Hybridization load introduced by Alpine ibex hybrid swarms

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, Turin, 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, United Kingdom

[7] Genomics Facility, Istituto Italiano di Tecnologia (IIT), Genoa, Italy

[8] Fondazione Clément Fillietroz ONLUS, Astronomical Observatory of Autonomous Region of Aosta Valley, Aosta, Italy

[9] CECAM-EPFL, Lausanne, Switzerland

[10] Non-coding RNA and RNA-based therapeutics, Center for Human Technology, Italian Institute of Technology (IIT), Genoa, Italy

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

[+] These authors equally contributed to the work

*co-corresponding authors (daniel.croll@unine.ch, luca.pandolfini@iit.it, christine.grossen@wsl.ch)

lead contact (christine.grossen@wsl.ch)

Summary

Species restoration efforts can be threatened by the accumulation of deleterious mutations and inbreeding depression associated with the historic population contraction. However, even successfully restored species could face deleterious mutation swamping from hybridization with an abundant and closely related species. Here, we analyze this risk for Alpine ibex (*Capra ibex*), a flagship species of large mammal restoration in the Alps. The Alpine ibex faced near-extinction two centuries ago, resulting in exceptionally low genome-wide diversity and increased inbreeding, which facilitated the purging of severe deleterious mutation load. For this, we produced a highly contiguous chromosome-level genome assembly of the Alpine ibex capturing structural divergence from domestic goat (*Capra hircus*) and mapping immune-relevant MHC genes. Analyses of eight recent ibex-goat hybrids from two swarms in Northern Italy, combined with 29 non-hybrid Alpine ibex and 22 domestic goats, identified 215 masked loss-of-function (LOF) mutations introduced via hybridization. Yet, we found no evidence for counter-selection in early backcrosses. This exposes Alpine ibex to further backcrosses compounding the deleterious mutation load of the species by a factor of up to two. Our work provides one of the first direct estimates of hybridization load and guides conservation efforts to preserve endangered species gene pools.

Keywords: conservation genetics, deleterious mutation load, anthropogenic hybridization, hybrid swarm, hybridization load, heterosis, Alpine ibex

Results and discussion

Human activity has profoundly affected species and populations through habitat degradation and over-exploitation leading to severe population reductions and extinctions [1]. A particular concern for species conservation is increased levels of inbreeding and the expression of deleterious mutations in small populations. Recent studies provided compelling evidence for changes in deleterious mutation load in natural populations (reviewed in Dussex et al. [2]; Robinson et al. [3]). As predicted by theory, population bottlenecks were repeatedly shown to lead to the selective removal (purging) of highly deleterious (usually recessive) mutations, while less severely deleterious mutations can accumulate (*e.g.,* Xue et al.[4], Grossen et al. [5], Dussex et al. [6]; reviewed in Dussex et al. [2]). Mutation load is defined as load resulting from newly introduced deleterious mutations [7]. Empirical assessments of fitness effects confirmed the detrimental impact of mutation load and inbreeding on both individual and population fitness [8,9]. However, the origin of deleterious mutations remains often unresolved. In addition to *de novo* mutations, a potentially significant source of mutation load could be deleterious variants introduced into a population through hybridization.

Hybridization in nature contributes to genetic variation and fosters adaptation [10,11]. Hybridization in bottlenecked species may introduce new adaptive genetic variation, compensating previously lost variation [12]. In addition, high levels of heterozygosity in hybrids can also initially mask deleterious mutations, thereby causing hybrid vigor (*i.e.*, heterosis) and mitigating inbreeding. Yet, this initial masking of deleterious mutations prevents their counter-selection in early generations. Hybridization also poses the risk of disrupting local adaptation through introgression of non-native alleles, in particular if the new species contact is caused by human activities [13–15]. Such introduction of deleterious mutations is defined as hybridization load [12]. Overall, deleterious mutations played a significant role in hybridization events in animals such as fish [16,17], plants including crops [18–20], and trees [21]. These studies indicate that while hybridization can introduce beneficial alleles, linked deleterious variants can reduce fitness and create partial barriers to introgression. The interplay of adaptive and maladaptive variation introduced through hybridization is likely complex and poorly understood.

As hybridization can introduce new adaptive genetic variation and alleviate the effects of inbreeding through hybrid vigour, genetically impoverished species may be particularly susceptible to the introduction of hybridization load. To quantify load in genomes and to predict effects of introgression, high-quality genomic resources are essential [22]. The Alpine ibex (*Capra*

86 *ibex*), a mountain ungulate endemic to the European Alps (Figure 1A, B), has experienced severe
87 bottlenecks since the early 19th century and repeated hybridization events [12,23,24]. However, the
88 evolution of deleterious mutation load driven by hybridization remains poorly understood.
89
90 Conservation efforts have restored the species to its original distribution area, with a current
91 population of about 55,000 individuals [25]. However, these bottlenecks have resulted in Alpine ibex
92 having one of the lowest levels of genetic variation among mammals [5,26]. This reduced genetic
93 diversity has led to increased inbreeding and inbreeding depression, with negative effects both at
94 the individual [27] and population level [28]. Severe genetic load was likely reduced through purging
95 of highly deleterious mutations, while less severe mutations accumulated [5]. Genome-wide
96 analyses showed that successful hybridizations with domestic goats (*Capra hircus*) likely occurred
97 during the severe historic bottleneck when ibex numbers were at their lowest [29,30]. Introgression
98 was particularly evident at immune-relevant genes, where genetic variability in Alpine ibex is very
99 low. Such introgressed variants may have been under positive selection due to heterozygote
100 advantage at immune loci [29]. However, hybridization is also expected to have detrimental
101 consequences for Alpine ibex, with potential loss of adaptations to the alpine environment and
102 the introduction of deleterious variants. Of particular concern are the recently discovered hybrid
103 swarms between Alpine ibex and domestic goats in Northern Italy [24,31]. Such populations of
104 hybrids have survived beyond the initial hybrid (F1) generation, and there is continued
105 backcrossing between hybrid individuals and parental species [24]. This extensive hybridization
106 raises concerns of maladaptation and loss of ibex identity. Here, we generated the first reference-
107 quality genome for Alpine ibex to quantify the hybridization load generated by high rates of
108 hybridization as observed in hybrid swarms.
109
110 **A high-quality chromosome-level assembly**
111 We assembled a chromosome-level genome of a wild female Alpine ibex (Odissea) from the Gran
112 Paradiso National Park population (Lauson Valley, Cogne, Italy). Using 130 Gb of high-quality
113 ONT data, we generated a 2.66 Gb draft genome with 1040 contigs (N50 58.7 Mb), polished with
114 short read sequencing data. The contigs showed high completeness, with BUSCO and k-mer-
115 based scores of 96.1% and 98.6%, respectively (Figure 1C). We generated Hi-C data to anchor
116 contigs into 30 pseudo-chromosomes (Figure S1) with a resulting N50 of 101.2 Mb and matching
117 the chromosome numbers of domestic goats [32] (Table S1, Figure 1D). The Alpine ibex assembly
118 also includes the full mitochondrial genome (16.7 Kb, Figure S2). The nuclear genome is
119 constituted of 43.4% repetitive elements (1.15 Gb), dominated by retroelements (38.22%),

including Long Interspersed Nuclear Elements (LINEs: 32.56%) and Long Terminal Repeats (LTRs: 4.87%) (Table S2). We integrated gene model evidence using transcriptome-assisted predictions as well as annotations lifted over from the sheep (*Ovis aries*) genome, yielding a BUSCO completeness of 99.7% at the transcript level (Table S3). The annotated gene set consisted of 32,783 protein-coding genes and 61,420 transcripts (Table S4).

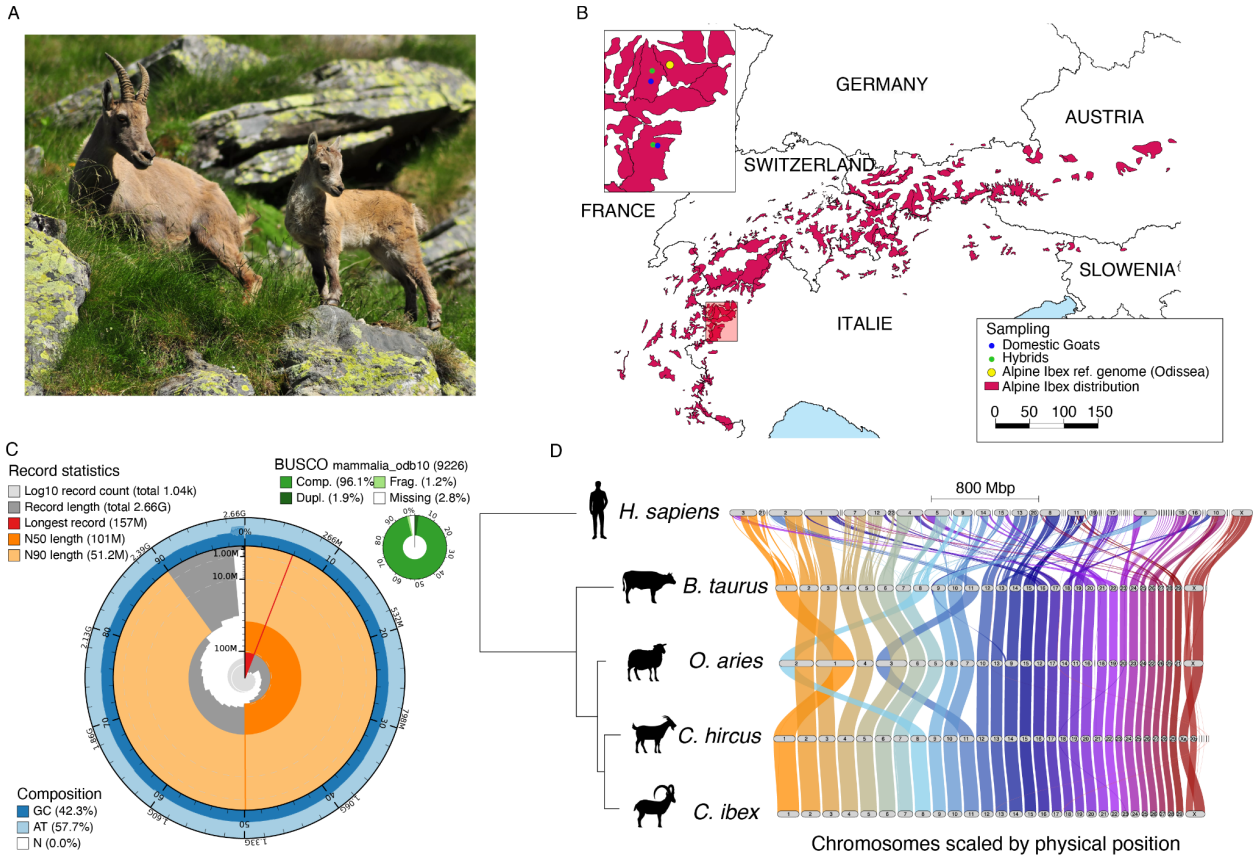

Figure 1: **Sampling and genome description.** A) Picture of a female Alpine ibex with her offspring (photo by Dario de Siena, Archivio PNGP). B) Map of species distribution and zoom in on the sampling location of Odissea (shown in yellow), the hybrids (green), and domestic goats (blue). C) Snail plot visualizing assembly statistics: scaffold statistics, sequence composition proportions, and BUSCO completeness scores. D) Riparian plot showing the results of the gene synteny analysis conducted on the cattle, sheep, domestic goat, and Alpine ibex genomes using GENESPACE. The human genome was used as an outgroup. See also Figures S1-S4.

Recent structural variants arising in Alpine ibex

137 We used gene order conservation analyses between Alpine ibex and three related domesticated
138 mammals, including cattle (*Bos taurus*), sheep (*Ovis aries*), and domestic goat, as well as the
139 human genome as an outgroup to detect chromosomal rearrangements (Figure 1D; Tables S5
140 and S6). Chromosomal synteny is largely conserved between the domestic goat and the Alpine
141 ibex. The sheep genome experienced three chromosome fusions, combining in each two distinct
142 chromosomes found in cattle and goats. The Alpine ibex genome includes a fully assembled X
143 chromosome compared to the domestic goat (contigs Xa and Xb; Figure 1D). In addition, the
144 Alpine ibex chromosomes 12 and 17 also include sequences matching to unplaced contigs in
145 domestic goat. Overall, domestic goat and Alpine ibex genomes share 2.4 Gb of syntenic regions
146 (~90.2%). Structural variant analyses with SyRI confirmed major rearrangements on
147 chromosomes 7, 18, and 23 (Figure S3, S4A), mainly involving indels (38 Mb, 1.4% of the
148 genome) and inversions (49 Mb, 1.8%), with fewer duplications and translocations (Figure S4B,
149 C). Highly divergent regions (HDR) covered a total of 43 Mb (1.6%; Figure S4B, C). Nearly half
150 (45%) of all insertions, deletions, and HDRs between domestic goat and Alpine ibex overlapped
151 with gene bodies, and the proportion was even higher for inversions (52%), duplications (54%),
152 and translocations (78%).
153
154 **MHC conservation and high heterozygosity in recent hybrids**
155 High levels of polymorphisms at the major histocompatibility complex (MHC) are typically favored
156 within species, and the locus is prone to duplications [33]. Consequently, assembling the MHC
157 regions can be challenging [see *e.g.,* challenges for human MHC assembly [34–36]]. Here, we report
158 a contiguous assembly of the Alpine ibex MHC. High-level synteny is well-conserved among
159 cattle, sheep, domestic goat, and Alpine ibex in both their main MHC regions (class I and II, Figure
160 2A). We detected a ~110 kb translocation between Alpine ibex and the domestic goat near the
161 MHC class II cluster that does not affect MHC coding sequences (Figure 2B). An additional
162 translocation of ~1.92 Mb occurred further away from the MHC class II on chromosome 23 (Figure
163 2B). In contrast to the human MHC, both the domestic goat and Alpine ibex MHC class II clusters
164 encode DR- and DQ-genes but no DP-genes [37,38]. The DR- and DQ-genes are generally
165 exceptionally polymorphic. Over 100 DRB1 alleles for sheep and dozens for goats have been
166 deposited in GenBank, reflecting their crucial role in immune diversity. Alpine ibex populations
167 have previously been shown to carry introgressed alleles from the domestic goat at the DRB gene
168 [30] and other immune-related genes [29].
169

170 To assess whether hybridization load may pose a risk to Alpine ibex, we generated whole-genome
171 sequencing data of eight recent hybrid individuals sampled from two hybrid swarms in Northern
172 Italy [24] and four domestic goats representing the local breed, and collected in the same areas as
173 the hybrids (Table S7). We analysed the new sequencing data together with publicly available
174 data representing 29 non-hybrid Alpine ibex as well as 18 domestic goats. Average individual
175 heterozygosity was consistently low in non-hybrid Alpine ibex individuals compared to hybrids
176 and domestic goats, including particularly striking contrasts in the DR- and DQ-gene regions
177 (25.40-25.75 Mb; Figure 2C). Diversity at the MHC is associated with disease resistance in a
178 number of species [39–45], including Alpine ibex [46]. The high diversity at the MHC in hybrids may
179 therefore confer a selective advantage under disease pressure. Hence, hybridization and MHC
180 introgression may be favored in the wild, posing a risk of disrupting local adaptation in wild Alpine
181 ibex.

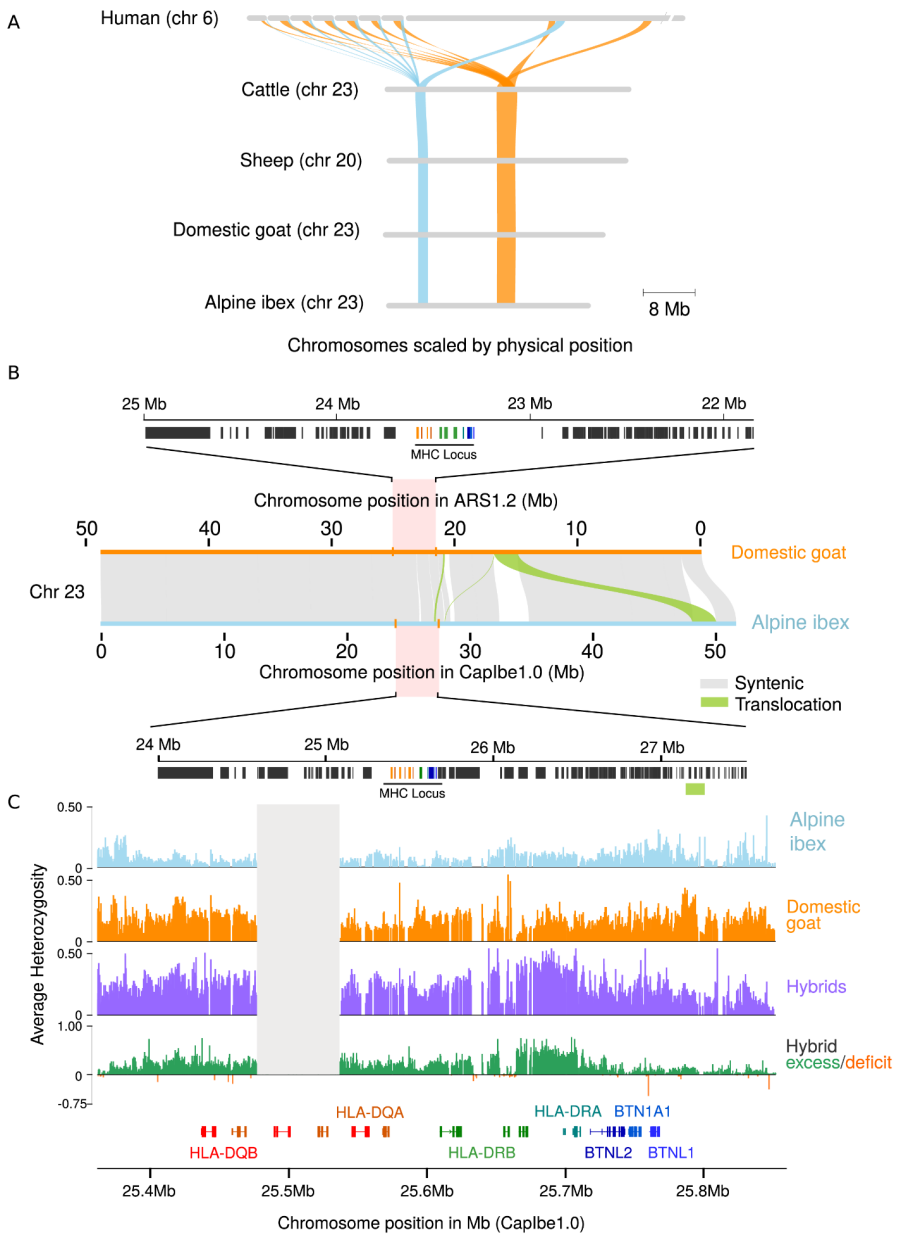

Figure 2: Conserved synteny at the MHC and high diversity in hybrids. A) Close-up of the synteny analysis, conducted with GENESPACE, highlighting two MHC regions within Alpine ibex chromosome 23 (light blue and orange, respectively).

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.

C) Average individual heterozygosity is shown as observed heterozygosity in Alpine ibex (blue), domestic goat (orange), and hybrids (purple). Heterozygosity excess and deficit in hybrids are shown with green (positive values) and orange (negative values), respectively. The grey box highlights a region corresponding to 25.47-25.55 Mb (encoding *DQA* and *DQB*), which showed a

192 local lack of coverage in domestic goats while being evenly covered in Alpine ibex (except for
193 Alpine ibex individuals with previously reported historic introgression in the MHC region, see
194 Figures S5-11 and Grossen et al. [30]).
195
196 **Emerging recombinant haplotypes in an Alpine ibex hybrid swarm**
197 The recent observation of hybrid swarms in Northern Italy renewed concerns about Alpine ibex
198 conservation efforts [24]. The high collinearity of Alpine ibex and domestic goat genomes should
199 facilitate recombination between haplotypes. At the phenotypic level, recent hybrids carry well-
200 recognizable domestic goat features such as horn shape and fur coloration [24] (Figure 3A, B). We
201 analyzed the whole-genome sequencing data to recapitulate introgression patterns in two hybrid
202 swarms. We furthermore sequenced four domestic goat individuals from the same two regions
203 (Figure 3, Table S7). Non-hybrid individuals of each species showed clear separation based on
204 a principal component analysis (PCA, n = 29 Alpine ibex, n = 22 domestic goats; Figure 3C)
205 performed on 86,650 biallelic SNPs. All eight hybrid swarm members showed clear evidence for
206 recent admixture (Figure 3C). Next, we performed chromosome painting using genome-wide
207 polymorphism of the hybrids to identify parental contributions. Individual 12 matched expectations
208 for a F1 hybrid with uninterrupted goat and Alpine ibex haplotypes over all chromosomes (Figure
209 3D). Hybrids that underwent backcrosses revealed evidence for recombination events as
210 interrupted haplotypes (Figure 3D). Genome-wide ancestry matched well with recombination
211 patterns and parental haplotype contributions. Individual 5 showed both the highest degree of
212 ancestry and the largest haplotypes of domestic goat. In contrast, individual 8 carried the longest
213 Alpine ibex haplotype blocks consistent with its strong Alpine ibex ancestry (Figure 3D). The
214 highly diverse set of recombinants in the Alpine ibex swarm raises significant concerns for further
215 introgression in the two respective regions and other populations with a high incidence of contacts
216 between domestic goat and Alpine ibex.
217

Figure 3

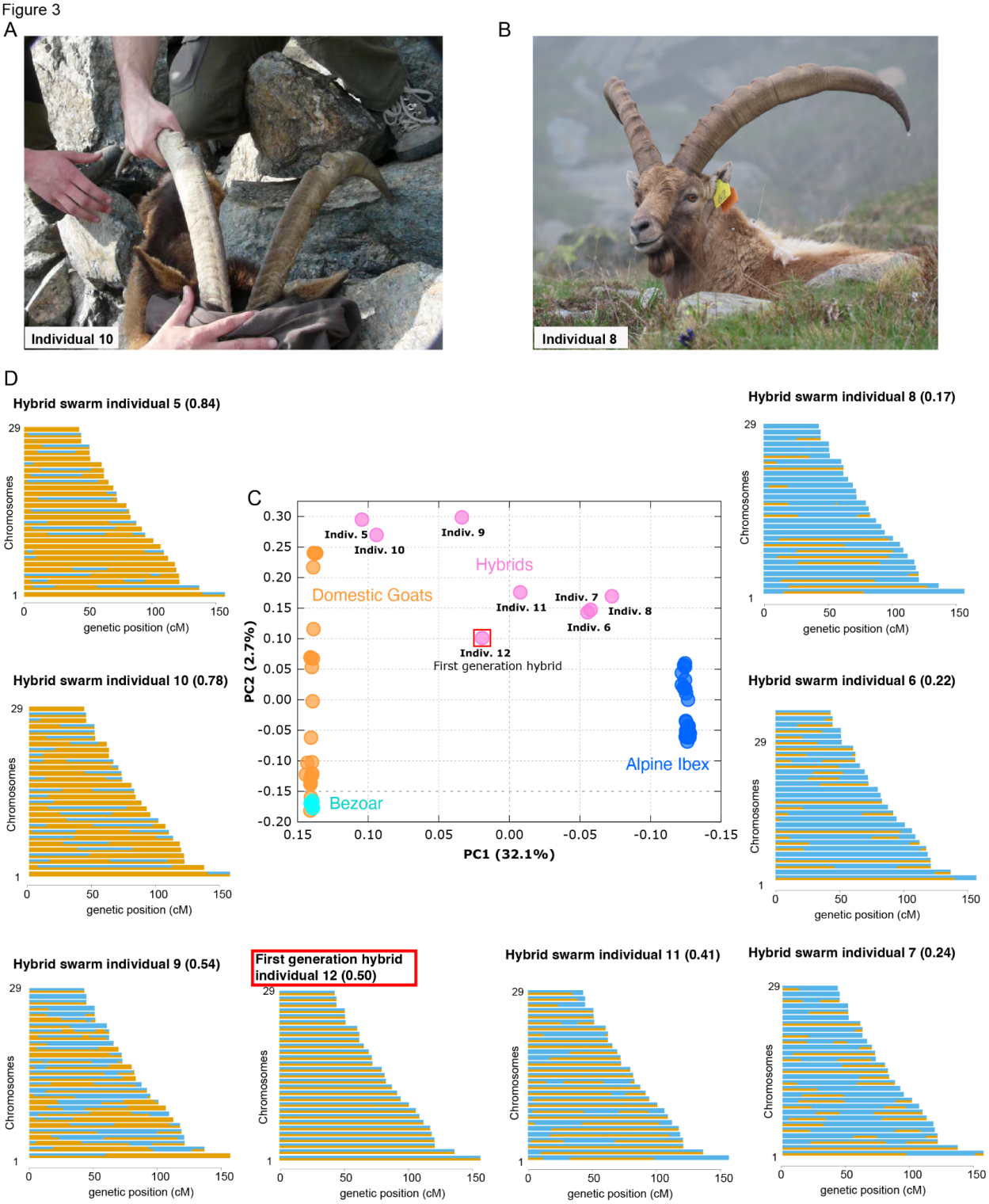

Figure 3: Genome-wide patterns of hybridization

A) Picture of hybrid individual 10. B) Picture of hybrid individual 8. C) Principal Component Analysis of biallelic autosomal SNPs showing the clustering of species and the distribution of hybrids along the conjunction between Alpine ibex and domestic goat/bezoar. D) Chromosome

223 paintings drawn with Mosaic, ordered according to the proportion of domestic goat ancestry
224 (decreasing proportion of goat genome shown in orange from left to right).
225
226 **Hybridization load in two Alpine ibex hybrid swarms**
227 We assessed the extent of hybridization load, which could be introduced into Alpine ibex through
228 a hybrid swarm. The cross-species polymorphism included 59,544,737 SNPs of which 0.009%,
229 0.311%, and 0.499% were classified as having putatively high, moderate, and low impacts on
230 encoded proteins, respectively (see also Tables S8 and S9). High-impact variants, such as those
231 causing premature stop codons or frameshifts, are predicted to severely disrupt protein function
232 (*e.g.*, loss-of-function, LOF variants). Moderate and low impact variants cover, respectively,
233 missense mutations that may alter protein structure without severe impairment and synonymous
234 changes with minimal expected effects on protein function. To evaluate the robustness of
235 predicted high-impact mutations, we tested whether these mutations were disproportionately
236 located in potential pseudogenes. However, we found that the presence of deleterious mutations
237 cannot be explained by pseudogene-like regions alone (Table S10).
238
239 We investigated deleterious mutations derived from either domestic goat or Alpine ibex as a proxy
240 for genetic load (Figure 4A). First, we analyzed candidate mutations underpinning heterosis in
241 hybrids. We defined these mutations as fixed derived LOF variants in Alpine ibex (fixed for the
242 ancestral allele in all other species). Seven LOF variants were identified as candidates for
243 heterosis (Table S11). Severely disruptive variants such as LOF are generally recessive; hence,
244 these should be masked in the heterozygous state, conferring increased fitness (heterosis) in
245 hybrids compared to non-hybrid Alpine ibex. We found that the analyzed hybrids each carried
246 between zero and all seven candidates for heterosis in the heterozygous state (in a backcrossed
247 individual into Alpine ibex and a F1 hybrid, respectively, Figure 4B). This suggests that the positive
248 effects of masking LOF variants in hybrids may quickly get lost through backcrossing. The
249 observed signs of hybrid vigour in Alpine ibex-domestic goat hybrids [24] may reflect a combination
250 of such masking effects alongside other genetic (*e.g.*, overdominance) or epigenetic factors [47].
251
252

Figure 4

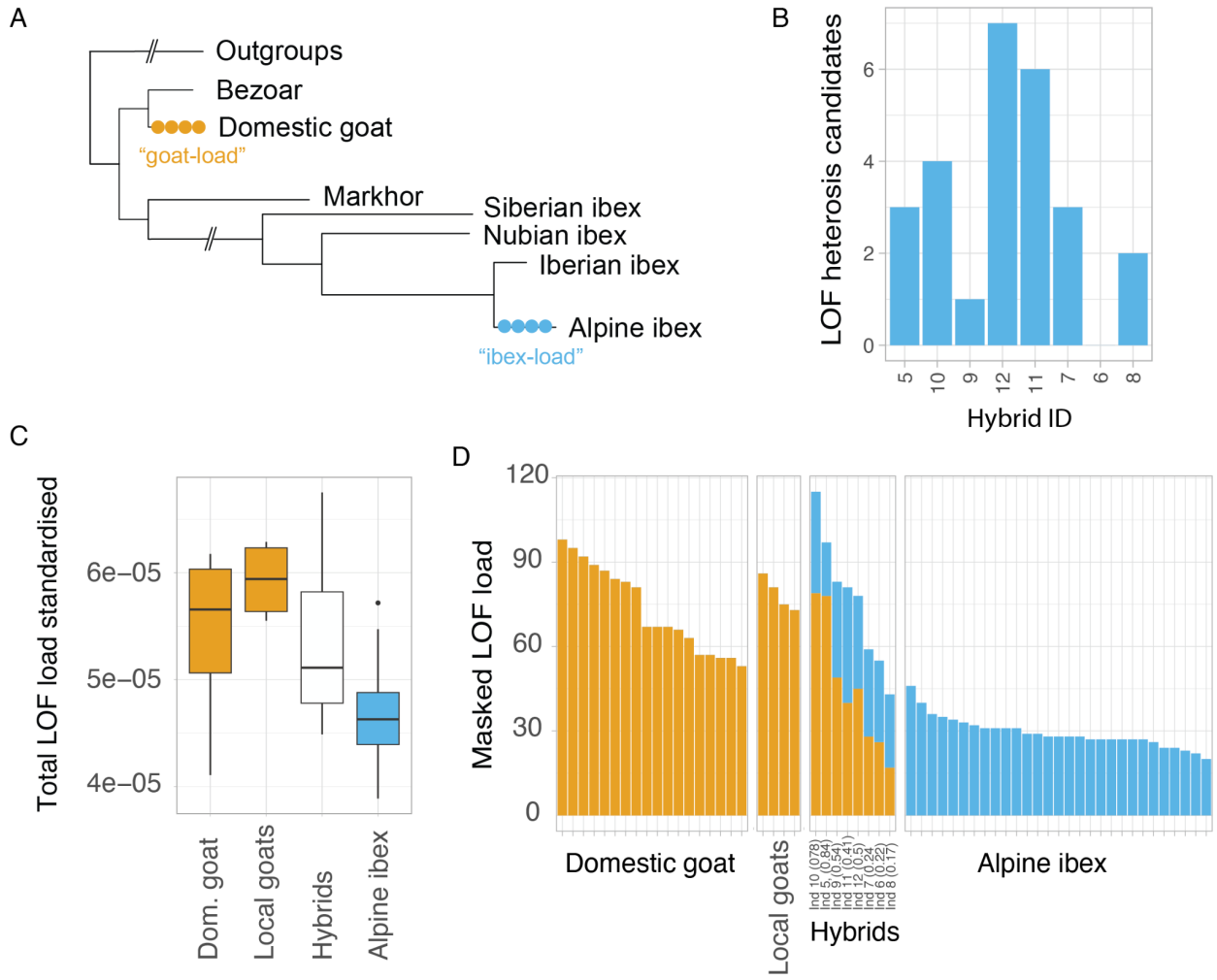

Figure 4: Deleterious mutation load in hybrids

A) Phylogenetic tree showing species relationships and the definition of "goat-load" and "ibex-load" defined as newly derived mutations private to the respective species. B) Counts of loss-of-function alleles found in the heterozygous state for each sequenced hybrid individual (ordered according to proportion of goat). C) Total counts of loss-of-function alleles standardized by the total counts of modifier alleles. Box plots per species. Local goats refer to domestic goat breeds kept in the region where hybrid swarms were detected. See also Figure S12. D) Masked loss-of-function load (heterozygote counts) per individual. Local goats, as in C.

We also investigated the extent to which hybridization load could affect Alpine ibex. We found that the mutation load for LOF variants (standardized by neutral variation) was higher in domestic goat than in Alpine ibex (Figure 4C, $p < 10^{-9}$). As a consequence, hybrids carried a total of 215 newly acquired LOF variants (Table S11), more than doubling the total number of LOF variants

in Alpine ibex (*n* = 177 LOF variants detected among 29 non-hybrid Alpine ibex). With an average of 76 masked LOF variants per hybrid individual compared to 29 in non-hybrid Alpine ibex, hybrids carried on average a 2.5 times higher severe masked load (Figure 4D). Hybrids backcrossed with Alpine ibex saw their mutation load decreased with each generation as expected (Figure 4D). Yet standardized LOF load did not decrease over the first few generations of back-crossing into Alpine ibex (Figure S12). Hence, there is no indication that selection acted to reduce mutation load. Such counterselection is expected to face less interference after further back-crossings (and recombination). Since the hybrid swarm acquired hybridization load (*i.e.,* masked mutation load), our data highlights that Alpine ibex are at risk of increasing their mutation load through leakage from the hybrid swarm.

Concluding remarks

The detrimental effects of deleterious variants are well-recognized as a risk factor in endangered species [2,3]. While a large number of recent studies investigated the evolution of deleterious mutation load as a consequence of small population size[4–6,48–52], much less consideration was given to the source of deleterious mutation load. Here, we investigated the role of hybridization with domestic goat as a source of mutation load for Alpine ibex. The high heterozygosity in hybrids may confer early-generation hybrids an advantage over non-hybrid individuals, compounding the risk of further introgression and a potential loss of local adaptation. Here, we show that high rates of hybridization can introduce deleterious mutations from a closely related species without evident counterselection. Given the increased rate of human-induced hybridization [53], tracking the deleterious effects of hybridizations should be a crucial element of species conservation actions.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the lead contact, Christine Grossen (Christine.Grossen@wsl.ch).

Materials availability

This study did not generate novel biological materials.

Data and code availability

Raw sequencing experiment data were submitted to the SRA archive under the project identifier PRJNA1104824 (reviewer link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA1104824?reviewer=kij4fipl6g0eheslrgertb0ak9). The final *C. ibex* genome assembly and high-confidence coding gene annotations will be publicly available in the NCBI repository XXXXXX (currently undergoing NCBI processing). The *CapIbe1.0* genome fasta file, full transcriptome models, gene functional annotation, and GERP scores relative to the novel genome can be downloaded from Zenodo (https://doi.org/10.5281/zenodo.15632776; https://doi.org/10.5281/zenodo.15632854; https://doi.org/10.5281/zenodo.15357984).
This paper does not report original code. The lead contact can provide any additional information required to reanalyze the data reported in this paper upon request.

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 Illumina whole-genome sequencing. We thank Sergio Decherchi, Alessandro Parodi, and Mattia Pini for their support for high-performance computing and data storage, and Dario de Siena 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, ACa, MV and LP 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) and continued their activity with a fellowship from Programma Regionale Fondo Sociale Europeo Plus 2021/2027 ("European Social Fund", ESF CUPJ51B24000170002).
This work is part of the "Technologies for Sustainability" Flagship program of IIT (L.P.).

Author contributions

AB conceived the study and managed DNA sample collection. AB, AP, GC, DW, JHM, and BB performed and coordinated the sampling. FGÇ, FL, DC, LP, and CG designed experiments and analyses. EH, YCCDS, and DV generated sequencing data under the supervision of FGÇ, LP, MV, and ST. FGÇ, FL, FF, LP, and CG performed analyses with input from DC. FGÇ, FL, AB, DC, LP, and CG wrote the manuscript with contributions from all authors. JMC, ACa, SG, AB, and CG provided funding and resources for the study.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Document S1 contains Figures S1–S12.

Figure S1: HiC contact map showing the data used for genome scaffolding, related to Figure 1.

Figure S2: Scheme of the Alpine ibex mitochondrial genome, related to Figure 1.

Figure S3: Structural variant calling between domestic goat and Alpine ibex (based on SyRI) across all 29 autosomes, related to Figure 1.

Figure S4: 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), related to Figure 1.

Figure S5: Read depth profiles of already available whole genome sequencing samples of Alpine ibex (N=29), domestic goat (N=18), Iberian ibex (N=4), and 12 newly sequenced individual samples (domestic goat, N=4; hybrids, N=8). The profiles are displayed for the genomic interval 25.47-25.55 Mb on chromosome 23, which includes portions of *DQA* and *DQB* loci, related to Figure 2.

Figure S6: Figure S5 continued, related to Figure 2.

Figure S7: Figure S5 continued, related to Figure 2.

Figure S8: Figure S5 continued, related to Figure 2.

Figure S9: Figure S5 continued, related to Figure 2.

Figure S10: Figure S5 continued, related to Figure 2.

Figure S11: Figure S5 continued, related to Figure 2.

Figure S12: Total counts of loss-of-function alleles standardized by the total counts of modifier alleles for each hybrid individual ordered by proportion of domestic goat (indicated in parentheses), related to Figure 4.

Document S2 contains Tables S1-S13.

Table S1: Genome contiguity statistics of the draft and scaffolded Alpine ibex assemblies, related to Figure 1.

Table S2: Summary of repeat annotations, related to Figure 1.

Table S3: BUSCO statistics for the scaffolded assembly and the protein-coding gene annotation of Alpine ibex, related to Figure 1.

Table S4: Genome annotation stats, related to Figure 1.

Table S5: Syntenic blocks identified in the GENESPACE analysis, related to Figure 1.

Table S6: Gene orthogroups identified in the GENESPACE analysis, related to Figure 1.

Table S7: Goat and goat/ibex hybrid individuals sequenced in the present study, related to Figures 2-4.

Table S8: Number of effects of genetic variants by type, related to Figure 4.

Table S9: Number of effects of genetic variants by region, related to Figure 4.

Table S10: Summary of deleterious mutations' PFAM annotations, related to Figure 4.

Table S11: Newly acquired loss-of-function (LOF) variants and candidates for heterosis, related to Figure 4.

Table S12: Summary of the ONT runs used for genome assembly, related to Figure 1.

Table S13: Species and IDs of the available datasets used in this study, related to Figures 3 and 4.

Document S3 Supplementary Text

Figure legends

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 the gene synteny analysis conducted on the cattle, sheep, domestic goat, and Alpine ibex genomes using GENESPACE. The human genome was used as an outgroup. See also Figures S1-S4.

399

400 **Figure 2: Conserved synteny at the MHC and high diversity in hybrids.** A) Close-up of the
401 synteny analysis, conducted with GENESPACE, highlighting two MHC regions within Alpine ibex
402 chromosome 23 (light blue and orange, respectively).
403 B) Structural variation along chromosome 23 based on SyRI, including annotations of domestic
404 goat (top) and Alpine ibex (bottom) in the range encompassing the major MHC II region.
405 C) Average individual heterozygosity is shown as observed heterozygosity in Alpine ibex (blue),
406 domestic goat (orange), and hybrids (purple). Heterozygosity excess and deficit in hybrids were
407 shown with green (positive values) and orange (negative values), respectively. The grey box
408 highlights a region corresponding to 25.47-25.55 Mb (encoding *DQA* and *DQB*), which showed a
409 local lack of coverage in domestic goats while being evenly covered in Alpine ibex (except for
410 Alpine ibex individuals with previously reported historic introgression in the MHC region, see
411 Figures S5-11 and Grossen et al. [30]).
412

413 **Figure 3: Genome-wide patterns of hybridization**
414 A) Picture of hybrid individual 10. B) Picture of hybrid individual 8. C) Principal Component
415 Analysis of biallelic autosomal SNPs showing the clustering of species and the distribution of
416 hybrids along the conjunction between Alpine ibex and domestic goat/bezoar. D) Chromosome
417 paintings drawn with Mosaic, ordered according to the proportion of domestic goat ancestry
418 (decreasing proportion of goat genome shown in orange from left to right).
419

420 **Figure 4: Deleterious mutation load in hybrids**
421 A) Phylogenetic tree showing species relationships and the definition of "goat-load" and "ibex-
422 load" defined as newly derived mutations private to the respective species. B) Counts of loss-of-
423 function alleles found in the heterozygous state for each sequenced hybrid individual (ordered
424 according to proportion of goat). C) Total counts of loss-of-function alleles standardized by the
425 total counts of modifier alleles. Box plots per species. Local domestic goat refers to breeds kept
426 in the region where hybrid swarms were detected. See also Figure S12. D) Masked loss-of-
427 function load (heterozygote counts) per individual. New goats, as in C.
428
429

430 # Methods

431 We collected fresh samples from thirteen individuals: one female Alpine ibex (Odissea), eight
432 hybrids, and four domestic goats. For the female Alpine ibex, we extracted multiple organs and
433 performed high molecular weight DNA extractions to enable Oxford Nanopore Technologies
434 (ONT; Oxford, UK) long-read (including ultra-long) and Illumina NovaSeq 6000 short-read
435 sequencing (Table S12). These data were used to produce a draft genome assembly, which was
436 then integrated with Hi-C sequencing data from the same individual's liver. The samples from the
437 other individuals were prepared for whole-genome paired-end sequencing on an Illumina
438 NovaSeq 6000. Detailed protocols are provided in the Supplementary Text.
439
440 **Genome assembly and evaluation**
441 We assembled the ONT data into a draft genome using Flye v2.9 [54], after which short-read data
442 were aligned and used for polishing with Pilon v1.23 [55]. We scanned the assembly for laboratory
443 or environmental contaminants using the NCBI Foreign Contamination Screen tool [56]. Following
444 quality filtering, we mapped Hi-C reads following the Arima Genomics pipeline
445 (https://github.com/ArimaGenomics/mapping_pipeline) and performed scaffolding with YaHS
446 v1.1 [57]. Finally, we evaluated the final assembly for contiguity, k-mer completeness, reference-
447 free QV, and expected gene content (further methodological details are in the Supplementary
448 Text). Genome assembly statistics were visualized with a snail plot in BlobToolKit v4.3.2 [58] (Figure
449 1C).
450
451 **Gene model prediction, curation, and functional annotation**
452 We softmasked the Alpine ibex genome by merging repeat coordinates identified via
453 RepeatModeler v2.0.3 [59] and TRF [60]. Then we incorporated RNA-seq and homology-based
454 evidence into the braker2 v2.1.6 pipeline [61–69] using pre-trained parameter sets for human (*Homo*
455 *sapiens)*, which is the evolutionarily closest taxon for Alpine ibex available within the software.
456 Subsequently, an annotation lift-over from the sheep reference genome [ARS-UI_Ramb_v2.0,
457 NCBI Genome GCF_016772045.1 [70,71]] was conducted with Liftoff v1.6.3 [72]. In order to reduce
458 the redundancy of the transcript set obtained, we employed an isoform pruning pipeline as
459 previously described [73]. Finally, gene models were annotated using eggNOG mapper v2.1.11 [74]
460 in DIAMOND-search mode. The details of used external evidence for gene prediction and the
461 detailed methodology can be found in the Supplementary Text. Functional annotation of the gene
462 models showed that 54,205 out of 61,420 (88.2%) transcripts had matches in the eggNOG protein
463 database (corresponding to 81.1% of genes).

464 The mitochondrial genome encodes 13 proteins, including seven complex I proteins (ND1-3,
465 ND4L, ND4-6), four complex IV proteins (COXI, COXII, COXIII, and CytB), and 2 ATP synthase
466 subunits (ATPase 6 and 8; Figure S2).
467
468 **Synteny analysis and structural rearrangement identification**
469 We ran GENESPACE v1.2.3 [75] with default parameters to examine synteny among the Alpine
470 ibex genome and the human (GRCh38.p14), cattle (ARS-UCD1.3), sheep (ARS-UI_Ramb_v2.0),
471 and domestic goat (ARS1.2, GCF_001704415.2) references, focusing on primary chromosome
472 scaffolds. We further explored syntenic regions and structural rearrangements between domestic
473 goat and Alpine ibex reference genomes with SyRI v1.6.3 [76] (see Supplementary Text). Results
474 were visualized using plotsr v1.1.1 [77].
475
476 **SNP genotyping and hybridization analyses**
477 Whole-genome sequencing data from eight newly sequenced hybrids and goats were combined
478 with available datasets for Alpine ibex (N=29), domestic goat (N=18), bezoar (N=4), Iberian ibex
479 (N=4), Nubian ibex (N=2), Siberian ibex (N=2), and markhor (N=1), plus four outgroup samples
480 (chamois, cattle, sheep, bighorn sheep; see Table S13) After alignment to the new Alpine ibex
481 genome following the *germline* pipeline of the NVIDIA Clara Parabricks v3.8 [78], variants were
482 called, and hard filtering was performed with GATK v4.1.0 [79]. Eventually, we retained SNPs
483 polymorphic among Capra spp. (see Supplementary Text for selection criteria), which were
484 autosomal, bi-allelic, with a minimal genotype quality of 20 and a minimal genotyping rate of 80%
485 using bcftools v1.2 [80]. We estimated average per-site heterozygosity with VCFtools [81] and used
486 PLINK v2.0 [82] for principal component analysis. After data phasing using beagle v22Jul22.46e [83],
487 the Mosaic R-toolbox [84] was employed to examine Alpine ibex and goat haplotypes in hybrids;
488 chromosome painting was visualized via the karyogram function. Finally, putative functional
489 changes were annotated with SnpEff v4.3t [85]. Individual (derived) genetic load was computed as
490 masked LOF load (number of heterozygotes) or total load (number of heterozygotes + 2* number
491 of homozygotes) and standardized by the respective numbers of MODIFIER genotypes. These
492 standardized estimates of LOF load in domestic goat vs. Alpine ibex were then compared using
493 a Welch two-sample t-test (see also Supplementary Text).
494

495 # References

496 1. 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. Preprint, https://doi.org/10.1111/padr.12283 https://doi.org/10.1111/padr.12283.

2. Dussex, N., Morales, H.E., Grossen, C., Dalén, L., and van Oosterhout, C. (2023). Purging and accumulation of genetic load in conservation. Trends Ecol. Evol. *38*, 961–969.

3. Robinson, J., Kyriazis, C.C., Yuan, S.C., and Lohmueller, K.E. (2023). Deleterious variation in natural populations and implications for conservation genetics. Annu. Rev. Anim. Biosci. *11*, 93–114.

4. Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., et al. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science *348*, 242–245.

5. Grossen, C., Guillaume, F., Keller, L.F., and Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat. Commun. *11*, 1001.

6. Dussex, N., van der Valk, T., Morales, H.E., Wheat, C.W., Díez-Del-Molino, D., von Seth, J., Foster, Y., Kutschera, V.E., Guschanski, K., Rhie, A., et al. (2021). Population genomics of the critically endangered kākāpō. Cell Genom. *1*, 100002.

7. Whitlock, M.C., and Davis, B. (2011). Genetic Load. eLS. https://doi.org/10.1002/9780470015902.a0001787.pub2.

8. Kardos, M., Zhang, Y., Parsons, K.M., A, Y., Kang, H., Xu, X., Liu, X., Matkin, C.O., Zhang, P., Ward, E.J., et al. (2023). Inbreeding depression explains killer whale population dynamics. Nat. Ecol. Evol. *7*, 675–686.

9. Stoffel, M.A., Johnston, S.E., Pilkington, J.G., and Pemberton, J.M. (2024). Purifying and balancing selection on embryonic semi-lethal haplotypes in a wild mammal. Evol. Lett. *8*, 222–230.

10. Genovart, M. (2009). Natural hybridization and conservation. Biodivers. Conserv. *18*, 1435–1439.

11. Mallet, J. (2005). Hybridization as an invasion of the genome. Trends Ecol. Evol. *20*, 229–237.

12. Moran, B.M., Payne, C., Langdon, Q., Powell, D.L., Brandvain, Y., and Schumer, M. (2021). The genomic consequences of hybridization. Elife *10*. https://doi.org/10.7554/eLife.69016.

13. Barton, N.H. (2001). The role of hybridization in evolution. Mol. Ecol. *10*, 551–568.

14. Burke, J.M., and Arnold, M.L. (2001). Genetics and the fitness of hybrids. Annu. Rev. Genet. *35*, 31–52.

15. Allendorf, F.W., Leary, R.F., Spruell, P., and Wenburg, J.K. (2001). The problems with hybrids: setting conservation guidelines. Trends Ecol. Evol. *16*, 613–622.

16. Leitwein, M., Cayuela, H., Ferchaud, A.-L., Normandeau, É., Gagnaire, P.-A., and

Bernatchez, L. (2019). The role of recombination on genome-wide patterns of local ancestry exemplified by supplemented brook charr populations. Mol. Ecol. *28*, 4755–4769.

17. Schumer, M., Xu, C., Powell, D.L., Durvasula, A., Skov, L., Holland, C., Blazier, J.C., Sankararaman, S., Andolfatto, P., Rosenthal, G.G., et al. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science *360*, 656–660.

18. Simon, M.F., Mendoza Flores, J.M., Liu, H.-L., Martins, M.L.L., Drovetski, S.V., Przelomska, N.A.S., Loiselle, H., Cavalcanti, T.B., Inglis, P.W., Mueller, N.G., et al. (2022). Phylogenomic analysis points to a South American origin of Manihot and illuminates the primary gene pool of cassava. New Phytol. *233*, 534–545.

19. Wang, L., Beissinger, T.M., Lorant, A., Ross-Ibarra, C., Ross-Ibarra, J., and Hufford, M.B. (2017). The interplay of demography and selection during maize domestication and expansion. Genome Biol. *18*, 215.

20. Janzen, G.M., Wang, L., and Hufford, M.B. (2019). The extent of adaptive wild introgression in crops. New Phytol. *221*, 1279–1288.

21. Liu, S., Zhang, L., Sang, Y., Lai, Q., Zhang, X., Jia, C., Long, Z., Wu, J., Ma, T., Mao, K., et al. (2022). Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across Populus genome. Mol. Biol. Evol. *39*. https://doi.org/10.1093/molbev/msac008.

22. Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P.R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G.R., Godoy, J.A., et al. (2023). How genomics can help biodiversity conservation. Trends Genet. *39*, 545–559.

23. Stüwe, M., and Nievergelt, B. (1991). Recovery of alpine ibex from near extinction: the result of effective protection, captive breeding, and reintroductions. Appl. Anim. Behav. Sci. *29*, 379–387.

24. Brambilla, A., Zehnder, N., Bassano, B., Rossi, L., and Grossen, C. (2024). Genetic evidence of a hybrid swarm between Alpine ibex (Capra ibex) and domestic goat (C. hircus). Evol. Appl. *17*, e13761.

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

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

27. Brambilla, A., Biebach, I., Bassano, B., Bogliani, G., and von Hardenberg, A. (2015). Direct and indirect causal effects of heterozygosity on fitness-related traits in Alpine ibex. Proc. Biol. Sci. *282*, 20141873.

28. Bozzuto, C., Biebach, I., Muff, S., Ives, A.R., and Keller, L.F. (2019). Inbreeding reduces long-term growth of Alpine ibex populations. Nat Ecol Evol *3*, 1359–1364.

29. Münger, X., Robin, M., Dalén, L., and Grossen, C. (2024). Facilitated introgression from domestic goat into Alpine ibex at immune loci. Mol. Ecol., e17429.

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

31. Moroni, B., Brambilla, A., Rossi, L., Meneguz, P.G., Bassano, B., and Tizzani, P. (2022). Hybridization between Alpine Ibex and Domestic Goat in the Alps: A Sporadic and Localized Phenomenon? Animals (Basel) *12*. https://doi.org/10.3390/ani12060751.

32. Bickhart, D.M., Rosen, B.D., Koren, S., Sayre, B.L., Hastie, A.R., Chan, S., Lee, J., Lam, E.T., Liachko, I., Sullivan, S.T., et al. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat. Genet. *49*, 643–650.

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

34. 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., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. *36*, 338–345.

35. 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., and Phillippy, A.M. (2018). De novo assembly of haplotype-resolved genomes with trio binning. Nat. Biotechnol. https://doi.org/10.1038/nbt.4277.

36. Chin, C.-S., Wagner, J., Zeng, Q., Garrison, E., Garg, S., Fungtammasan, A., Rautiainen, M., Aganezov, S., Kirsche, M., Zarate, S., et al. (2020). A diploid assembly-based benchmark for variants in the major histocompatibility complex. Nat. Commun. *11*, 4794.

37. Beck, S., and Trowsdale, J. (1999). Sequence organisation of the class II region of the human MHC. Immunol. Rev. *167*, 201–210.

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

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

40. Westerdahl, H. (2007). Passerine MHC: genetic variation and disease resistance in the wild. J. Ornithol. *148*, 469–477.

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

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

43. Paterson, S., Wilson, K., and 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*, 3714–3719.

44. Langefors, A., Lohm, J., Grahn, M., Andersen, O., and von Schantz, T. (2001). Association between major histocompatibility complex class IIB alleles and resistance to Aeromonas salmonicida in Atlantic salmon. Proc. Biol. Sci. *268*, 479–485.

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

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

47. Chen, Z.J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. Nat. Rev. Genet. *14*, 471–482.

48. Garcia-Erill, G., Liu, S., Le, M.D., Hurley, M.M., Nguyen, H.D., Nguyen, D.Q., Nguyen, D.H., Santander, C.G., Sánchez Barreiro, F., Gomes Martins, N.F., et al. (2025). Genomes of critically endangered saola are shaped by population structure and purging. Cell. https://doi.org/10.1016/j.cell.2025.03.040.

49. Henn, B.M., Botigué, L.R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B.K., Martin, A.R., Musharoff, S., Cann, H., Snyder, M.P., et al. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. U. S. A. *113*, E440–E449.

50. Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., Zachariah, A., Jhala, Y., Kardos, M., et al. (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. bioRxiv. https://doi.org/10.1101/2021.05.18.444660.

51. Ochoa, A., and Gibbs, H.L. (2021). Genomic signatures of inbreeding and mutation load in a threatened rattlesnake. Mol. Ecol. *30*, 5454–5469.

52. Kleinman-Ruiz, D., Lucena-Perez, M., Villanueva, B., Fernández, J., Saveljev, A.P., Ratkiewicz, M., Schmidt, K., Galtier, N., García-Dorado, A., and Godoy, J.A. (2022). Purging of deleterious burden in the endangered Iberian lynx. Proc. Natl. Acad. Sci. U. S. A. *119*, e2110614119.

53. Ottenburghs, J. (2021). The genic view of hybridization in the Anthropocene. Evol. Appl. *14*, 2342–2360.

54. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546.

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

56. Astashyn, A., Tvedte, E.S., Sweeney, D., Sapojnikov, V., Bouk, N., Joukov, V., Mozes, E., Strope, P.K., Sylla, P.M., Wagner, L., et al. (2024). Rapid and sensitive detection of genome contamination at scale with FCS-GX. Genome Biol. *25*, 60.

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

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

59. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. *117*, 9451–9457.

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

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

62. Hoff, K.J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-genome annotation with BRAKER. Methods Mol. Biol. *1962*, 65–95.

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

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

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

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

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

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

69. Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. *42*, e119.

70. 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., et al. (2010). The sheep genome reference sequence: a work in progress. Anim. Genet. *41*, 449–453.

691 71. Li, R., Yang, P., Li, M., Fang, W., Yue, X., Nanaei, H.A., Gan, S., Du, D., Cai, Y., Dai, X., et
692 al. (2021). A Hu sheep genome with the first ovine Y chromosome reveal introgression
693 history after sheep domestication. Sci. China Life Sci. *64*, 1116–1130.

694 72. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations.
695 Bioinformatics *37*, 1639–1643.

696 73. Ivankovic, M., Brand, J.N., Pandolfini, L., Brown, T., Pippel, M., Rozanski, A., Schubert, T.,
697 Grohme, M.A., Winkler, S., Robledillo, L., et al. (2023). A comparative analysis of planarian
698 genomes reveals regulatory conservation in the face of rapid structural divergence. bioRxiv.
699 https://doi.org/10.1101/2023.12.22.572568.

700 74. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J.
701 (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
702 Prediction at the Metagenomic Scale. Mol. Biol. Evol. *38*, 5825–5829.

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

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

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

712 78. Franke, K.R., and Crowgey, E.L. (2020). Accelerating next generation sequencing data
713 analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms.
714 Genomics Inform. *18*, e10.

715 79. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A.,
716 Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis
717 Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.
718 Genome Res. *20*, 1297–1303.

719 80. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A.,
720 Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and
721 BCFtools. Gigascience *10*, giab008.

722 81. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker,
723 R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and
724 VCFtools. Bioinformatics *27*, 2156–2158.

725 82. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015).
726 Second-generation PLINK: rising to the challenge of larger and richer datasets.
727 Gigascience *4*, 7.

728 83. Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of
729 large-scale sequence data. Am. J. Hum. Genet. *108*, 1880–1890.

730 84. Salter-Townshend, M., and Myers, S. (2019). Fine-Scale Inference of Ancestry Segments
731 Without Prior Knowledge of Admixing Groups. Genetics *212*, 869–889.

732 85. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.,
733 and Ruden, D.M. (2012). A program for annotating and predicting the effects of single
734 nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain
735 w1118; iso-2; iso-3. Fly *6*, 80–92.