4, Noel Zehnder and Luca Rossi. We thank Sara Gottardo, Francesca Groppo, and the sequencing facility of CMP^{3}VdA in Aosta for technical assistance with the whole-genome Illumina sequencing. We thank Sergio Decherchi, Alessandro Parodi and Mattia Pini for their support for high-performance computing and data storage and Dario deSiena and Raffaele Turvani for the beautiful ibex pictures

F. Gözde Çilingir was funded by the Swiss National Science Foundation (IZCOZ0_198147 to A.B. and C.G. and 31003A_182343 to C.G.).

This work was supported by the 5000genomi@VdA Project co-founded by "Fondo Europeo di Sviluppo Regionale (FESR CUPB68H19005520007) to SG and AC; FL, ST, AC, and MV were supported by "Fondo Europeo di Sviluppo Regionale (FESR CUPB68H19005520007); DC and FF were recipient of a fellowship from Programma Investimenti per la crescita e l'occupazione 2014/20" (European Social Fund, ESF CUP B65F19001200009).

References

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622.

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Astashyn, A., Tvedte, E. S., Sweeney, D., Sapojnikov, V., Bouk, N., Joukov, V., Mozes, E., Strope, P. K., Sylla, P. M., Wagner, L., Bidwell, S. L., Brown, L. C., Clark, K., Davis, E. W., Smith-White, B., Hlavina, W., Pruitt, K. D., Schneider, V. A., & Murphy, T. D. (2024). Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biology*, *25*(1), 60.

Bao, Z., & Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*, *12*(8), 1269–1276.

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, *27*(12), 1691–1692.

Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*(3), 551–568.

Beck, S., & Trowsdale, J. (1999). Sequence organisation of the class II region of the human MHC. *Immunological Reviews*, *167*, 201–210.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580.

Biebach, I., & Keller, L. F. (2010). Inbreeding in reintroduced populations: the effects of early reintroduction history and contemporary processes. *Conservation Genetics*, *11*(2), 527–538.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bongaarts, J. (2019). IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In *Population and Development Review* (Vol. 45, Issue 3, pp.

680–681). https://doi.org/10.1111/padr.12283

Bozzuto, C., Biebach, I., Muff, S., Ives, A. R., & Keller, L. F. (2019). Inbreeding reduces long-term growth of Alpine ibex populations. *Nature Ecology & Evolution*, *3*(9), 1359–1364.

Brambilla, A., Biebach, I., Bassano, B., Bogliani, G., & von Hardenberg, A. (2015). Direct and indirect causal effects of heterozygosity on fitness-related traits in Alpine ibex. *Proceedings of the Royal Society B*, *282*(1798), 20141873.

Brambilla, A., Keller, L., Bassano, B., & Grossen, C. (2018). Heterozygosity-fitness correlation at the major histocompatibility complex despite low variation in Alpine ibex (*Capra ibex*). *Evolutionary Applications*, *11*(5), 631–644.

Brambilla, A., Von Hardenberg, A., Nelli, L., & Bassano, B. (2020). Distribution, status, and recent population dynamics of Alpine ibex *Capra ibex* in Europe. *Mammal Review*, *50*(3), 267–277.

Brambilla, A., Zehnder, N., Bassano, B., Rossi, L., & Grossen, C. (2023). Genetic evidence of a hybrid swarm between Alpine ibex and domestic goat. *bioRxiv*, https://doi.org/10.1101/2023.12.22.572967

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, *10*(11). https://doi.org/10.3390/genes10110846

Broad Institute. (2019). Picard Toolkit. https://broadinstitute.github.io/picard/

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*(1), lqaa108.

Brůna, T., Lomsadze, A., & Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics*, *2*(2), lqaa026.

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60.

Burke, J. M., & Arnold, M. L. (2001). Genetics and the fitness of hybrids. *Annual Review of Genetics*, *35*, 31–52.

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021).

eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829.

Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit--interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*, *10*(4), 1361–1374.

Chin, C.-S., Wagner, J., Zeng, Q., Garrison, E., Garg, S., Fungtammasan, A., Rautiainen, M., Aganezov, S., Kirsche, M., Zarate, S., Schatz, M. C., Xiao, C., Rowell, W. J., Markello, C., Farek, J., Sedlazeck, F. J., Bansal, V., Yoo, B., Miller, N., Zhou, X., Carroll, A., Barrio, A. M., Salit, M., Marschall, T., Dilthey, A. T., Zook, J. M. (2020). A diploid assembly-based benchmark for variants in the major histocompatibility complex. *Nature Communications*, *11*(1), 4794.

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: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, *3*(1), 99–101.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451–9457.

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svardal, H., Theofanopoulou, C., de Vries, J., Waldvogel, A.-M.,

Zhang, G., Mazzoni, C. J., Jarvis, E. D., Bálint, M., & European Reference Genome Atlas (ERGA) Consortium.(2022). The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution*, *37*(3), 197–202.

Franke, K. R., & Crowgey, E. L. (2020). Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. *Genomics & Informatics*, *18*(1), e10.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152.

Fu, M., Eimes, J. A., Kong, S., Lamichhaney, S., & Waldman, B. (2023). Identification of major histocompatibility complex genotypes associated with resistance to an amphibian emerging infectious disease. *Infection, Genetics and Evolution*, *113*, 105470.

Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M., & Stanke, M. (2021). TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*, *22*(1), 566.

Genovart, M. (2009). Natural hybridization and conservation. *Biodiversity and Conservation*, *18*(6), 1435–1439.

Goel, M., & Schneeberger, K. (2022). plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*, *38*(10), 2922–2926.

Goel, M., Sun, H., Jiao, W.-B., & Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, *20*(1), 277.

Gotoh, O., Morita, M., & Nelson, D. R. (2014). Assessment and refinement of eukaryotic gene structure prediction with gene-structure-aware multiple protein sequence alignment. *BMC Bioinformatics*, *15*, 189.

Gowane, G., Vandre, R., Nangre, M., & Sharma, A. (2014). Major histocompatibility complex (MHC) of bovines: an insight into infectious disease resistance. *Livestock Research International*, 1(2), 46-57.

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking Bbarriers: Causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution*, *33*(3), 198–212.

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications*, *11*(1), 1001.

Grossen, C., Keller, L., Biebach, I., International Goat Genome Consortium, & Croll, D. (2014). Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. *PLoS Genetics*, *10*(6), e1004438.

Hedrick, P. W., Kim, T. J., & Parker, K. M. (2001). Parasite resistance and genetic variation in the endangered Gila topminnow. *Animal Conservation Forum*, *4*(2), 103–109.

Hedrick, P. W., Parker, K. M., Miller, E. L., & Miller, P. S. (1999). Major histocompatibility complex variation in the endangered Przewalski's horse. *Genetics*, *152*(4), 1701–1710.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*(5), 767–769.

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods in Molecular Biology*, *1962*, 65–95.

International Sheep Genomics Consortium, Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., Maddox, J. F., McEwan, J. C., Hutton Oddy, V., Raadsma, H. W., Wade, C., Wang, J., Wang, W., & Xun, X. (2010). The sheep genome reference sequence: a work in progress. *Animal Genetics*, *41*(5), 449–453.

Ivankovic, M., Brand, J. N., Pandolfini, L., Brown, T., Pippel, M., Rozanski, A., Schubert, T., Grohme, M. A., Winkler, S., Robledillo, L., Zhang, M., Codino, A., Gustincich, S., Vila-Farré, M., Zhang, S., Papantonis, A., Marques, A., & Rink, J. C. (2023). A comparative analysis of planarian genomes reveals regulatory conservation in the face of rapid structural divergence. *bioRxiv*. https://doi.org/10.1101/2023.12.22.572568

Iwata, H., & Gotoh, O. (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Research*, *40*(20), e161.

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., Snutch, T. P., Tee, L., Paten, B., Phillippy, A. M., Simpson, J. T., Loman, N. J., & Loose, M.(2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*, W5–W9.

Kelley, J., Walter, L., & Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics*, *56*(10), 683–695.

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546.

Koren, S., Rhie, A., Walenz, B. P., Dilthey, A. T., Bickhart, D. M., Kingan, S. B., Hiendleder, S., Williams, J. L., Smith, T. P. L., & Phillippy, A. M. (2018). De novo assembly of haplotype-resolved genomes with trio binning. *Nature Biotechnology*. https://doi.org/10.1038/nbt.4277

Korman, A. J., Boss, J. M., Spies, T., Sorrentino, R., Okada, K., & Strominger, J. L. (1985). Genetic complexity and expression of human class II histocompatibility antigens. *Immunological Reviews*, *85*, 45–86.

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *47*(D1), D807–D811.

Kuhner, M. K., & Peterson, M. J. (1992). Genetic exchange in the evolution of the human MHC class II loci. *Tissue Antigens*, *39*(4), 209–215.

Langefors, A., Lohm, J., Grahn, M., Andersen, O., & von Schantz, T. (2001). Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon. *Proceedings. Biological Sciences / The Royal Society*, *268*(1466), 479–485.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Li, R., Yang, P., Li, M., Fang, W., Yue, X., Nanaei, H. A., Gan, S., Du, D., Cai, Y., Dai, X., Yang, Q., Cao, C., Deng, W., He, S., Li, W., Ma, R., Liu, M., & Jiang, Y. (2021). A Hu sheep genome with the first ovine Y chromosome reveal introgression history after sheep domestication. *Science China. Life Sciences*, *64*(7), 1116–1130.

Lomsadze, A., Burns, P. D., & Borodovsky, M. (2014). Integration of mapped RNA-seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, *42*(15), e119.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, *33*(20), 6494–6506.

Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, *11*. https://doi.org/10.7554/eLife.78526

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237.

Marsh, S. G., & Bodmer, J. G. (1993). HLA class II nucleotide sequences, 1992. *European Journal of Immunogenetics*, *37*(2), 79-94.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150.

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *eLife*, *10*, e69016.

Moroni, B., Brambilla, A., Rossi, L., Meneguz, P. G., Bassano, B., & Tizzani, P. (2022). Hybridization between Alpine ibex and domestic goat in the Alps: A sporadic and localized phenomenon? *Animals*, *12*(6), 751.

Münger, X., Robin, M., Dalén, L., & Grossen, C. (2024). Facilitated introgression from domestic goat into Alpine ibex at immune loci. *Molecular Ecology*, e17429.

Ottenburghs, J. (2021). The genic view of hybridization in the Anthropocene. *Evolutionary Applications*, *14*(10), 2342–2360.

Paez, S., Kraus, R. H. S., Shapiro, B., Gilbert, M. T. P., Jarvis, E. D., Vertebrate Genomes Project Conservation Group, Al-Ajli, F. O., Ceballos, G., Crawford, A. J., Fedrigo, O., Johnson, R. N., Johnson, W. E., Marques-Bonet, T., Morin, P. A., Mueller, R. C., Ryder, O. A., Teeling, E. C., & Venkatesh, B. (2022). Reference genomes for conservation. *Science*, *377*(6604), 364–366.

Paterson, S., Wilson, K., & Pemberton, J. M. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (Ovis aries L.). *Proceedings of the National Academy of Sciences*, *95*(7), 3714–3719.

Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, *9*. https://doi.org/10.12688/f1000research.23297.2

Piertney, S. B., & Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, *96*(1), 7–21.

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, *21 Suppl 1*, i351–i358.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1), 245.

Salter-Townshend, M., & Myers, S. (2019). Fine-scale inference of ancestry segments without prior knowledge of admixing groups. *Genetics*, *212*(3), 869–889.

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, *1962*, 227–245.

Shumate, A., & Salzberg, S. L. (2021). Liftoff: accurate mapping of gene annotations. *Bioinformatics*, *37*(12), 1639–1643.

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644.

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, *7*, 62.

Stüwe, M., & Nievergelt, B. (1991). Recovery of alpine ibex from near extinction: the result of effective protection, captive breeding, and reintroductions. *Applied Animal Behaviour Science*, *29*(1), 379–387.

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G. R., Godoy, J. A., Jentoft, S., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Secomandi, S., Svardal, H., Theofanopoulou, C., de Vries, J., Waldvogel, A.-M., Zhang, G., Jarvis, E. D., Bálint, M., Ciofi, C., Waterhouse, R. M., Mazzoni, C. J., Höglund, J., & European Reference Genome Atlas Consortium. (2023). How genomics can help biodiversity conservation. *Trends in Genetics*, *39*(7), 545–559.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One*, *9*(11), e112963.

Westerdahl, H. (2007). Passerine MHC: genetic variation and disease resistance in the wild. *Journal of Ornithology*, *148*(2), 469–477.

Zhou, C., McCarthy, S. A., & Durbin, R. (2023). YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*, *39*(1), btac808.