Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/21/eaaz5216/DC1

Supplementary Materials for

The origin of domestication genes in goats

Zhuqing Zheng, Xihong Wang, Ming Li, Yunjia Li, Zhirui Yang, Xiaolong Wang, Xiangyu Pan, Mian Gong, Yu Zhang, Yingwei Guo, Yu Wang, Jing Liu, Yudong Cai, Qiuming Chen, Moses Okpeku, Licia Colli, Dawei Cai, Kun Wang, Shisheng Huang, Tad S. Sonstegard, Ali Esmailizadeh, Wenguang Zhang, Tingting Zhang, Yangbin Xu, Naiyi Xu, Yi Yang, Jianlin Han, Lei Chen, Joséphine Lesur, Kevin G. Daly, Daniel G. Bradley, Rasmus Heller, Guojie Zhang, Wen Wang, Yulin Chen*, Yu Jiang*

*Corresponding author. Email: yu.jiang@nwafu.edu.cn (Y.J.); chenyulin@nwafu.edu.cn (Y.C.)

Published 20 May 2020, *Sci. Adv.* **6**, eaaz5216 (2020)
DOI: 10.1126/sciadv.aaz5216

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/21/eaaz5216/DC1)

Supplementary Materials and Methods

Section 1. Read alignment

1.1 Modern genome sequence data

To obtain high-quality reads and minimize false genotyping due to low-quality reads, we implemented the following quality control procedures to filter the reads using Trimmomatic v0.36 (*60*). First, leading or trailing stretches of Ns and bases with quality below 3 were trimmed. Second, the reads were scanned with a 4-base-wide sliding window, cutting when the average quality per base dropped below 15. Finally, only reads with 40 nucleotides or longer were kept.

High-quality paired reads were aligned to the latest goat reference genome (GCF_001704415.1) (*61*) using BWA-MEM v0.7.15 (*62*) with default parameters except that "-M" was enabled. The alignment BAM files were then processed to sort reads, merge read groups belonging to the same sample, and mark duplicates using Picard v2.1 (https://broadinstitute.github.io/picard/). We then estimated the coverage distribution at each called site for each sample only using reads with a mapping quality above 20 using QualiMap v2.2 (*63*).

1.2 Ancient genome sequence data

We collected a total of five ancient goat samples (table S3). Sample SMG07 (a mandibula) and SMG11 (a humerus) (fig. S3A) were excavated from the Shimao site, an important Neolithic walled settlement in Shenmu County (Shaanxi province, China) in the northern part of the Loess Plateau (*14*). These two remains have been dated as early as 3,975-3,835 cal BP based on radiocarbon dating of cattle bones which were excavated jointly with these samples (*64*). Sample WDH06S (a tooth, dating to approximately 2,500 cal BP) (fig. S3A) was obtained from the Wangdahu site located in Pengyang county (Ningxia province, China). Sample KA01G from Northern Caucasus was dated to 1,296-1,270 cal BP. Sample YJL02G (a humerus) (fig. S3A), excavated from Yanjialiang site (Inner Mongolia Autonomous Region, China) was dated to 700-600 cal BP (Yuan Dynasty) with 95% confidence interval using radiocarbon dating conducted by Beta Analytic Radiocarbon Dating Laboratory (Miami, FL, USA) (*65*).

The dust and clay on the outer surface of teeth or bones were cleaned with a fur brush. Subsequently, the cleaned samples were cut into small pieces and soaked in 10% bleach for 20 min, rinsed with ethanol and distilled water, and then subjected to UV-irradiation for 30 min on each side. Finally, the samples were powdered under liquid nitrogen using a 6850 Freezer Mill (SPEX CertiPrep, Methucen, NJ, USA).

Ancient DNA was extracted from the sample powder by using a modified silica-based spin column method (*66*) in a dedicated ancient DNA laboratory at Jilin University. Briefly, 200 mg of the powder was incubated overnight with 3 ml of lysis buffer (0.5 M EDTA pH 8.0 and 0.5 mg/ml Proteinase K) in a rotating hybridization oven at 50℃(220 rpm/min). After centrifugation, the supernatant was transferred into an Amicon® Ultra-4 centrifugal filter device (Merck Millipore Ltd, 10000 Nominal Molecular Weight Limit), reduced to less than 100 ul, and purified with QIAquick® PCR Purification Kit (QIAGEN).

Genomic DNA libraries for the Illumina platform were prepared from 55.5 μl of ancient DNA using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs

Inc.) following the manual, with some minor modifications as described below. The adaptor ligated DNA fragments without size selection were cleaned with the MinElute® PCR Purification Kit (QIAGEN) following the manual. PCR amplification of the adaptor ligated DNA fragments was then purified with 1.8× AMPure XP Beads (Beckman Coulter). After DNA library preparation, the genomic DNA libraries were quantified with Qubit® dsDNA HS Assay Kits in Qubit® 2.0 Fluorometer (Life Technologies), and sent to Novogene for Paired-End sequencing (2×150 bp) on a HiSeq X Ten Platform.

Adapter sequences and low-quality bases were removed from the reads using AdapterRemoval v2.2 (*67*), followed by read mapping using BWA v0.7.15 (*68*) with the seed option ("-l 1024") disabled. Alignments showing mapping qualities lower than 20 were discarded. PCR duplicates were then removed on 5' read coordinates for single-end sequencing reads and both start and end for collapsed paired-end data using SAMtools v1.3 (*69*). Finally, reads were realigned around indels using the IndelRealigner procedure from GATK v3.7.0 (*46*). The presence of nucleotide misincorporation profiles typical of ancient DNA data was verified using mapDamage2 (*70*) (figs. S2B to S2F). Genome coverage was calculated via mosdepth (*71*). Sex determination was performed by comparing coverage of the X chromosome versus coverage of autosomal chromosomes.

1.3 Historical genome sequence data

Genomic sequence reads from *C. caucasica* were first trimmed for adapter sequences and low-quality bases with AdapterRemoval v2.2 (*67*). To account for the evolutionary divergence between the reference genome and the historical samples, we used relaxed alignment settings (-l 1024 -n 0.01 -o 2) with BWA v0.7.15 (*68*). We excluded reads with a MapQuality score below 20 and removed duplicate reads with DeDup v0.12.3 (*72*). Damage patterns assessment using mapDamage2 (*70*), showed no signatures of increased damage in the nucleotide positions at the reads terminals (fig. S2A). The mtDNA haplotype of the historical sample was confirmed by examining the complete mtDNA consensus sequence generated on the Illumina platform (fig. S1D).

We reanalyzed the ancient goats with >0.01× mean coverage from ref. (*6*). After removing the adapter sequences using cutadapt v1.16 (*73*) (cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -O 1 -m 30), the short read alignment was performed using BWA v0.7.15 (*68*) with the seed option ("-l 1024") disabled. Alignments showing mapping qualities lower than 20 were discarded. PCR duplicates were then removed on 5' read coordinates for single-end sequencing reads and both start and end for collapsed paired-end data using SAMtools v1.3 (*69*). Finally, reads were realigned around indels using the IndelRealigner procedure from GATK v3.7.0 (*46*).

Section 2. Population structure and phylogenetic analysis

The following nuclear genome analyses were performed using all modern samples, including those of both high and low genomic coverage.

2.1 Phylogenetic tree

For phylogenetic reconstruction, genetic distances were calculated between all individuals

using plink v1.9 (*74*). The distance matrix was subsequently used to construct a neighbor-joining (NJ) tree as implemented in MEGA v6.0 (*75*). The final tree topology was visualized using iTOL (*76*), and the tree was rooted at the branch of sibling *Capra* species. We also used all of the 5,043,096 fourfold degenerate (4d) sites to construct a maximum likelihood (ML) phylogenetic tree (fig. S4). Sites containing heterozygous SNPs were represented using the standard International Union of Pure and Applied Chemistry chemical nomenclature (IUPAC) codes (*77*). The concatenated sequences were used to build a maximum likelihood (ML) tree using RAxML v8.2.9 (*78*) with the following parameters: -f a -x 123 -p 23 -# 100 -k -m GTRGAMMA.

2.2 Principal component analysis (PCA)

We used linkage disequilibrium (LD)-pruned (plink: --indep-pairwise 50 10 0.2) unphased data to perform PCA using the smartpca program in the package of EIGENSOFT v6.1 (*79*) with default parameters and the settings numoutlieriter = 0 and numchrom = 29. The significance level of the eigenvectors was determined by a Tracy-Widom test.

2.3 ADMIXTURE software clustering

Population structure analysis and individual clustering were carried out using ADMIXTURE v1.3 (*80*) for *k* values from 2 to 7, using a 5-fold cross-validation procedure to test the fit. We ran ADMIXTURE 20 times per *k* and calculated the mean cross-validation error for each *k* across runs. The results were plotted using R (*81*). However, ADMIXTURE accuracy can be affected by sample size (*82*). For example, the genetic difference between bezoar and domestic goat populations here is greater than that estimated between EUR and EAS by pairwise genome-wide fixation index (F_{ST}) (table S6), phylogenetic analysis and PCA (Fig. 1). However, at $k = 2$, ADMIXTURE analysis suggested the differentiation between western (Europe and Africa) and eastern (South Asia and East Asia) populations (fig. S5). This inconsistency may result from an inappropriate sampling scheme due to the presence of only a small number of bezoars. To minimize the effect of sample size variation, we randomly reduced the sample size in the different domestic populations (table S1). Following this strategy, the analysis revealed a clear structure between bezoar and domestic goat populations at $k = 2$, which was different from that detected in the full data-set (fig. S6). This discordance provides some hints on how sample size can substantially affect clustering and ancestral population inference.

To assess the effect of sample size described above, we also adopted two different sampling strategies based on simulations using fastsimcoal2 (*83*). We simulated SNP data with a model including three populations (POP1, POP2, and POP3) derived from the PCA and phylogenetic analyses. In this model, we assumed that POP2 and POP3 diverged from POP1 15,000 generations ago and that POP2 and POP3 diverged from each other 5,000 generations ago. The first strategy included an even sampling from all three populations and found that ADMIXTURE was indeed able to recover the correct population structure when sampling was even. The second strategy also involved three populations but in this case sampling across populations was uneven. The total number of samples used for these two sampling strategies was kept constant (n = 210). The result of this analysis showed that, when samples were unevenly drawn from the three populations, at $k = 2$, the results of ADMIXTURE matched to an even lesser extent with the known three-population structure (fig. S7).

2.4 TreeMix analysis

To confirm the population structure constructed by phylogenetic tree, PCA and admixture, we constructed a ML tree using TreeMix v1.12 (*84*) accounting for LD by grouping sites in blocks of 500 SNPs (-k 500). To allow for the geographic structure of the sampled bezoars, we divided bezoars into three regional populations (Azerbaijan, Alborz, and Zagros) based on initial PCA and ADMIXTURE analysis. The confidence of the inferred tree topology was evaluated through 100 replicates. The inferred trees and corresponding residuals were visualized with the in-built R script plotting functions of TreeMix software (fig. S8).

2.5 ChromoPainter and fineSTRUCTURE analysis

In addition to the approaches described above, we also investigated the population structure and relationships between different populations using ChromoPainter/fineSTRUCTURE v2.1.3 (*85*), which can explicitly model the correlation between nearby SNPs and use extended multi-marker haplotypes. All haplotypes of the 164 modern goats and 24 modern bezoars extracted from the BEAGLE-phased data were analyzed using the ChromoPainter linked model. We performed expectation maximization (EM) inference using 50 EM steps to estimate the effective population size from our data and then used this estimated parameter in ChromoPainter. To perform Markov Chain Monte Carlo (MCMC) analysis, we used 1,000,000 burn-in iterations and 1,000,000 sample iterations with a thinning interval of 10,000. Visualization of the posterior distribution of clusters was then performed using the tree-building algorithm of fineSTRUCTURE (fig. S9). Since the results may be affected significantly by sample size of each population, the exact populations and breeds sampled etc., they should be treated as an approximate guide to genetic similarity, rather than as a full population history.

2.6 LD analysis

The LD coefficient (r^2) was calculated pairwise between high-quality SNPs with minor allele frequencies greater than 0.05 using Haploview (*86*). The parameters were set at "-minMAF 0.05 -hwcutoff 0.001". To minimize bias due to sample size, we randomly reduced the size of EUR, AFR, SWA, and EAS to 16 (fig. S10A).

Section 3. Demographic reconstruction

3.1 Estimating mutation rate and generation time

The mutation rate (u) for goats was estimated using the homologous DNA sequences from goats and sheep (*Ovis aries*) (*52*). The sequence divergence (D) between two species was estimated to be 0.022848. The divergence time (T) was estimated to be approximately 5.29 million years (*52*), and the mean generation time (g) for goats was set to 2 years (*51*). Therefore, the *u* was 4.32×10^{-9} per generation per site for goats, which was estimated by the formula $u = (D \times g)/(2 \times T)$. Our estimation is similar to that obtained (a rate of 2.23×10^{-9} per year per site for *Capra ibex*) based on phylogenetic comparisons of diverse ruminant taxa (*52*).

Note that the use of phylogenetic comparisons can generally avoid the influence of generation time and mutation rate on the population divergence time. The program we used

(see below) all output the scaled times which are given in units of the per-generation mutation rate. This means that in order to convert scaled times to generation, divide them by the mutation rate (u, site/gen). To convert generations into years, multiply by the generation time (g, years). Thus, population divergence time is affected by the ratio between g and u. According to the above formula, we can find that g/u is only related to D and T. If we use a generation time of four years, the u will also be doubled; however, g/u will remain constant.

3.2 MSMC analysis

A Multiple Sequential Coalescent Markovian model (MSMC2, an updated version of MSMC) (*87*) was used to reconstruct the effective population size and split history of bezoar and domestic populations over time. To ensure that heterozygous loci were called, we only used high-coverage data (ranging from 11× to 43×) from each population. We also applied the genome mask as recommended in the documentation of the software. The alignability of each base in the reference genome was evaluated using ComputeGenomeMask, which is a part of Genome STRiP2.0 (*88*), and a mask file containing the sites with depth between half and twice of mean depth generated with custom scripts from msmc-tools (https://github.com/stschiff/msmc-tools). Two individuals (4 phased haplotypes) from each population were used to infer the effective population size. For relative cross-coalescence rate (RCCR) inference, we used one pair of individuals from each population. To take into account the uncertainty about the parameters of mutation rate and generation time, we used the time points with a RCCR of 0.25 and 0.75 to provide a time range when the population split might have occurred.

3.3 SMC++ analysis

Because the accuracy of the MSMC method is sensitive to haplotype phasing quality, we also estimated population size histories and split time using SMC++ (*89*), which does not rely on haplotypic phase information. However, when making inferences about times of divergence, the SMC++ assumes a 'clean split' model, in which no gene flow occurs after the populations split (*89*). We used 11 genomes (table S1) from each group of EUR, AFR, SWA, and EAS for this inference and excluded the low-alignability regions identified by Genome STRiP2.0. Due to the limited sample size and sequencing depth for three substructured bezoar populations, we used 7 genomes for Azerbaijan, 6 genomes for Alborz and 3 genomes for Zagros bezoar. We measured the variance of the estimated results by a bootstrapping strategy, which was performed by breaking up the genome into 5-Mb segments and then randomly sampling with replacement.

3.4 ∂a∂i analysis

To further derive a more detailed demographic model, we analyzed the joint allele frequency spectra with diffusion approximation for demographic inference (∂a∂i) (*90*). Based on the population structure and the model-based assignment of the individual genomes, we constructed four domestic groups: EUR, AFR, SWA, and EAS. To minimize the effect of selection sweeps, we only considered genomic segments located at least 10,000 bp away from any coding locus. We also excluded the low-alignability regions identified by Genome STRiP2.0 and uncalled regions obtained from GATK. Finally, contiguous genomic regions of

at least 1 kb spanning a total of 454,292,184 bp were used for the demographic analysis. We estimated two-dimensional site frequency spectra (SFS) using the doSaf function within ANGSD to estimate per-site allele frequencies combined with the realSFS (*91*) program to optimize the genome-wide SFS. To minimize potential biases introduced by determining the ancestral allelic states, we used the folded SFS. As suggested in ref. (*92*), we specified simple models first and gradually fitted the models with increasing complexity (fig. S13). The likelihood and Akaike information criteria were used to optimize the model selection, with the best model shown in fig. S14. We also performed nonparametric bootstrapping (100 replicates) to determine the confidence interval of each parameter.

Section 4. Uniparental markers

4.1 Mitochondrial DNA

To assemble complete mitochondrial genomes (mtDNA), we mapped the clean paired reads to the mitochondrial genome (GenBank: GU068049.1). Given that mitochondrial genomes are circular, we added 300 of the first base pairs to the end of the reference to assure equal coverage of the sequences across the mtDNA. For each sample, reads showing unique hits after removing duplicates were included for subsequent analysis. To obtain highly accurate haplotype information, four individuals (FRCH05, CNSCH09, FRCH06, and NLCH03) were filtered out due to low sequencing coverage (100×). The filtered sequences were then aligned to an unmodified reference using MIA (*93*) (https://github.com/mpieva/mapping-iterative-assembler; parameters: -H 1 -i -c). Additional whole mtDNA sequences corresponding to individuals of known haplogroup affiliation were retrieved from GenBank. ML phylogenetic tree was constructed from the alignment (*94*) of all of the filtered sequences (fig. S29A).

4.2 Y chromosome

In the absence of a reference sequence for the goat Y-chromosome, we used a read depth-based method implemented in CNVcaller (*95*) to identify the putative Y-linked scaffolds by comparing the average copy number (CP) between females ($CP < 0.1$, sample size = 112) and males ($0.25 < CP < 0.75$, sample size = 83). In this way, we identified 345 Y-chromosomal scaffolds summing up to 12,137,976 bp (Data file S7).

We first called the putative variant sites within the identified Y-chromosome scaffolds using GATK HaplotypeCaller (*46*). Female individuals were used as control. No more than three females should have high-quality mapped reads in the putative variation sites. To construct a preliminary consensus call set, the list of the putative sites was used as a "-sites" file input for ANGSD for all male individuals, including one *C. sibirica* (80 individuals in total). We then applied five filters to obtain confident Y-chromosome SNPs: (a) keep only biallelic SNPs; (b) no individual should have maximum-likelihood genotype state as heterozygous; (c) the filtered-read depth across all individuals should be between 300 and 700; (d) number of individuals with zero high-quality reads mapping to the site should be below three; and (e) *P* value $< 1 \times 10^{-6}$. Genotypes were then called using BEAGLE, which yielded a total of 61,934 SNPs. We then used this call set to construct the ML phylogenetic tree using RAxML (fig. S29B). We also used the 18,232 SNPs showing polymorphism in bezoars and domestic goats

to build a Minimum Spanning Network using pegas (*96*) (fig. S30A). Each individual was also assigned to the two Y-chromosome haplogroups (Y1 and Y2) that have been previously described (*97*).

For Y chromosome haplogroup calling in ancient samples, we filtered out reads with mapping quality below 25, bases with base quality below 20, and restricted the analysis to positions covered at least 2-fold. We used 6,916 Y-chromosome SNPs showing $F_{ST} = 1$ between haplogroups Y1 and Y2 to obtain haplogroup calls for each sample with ANGSD. The heatmap based on the haplogroups is presented in fig. S31. We found that early domestic goat haplogroups are highly structured, and this pattern has continued in the modern samples.

4.3 Estimating the divergence time of the paternal lineages

The Y-linked scaffolds were used to estimate the split time between the different paternal lineages, using argali (*Ovis ammon*) as an outgroup. Y consensus sequences were generated using ANGSD for the outgroup and 8 bezoar individuals representing four clades. Called positions were required to have a depth of coverage ≥ 2, and only bases with quality ≥ 20 were considered. The resulting FASTA file for every individual was then concatenated to a single sequence. Then, data from all individuals were combined to obtain a multiple alignment file. Additionally, the alignment file was manually inspected to remove sites containing ambiguous sites and gaps. This strict filtering yielded an alignment file 5,108,492 bp in size. We converted the FASTA alignment file to NEXUS format. The NEXUS formatted file was used to generate the BEAST XML input file for the BEAST v2.4.8 program (*98*). To estimate split time among paternal lineages, we used a strict clock model, a Yule tree model and a log normal prior of 5.29 Mya (mean = 1.665, SD = 0.08) representing goat-sheep divergence (*52*). We chose TVM as the best substitution model, gamma-corrected to account for site heterogeneity, as indicated by the Bayesian Information Criterion in jModelTest v2.1.10 (*99*). We set the number of generations to 100 million. The log output files were obtained by running BEAST software. Tracer v1.6 (https://github.com/beast-dev/tracer) was used to analyze the output file and estimate split time, and TreeAnnotator v2.4.8 was used to obtain the maximum credibility tree topology. Finally, we visualized the tree in FigTree v1.4.2 (fig. S30B).

Section 5. Gene flow analysis

5.1 f_3 statistic

The subsampled modern individuals (table S1) and ancient samples with average coverage >3× (table S4) were included in the variant call set. As *qp3pop* included in the ADMIXTOOLs package (*100*) requires genotype calls, we randomly sampled genotypes according to the posterior probabilities as described in ref. (*101*). Only positions sequenced at least once in each individual were considered, resulting in a total of 36,435,593 polymorphic sites. For analyses involving ancient samples, only the 8,604,288 transversion sites were considered to reduce biases introduced by post-mortem DNA damage. We ran f_3-statistics on all possible triplets of modern populations. The results showed gene flows from AFR into SWA population (table S9). While investigating the admixture history of *C. caucasica* in the form of (modern bezoar, *C. caucasica*; target), we discovered a clear signal of admixture involving *C. caucasica* and modern bezoar from Azerbaijan as source populations and Hovk1 (>47,000 BP Armenian

bezoar) as target (table S9).

5.2 *D*-statistic (ABBA/BABA test)

D-statistic was used to investigate population relatedness and test for gene flow between wild *Capra* species and domestic goat populations at the group level using doAbbababa2 in ANGSD (*102*). Bezoars and domestic goats were grouped as described in table S1. We calculated the *D*-statistics for the tree (((H1, H2) H3) argali), considering only the autosomal regions with a minimum sequencing base quality of 20 and a mapping quality of 25. H1 and H2 denoted two different domestic goat populations and H3 denoted *C. aegagrus* or other wild *Capra* species. To assess statistical significance, the *D*-statistics was represented as a Z score, by applying a jackknife procedure using a nonoverlapping 5 Mb sliding windows. An absolute value of Z score higher than 3 was considered to be significant. If H1 shared more alleles with H3 than H2 does, the *D*-statistic would be negative and vice versa.

5.3 Haplotype analysis of *MUC6* locus

We performed a haplotype analysis using the *MUC6* non-repeat region (29:46,258,000-46,268,000). The genotype likelihoods (GLs) of variant sites for all modern individuals were extracted via ANGSD (*45*), applying the same criteria described in methods (Read alignment and variant calling). Then, the GLs were converted into hard-called genotypes using BEAGLE (*48, 49*). This yielded a total of 304 polymorphic sites. We constructed a haplotype network including wild *Capra* species, modern bezoar, and domestic goats for the 10 kb *MUC6* non-repeat region using pegas (*96*). The observed haplotype structure clearly showed that the overwhelming majority of domestic haplotypes were more distant from their wild progenitor than from the *C. caucasica* haplotypes (Fig. 2F). We also estimated the time to the most common ancestor (TMRCA) based on the frequentist estimator (*103*): TMRCA = $d_{ij}/2ul$, where d_{ij} is the number of nucleotide differences between any two sequences (haplotypes) *i* and *j*, *u* is the mutation rate (2.16×10^{-9} per site per year), and *l* denotes the sequence length (10,000 bp). The TMRCA for the highly divergent haplotypes was estimated to be >1.5 million years (fig. S18A).

5.4 Simulations, selection on a *de novo* mutation and on standing variation

We used msms (*104*) to simulate SNP variants for a population of constant size with mutation, recombination and positive selection affecting a single site (command: -N 15000 -ms 328 10000 -s 86 -r Rec 10001 -Smu 0.0002592 -SAA Sel -SaA Sel/2 -SI 0.083333 1 Frq). In modern domestic goats, we found that 48 SNPs with derived allele frequency >= 0.95 and were highly differentiated with bezoars (had frequency 0 in bezoars) in the non-repeat region of *MUC6* (29:46,258,000-46,268,000). We performed simulations conditioning on 86 segregating sites (minor allele frequency >= 0.01 in domestic goats) in the 10 kb non-repeat region. Effective population size (*N*) for domestic goats was estimated to be ~15,000 (fig. S11). We assumed three different selection strengths for the homozygote ($2Ns$ = 200, 500, and 1,000, where *s* is the selection coefficient of the beneficial mutations. *s* for the heterozygote is half that for the homozygote). The recombination rate of the *MUC6* locus was $rho = 4Nr = 0$ (Fig. 3C), as estimated by FastEPRR (*105*). we also set two different recombination rates ($4Nr$ = 10 and

100, where *r* is the probability of cross-over per generation between the ends of the locus being simulated). We set the mutation rate to 4.32×10^{-9} per site per generation and generation time to 2 years, respectively (see section 3.1). For simulating selection, we assumed three varying initial frequencies (0, 0.01, and 0.1) for the beneficial mutation when selection started. A total of 27 conditions were simulated and each condition was simulated 10,000 times. For a rough comparison of the number of beneficial mutations in observed and simulated data, we counted the number of beneficial mutations which had a frequency >= 0.95. We find that the observed number of nearly fixed beneficial mutations is significantly higher than what is expected by simulations under any of models explored (fig. S18B).

Section 6. Selective sweep analysis

6.1 Genome-wide patterns of heterozygosity and neutrality tests

The nucleotide diversity (π), population genetic differentiation (F_{ST}), Tajima's D and Theta Watterson (θ_w) were calculated using a sliding window approach with windows of 50 kb and a step of 20 kb (*106*) (fig. S10B). To solve the problems of various sequencing depth and missing data, and to avoid ascertainment problems introduced by the SNP discovery process (i.e., incorporate uncertainty in the genotypes through direct analyses of the GL), an empirical Bayes approach was used to calculate site-specific posterior probabilities for the sample frequency spectrum using a ML estimation of the SFS, which records the proportions of sites at different allele frequencies as a prior. The SFS is typically computed for each population separately using program realSFS (*91*). Only the set of overlapping sites was considered robust for the analysis.

6.2 Screening for selective sweeps during domestication

To uncover genetic changes that may have been subject to selection during domestication, we combined all domestic goat populations into a single domestic gene pool, which can considerably reduce the confounding effects of population-specific genetic drift. We then searched for genomic regions with the highest differences in genetic diversity (π ln ratio bezoars/domestic goats) and exceptionally differentiated in allele frequency (F_{ST}) between modern bezoars and modern domestic goats. The π and F_{ST} were calculated using a 50 kb window with a 20 kb step across the autosomal chromosomes. The π log-ratio was calculated as $\ln(\pi_W)$-$\ln(\pi_D)$, where π_W and π_D are the nucleotide diversity values for modern bezoars and modern domestic goats, respectively. We also performed the cross-population extended haplotype homozygosity (XP-EHH) test for every SNP using the default settings of the selscan v1.1 (*107*). For the XP-EHH selection scan, our test statistic was the average normalized XP-EHH score in each 50 kb region. An XP-EHH score is directional: a positive score suggests that selection is likely to have happened in domestic goats, whereas a negative score suggests the same about bezoars. To filter for candidate windows, we defined a significance level of $P < 0.005$ (Z test, with $F_{ST} > 0.195$, π ln ratio > 0.395 and XP-EHH > 2.1) (fig. S19). To provide better insight into the selective sweep, we also performed selective sweep analysis between modern bezoars and the four domestic populations (EUR, AFR, SWA-SAS, and EAS) defined by population structure analysis, separately (Data file S2). Moreover, to further evaluate the hypothesis of a selective sweep, Tajima's D and the composite likelihood ratio test

(implemented in SweepFinder2) (*108*) were applied to domestic goats using the same sliding window approach.

6.3 Selection sweep analysis on chromosome X

We analyzed chromosome X separately due to its difference from autosomes under several aspects, including a reduction in effective population size and recombination rate. In fact, chromosome X is more sensitive to genetic drift because of its reduced effective population size and different mutation rate, which can affect its genetic diversity. We calculated the F_{ST} and θπ using a 50 kb window with a 20 kb step across chromosome X with VCFtools (*47*). We used the same method to define candidate domestication regions as described previously for the autosomes. We identified a total of 24 candidate selected regions on chromosome X (fig. S20 and table S11). One putative sweep region at NW_017189516.1:13,720,001-14,450,000 had the most extreme signal in terms of markedly higher F_{ST} (0.748) and ln (θπ ratio) (3.081). This region harbors the *AR* gene that encodes the androgen receptor, which plays a crucial role in a wide range of developmental and physiological responses (*109*).

Supplementary Text

Text S1. Historical and ancient genomes analyses

With the objective of clarifying the ancestry of the specimen, we first investigated the horn morphology. Based on the morphological analysis, the specimen was assigned to the general ibex morphotype which is shared by *C. ibex*, *C. nubiana*, *C. sibirica*, and *C. caucasica* (*7*). The horn of this specimen displays a subtriangular equilateral horn cross-section without a well-defined frontal surface and with less prominent transverse knobs (fig. S1A), suggesting that morphological similarity to *C. caucasica* (*110*).

We then assessed its ancestry by molecular evidence. Phylogenetic analysis of the whole-genome and Y-chromosome data revealed that the specimen unambiguously grouped together with the ibex-like species (figs. S1C and S1E). Based on whole-genome data, the specimen was placed very close to *C. ibex*, well outside of other examined ibex-like species (fig. S1C). Furthermore, analysis of 11 diagnostic SNP markers (*111*) in our sequenced specimen and *C. ibex* genomes indicated that the specimen is genetically distinct from *C. ibex* (fig. S1B). Finally, maximum likelihood analyses of the whole mitochondrial genome revealed that the specimen and the available data from *C. caucasica* (GenBank: JN632609.1) belong to one clade (100% bootstrap) (fig. S1D). To rule out the possibility of recent hybridization with domestic goat, the absolute divergence (Dxy) calculated in 1 Mb sliding windows between the specimen and *C. ibex*, and contrasted it with Dxy between the specimen and domestic goat. Over 99.7% of the windows show higher similarity with *C. ibex* than with domestic goat (fig. S1F), suggesting that recent hybridization between this specimen and domestic goat is highly unlikely. Moreover, a TreeMix analysis revealed that there was gene flow from this specimen to the ancestors of the goats and bezoars (confirming results from other analyses in the study), and suggested that the specimen is not the result of a recent hybrid origin (fig. S8). Together with the fact that the specimen was not captive-born, our results based on both molecular and morphology data confirm that the specimen belongs to *C. caucasica* and shows no signs of recent genetic admixture with other *Capra* species.

To investigate the temporal context of selection on specific genetic loci identified as differentiating modern domestic goats from their extant wild ancestors, we carried out direct shotgun sequencing with 0.04 to 13.44-fold coverage for five ancient domestic goat remains from North China and North Caucasus (fig. S3A and table S3), dating to between ~3,900 and ~600 years ago. For each ancient genome, the mitochondrial sequence was first assembled using MIA (*93*). After adding sequences representing the known goat mtDNA haplogroups, we then reconstructed the ML phylogenetic tree from the full multi-FASTA sequence alignment to assess the most likely phylogenetic placement of the five ancient samples. A total of three haplotypes were identified among the ancient specimens: three belonging to haplogroup A, one belonging to haplogroup B and one belonging to haplogroup C (fig. S3C).

The genetic affinities between the five ancient samples and the modern goats were assessed. First, a phylogenetic reconstruction based on pairwise genetic distances was carried out. We used ngsDist (*112*) to compute a pairwise genetic distance matrix, upon which a NJ tree was built with FastME (*113*). The tree was rooted at the branch of *C. sibirica* and *C. caucasica*. Clade support values were based on 100 bootstrap pseudo-replicates. Second, we also used the *D*-statistic (argali, ancient; H1, H2) (implemented in ANGSD, -doAbbababa 1) to evaluate the relationship without transitions. We only considered the positions with a minimum coverage of two-fold in the ancient sample. According to the NJ tree based on the pairwise genetic distances calculated with ngsDist, all four ancient goats from North China clustered tightly with present-day East Asian goats, while the ~1,300 BP sample (KAG01) from North Caucasus showed affinity to present-day European domestic goats (fig. S3B). Additional *D*-statistic also supported this pattern (fig. S3D).

Text S2. Demographic history

We inferred the population size of modern goats and bezoars using MSMC2 and SMC++. All domestic populations showed a remarkably similar history of a population decline from ~50,000-60,000 years ago until ~10,000-20,000 years ago, followed by population increase, suggesting that the ancestors of the sampled domestic goats originated from a limited population (figs. S11B and S11D). In the last 10,000 years, a somewhat different pattern emerged for the EAS population, reaching a nadir approximately 6,500 years ago, followed by population growth (fig. S11D). It is also notable that, starting from ~120,000 years ago, three distinct demographic histories were inferred for the bezoar populations native to different geographic areas (figs. S11A and S11C).

We then used MSMC2 and SMC++ to compute divergence times as a means to assess the time frame of the shared population history amongst domestic goats (fig. S12). We found that the time ranges estimated from MSMC2 and SMC++ are overlapped, however SMC++ provides a much narrower and more recent range (fig. S12B). This is partly due to the fact that SMC++ assumes a clean split model without subsequent gene flow (*89*). These analyses also suggested that a much older spit time between eastern (EAS and SWA) and western (AFR and EUR) populations than did within each of them (fig. S12B). In particular, the split time between EAS and EUR inferred from both MSMC2 and SMC++ preceded the domestication event dated on archaeological evidence (~11,000 years ago). We also tested whether the RCCR patterns observed amongst EUR, AFR, and EAS could be the result of admixture with the modern bezoar. To test this, we computed their RCCRs without “bezoar-like” segments in the

genome. In an attempt to map "bezoar-like" segments, we used refined IBD (*114*) to identify segments inherited from a common ancestor in goat-bezoar pairs (segments with a LOD score < 3 and a length shorter than 50 kb were excluded). This resulted in 137-Mb of bezoar-like segments in domestic goat genomes. We then masked these segments and re-estimated RCCR with MSMC2. The results show that the observed RCCR patterns cannot be attributed to gene flow from modern bezoar (fig. S12B).

We further derived a more detailed demographic model using ∂a∂i. The results showed that the initial divergence occurred 42,432 years ago between EUR and AFR (bootstrap 95% CI 30,087 to 54,929 years ago), 191,186 years ago between EUR and SWA (bootstrap 95% CI 158,396 to 242,618 years ago), and 159,086 years ago between AFR and SWA (bootstrap 95% CI 132,658 to 197,505 years ago), which are much earlier compared to the model-free (MSMC2 and SMC++) estimation. Given a complex picture of the mosaic origin of domestic goats, it is perhaps unsurprising that different methods demonstrate variability in their inferred population split times. However, our results from different methods consistently suggested an older divergence time between modern East Asian and European goat populations. In contrast, the initial divergence time between EAS and SWA was 10,374 years ago (bootstrap 95% CI 8,720 to 11,705 years ago), roughly coinciding with the estimations from MSMC2 and SMC++. We noted that the population pairs whose initial divergence was estimated to have occurred deeper in the past had a gradual decline in RCCR and a larger effective ancestral population size, perhaps reflecting complex ancestral structure with admixture. Additionally, we also observed that the ancestral effective population size was estimated to be large for these population pairs (table S7). A previous study on the evolutionary history of Tibetans found a similar discrepancy in estimated Han-Tibetan divergence time by means of MSMC and ∂a∂i, which was explained by population structure between the ancestral Han and Tibetan subpopulations but with high rates of gene flow (*115*). In addition, *D* statistics show that domestic goat populations have different levels of allele sharing with three regional samples of modern bezoar populations (fig. S15A). Thus, our demographic history estimates from whole-genome data suggest that differential admixture from multiple divergent populations of wild goats could have contributed to the modern goat populations.

Of the six mitochondrial haplogroups so far identified in modern domestic goats, haplogroup A is overwhelmingly predominant (figs. S29A and S29C). Interestingly, we also identified two well-defined Y-chromosome haplogroups, Y1 and Y2, which diverged ~297,500 years ago (95% HPD interval: 252,600-346,300 years ago), using a high-resolution Y-chromosome haplotype network based on 18,232 SNPs (fig. S30). A recent study suggests that the frequency of mtDNA haplogroup A has changed dramatically during the Neolithic Age (*6*). By contrast, the frequencies of the Y-chromosome haplogroups remained relatively stable (Fig. 5B). Altogether, the spatial and temporal uniparental data suggest that mainly females participated in the migrations.

Text S3. Comparison with previously identified candidate domestication loci

Several studies have attempted to identify regions that are strongly differentiated between domestic goats and bezoars, with the goal of identifying candidate targets of selection during domestication. Alberto et al. (*5*) found 44 candidate regions under selection. Out of these 44 regions, 27 regions show directional positive or stabilizing selection in domestic goats. Daly et

al. (*6*) identified 19 loci that underwent selective sweeps in either six eastern Neolithic genomes or four western Neolithic genomes by comparing each population to modern bezoar genomes. When compared to previous studies, we found 17 out of 27 loci from Alberto et al. and two out of 19 loci from Daly et al. passed our filtrations (Data file S2). Notably, the *KITLG* and *KIT* loci, which are the only two selected regions shared among Neolithic Balkan and Iranian populations, are also detected in our study. Regarding the reason that Daly et al. didn't detect the two highlighted positively selected loci (*STIM1-RRM1* locus and *MUC6* locus) in domestic goats identified in our study, the candidate adaptive variants of these two loci started with a very low frequency (0.07 and 0) during the Neolithic period, and increased gradually in frequency until ~6,500 years ago (Fig. 5A). This may explain why Daly et al. missed these two loci using samples only from the Neolithic Age.

Furthermore, our approach to identify candidate targets of positive selection during domestication differs from previous studies in several ways. First, we used the latest goat reference genome (ARS1) and annotation, which is one of the most contiguous assemblies among vertebrates. In contrast, Alberto et al. and Daly et al. used the previous goat assembly (CHIR1.0) (*116*). In CHIR1.0, the *MUC6* gene resided in a contig (NW_005101268.1) which was not assigned to any chromosome, thus could not be identified by sliding window across chromosomes. Second, we used a different method (intersection of F_{ST}, π ratio, and XP-EHH) to define selection sweeps compared to Alberto et al. (hapFLK and π ratio) and Daly et al. (F_{ST} and π ratio). Third, our study covers a more complete and global representation of samples, therefore, which enable us to detect common selection signals during domestication rather than a potential result of geographically restricted selection. As more modern worldwide domestic goats and ancient samples are sequenced, it is likely that these candidate sweeps will be refined and narrowed.

A

B

		ZP2 exon8–exon9	*ZP2* exon11–exon13						*ZP3* exon3–exon4			
		25:18876528	25:18874159	25:18874181	25:18874208	25:18874262	25:18874471	25:18874591	25:35365072	25:35365174	25:35365188	25:35365216
Slave Jordan et al. 2006	*C. caucasica*	2:A,1:T	1:C,1:C/T	1:A,1:A/G	1:A,1:A/C	1:C,1:C/T	1:G,1:A/G	1:C,1:C/T	2:C,1:A/C	2:A,1:A/G	3:T	2:C,1:C/T
	C. cylindricornis	2:A/T	2:C/T	2:A/G	2:A/C	2:C/T	2:A/G	2:C/T	2:A/C	2:A/G	1:T,1:G/T	2:C/T
	C. ibex	2:A	2:C	2:A	2:A	2:C	2:G	2:C	2:C	2:A	2:T	2:C
	C. pyrenaica	1:A	1:C	1:A	1:A	1:C	1:G	1:C	1:C	1:A	1:T	1:C
	C. nubiana	9:A	8:C	8:A	8:A	9:C	9:G	9:C	9:C	9:A	9:T	9:C
	C. sibirica	7:A	2:C,3:T,1:C/T	2:A,3:G,1:A/G	2:A,3:C,1:A/C	6:C	6:A	2:C,3:T,1:C/T	7:C	7:G	7:G	7:C
This Study	*C. caucasica*	1:A/T	1:C/T	1:G	1:C	1:C/T	1:A	1:C/T	1:A	1:A/G	1:G/T	1:C/T
	C. ibex	1:A	1:C	1:A	1:A	1:C	1:G	1:C	1:C	1:A	1:T	1:C
	C.sibirica	3:A	3:T	3:G	3:C	3:C	3:A	3:T	3:C	3:G	3:G	3:C
	C.falconeri	3:A	1:T,2:C/T	1:G,2:A/G	1:C,2:A/C	3:C	1:A,2:A/G	1:T,2:C/T	3:C	3:G	3:G	3:C

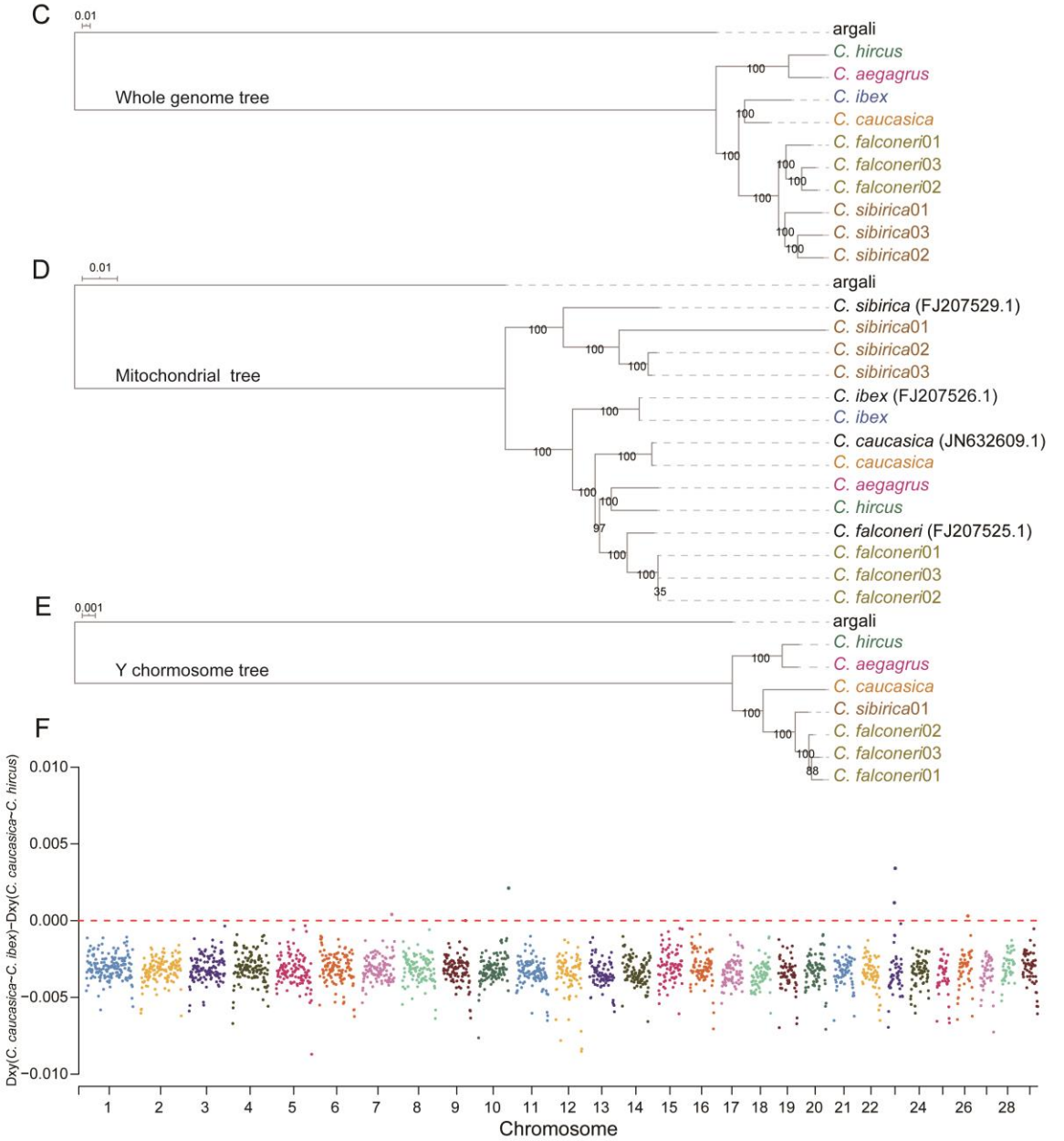


Fig. S1.
Morphological and genetic analysis of historical *Capra caucasica* sample. (**A**) *C. caucasica* horn and skull morphology. (1) View from the top. (2) View from the back side.

(3) View from the right. (4) View from the left. Photo Credit: Joséphine Lesur, Muséum National d'Histoire Naturelle. (**B**) Diagnostic SNPs used to verify species status of the specimen. 11 SNPs across three regions, including parts of the *zona pellucida 2* (*ZP2*) gene and the *zona pellucida 3* (*ZP3*) gene. The genotype is noted below each column, and variants from that base, as well as their frequency, are noted within the column. (**C**) Neighbor-joining tree (100 bootstrap replicates) based on pairwise genetic distance calculated with ~68.2 million autosomal SNPs. (**D**) Maximum likelihood phylogeny (100 bootstrap replicates) of the mitochondrial genome. Species with an accession number were obtained from GenBank. (**E**) Maximum likelihood phylogeny (100 bootstrap replicates) of the Y chromosome. (**F**) Sequence absolute divergence (Dxy) calculated between *C. caucasica* and *C. ibex*, contrasted with Dxy between *C. caucasica* and *C. hircus*. Negative values indicate greater similarity with *C. ibex* than with *C. hircus*.

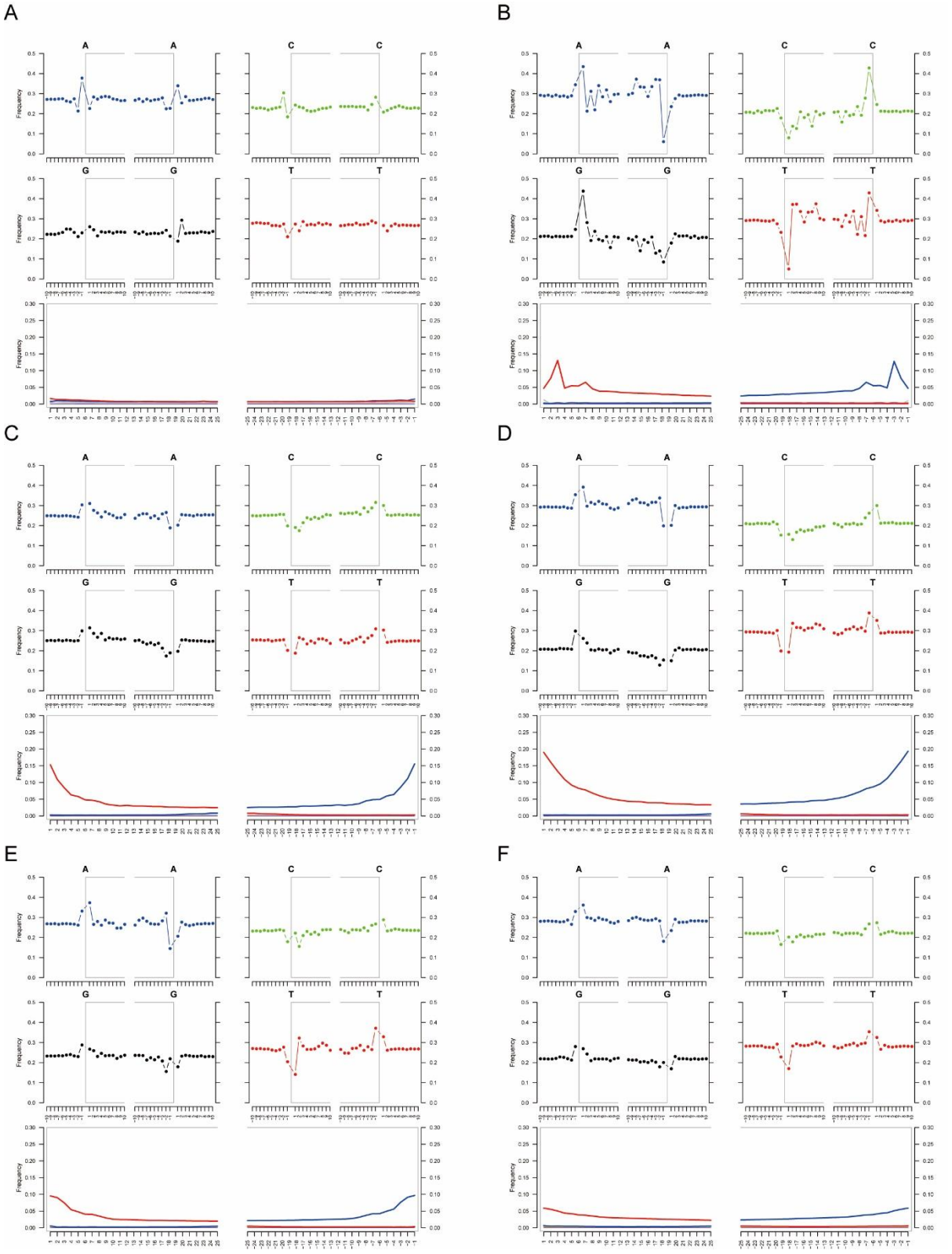


Fig. S2.
DNA fragmentation and nucleotide mis-incorporation profiles for the historical and ancient genomes sequenced in this study. (**A**) Tur1 (*Capra caucasica*). (**B**) KA01G. (**C**) SMG07. (**D**) SMG11. (**E**) WDH06S. (**F**) YJL02G.

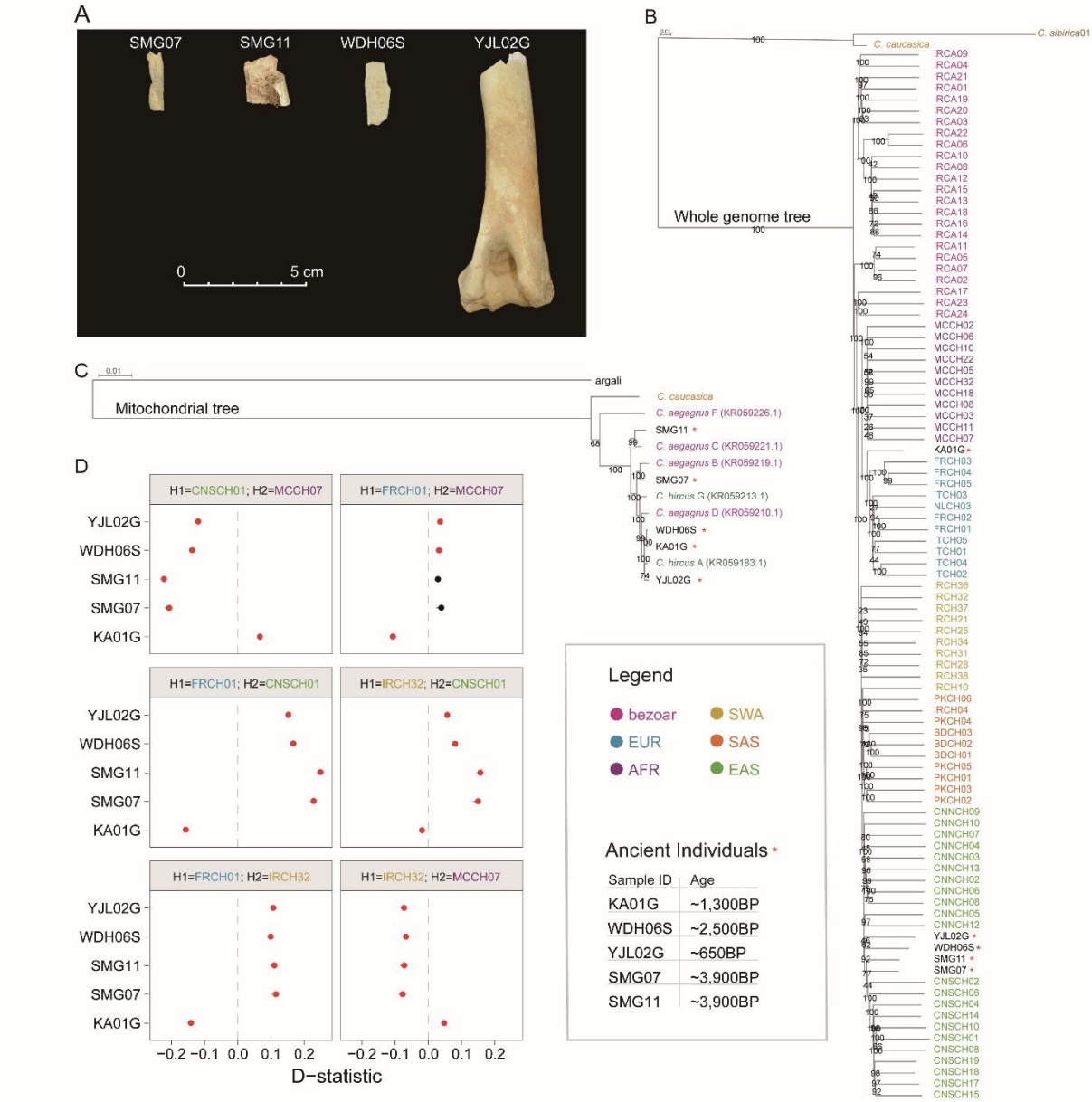


Fig. S3.

Genetic analysis and genealogy of ancient bone samples. (**A**) Bones of the four ancient goats from North China. Mandibula (SMG07), humeri (SMG11 and YJL02G), and tooth (WDH06S). Photo Credit: Dawei Cai, Jilin University. (**B**) Neighbor-joining tree (100 bootstrap replicates) based on pairwise genetic distance calculated with ~13.7 million autosomal transversion sites. Labels are color-coded according to their respective populations. The ancient samples are highlighted by red stars. (**C**) Maximum likelihood phylogeny (100 bootstrap replicates) of the ancient mitochondrial genomes. Samples with an accession number were obtained from GenBank. (**D**) *D*-statistics in the form of (argali, ancient; H1, H2). Analyses have been carried out disregarding transitions. The names of the ancient goats are reported on the Y-axis. Positive values support a close relationship between H2 and the ancient goat, while negative values support a close relationship between H1 and the ancient goat. Red dots depict significant tests, defined by $|Z\text{-scores}| \geq 3$.

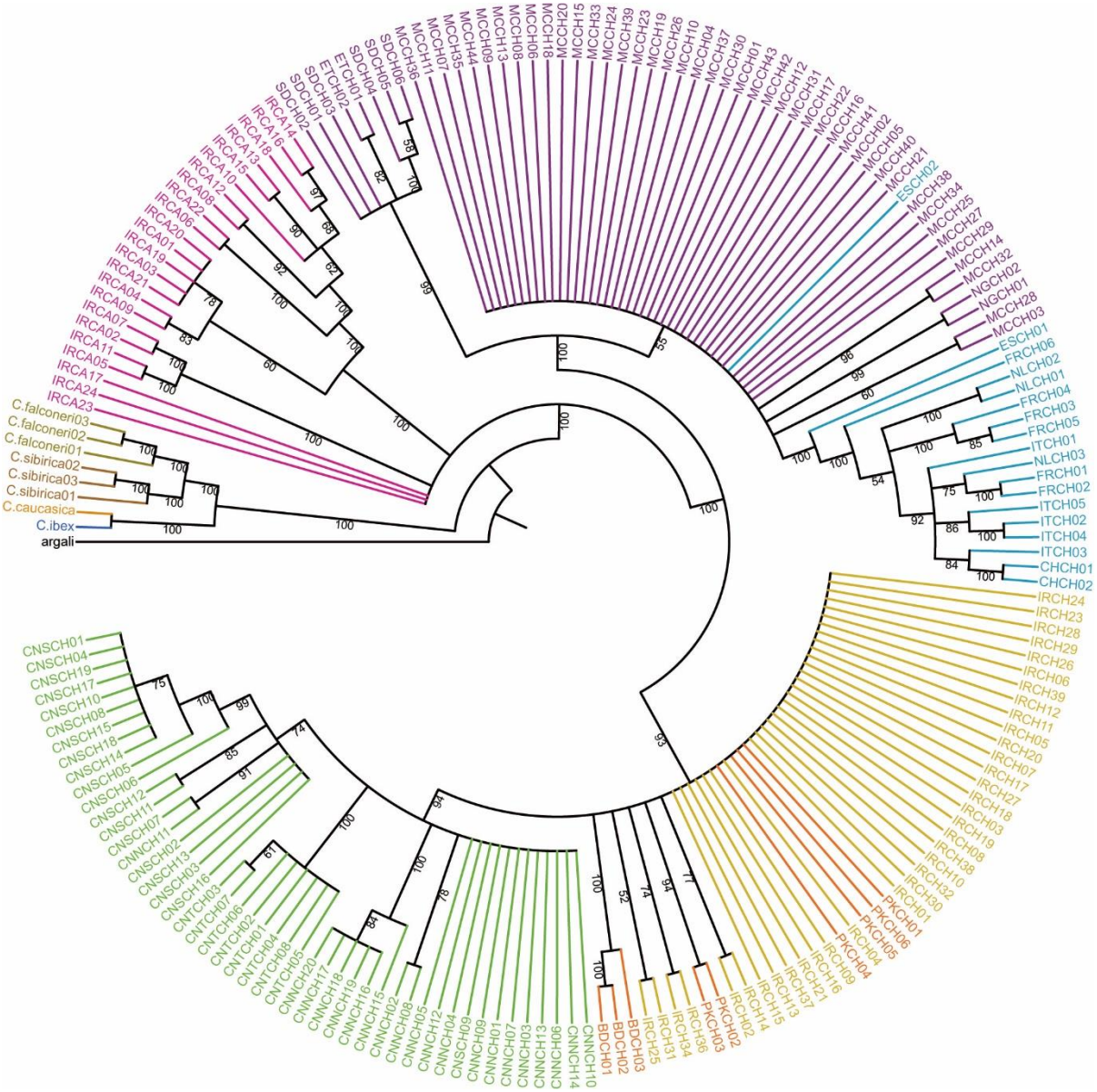


Fig. S4.
Phylogenetic tree obtained using a set of 5,043,096 fourfold degenerate sites using RAxML. The numbers at branches indicate bootstrap support (100 replicates). Branch color indicates membership in different geographical populations. This analysis also confirms the population structure observed in other analyses.

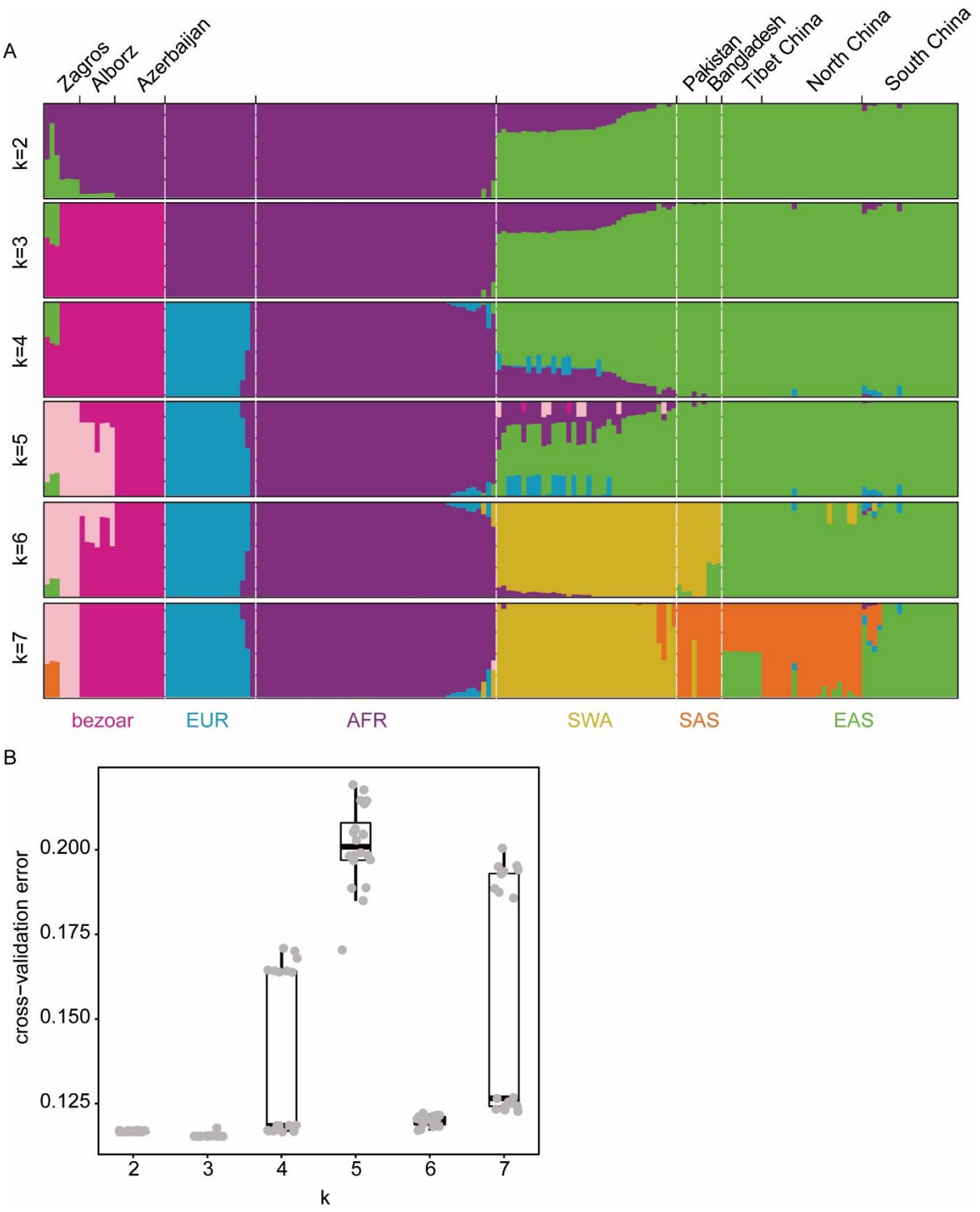


Fig. S5.
ADMIXTURE analysis for SNPs from modern bezoars and domestic goats after LD-pruning. (**A**) ADMIXTURE results for $k = 2$ to $k = 7$. The run with the lowest cross-validation (CV) error (out of 20 replicates) is plotted. The population names are at the bottom, and the black lines at the top of the figure denote the geographic locations of bezoar, SAS, and EAS. (**B**) CV error for varying k in the ADMIXTURE analysis.

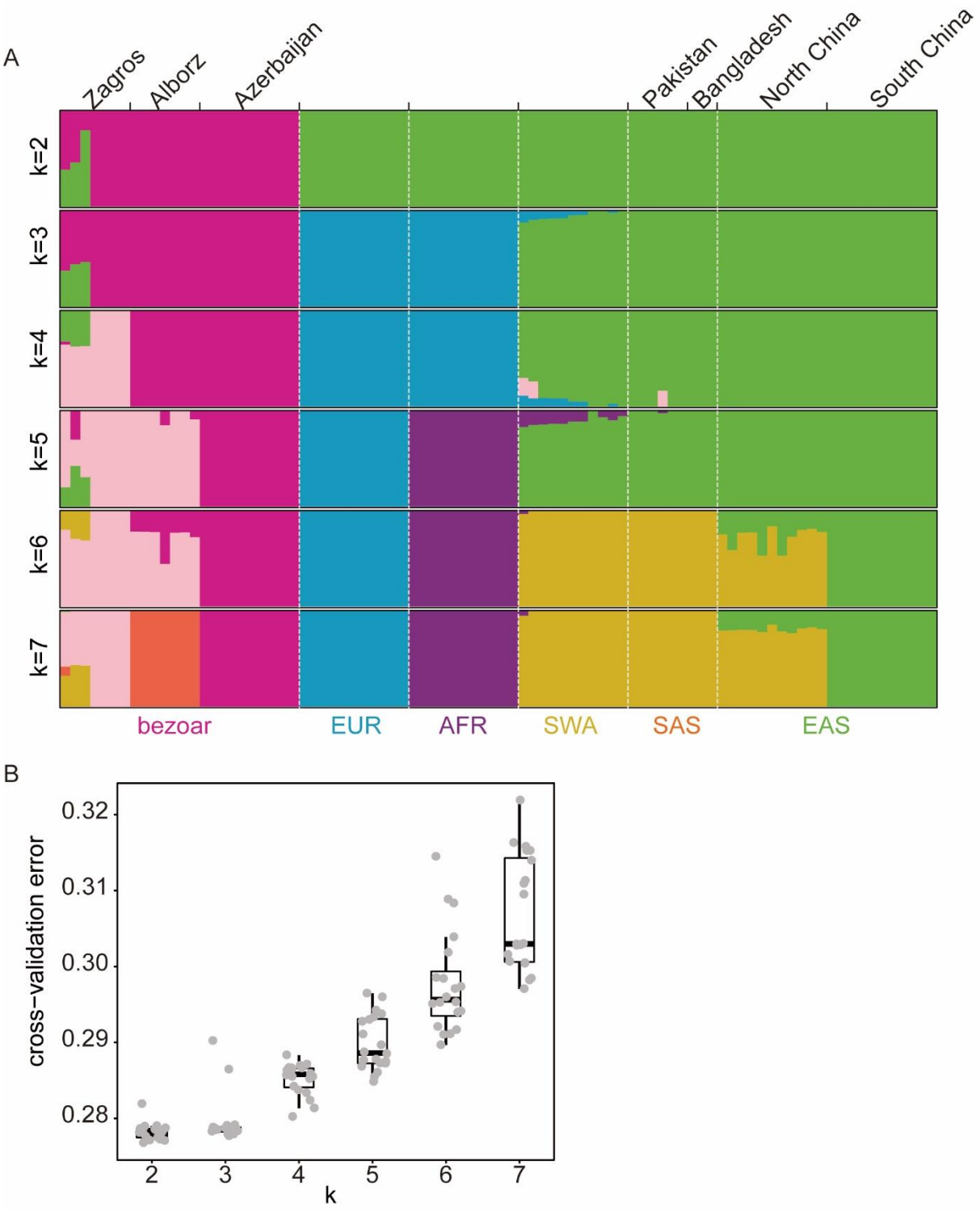


Fig. S6.
ADMIXTURE analysis on the subset of 88 samples. (**A**) ADMIXTURE results for $k = 2$ to $k = 7$. The run with the lowest cross-validation error (out of 20 replicates) is plotted. The group names are at the bottom, and the black lines at the top of the figure denote the geographic locations of bezoar, SAS, and EAS. (**B**) CV error for varying k in the ADMIXTURE analysis.

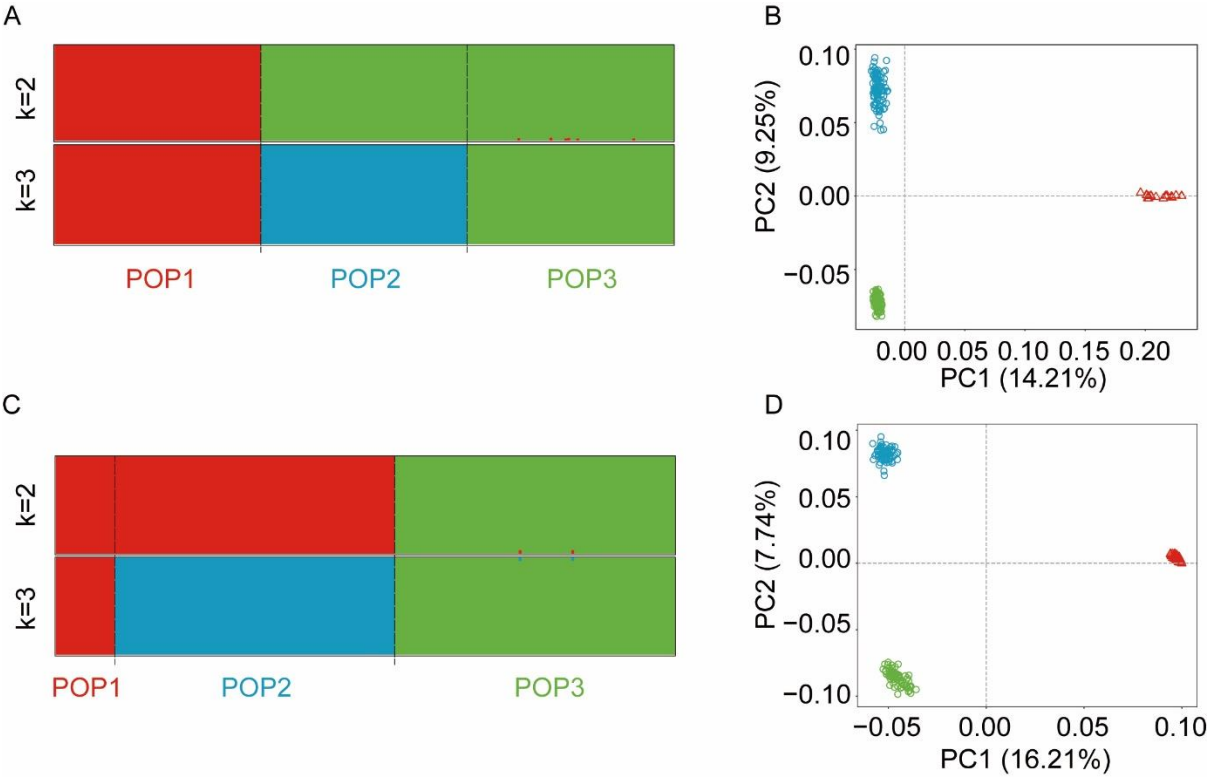


Fig. S7.
Results from running ADMIXTURE and PCA on the simulated data. Models *k* = 2 and *k* = 3 are shown. (**A**) Clustering of individuals following the even sampling strategy. (**B**) Analysis of the first two PCs of the even sampling strategy. (**C**) Clustering of individuals following the uneven sampling strategy. (**D**) Analysis of the first two PCs of the uneven sampling strategy.

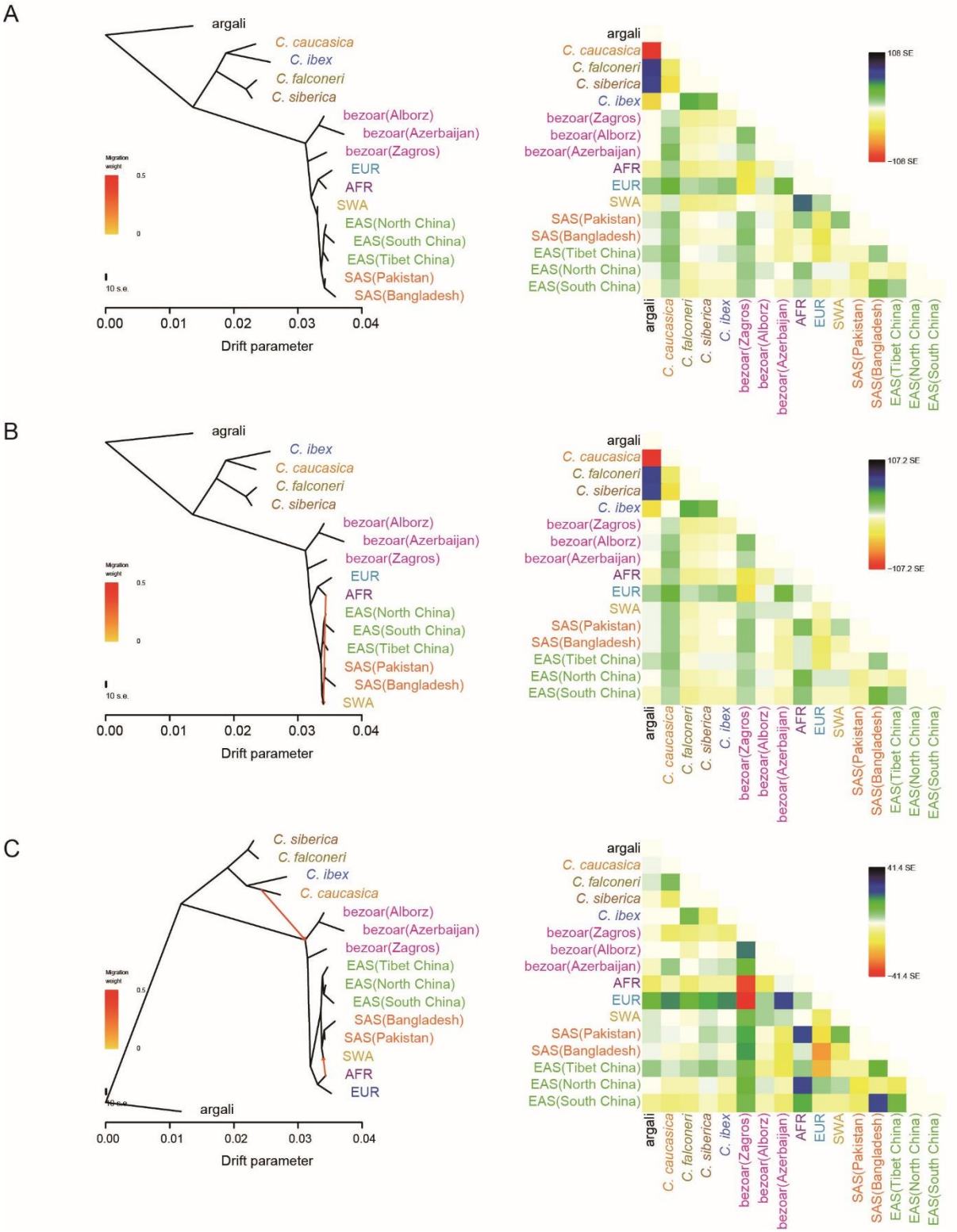


Fig. S8.

Inference of population splits and mixtures by means of TreeMix. Maximum likelihood (ML) trees (left panel) and corresponding model residuals (right panel). All nodes have 100% support (100 bootstrap replicates). (**A**) ML tree with no migrations explaining 99.82% of the variance. (**B**) ML tree with one migration event from AFR to SWA explaining 99.84% of the variance. (**C**) ML tree with adding a migration event from *C. caucasica* to the node leading to bezoars and domestic goats explaining 99.95% of the variance.

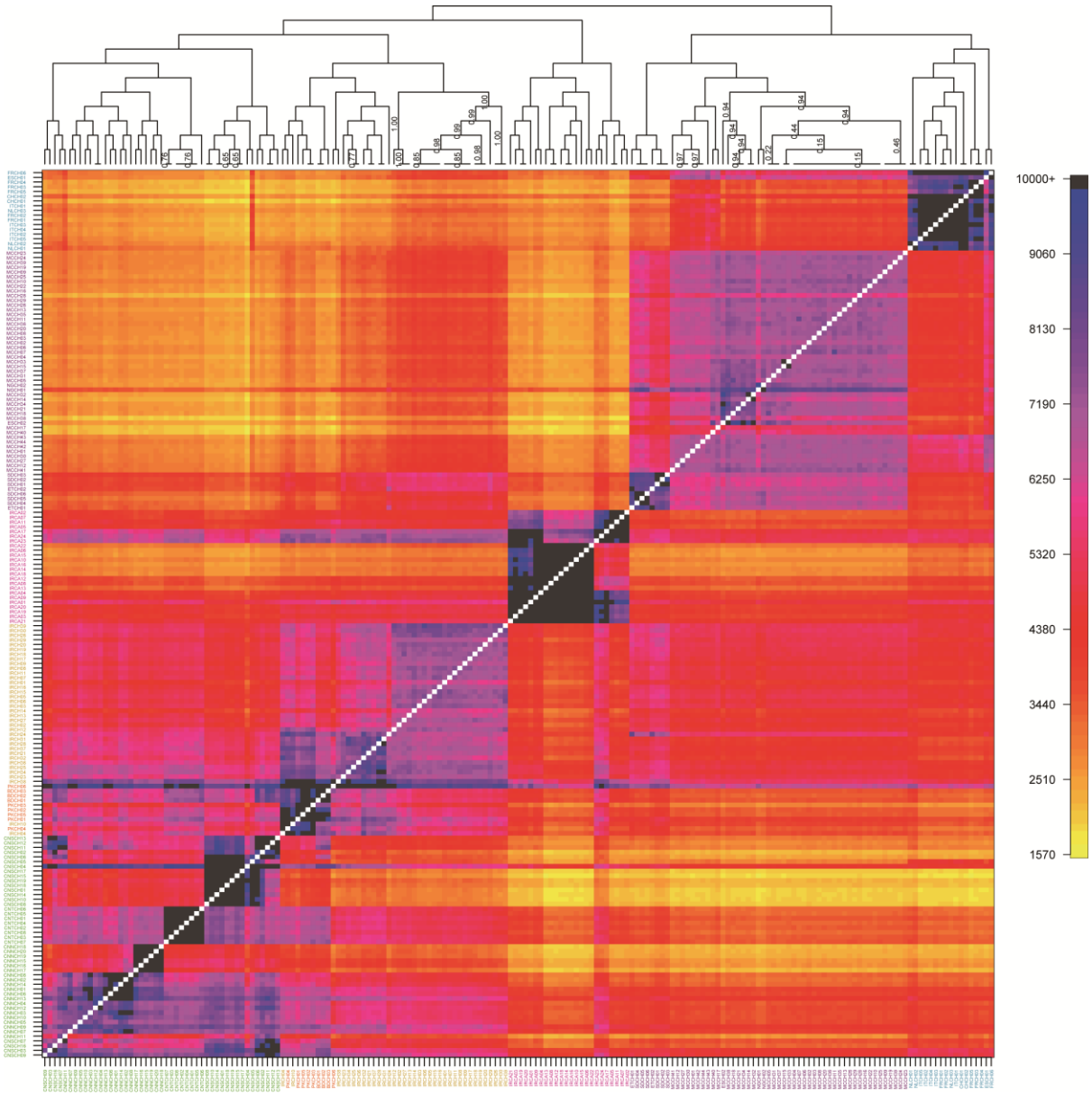


Fig. S9.

Coancestry heatmap. ChromoPainter and fineSTRUCTURE results, showing the underlying number of discrete "haplotypes" that an individual (rows) receives from other donor individuals (columns). The dendrogram shows the clustering of the analyzed individuals. There is no underlying historical or evolutionary model assumed by this representation. The numbers on the dendrogram give the proportion of MCMC iterations for which each population split is observed. All the "unreported" values were 1. Darker colors on the heatmap represent greater haplotype sharing. fineSTRUCTURE distinguished between the different populations present in the bezoar population. Further structure mirrored genetic drift in AFR, EUR, SWA, SAS, and EAS.

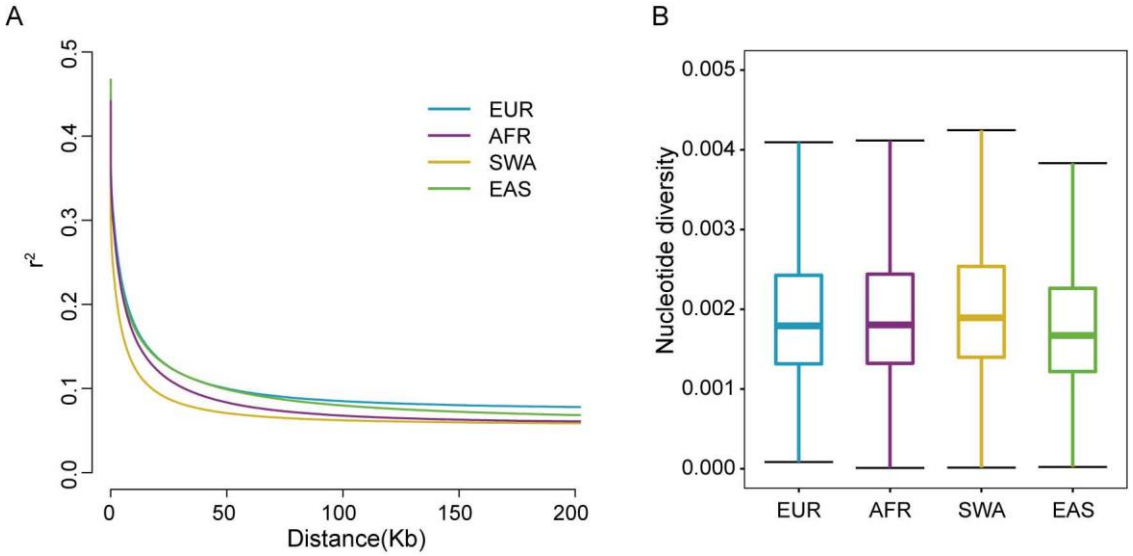


Fig. S10.
Linkage disequilibrium decay and genetic diversity of four domestic populations. For EAS, we consider only samples from South China. (**A**) The decay of linkage disequilibrium measured as the squared correlation coefficient by pairwise physical distance in four domestic populations. To minimize the effect of sample size variation, we randomly reduced the sample size to 16 for each population. (**B**) Boxplots of nucleotide diversity, calculated in 50 kb sliding window with 20 kb increments across the genome.

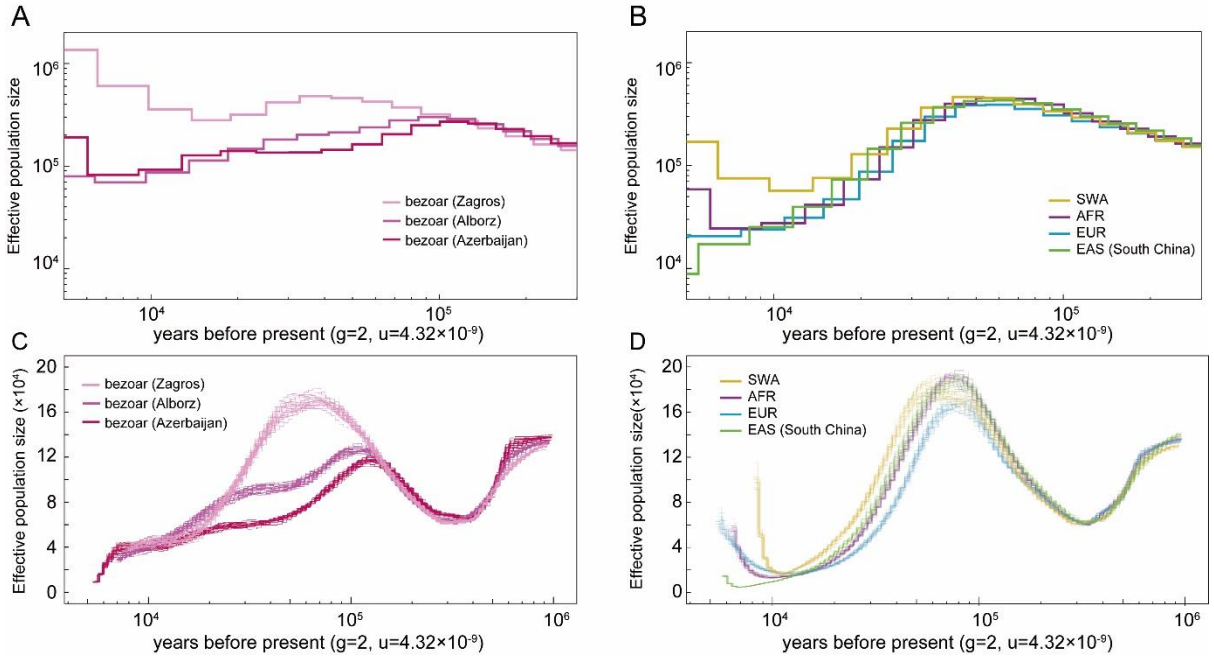


Fig. S11.
Inference of population size from whole-genome sequences. For EAS, we consider only samples from South China. (**A**, **B**) Effective population size histories estimated using MSMC2 from four haplotypes (two phased individuals) for each of seven populations. (**C**, **D**) Effective population size histories inferred using SMC++ with 20 bootstrap replicates.

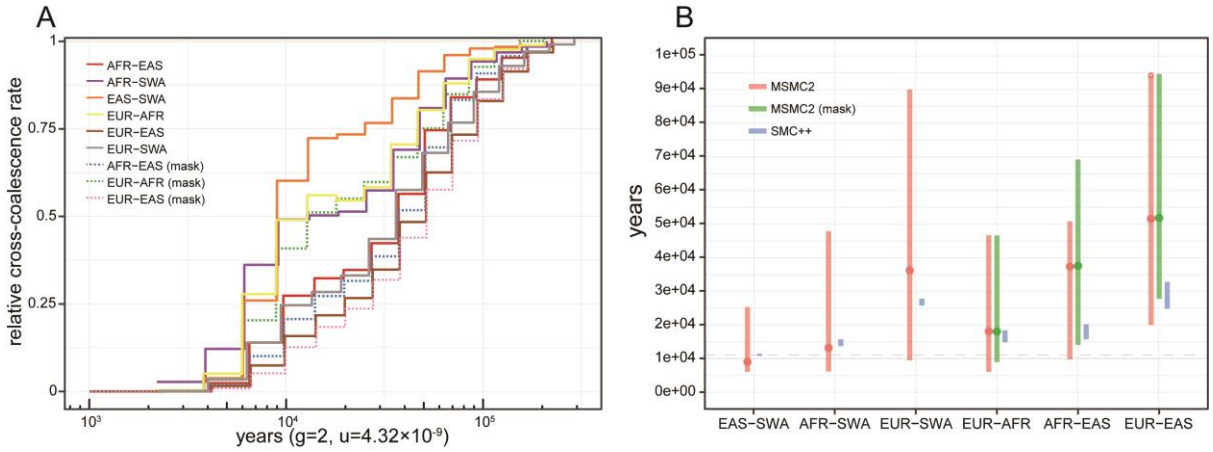


Fig. S12.

Genetic separation between population pairs. (**A**) Relative cross coalescence rates between domestic goat populations. Values close to 1 indicate that the two populations have not yet diverged. Values close to 0 indicate that the populations have completely diverged. (**B**) Split times inferred using MSMC2 and SMC++. The red lines represent the time inferred from MSMC2. The green lines represent the time inferred from MSMC2 after masking out the recent segments of bezoar ancestry. Dots, lower and upper bar represents the time at which cross coalescence rate dropped below, 50%, 25%, and 75% respectively. The blue lines represent the time inferred using SMC++ on 11 genomes per group. The estimation is based on the results of 20 bootstrap sets.

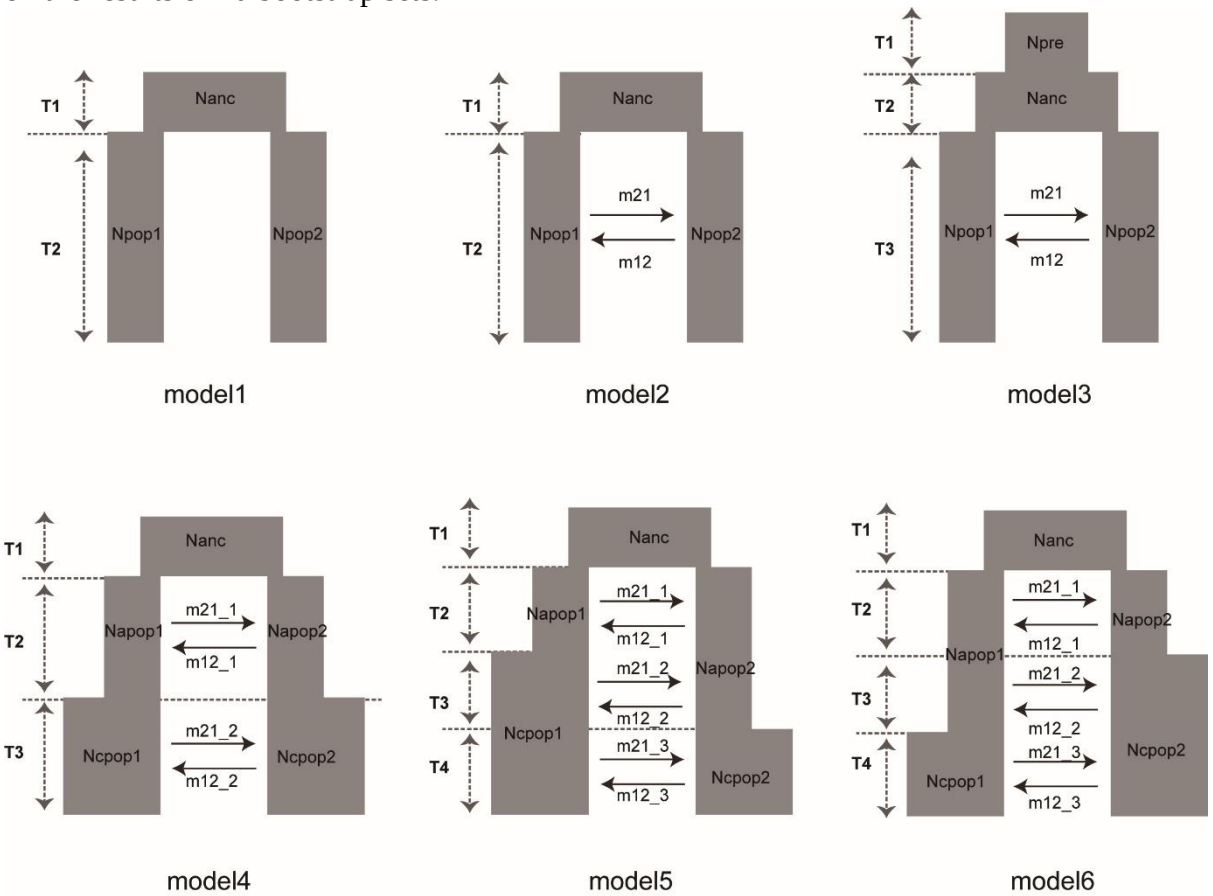


Fig. S13.

Six tested demographic models applied to pairs of populations. The description of the parameters is given in table S7.

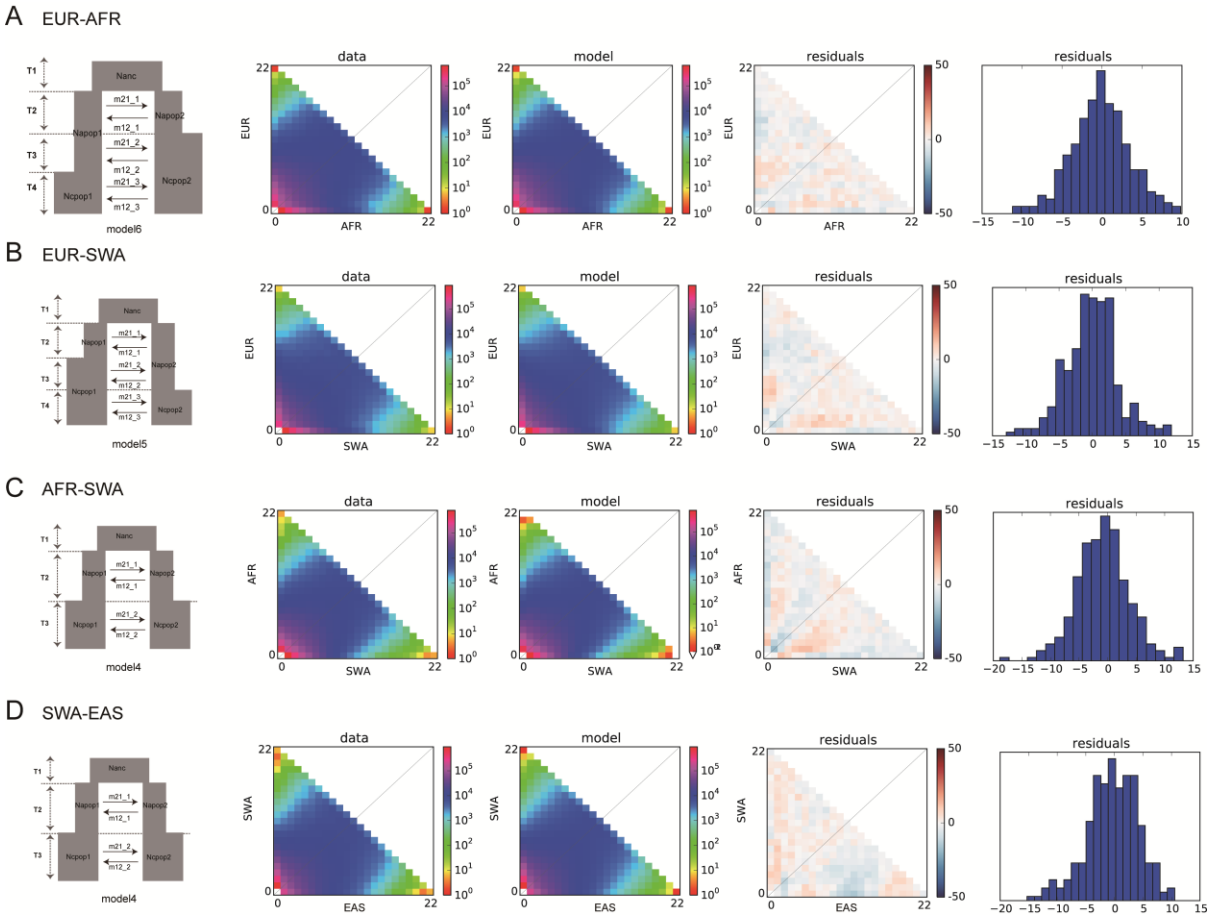


Fig. S14.
Demographic modeling for pairs of populations. (**A**) EUR and AFR. (**B**) EUR and SWA. (**C**) AFR and SWA. (**D**) SWA and EAS. A simplified graphic of the best-fit model is depicted, along with comparisons of the two-dimensional site frequency spectrum for the data, the model and resulting residuals. Parameter values are provided in table S7.

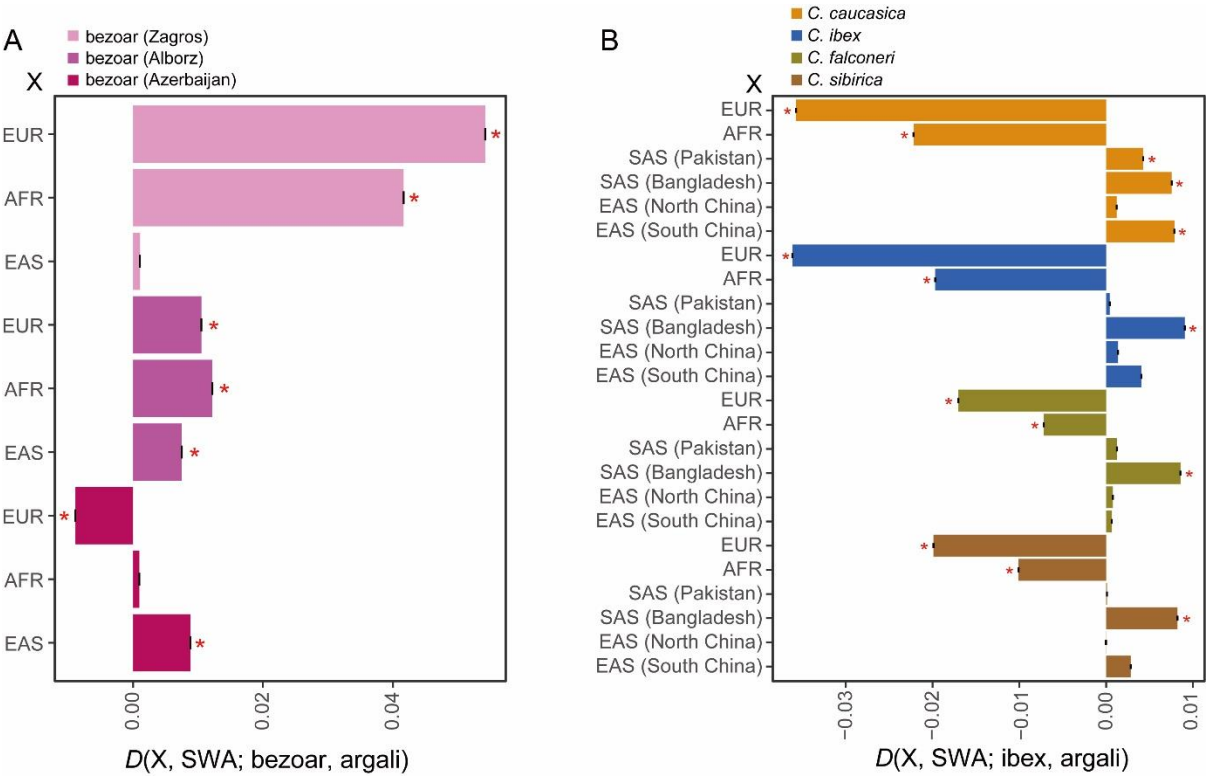
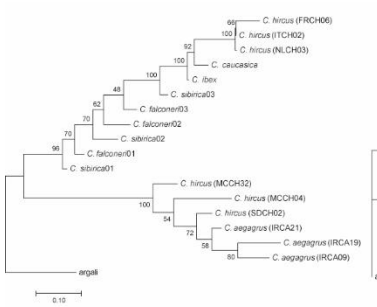
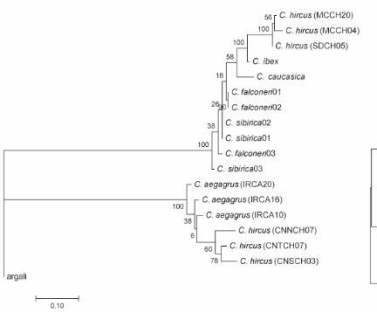


Fig. S15.

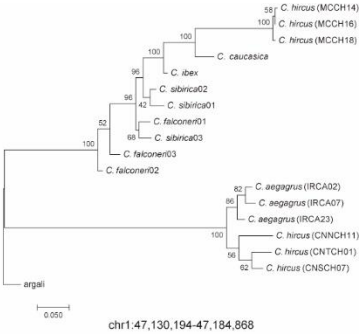
Allele sharing between domestic goats and wild goats. (**A**) Allele sharing between domestic goats and bezoars. (**B**) Allele sharing between domestic goats and ibex-like species. Statistically significant results, defined by |Z-scores| ≥ 3, are marked with a red asterisk. Negative values were obtained if wild goats were closer to X, and positive values if wild goats were closer to SWA.



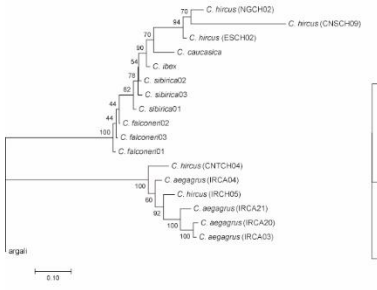
chr1:45,158,344-45,208,111



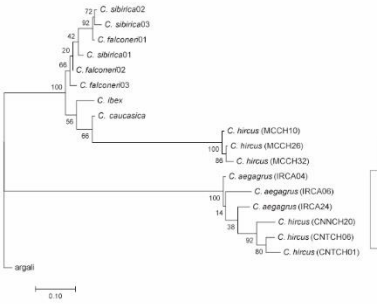
chr1:45,839,068-45,864,135



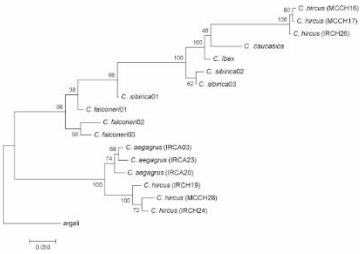
chr1:47,130,194-47,184,868



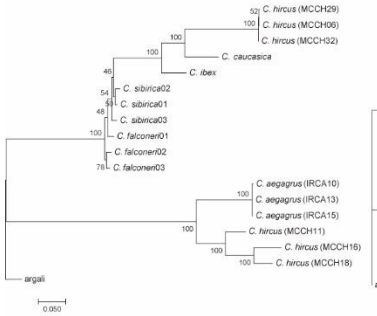
chr1:62,891,663-62,969,963



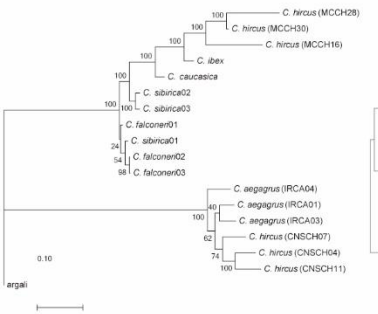
chr1:66,270,143-66,309,645



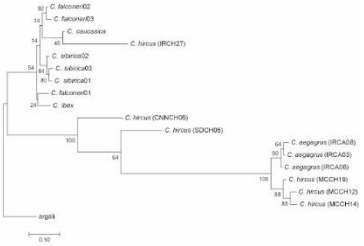
chr1:76,014,923-76,049,022



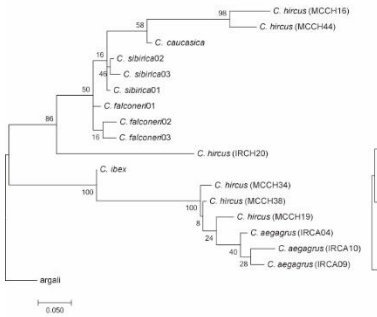
chr1:104,150,173-104,349,720



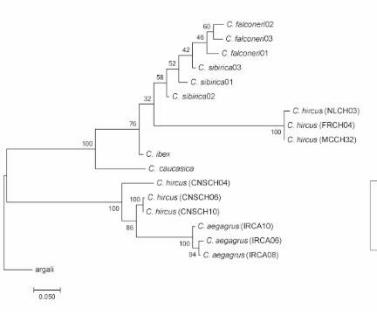
chr1:116,250,672-116,409,358



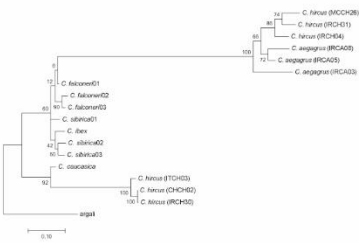
chr1:118,433,695-118,469,862



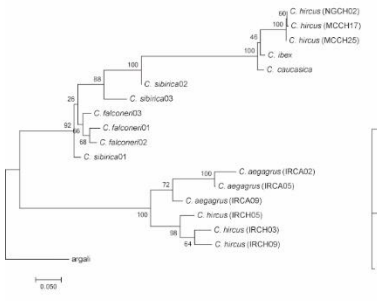
chr1:121,111,765-121,139,821



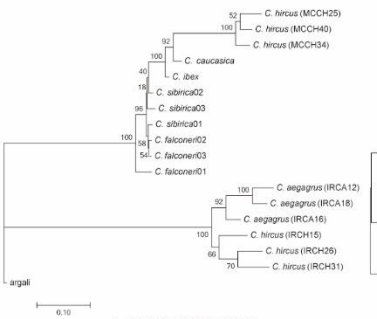
chr1:124,031,076-124,067,980



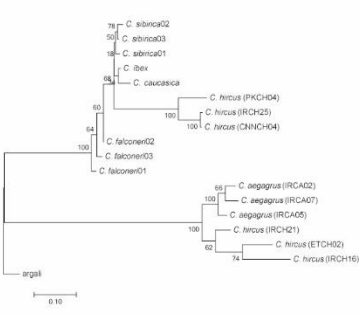
chr2:20,433,559-20,469,857



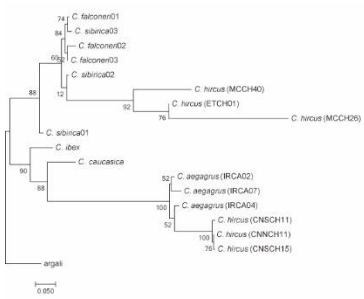
chr2:24,324,410-24,369,675



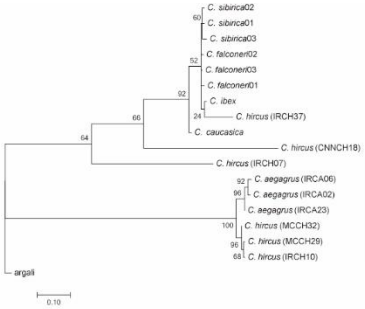
chr2:25,015,487-25,107,656



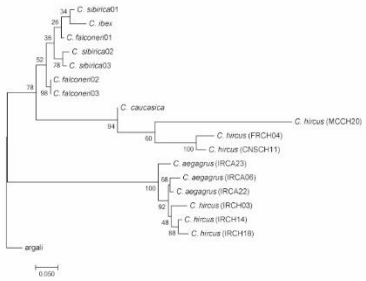
chr2:39,223,705-39,269,620



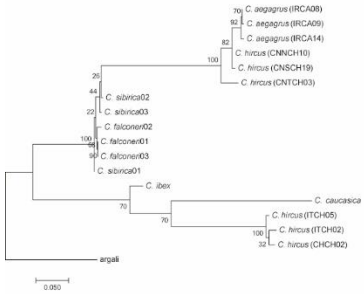
chr2:71,813,208-71,848,112



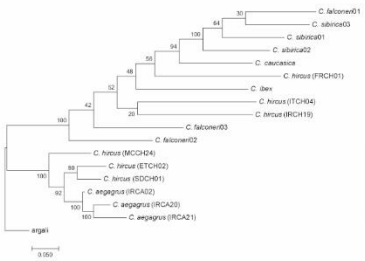
chr2:78,444,113-78,463,513



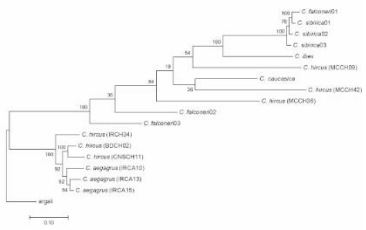
chr2:94,099,862-94,128,976



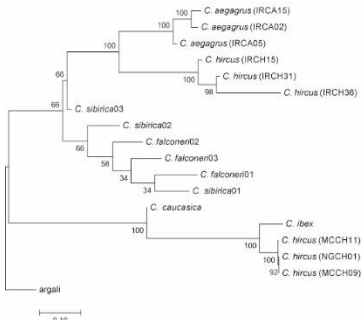
chr3:17,070,266-17,109,926



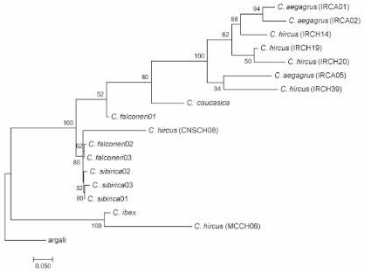
chr3:66,410,139-66,489,579



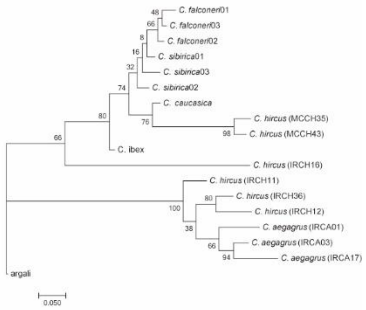
chr3:66,602,349-66,769,681



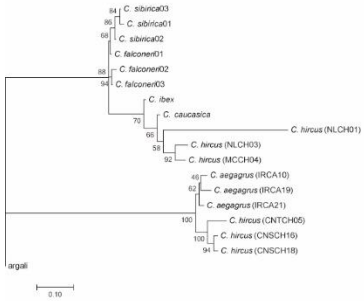
chr3:108,170,029-108,228,388



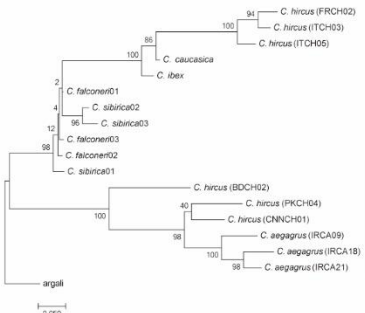
chr3:108,332,708-108,369,750



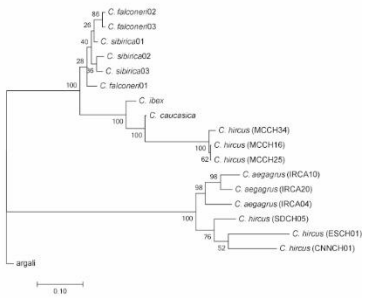
chr3:111,872,072-111,908,509



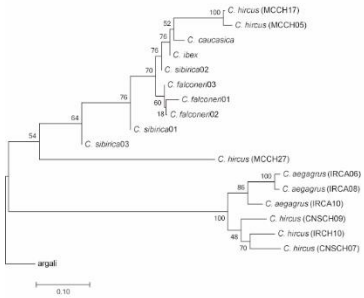
chr4:13,670,435-13,708,894



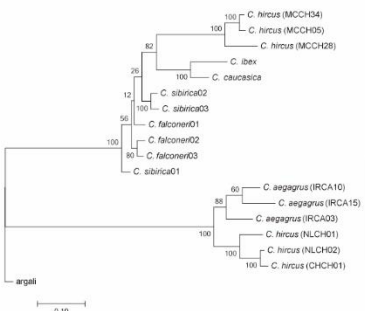
chr4:64,380,436-64,428,804



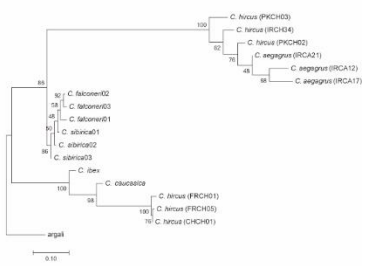
chr4:67,335,251-67,448,678



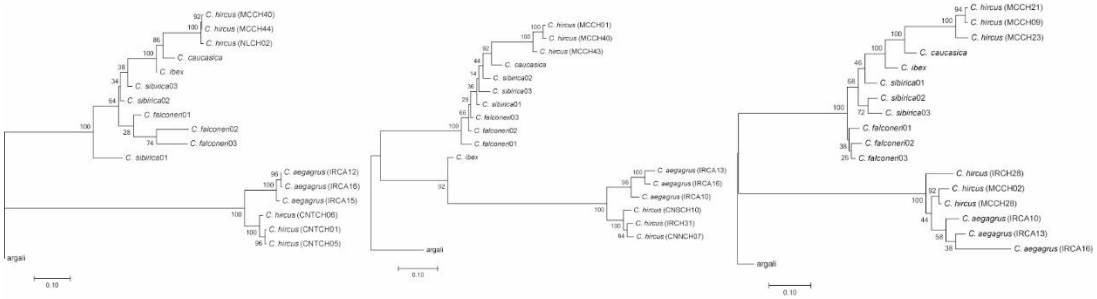
chr4:96,990,515-97,024,836

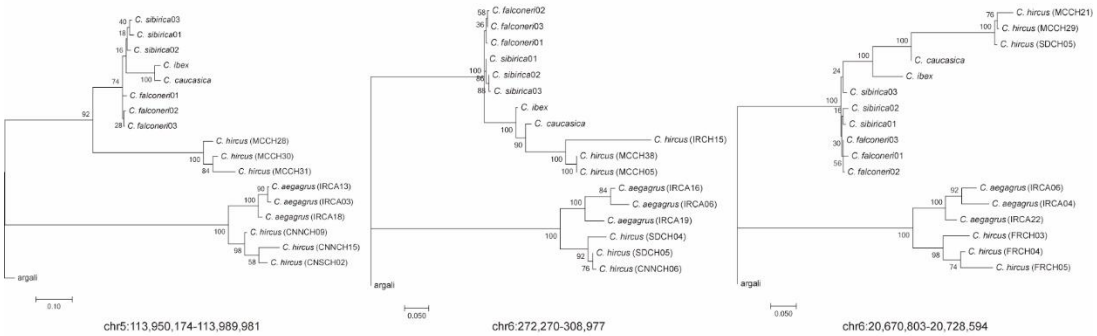


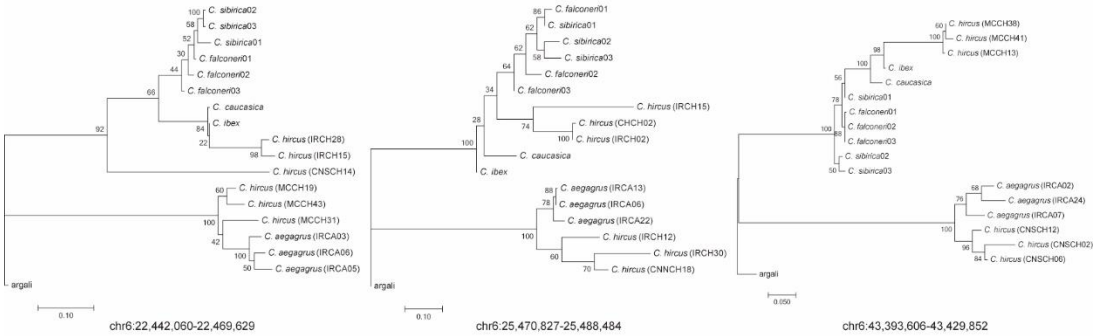
chr5:17,970,193-18,229,602

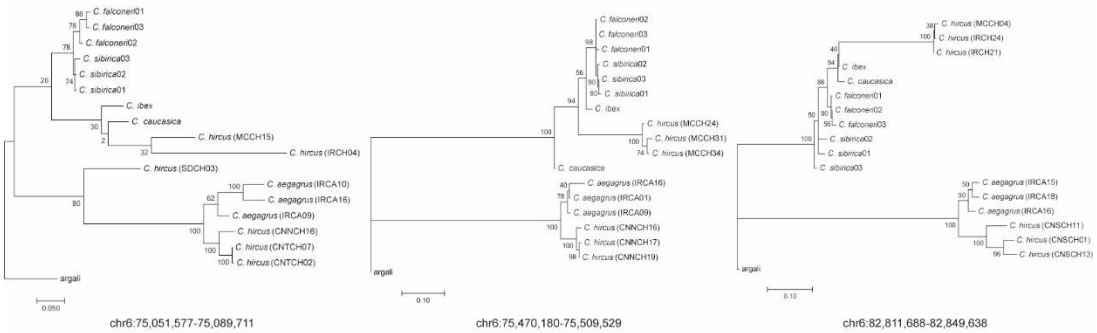


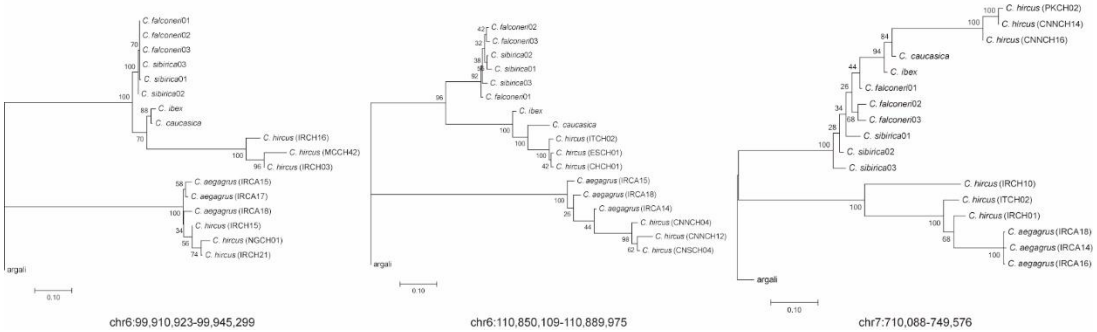
chr5:26,694,246-26,728,907

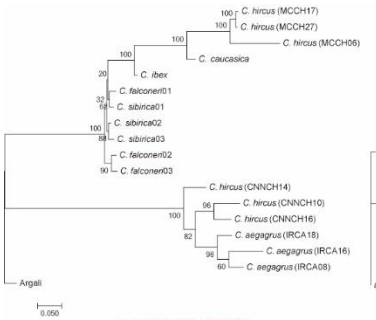

C. hircus (MCCH40)
C. hircus (MCCH44)
C. hircus (NLCH02)
C. caucasica
C. ibex
C. sibirica03
C. sibirica02
C. falconeri01
C. falconeri02
C. falconeri03
C. sibirica01
C. aegagrus (IRCA12)
C. aegagrus (IRCA16)
C. aegagrus (IRCA15)
C. hircus (CNTCH06)
C. hircus (CNTCH01)
C. hircus (CNTCH05)
argali
0.10
C. hircus (MCCH01)
C. hircus (MCCH40)
C. hircus (MCCH43)
C. caucasica
C. sibirica02
C. sibirica03
C. sibirica01
C. falconeri02
C. falconeri03
C. falconeri01
C. ibex
C. aegagrus (IRCA13)
C. aegagrus (IRCA16)
C. aegagrus (IRCA10)
C. hircus (CNSCH10)
C. hircus (IRCH31)
C. hircus (CNNCH07)
argali
0.10
C. hircus (MCCH21)
C. hircus (MCCH09)
C. hircus (MCCH23)
C. caucasica
C. ibex
C. sibirica01
C. sibirica02
C. sibirica03
C. falconeri01
C. falconeri02
C. falconeri03
C. hircus (IRCH28)
C. hircus (MCCH02)
C. hircus (MCCH28)
C. aegagrus (IRCA10)
C. aegagrus (IRCA13)
C. aegagrus (IRCA16)
argali
0.10
chr5:80,922,357-80,949,324
chr5:94,651,258-94,688,986
chr5:113,696,707-113,745,333


C. sibirica03
C. sibirica01
C. sibirica02
C. ibex
C. caucasica
C. falconeri01
C. falconeri02
C. falconeri03
C. hircus (MCCH28)
C. hircus (MCCH30)
C. hircus (MCCH31)
C. aegagrus (IRCA13)
C. aegagrus (IRCA03)
C. aegagrus (IRCA18)
C. hircus (CNNCH09)
C. hircus (CNNCH15)
C. hircus (CNSCH02)
argali
0.10
C. falconeri02
C. falconeri03
C. falconeri01
C. sibirica01
C. sibirica02
C. sibirica03
C. ibex
C. caucasica
C. hircus (IRCH15)
C. hircus (MCCH38)
C. hircus (MCCH05)
C. aegagrus (IRCA16)
C. aegagrus (IRCA06)
C. aegagrus (IRCA19)
C. hircus (SDCH04)
C. hircus (SDCH05)
C. hircus (CNNCH06)
argali
0.050
C. hircus (MCCH21)
C. hircus (MCCH29)
C. hircus (SDCH05)
C. caucasica
C. ibex
C. sibirica03
C. sibirica02
C. sibirica01
C. falconeri03
C. falconeri01
C. falconeri02
C. aegagrus (IRCA06)
C. aegagrus (IRCA04)
C. aegagrus (IRCA22)
C. hircus (FRCH03)
C. hircus (FRCH04)
C. hircus (FRCH05)
argali
0.050
chr5:113,950,174-113,989,981
chr6:272,270-308,977
chr6:20,670,803-20,728,594

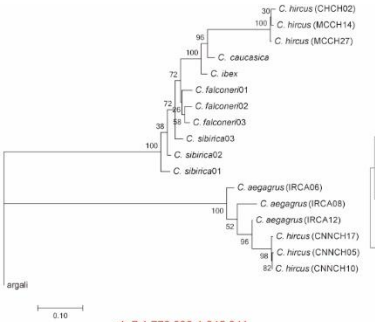

C. sibirica02
C. sibirica03
C. sibirica01
C. falconeri01
C. falconeri02
C. falconeri03
C. caucasica
C. ibex
C. hircus (IRCH28)
C. hircus (IRCH15)
C. hircus (CNSCH14)
C. hircus (MCCH19)
C. hircus (MCCH43)
C. hircus (MCCH31)
C. aegagrus (IRCA03)
C. aegagrus (IRCA06)
C. aegagrus (IRCA05)
argali
0.10
C. falconeri01
C. sibirica01
C. sibirica02
C. sibirica03
C. falconeri02
C. falconeri03
C. hircus (IRCH15)
C. hircus (CHCH02)
C. hircus (IRCH02)
C. caucasica
C. ibex
C. aegagrus (IRCA13)
C. aegagrus (IRCA06)
C. aegagrus (IRCA22)
C. hircus (IRCH12)
C. hircus (IRCH30)
C. hircus (CNNCH18)
argali
0.10
C. hircus (MCCH38)
C. hircus (MCCH41)
C. hircus (MCCH13)
C. ibex
C. caucasica
C. sibirica01
C. falconeri01
C. falconeri02
C. falconeri03
C. sibirica02
C. sibirica03
C. aegagrus (IRCA02)
C. aegagrus (IRCA24)
C. aegagrus (IRCA07)
C. hircus (CNSCH12)
C. hircus (CNSCH02)
C. hircus (CNSCH06)
argali
0.050
chr6:22,442,060-22,469,629
chr6:25,470,827-25,488,484
chr6:43,393,606-43,429,852


C. falconeri01
C. falconeri03
C. falconeri02
C. sibirica03
C. sibirica02
C. sibirica01
C. ibex
C. caucasica
C. hircus (MCCH15)
C. hircus (IRCH04)
C. hircus (SDCH03)
C. aegagrus (IRCA10)
C. aegagrus (IRCA16)
C. aegagrus (IRCA09)
C. hircus (CNNCH16)
C. hircus (CNTCH07)
C. hircus (CNTCH02)
argali
0.050
C. falconeri02
C. falconeri03
C. falconeri01
C. sibirica02
C. sibirica03
C. sibirica01
C. ibex
C. hircus (MCCH24)
C. hircus (MCCH31)
C. hircus (MCCH34)
C. caucasica
C. aegagrus (IRCA16)
C. aegagrus (IRCA01)
C. aegagrus (IRCA09)
C. hircus (CNNCH16)
C. hircus (CNNCH17)
C. hircus (CNNCH19)
argali
0.10
C. hircus (MCCH04)
C. hircus (IRCH24)
C. hircus (IRCH21)
C. ibex
C. caucasica
C. falconeri01
C. falconeri02
C. falconeri03
C. sibirica02
C. sibirica01
C. sibirica03
C. aegagrus (IRCA15)
C. aegagrus (IRCA18)
C. aegagrus (IRCA16)
C. hircus (CNSCH11)
C. hircus (CNSCH01)
C. hircus (CNSCH13)
argali
0.10
chr6:75,051,577-75,089,711
chr6:75,470,180-75,509,529
chr6:82,811,688-82,849,638

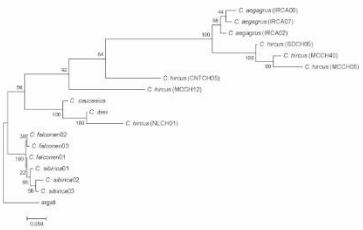

C. falconeri01
C. falconeri02
C. falconeri03
C. sibirica03
C. sibirica01
C. sibirica02
C. ibex
C. caucasica
C. hircus (IRCH16)
C. hircus (MCCH42)
C. hircus (IRCH03)
C. aegagrus (IRCA15)
C. aegagrus (IRCA17)
C. aegagrus (IRCA18)
C. hircus (IRCH15)
C. hircus (NGCH01)
C. hircus (IRCH21)
argali
0.10
C. falconeri02
C. falconeri03
C. sibirica02
C. sibirica01
C. sibirica03
C. falconeri01
C. ibex
C. caucasica
C. hircus (ITCH02)
C. hircus (ESCH01)
C. hircus (CHCH01)
C. aegagrus (IRCA15)
C. aegagrus (IRCA18)
C. aegagrus (IRCA14)
C. hircus (CNNCH04)
C. hircus (CNNCH12)
C. hircus (CNSCH04)
argali
0.10
C. hircus (PKCH02)
C. hircus (CNNCH14)
C. hircus (CNNCH16)
C. caucasica
C. ibex
C. falconeri01
C. falconeri02
C. falconeri03
C. sibirica01
C. sibirica02
C. sibirica03
C. hircus (IRCH10)
C. hircus (ITCH02)
C. hircus (IRCH01)
C. aegagrus (IRCA18)
C. aegagrus (IRCA14)
C. aegagrus (IRCA16)
argali
0.10
chr6:99,910,923-99,945,299
chr6:110,850,109-110,889,975
chr7:710,088-749,576



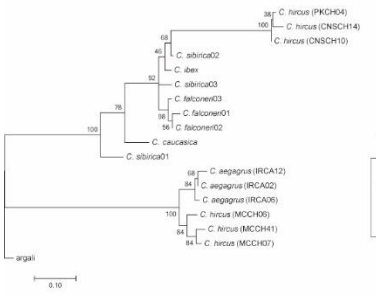
chr7:1,650,338-1,729,960



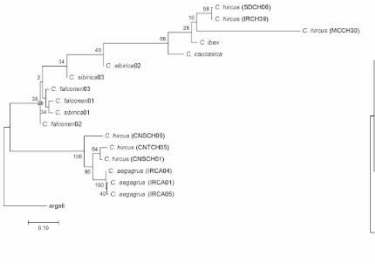
chr7:1,770,093-1,845,841



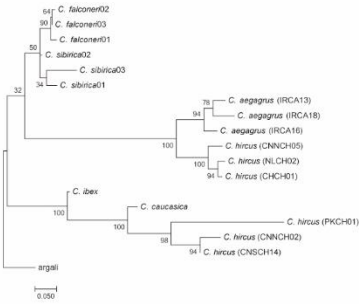
chr7:24,370,068-24,409,932



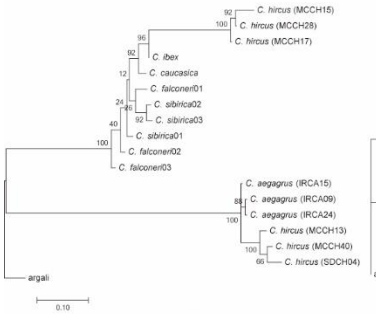
chr7:67,695,478-67,723,558



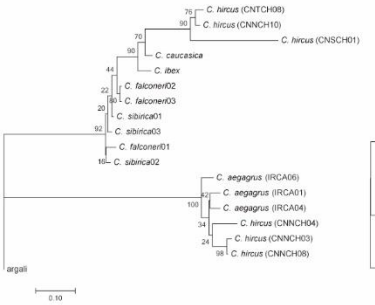
chr8:2,107,219-2,129,797



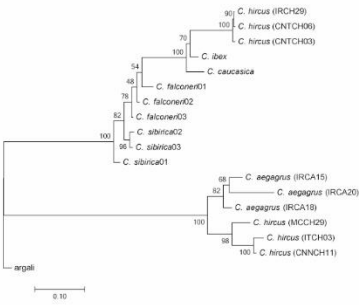
chr8:17,010,008-17,048,002



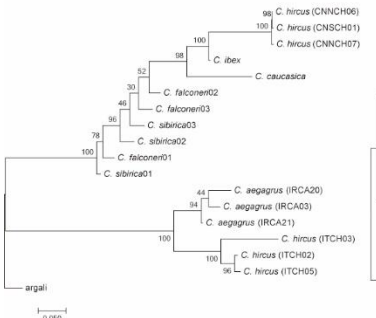
chr8:58,031,254-58,089,635



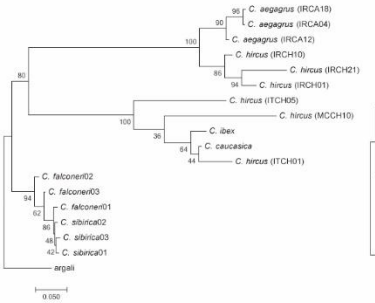
chr9:512,974-542,712



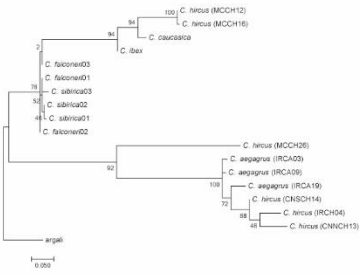
chr9:12,371,236-12,449,664



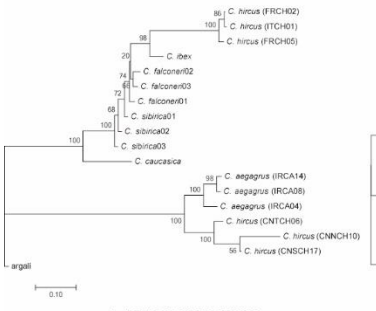
chr9:41,531,672-41,609,811



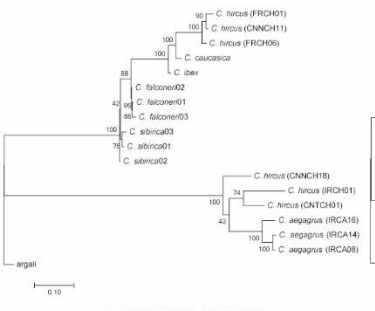
chr9:57,750,453-57,789,985



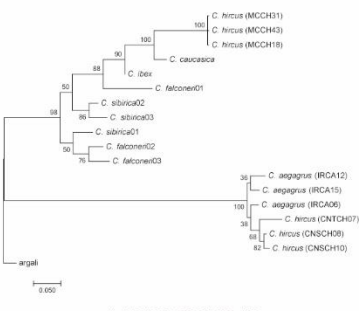
chr9:73,697,723-73,729,986



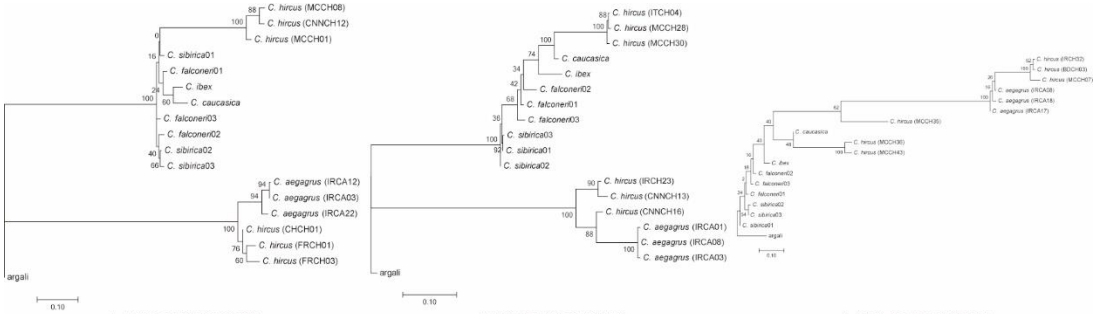
chr10:54,910,317-55,067,477



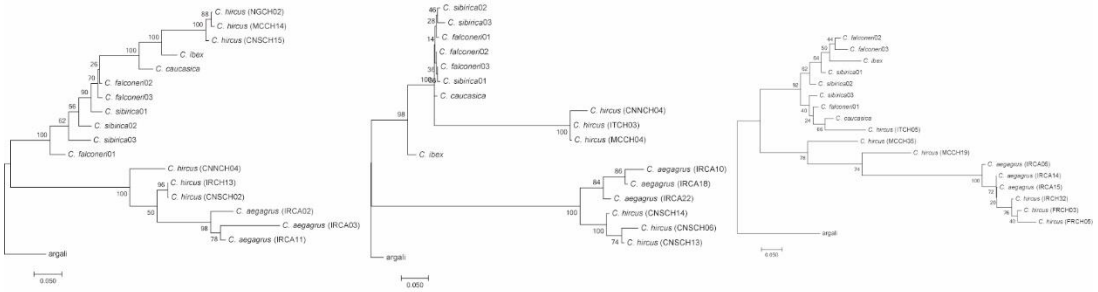
chr10:100,831,506-100,887,335



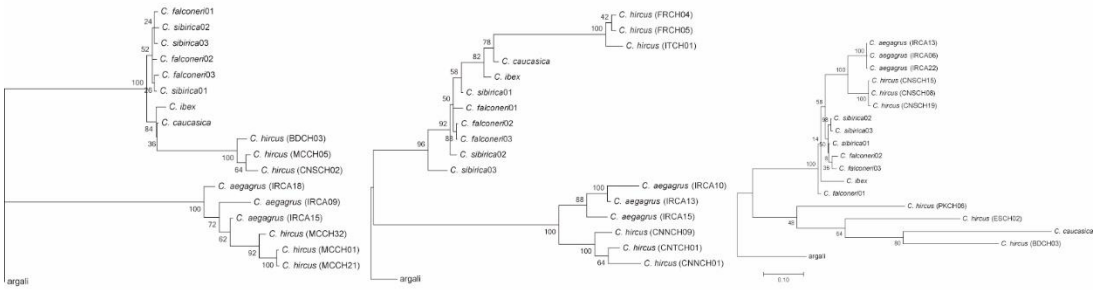
chr11:34,599,880-34,649,117



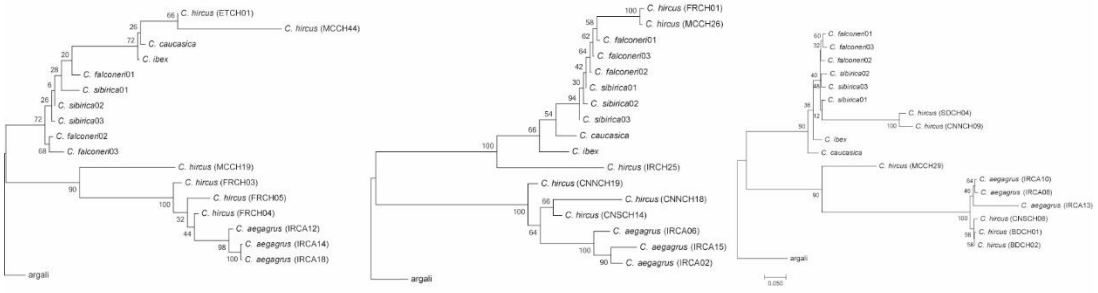
chr11:66,897,539-66,929,689 chr11:68,070,443-68,128,333 chr11:78,464,458-78,495,561



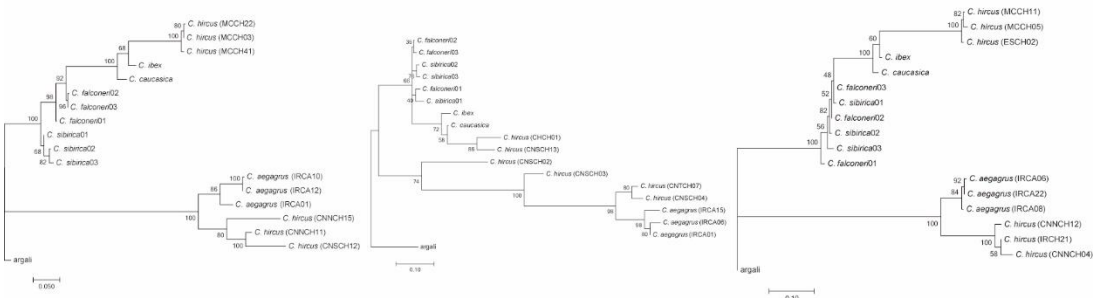
chr12:28,170,650-28,225,631 chr12:44,850,093-44,889,747 chr12:54,130,985-54,172,473



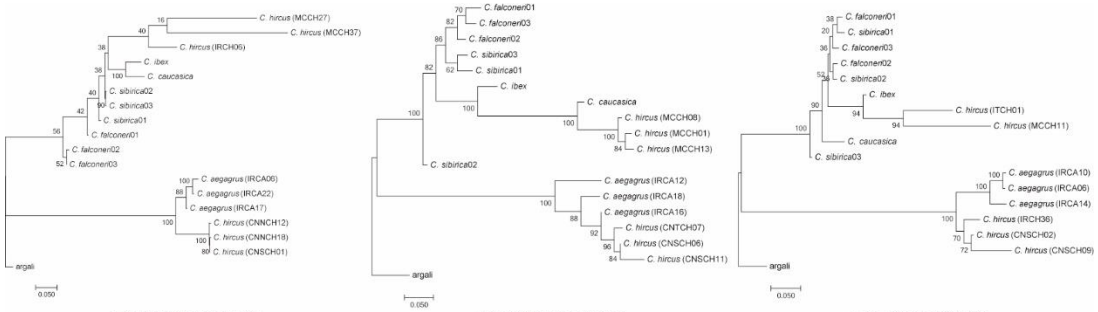
chr12:79,110,590-79,168,979 chr12:83,010,911-83,068,040 chr13:66,710,508-66,749,824



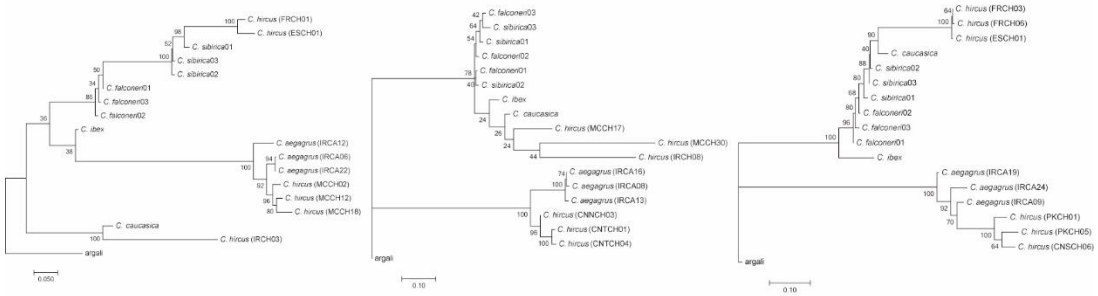
chr14:48,611,708-48,647,670 chr14:77,550,799-77,586,679 chr14:84,877,876-84,909,025



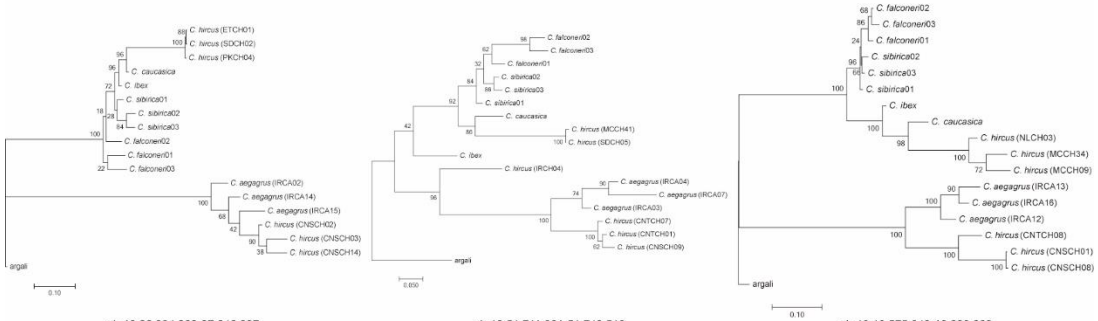
chr14:86,957,811-86,989,811 chr15:55,370,161-55,409,396 chr15:59,956,713-59,996,288



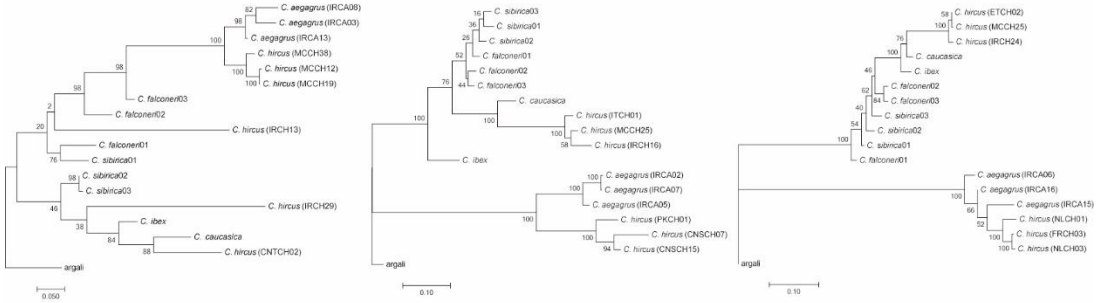
chr15:71,070,083-71,109,191 chr15:73,690,525-73,729,917 chr16:7,950,015-7,989,109



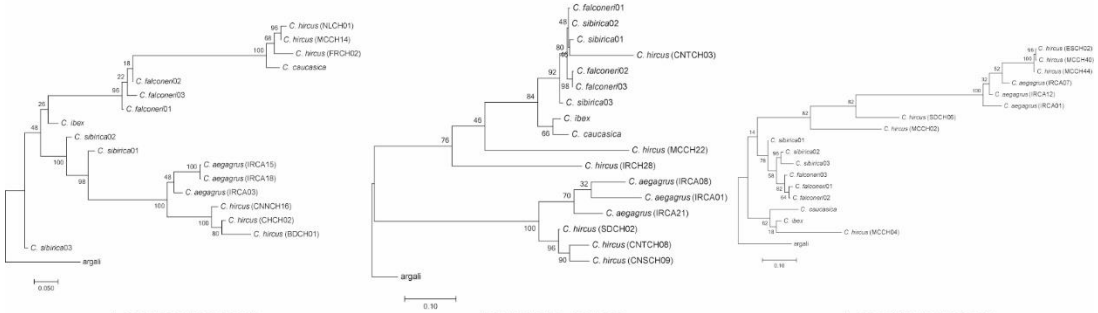
chr16:10,590,301-10,708,790 chr16:55,795,461-55,829,191 chr17:46,651,112-46,692,346



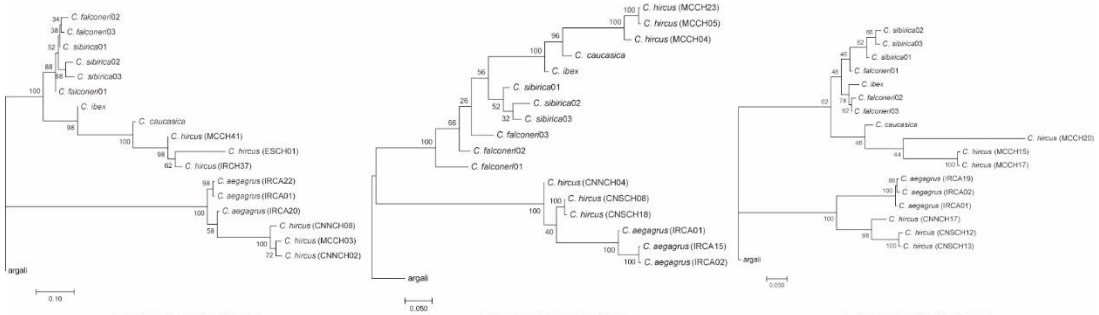
chr18:26,994,268-27,046,087 chr18:51,711,061-51,749,510 chr19:13,575,349-13,629,669



chr20:15,770,044-15,799,982 chr20:31,710,371-31,844,323 chr20:34,450,869-34,486,606



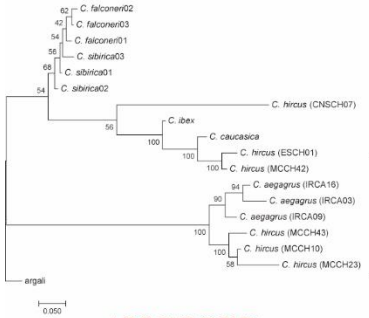
chr20:37,853,810-37,887,311 chr20:43,010,176-43,047,836 chr20:64,770,119-64,809,512



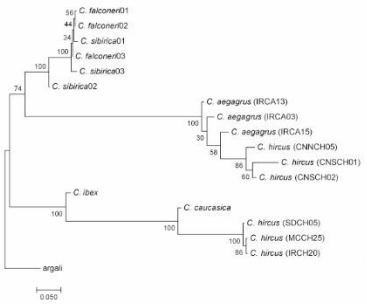
chr21:26,951,414-26,989,920

chr21:50,851,815-50,909,897

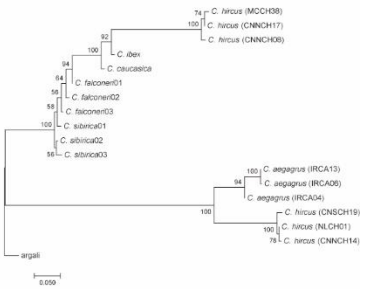
chr23:12,670,017-12,705,980



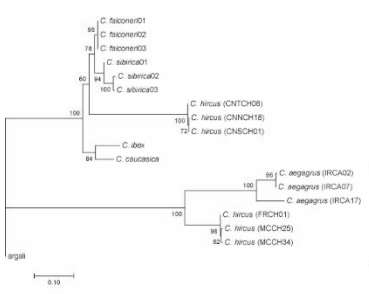
chr23:28,194,061-28,269,721



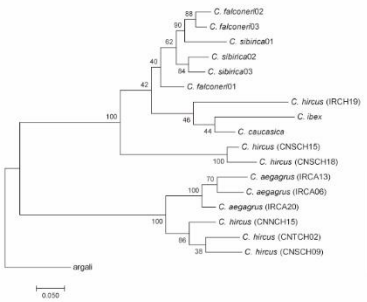
chr24:9,910,474-9,949,908



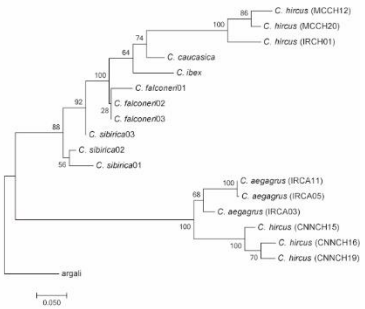
chr24:15,430,619-15,469,648



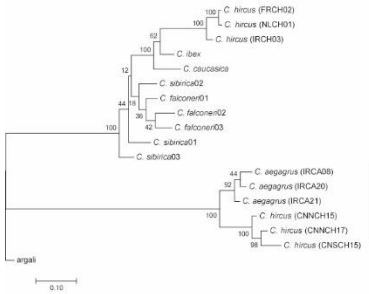
chr24:25,870,350-25,909,099



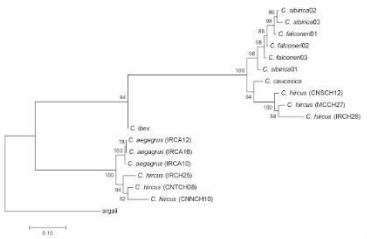
chr24:62,030,118-62,068,556



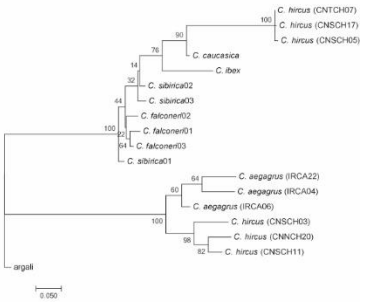
chr25:17,651,855-17,685,183



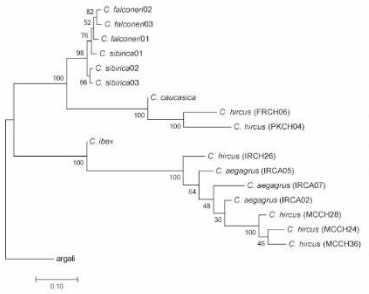
chr25:37,133,196-37,187,274



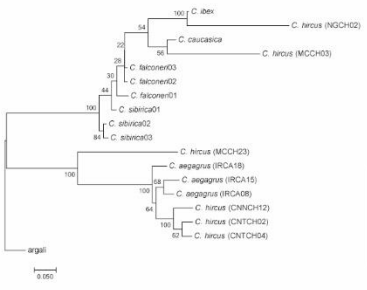
chr26:9,415,069-9,455,900



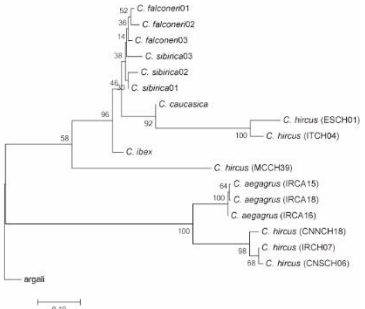
chr26:19,590,005-19,625,950



chr26:22,054,245-22,168,896



chr27:310,395-389,664



chr27:10,590,954-10,628,687

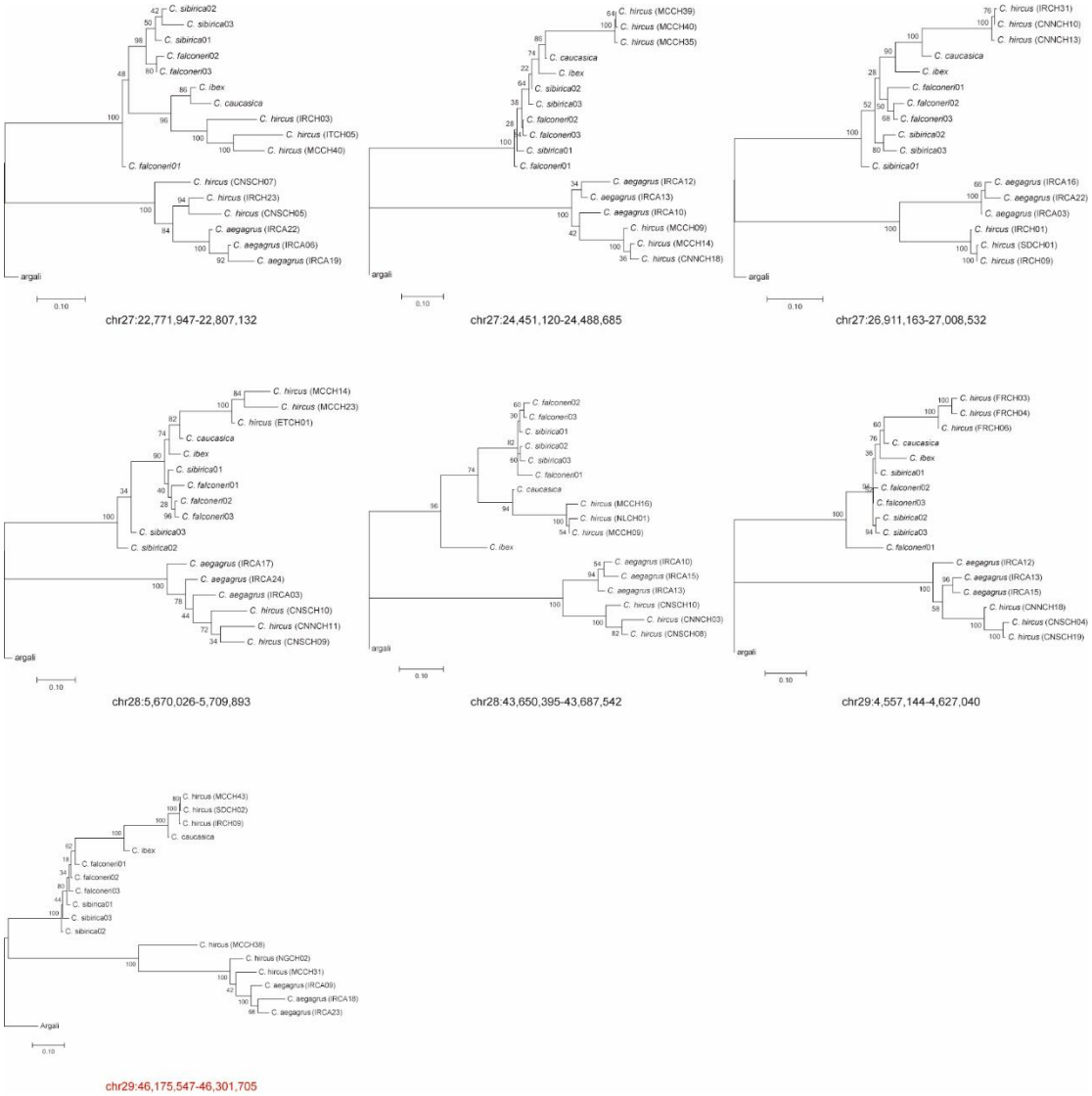


Fig. S16.

Phylogenetic trees were built using pseudo-haplotypes covering the candidate introgressed regions (Data file S1). The pseudo-haplotype was created by randomly calling one allele at the polymorphic sites in ibex-like species and bezoar-goat branch. The immune-related loci are highlighted by red color.

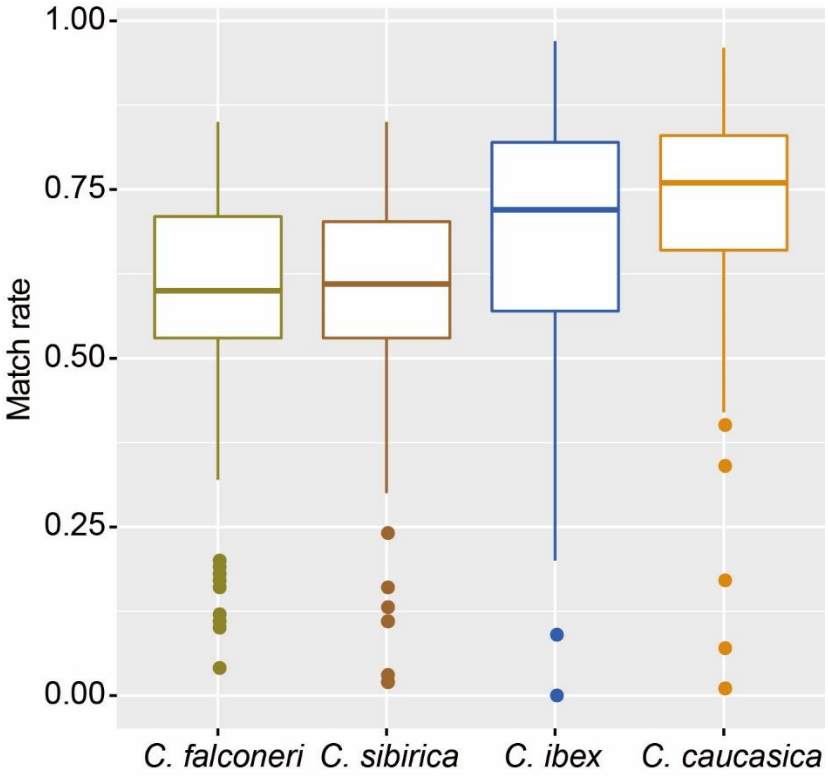


Fig. S17.
The distribution of match rates of the 112 putative introgressed segments to ibex-like genomes. For *C. sibirica* and *C. falconeri*, the averaged match rates are shown (see also Data file S1). Rates of matching of putatively introgressed alleles to ibex-like genomes indicate the degree of divergence between the introgressing and sequenced ibex-like individuals. Match rates suggest that *C. caucasica* shows the highest similarity with the introgressed alleles among the four ibex-like species.

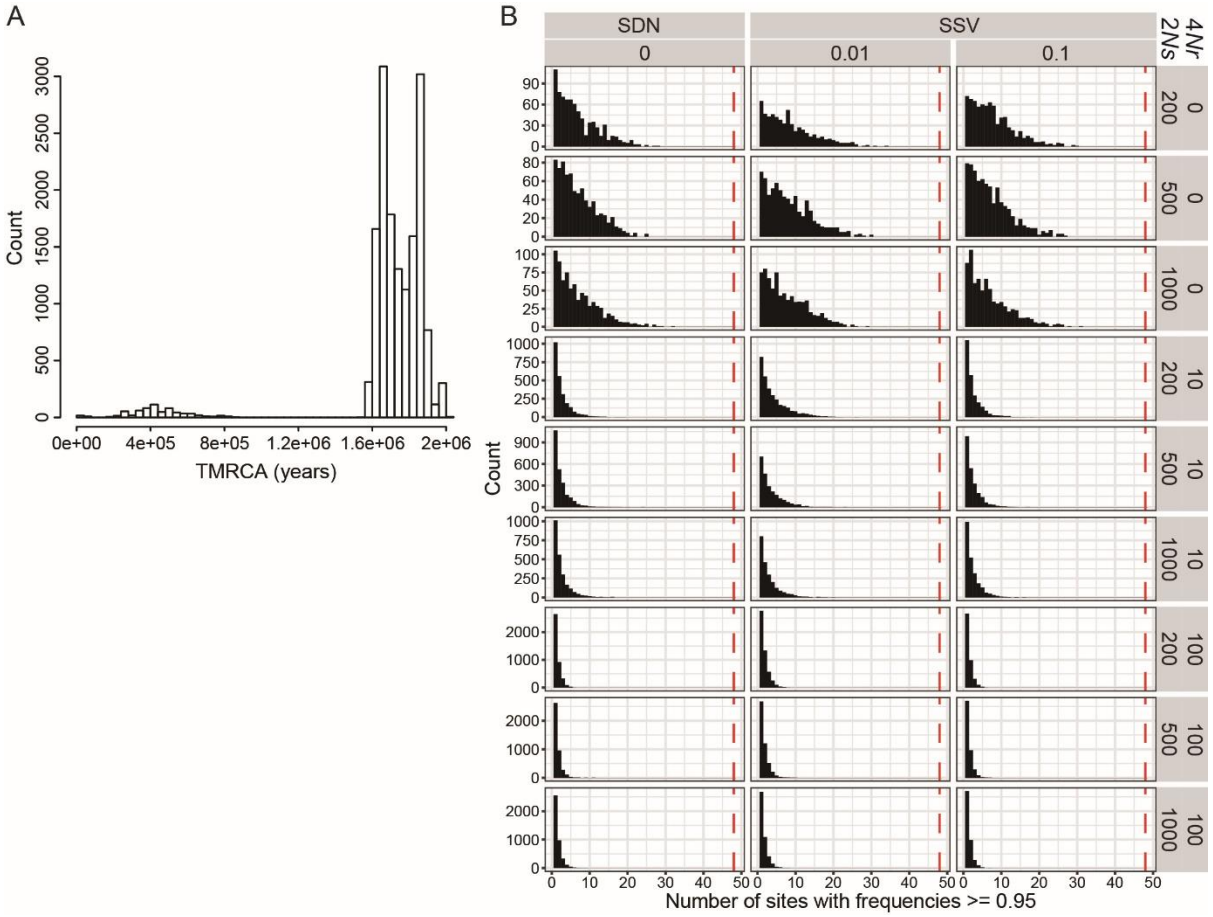


Fig. S18.

Introgression may be responsible for the divergent haplotype in the *MUC6* locus. (A) Distribution of the time to the most recent common ancestor (TMRCA) for all possible pairwise haplotypes between modern bezoars and domestic goats covering the non-repeated region (29:46,258,000-46,268,000) of *MUC6*. (**B**) Distribution of highly differentiated sites observed under two assumed selection models, selection on a *de novo* mutation (SDN) and selection on standing variation (SSV). The initial frequencies of the selected allele in the SSV model are 1% and 10%. Each row of panels corresponds to the combination of the different selection strengths (2*Ns*) from 200 to 1,000 and different recombination rates (4*Nr*) from 0 to 100. The red dashed lines mark the number of highly differentiated sites observed in the real data across the non-repeated region of *MUC6*.

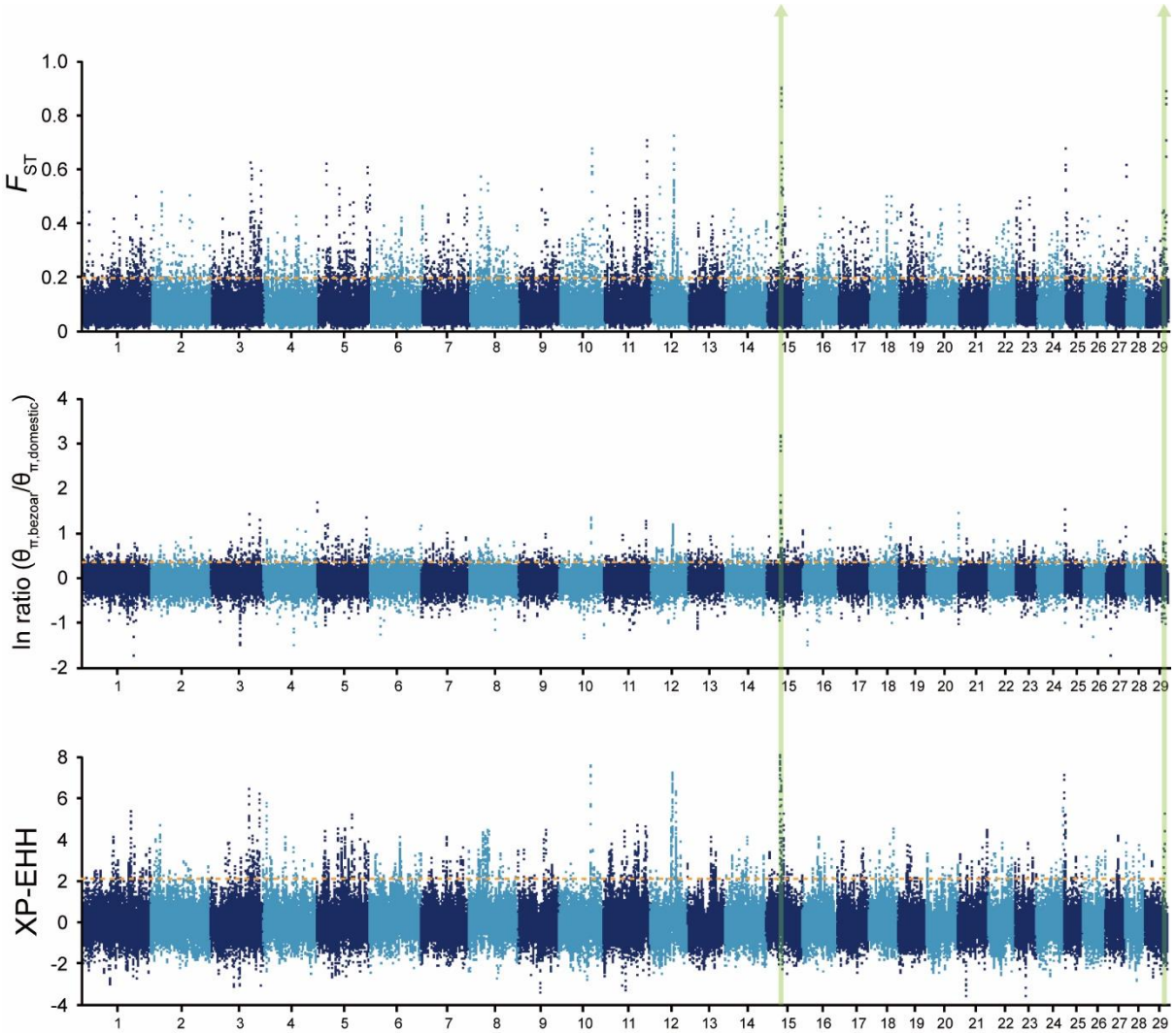


Fig. S19.
Selective sweep analysis by comparing genomes between bezoars and domestic goats. Pairwise fixation index (F_{ST}) (top panel), π ln ratio (middle panel) and normalized XP-EHH scores (bottom panel) calculated between bezoars and domestic goats in a 50 kb sliding window with a 20 kb step across all autosomes. The dashed horizontal lines indicate the significance threshold (corresponding to *Z* test $P < 0.005$, where $F_{ST} > 0.195$, π ln ratio > 0.395 and XP-EHH > 2.1) used for extracting outliers. Two loci with the highest F_{ST} are highlighted by a shaded green column on chromosome 15 and 29.

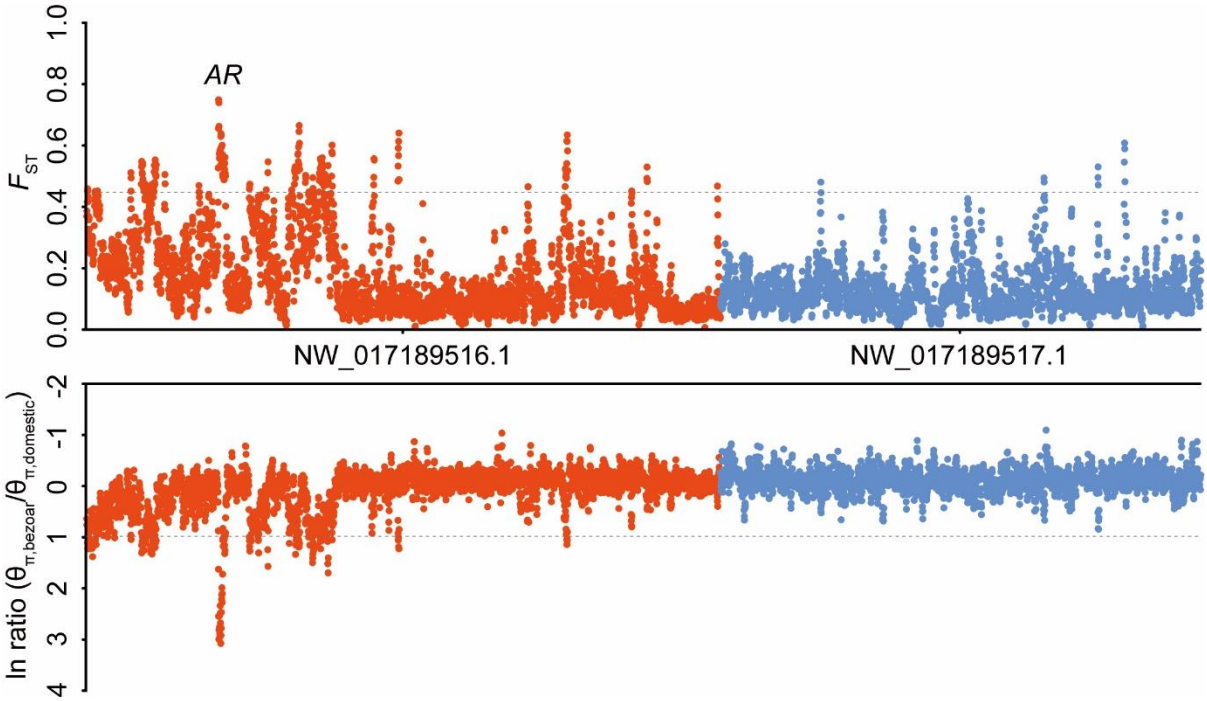


Fig. S20.
Distribution of F_{ST} and ln (θπ ratio) between bezoars and domestic goats across chromosome X. Dashed horizontal lines show F_{ST} > 0.448 and ln (θπ ratio) > 0.981, respectively. Genes residing in the top one window are indicated by their symbols.

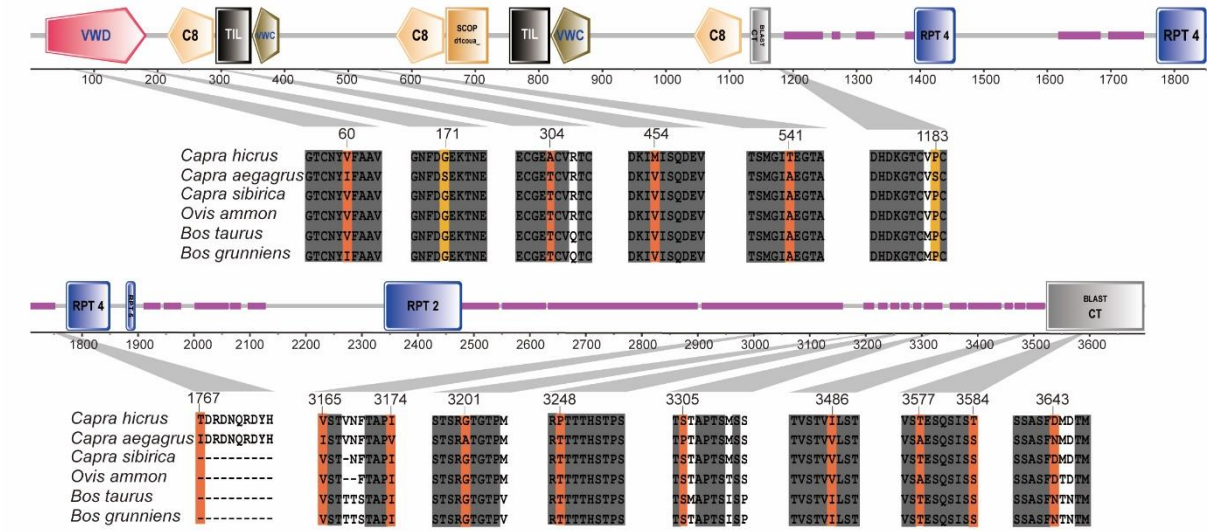


Fig. S21.
Schematic representations of the domain architecture of the MUC6 protein produced using SMART (*117*). Purple lines show low-complexity segments, and regions of protein without any predicted features are marked with gray bars. The lower panel displays the amino acid sequence alignment of MUC6. The orange and yellow shadows highlight the 16 highly differentiated missense mutations in the bezoar-domestic comparison. The yellow shadows indicate two deleterious mutations discovered by Ensembl VEP tool.

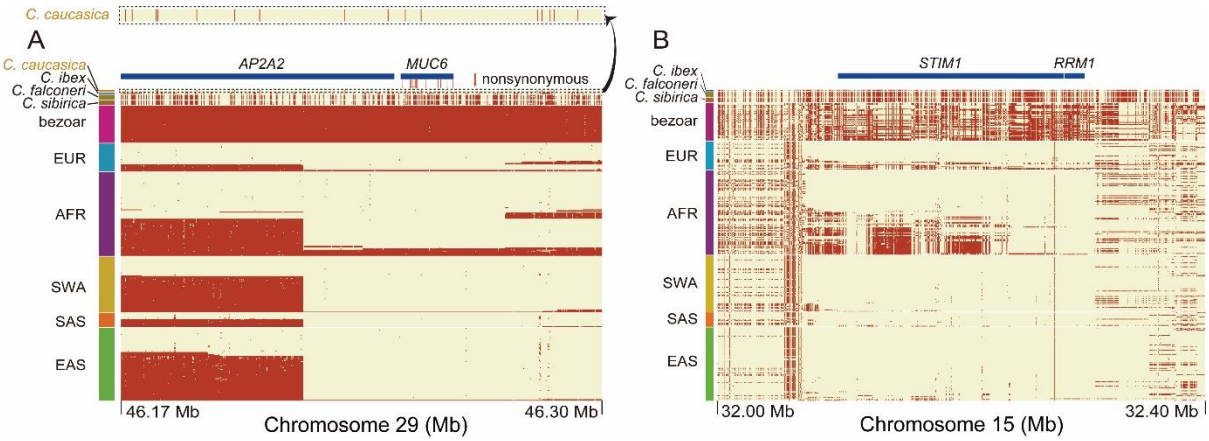


Fig. S22.

Characterization of the genomic regions exhibited strong selective sweep signals. (**A**) Haplotype pattern at the *MUC6* locus defined by putatively introgressed variants (yellow, predicted introgressed allele; red, predicted non-introgressed allele). (**B**) The degree of haplotype sharing across the goat population at the *STIM1-RRM1* locus. The reference/alternative allele is indicated in light yellow/red.

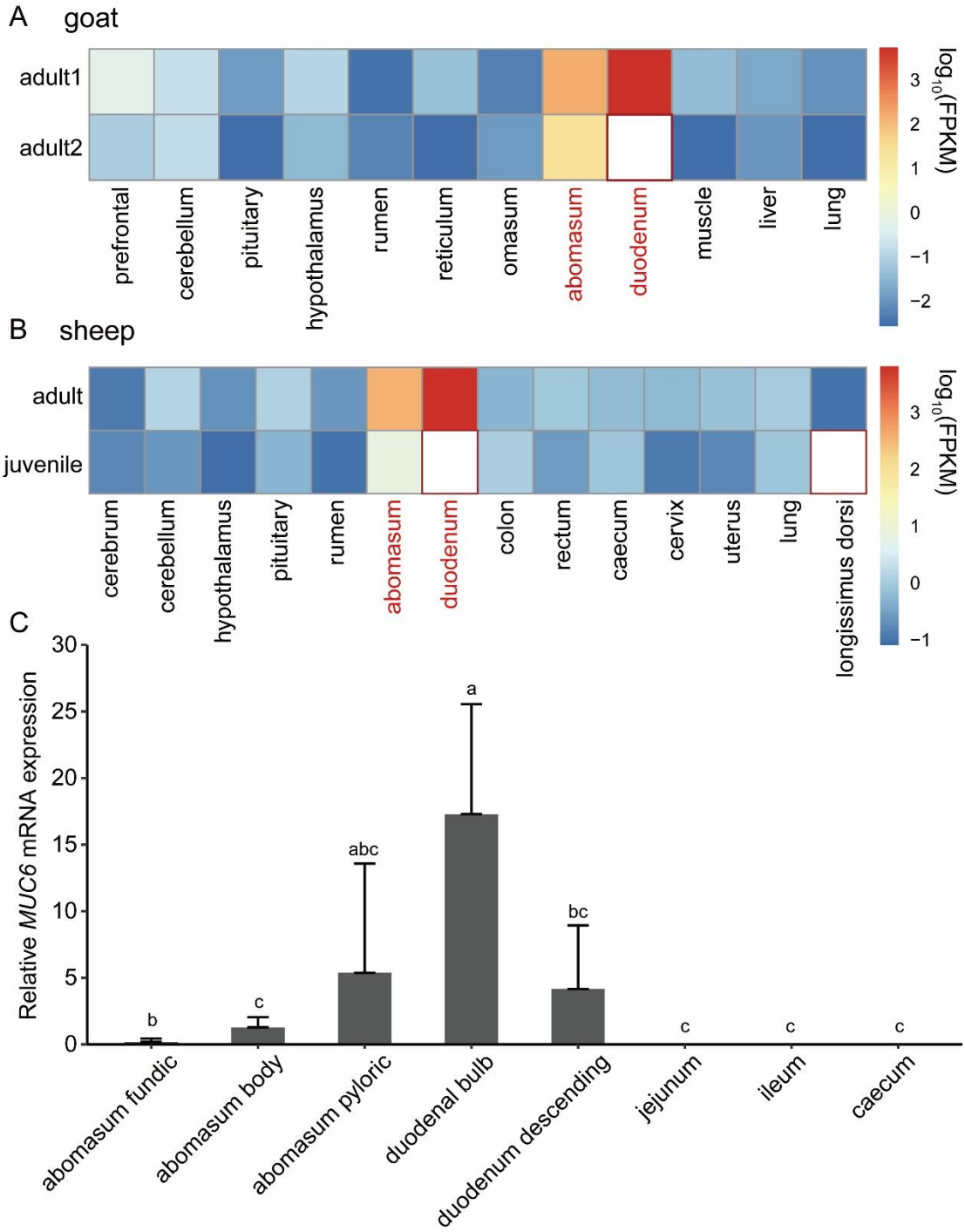


Fig. S23.

Expression analysis of *MUC6* gene. (**A** and **B**) *MUC6* is expressed across different tissues and is specific and highly expressed in the duodenum and abomasum in goat (**A**) and sheep (**B**) (*118*). Cells surrounded by a red frame indicate missing samples. (**C**) q-PCR analysis of goat *MUC6* expression in the gastrointestinal tract. Error bars represent the standard deviations (n = 3). Different letters represent statistically significant differences between tissues (One Way ANOVA post-hoc tests (Dunnett's T3), $P < 0.05$).

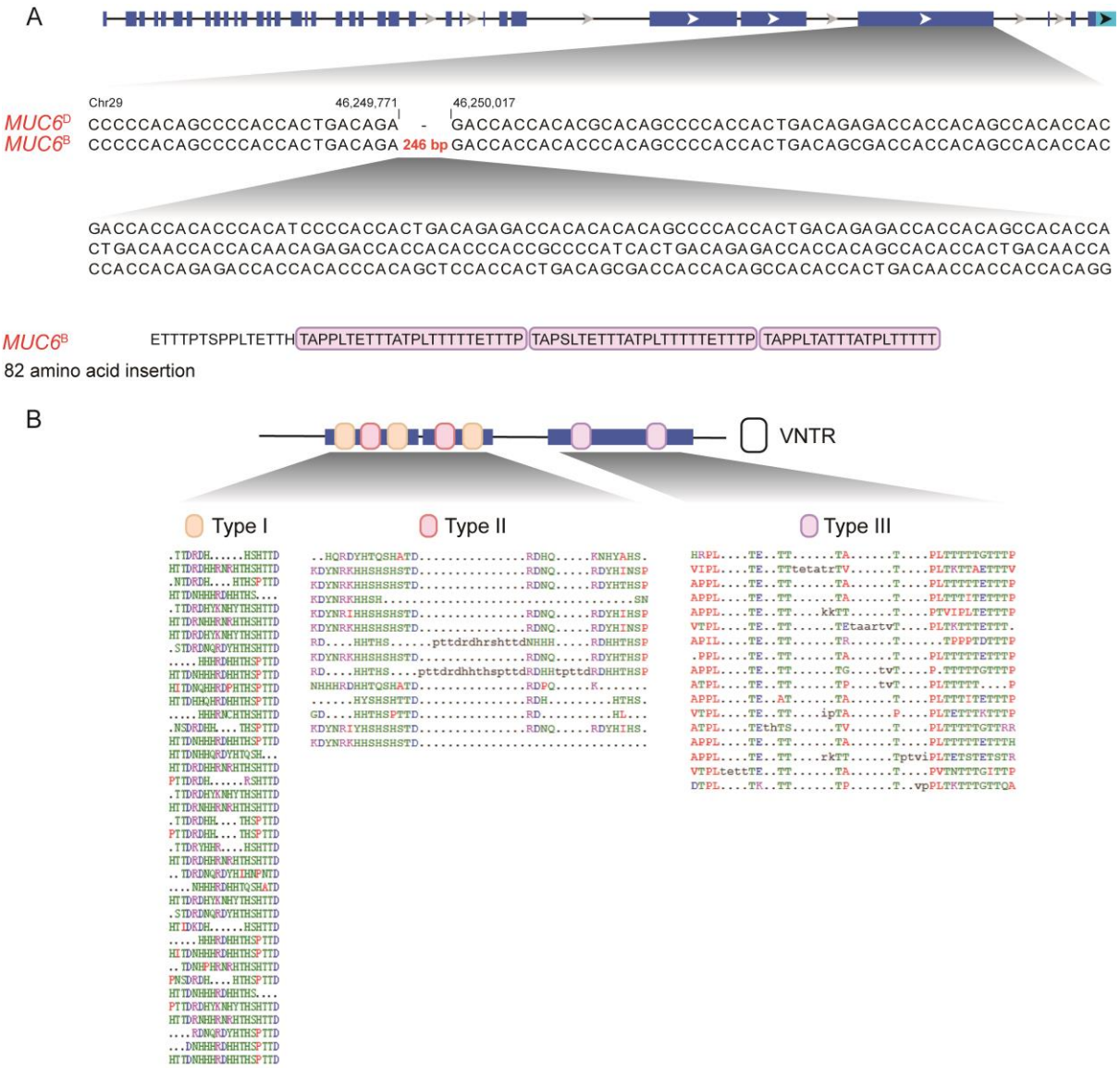


Fig. S24.

Gene structure of *MUC6* gene. (**A**) We found a 246 bp deletion in the *MUC6*^{D} compared to *MUC6*^{B} which located at the 32nd exon in chromosome 29:46,249,771-46,250,017. (**B**) We predicted three distinct repeat units with different copies in the *MUC6*^{D} haplotype. This deletion encodes 82 amino acid which contains three copies of type III VNTR unit.

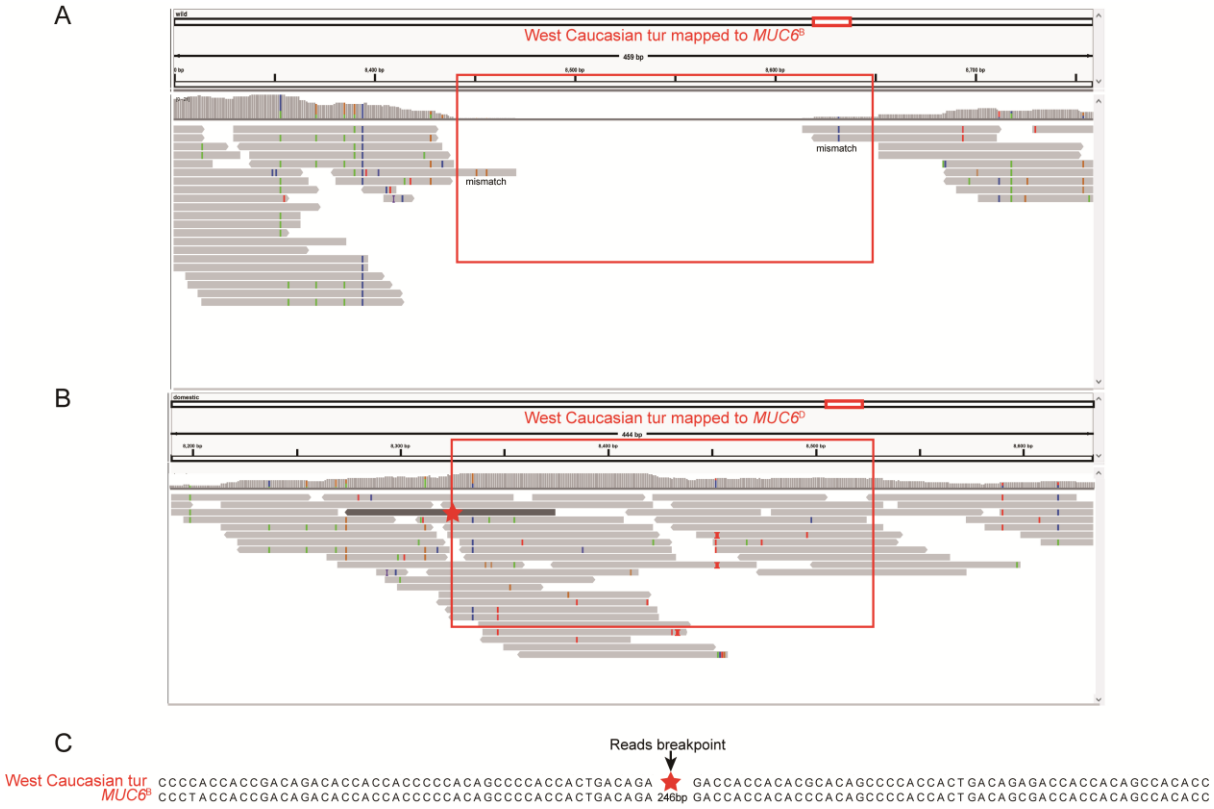


Fig. S25.

The 246 bp deletion of *MUC6*^{D} in West Caucasian tur. IGV snapshot showing West Caucasian tur reads mapped to (**A**) *MUC6*^{B} mRNA sequence with no reads mapping to the 246 bp deletion except three reads with mismatches (**B**) *MUC6*^{D} mRNA sequence and the reads with no mismatch could across the boundary of the deletion was filled dark grey. The red rectangular circled the region of the 246 bp deletion. (**C**) We further checked the West Caucasian tur reads by aligning it to *MUC6*^{B} mRNA sequence by MEGA6 and the red star pointed the position of the breakpoint.

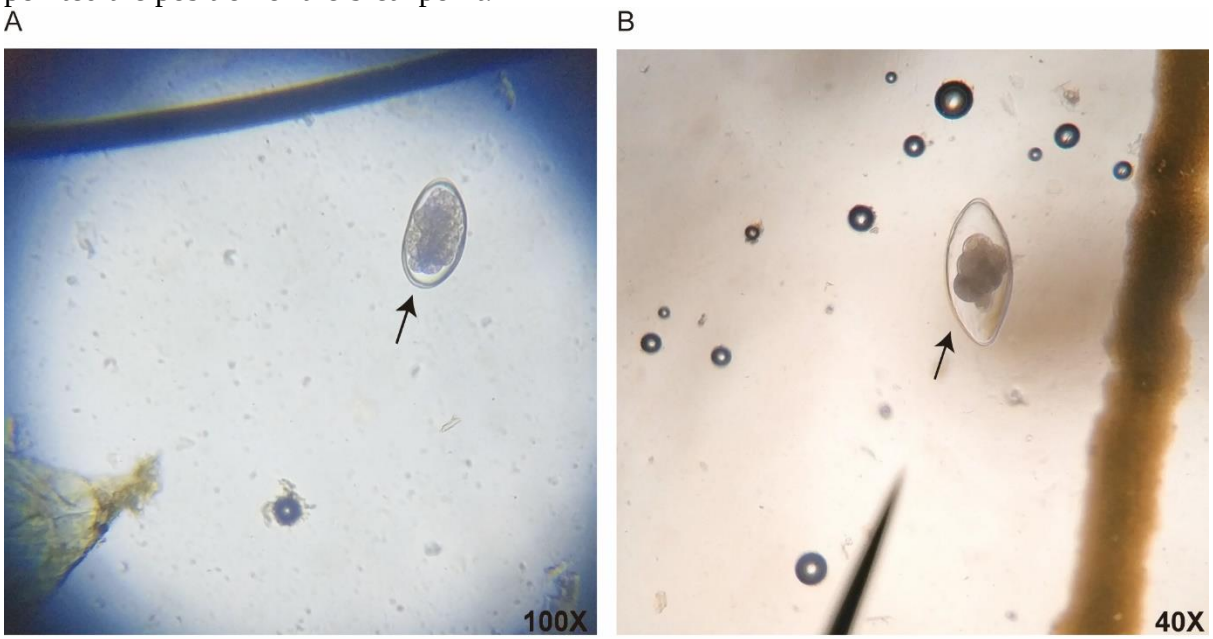


Fig. S26.

Common gastrointestinal nematode eggs in the trial. (**A**) *Hemonchus contortus* and (**B**) *Nematodirus sp*.

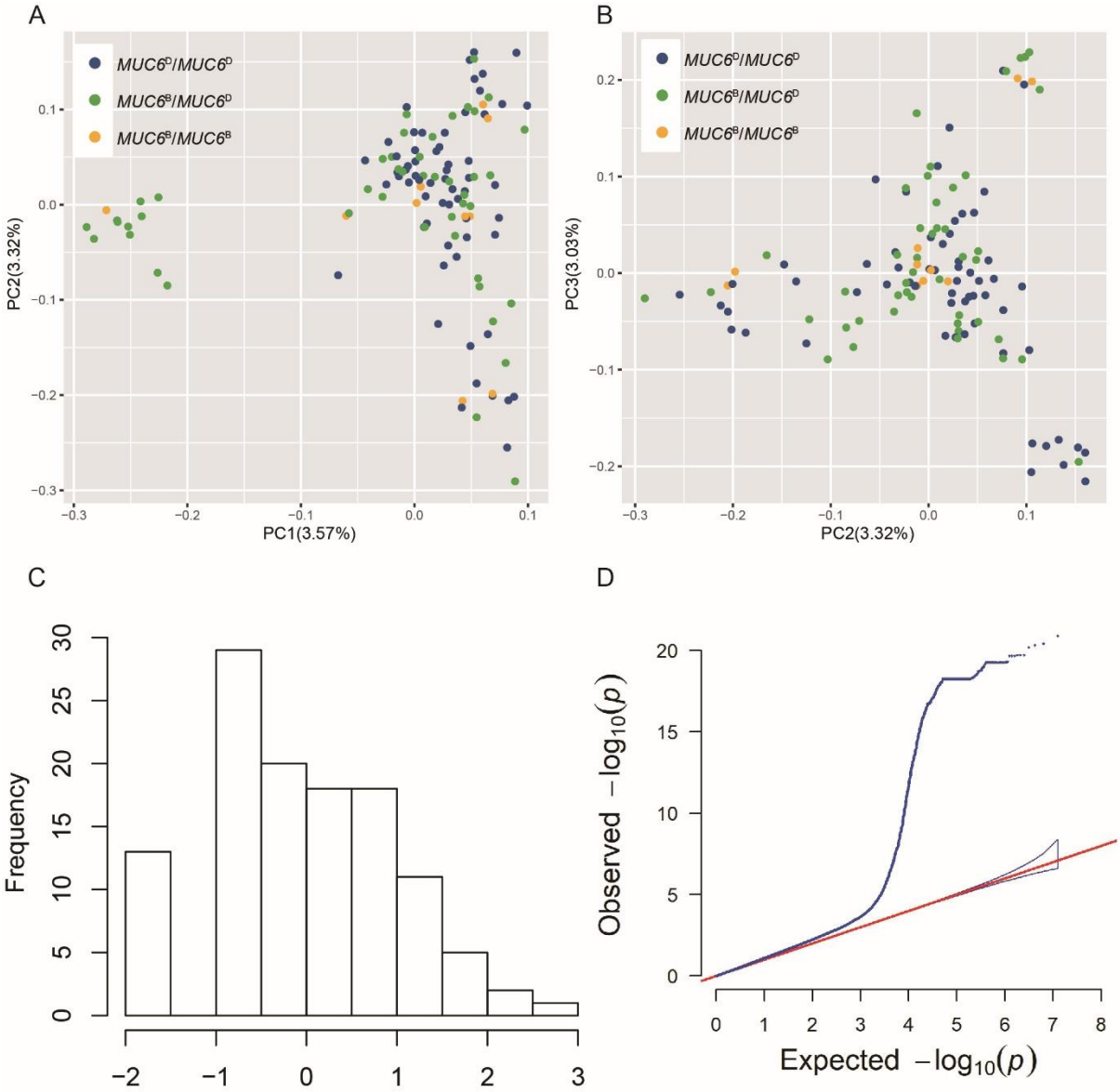


Fig. S27.

Genome wide association study for fecal egg counts. (**A** and **B**) Principal component analysis used for genome wide association data. Colors of dots indicate the genotype at *MUC6* locus. (**C**) The distribution of rank-based transformed fecal egg counts. (**D**) Quantile-quantile plot for fecal egg counts. The 95% confidence interval is shaded in blue.

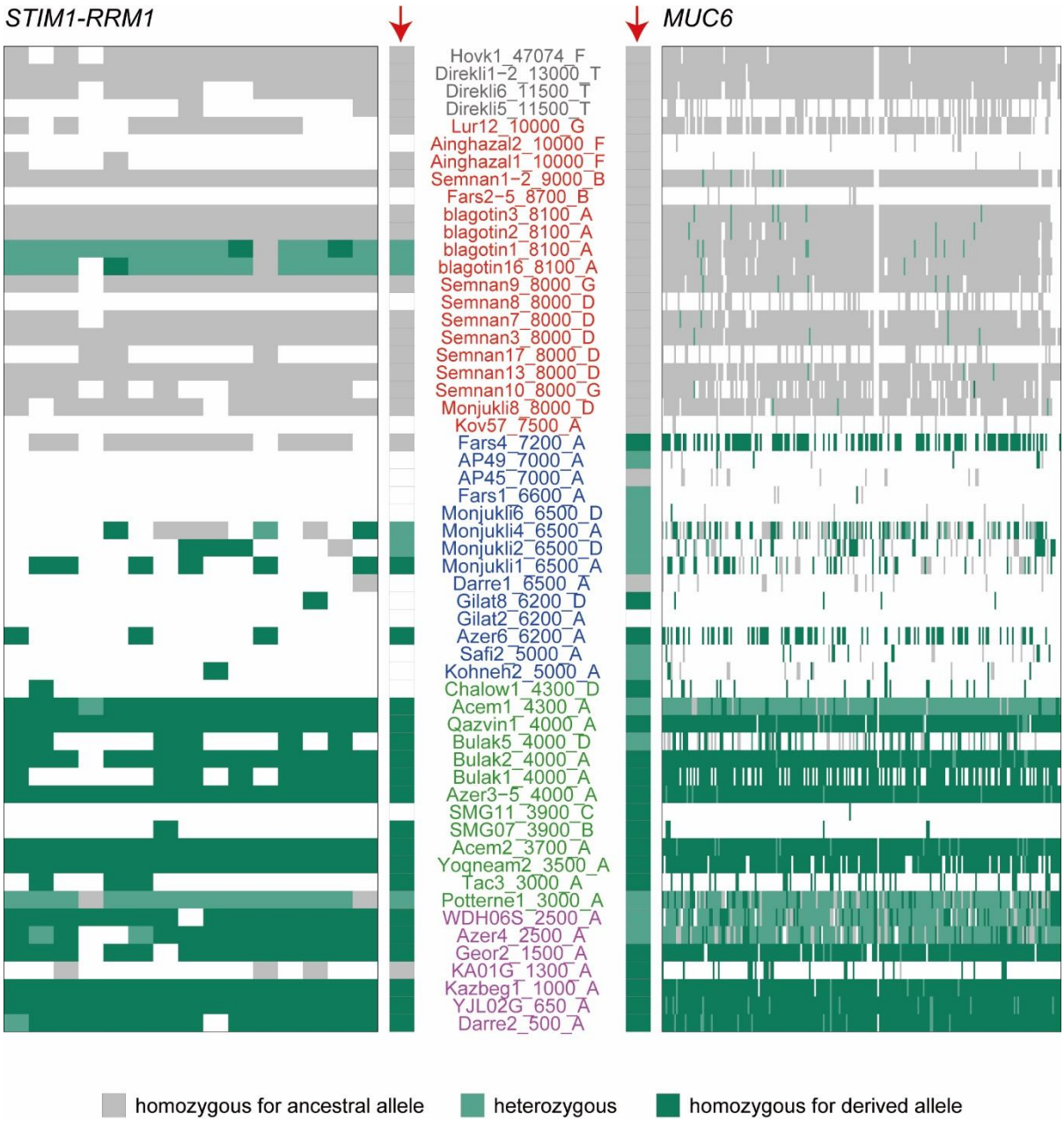


Fig. S28.

Genotyping information at the *STIM1-RRM1* and *MUC6* loci for the ancient goats. 15 SNPs within *STIM1-RRM1* locus and 228 SNPs within *MUC6* locus were used (Data file S3). The red arrows indicate the "genotype" for each individual. The presence of homozygosity and heterozygosity is shown in green and intermediate green, respectively. The absence of the derived allele is depicted in gray. Non-genotyped positions or individuals are indicated in white. The text in the middle indicates sample name, approximate age (years ago) and mitochondrial haplogroup, and is marked by colors which represent different periods from Paleolithic to Medieval contexts.

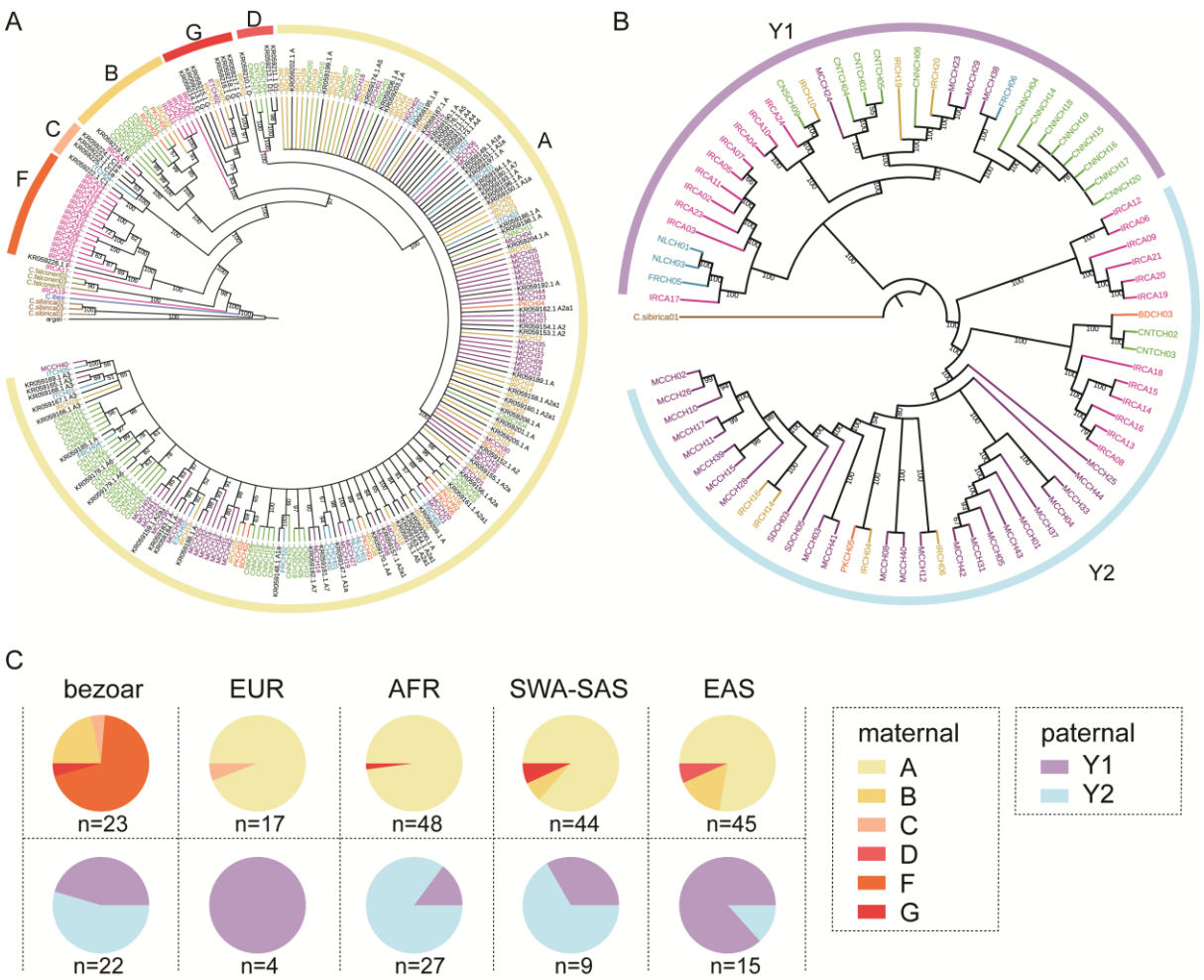


Fig. S29.

Mitochondrial genome and Y-chromosome haplotype analysis. (**A**) Maximum likelihood phylogeny of the mitochondrial genomes. The majority of bezoars samples fell in haplogroup F. Sample IRCA19 fell into a more divergent clade showing similarity to the West Caucasian tur (*Capra caucasica*). (**B**) Maximum likelihood phylogeny of the Y-chromosome. Node labels show bootstrap support values. Population labels are color-coded as in Fig. 1. Labels without an accession number refer to samples from this study, and each node is associated with a color strip representing different haplogroups. (**C**) Piechart plot representing the proportion of mitogenome and Y chromosome haplogroups in bezoars and different domestic goat populations.

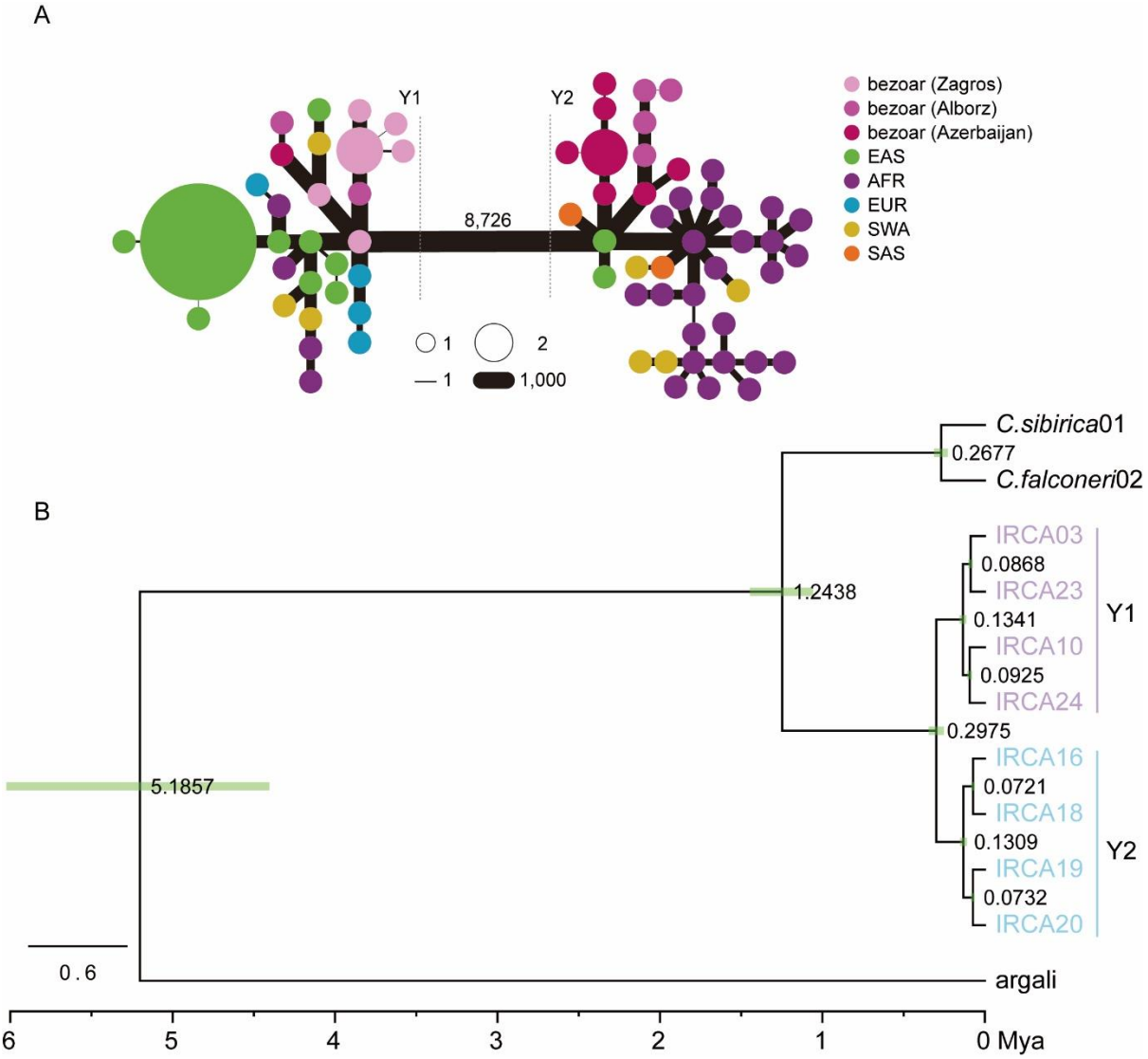


Fig. S30.

Network and time-calibrated Bayesian phylogeny of Y-chromosome haplogroups. (**A**) Y-chromosome haplotype network. The width of the edges is proportional to the number of pairwise differences between the joined haplotypes. (**B**) Time-calibrated Bayesian phylogeny obtained with BEAST. The number indicates estimated divergence time for selected clades including 95% confidence intervals.

Fig. S31.
Y-chromosome haplogroups in ancient samples. Sites with a sequencing coverage lower than 2× are represented in white.

Table S1.

Summary statistics of the modern individuals aligned in this study. Coverage is given relative to the goat reference nuclear genome. NHOM=Number of homozygous non-reference sites; NHET=Number of heterozygous.

Sample ID	accession ID	Coverage (×)	gender	genetic group	Sampling country	Autosome			Chromosome X			Subsample[1]
						NSNP	NHOM	NHET	NSNP	NHOM	NHET	
CNNCH01	this study	12.80	female	EAS	China	7,764,016	2,592,294	5,171,722	236,457	78,399	158,058	
CNNCH02	this study	12.99	female	EAS	China	7,633,182	2,921,719	4,711,463	224,268	92,948	131,320	yes
CNNCH03	this study	10.62	female	EAS	China	7,625,534	2,788,236	4,837,298	227,812	84,820	142,992	yes
CNNCH04	this study	11.76	male	EAS	China	7,597,010	2,859,356	4,737,654	169,512	135,410	34,102	yes
CNNCH05	this study	10.87	female	EAS	China	7,639,388	2,821,874	4,817,514	226,594	87,568	139,026	yes
CNNCH06	this study	9.52	male	EAS	China	7,167,138	2,916,261	4,250,877	158,723	124,843	33,880	yes
CNNCH07	this study	10.63	female	EAS	China	7,351,677	2,831,260	4,520,417	219,685	85,227	134,458	yes
CNNCH08	this study	10.99	female	EAS	China	7,432,754	2,922,381	4,510,373	218,498	94,833	123,665	yes
CNNCH09	this study	9.21	female	EAS	China	7,280,345	2,729,919	4,550,426	219,608	82,650	136,958	yes
CNNCH10	this study	10.94	female	EAS	China	7,531,516	2,846,607	4,684,909	223,482	84,339	139,143	yes
CNNCH11	this study	11.77	female	EAS	China	6,682,673	3,758,226	2,924,447	178,368	137,321	41,047	
CNNCH12	this study	13.04	female	EAS	China	7,725,920	2,689,658	5,036,262	229,732	80,316	149,416	yes
CNNCH13	this study	10.09	female	EAS	China	7,234,063	2,707,696	4,526,367	212,174	82,297	129,877	yes
CNNCH14	this study	12.33	male	EAS	China	7,698,821	2,885,255	4,813,566	174,514	144,490	30,024	
CNNCH15	this study	10.88	male	EAS	China	7,429,440	3,137,526	4,291,914	168,441	138,605	29,836	
CNNCH16	this study	11.23	male	EAS	China	7,566,471	2,979,773	4,586,698	170,545	141,760	28,785	
CNNCH17	this study	11.25	male	EAS	China	7,407,935	3,189,998	4,217,937	169,576	141,384	28,192	
CNNCH18	this study	11.53	male	EAS	China	7,399,435	3,218,693	4,180,742	165,815	146,096	19,719	
CNNCH19	this study	10.26	male	EAS	China	7,469,922	3,064,650	4,405,272	162,110	142,006	20,104	
CNNCH20	this study	11.49	male	EAS	China	7,486,768	3,128,848	4,357,920	170,558	140,502	30,056	
CNSCH01	SRS309478	43.24	female	EAS	China	7,406,216	3,292,623	4,113,593	222,281	102,540	119,741	yes
CNSCH02	this study	12.90	female	EAS	China	6,855,717	3,685,390	3,170,327	185,274	131,261	54,013	yes
CNSCH03	this study	15.77	female	EAS	China	7,311,824	2,667,917	4,643,907	211,038	90,897	120,141	
CNSCH04	this study	14.85	female	EAS	China	7,355,257	2,835,384	4,519,873	215,726	92,804	122,922	yes
CNSCH05	this study	11.96	female	EAS	China	7,545,844	2,921,086	4,624,758	219,664	100,973	118,691	
CNSCH06	this study	12.85	female	EAS	China	7,295,432	3,177,334	4,118,098	157,952	144,788	13,164	yes

CNSCH07	this study	12.32	female	EAS	China	7,583,199	2,855,644	4,727,555	223,086	88,314	134,772	
CNSCH08	this study	12.69	female	EAS	China	7,171,300	3,390,615	3,780,685	215,901	107,527	108,374	yes
CNSCH09	this study	8.15	male	EAS	China	6,992,731	2,825,075	4,167,656	150,419	120,327	30,092	
CNSCH10	this study	12.80	female	EAS	China	7,018,716	3,578,536	3,440,180	220,172	104,436	115,736	yes
CNSCH11	this study	10.66	female	EAS	China	7,141,271	3,154,724	3,986,547	211,397	100,330	111,067	
CNSCH12	this study	12.86	female	EAS	China	7,387,635	3,084,585	4,303,050	222,342	97,629	124,713	
CNSCH13	this study	8.71	female	EAS	China	6,693,213	3,432,988	3,260,225	211,560	94,682	116,878	
CNSCH14	this study	18.68	female	EAS	China	7,222,508	3,463,890	3,758,618	214,114	113,753	100,361	yes
CNSCH15	this study	25.56	female	EAS	China	7,379,823	3,340,508	4,039,315	223,474	102,434	121,040	yes
CNSCH16	this study	12.15	female	EAS	China	7,607,329	2,868,735	4,738,594	227,449	94,752	132,697	
CNSCH17	this study	13.46	female	EAS	China	7,124,274	3,467,064	3,657,210	215,088	109,045	106,043	yes
CNSCH18	this study	10.86	female	EAS	China	6,796,583	3,585,190	3,211,393	216,465	101,635	114,830	yes
CNSCH19	this study	17.75	female	EAS	China	7,114,111	3,469,349	3,644,762	220,593	96,445	124,148	yes
CNTCH01	this study	9.02	male	EAS	China	7,488,932	2,977,153	4,511,779	164,297	135,973	28,324	
CNTCH02	this study	9.33	male	EAS	China	7,526,938	2,968,359	4,558,579	166,015	138,408	27,607	
CNTCH03	this study	9.50	male	EAS	China	7,516,524	2,999,015	4,517,509	168,203	139,711	28,492	
CNTCH04	this study	10.19	male	EAS	China	7,572,738	2,965,445	4,607,293	170,042	139,262	30,780	
CNTCH05	this study	10.05	male	EAS	China	7,517,364	3,044,842	4,472,522	171,100	142,296	28,804	
CNTCH06	this study	11.84	female	EAS	China	7,655,700	2,963,002	4,692,698	227,935	88,378	139,557	
CNTCH07	this study	12.52	female	EAS	China	7,399,693	3,018,908	4,380,785	218,883	87,925	130,958	
CNTCH08	this study	11.23	female	EAS	China	7,572,433	2,981,135	4,591,298	227,416	85,175	142,241	
BDCH01	this study	9.27	female	SAS	Bangladesh	7,048,223	3,019,440	4,028,783	211,016	93,155	117,861	yes
BDCH02	this study	9.21	female	SAS	Bangladesh	7,094,294	3,075,834	4,018,460	213,204	93,367	119,837	yes
BDCH03	this study	9.91	male	SAS	Bangladesh	6,843,129	3,181,176	3,661,953	150,140	127,107	23,033	yes
PKCH01	this study	13.29	female	SAS	Pakistan	7,744,734	2,611,474	5,133,260	229,497	81,740	147,757	yes
PKCH02	this study	11.53	female	SAS	Pakistan	7,532,455	2,921,097	4,611,358	227,448	89,433	138,015	yes
PKCH03	this study	13.07	female	SAS	Pakistan	6,911,086	3,616,423	3,294,663	175,851	138,827	37,024	yes
PKCH04	this study	12.46	female	SAS	Pakistan	7,327,290	3,248,816	4,078,474	211,482	104,247	107,235	yes
PKCH05	this study	12.10	male	SAS	Pakistan	7,622,119	2,879,240	4,742,879	167,518	140,575	26,943	yes
PKCH06	this study	7.76	female	SAS	Pakistan	5,631,741	2,443,470	3,188,271	159,401	72,262	87,139	yes
IRCH01	ERR297229	12.80	female	SWA	Iran	7,085,338	3,349,431	3,735,907	187,072	115,644	71,428	
IRCH02	ERR299449	12.81	female	SWA	Iran	7,382,642	3,034,086	4,348,556	231,838	78,578	153,260	

IRCH03	ERR299456	12.63	female	SWA	Iran	7,713,126	2,682,648	5,030,478	230,925	78,307	152,618	
IRCH04	ERR313197	11.85	male	SWA	Iran	7,669,646	2,833,878	4,835,768	170,054	142,453	27,601	yes
IRCH05	ERR313198	11.76	female	SWA	Iran	7,733,032	2,671,560	5,061,472	226,549	77,249	149,300	
IRCH06	ERR313199	13.44	male	SWA	Iran	7,612,216	2,803,962	4,808,254	169,941	138,908	31,033	
IRCH07	ERR313200	13.49	female	SWA	Iran	7,799,722	2,674,415	5,125,307	228,371	79,456	148,915	
IRCH08	ERR313202	12.85	female	SWA	Iran	7,728,454	2,684,498	5,043,956	225,140	86,193	138,947	
IRCH09	ERR313204	12.71	female	SWA	Iran	7,479,543	2,932,865	4,546,678	201,739	103,995	97,744	
IRCH10	ERR313206	11.77	male	SWA	Iran	7,685,202	2,791,637	4,893,565	167,449	138,805	28,644	yes
IRCH11	ERR313207	13.24	female	SWA	Iran	7,687,335	2,765,565	4,921,770	229,761	81,089	148,672	
IRCH12	ERR313209	11.78	female	SWA	Iran	7,735,407	2,690,359	5,045,048	229,071	81,050	148,021	
IRCH13	ERR313210	12.98	female	SWA	Iran	7,449,835	3,011,097	4,438,738	209,013	98,571	110,442	
IRCH14	ERR313211	12.72	male	SWA	Iran	7,006,847	3,461,821	3,545,026	170,322	140,058	30,264	
IRCH15	ERR313212	12.80	female	SWA	Iran	7,751,422	2,678,828	5,072,594	228,649	79,704	148,945	
IRCH16	ERR313213	12.88	male	SWA	Iran	7,582,483	2,877,784	4,704,699	166,634	137,339	29,295	
IRCH17	ERR313215	12.41	female	SWA	Iran	7,636,896	2,780,865	4,856,031	230,648	78,694	151,954	
IRCH18	ERR340332	11.21	female	SWA	Iran	7,674,970	2,708,378	4,966,592	230,980	80,804	150,176	
IRCH19	ERR340337	10.86	male	SWA	Iran	7,643,341	2,724,586	4,918,755	162,479	134,161	28,318	
IRCH20	ERR340339	11.35	male	SWA	Iran	7,683,661	2,693,103	4,990,558	163,826	135,637	28,189	
IRCH21	this study	13.25	female	SWA	Iran	7,582,163	2,763,983	4,818,180	227,004	80,323	146,681	yes
IRCH23	this study	11.82	female	SWA	Iran	7,429,074	2,742,284	4,686,790	216,751	80,769	135,982	
IRCH24	this study	10.26	female	SWA	Iran	7,298,436	2,661,151	4,637,285	212,424	78,808	133,616	
IRCH25	this study	8.68	female	SWA	Iran	7,273,403	2,794,216	4,479,187	218,824	88,775	130,049	yes
IRCH26	this study	11.83	female	SWA	Iran	7,653,126	2,684,076	4,969,050	227,049	76,290	150,759	
IRCH27	this study	15.18	female	SWA	Iran	7,724,579	2,722,809	5,001,770	231,210	78,404	152,806	
IRCH28	this study	12.44	female	SWA	Iran	7,558,682	2,833,825	4,724,857	226,357	82,818	143,539	yes
IRCH29	this study	16.00	female	SWA	Iran	7,831,222	2,651,724	5,179,498	237,375	80,087	157,288	
IRCH30	this study	9.74	female	SWA	Iran	7,520,621	2,708,712	4,811,909	225,645	79,516	146,129	
IRCH31	this study	10.87	female	SWA	Iran	7,489,457	2,904,654	4,584,803	222,132	87,650	134,482	yes
IRCH32	this study	13.51	female	SWA	Iran	7,565,674	2,875,569	4,690,105	226,795	82,456	144,339	yes
IRCH34	this study	7.26	female	SWA	Iran	7,071,485	3,012,863	4,058,622	207,788	92,943	114,845	yes
IRCH36	this study	10.49	female	SWA	Iran	7,241,016	2,914,343	4,326,673	219,902	76,400	143,502	yes
IRCH37	this study	9.12	female	SWA	Iran	7,537,910	2,794,331	4,743,579	225,722	83,738	141,984	yes

IRCH38	this study	6.23	female	SWA	Iran	5,915,690	2,988,244	2,927,446	176,115	84,511	91,604	yes
IRCH39	this study	9.74	female	SWA	Iran	7,406,295	2,791,730	4,614,565	220,524	80,413	140,111	
ETCH01	this study	11.03	female	AFR	Ethiopia	7,257,110	2,854,724	4,402,386	207,348	82,716	124,632	
ETCH02	this study	9.74	female	AFR	Ethiopia	7,233,827	2,792,007	4,441,820	208,017	81,484	126,533	
MCCH01	ERR219543	12.67	male	AFR	Morocco	7,346,216	2,580,059	4,766,157	152,967	125,364	27,603	
MCCH02	ERR219546	13.48	male	AFR	Morocco	7,023,881	2,960,878	4,063,003	151,635	123,250	28,385	yes
MCCH03	ERR229471	13.22	male	AFR	Morocco	7,077,277	2,888,355	4,188,922	153,016	127,498	25,518	yes
MCCH04	ERR229474	12.72	male	AFR	Morocco	6,818,630	3,083,415	3,735,215	139,624	121,388	18,236	
MCCH05	ERR229476	13.35	male	AFR	Morocco	7,076,086	2,845,637	4,230,449	148,866	117,333	31,533	yes
MCCH06	ERR229478	13.51	female	AFR	Morocco	7,339,483	2,634,557	4,704,926	200,082	79,672	120,410	yes
MCCH07	ERR229479	12.03	female	AFR	Morocco	7,278,077	2,657,545	4,620,532	202,619	72,268	130,351	yes
MCCH08	ERR229481	13.50	male	AFR	Morocco	7,347,826	2,636,145	4,711,681	154,208	126,885	27,323	yes
MCCH09	ERR229484	15.05	female	AFR	Morocco	7,414,975	2,630,026	4,784,949	211,071	74,063	137,008	
MCCH10	ERR229485	12.84	male	AFR	Morocco	7,283,645	2,616,435	4,667,210	156,569	129,826	26,743	yes
MCCH11	ERR229487	13.14	male	AFR	Morocco	7,338,072	2,652,133	4,685,939	156,425	126,494	29,931	yes
MCCH12	ERR232492	12.26	male	AFR	Morocco	7,309,614	2,592,036	4,717,578	151,780	123,707	28,073	
MCCH13	ERR234304	12.15	female	AFR	Morocco	7,024,848	2,927,486	4,097,362	210,797	72,770	138,027	
MCCH14	ERR234305	13.44	female	AFR	Morocco	6,930,851	3,013,168	3,917,683	198,600	73,179	125,421	
MCCH15	ERR234315	13.55	male	AFR	Morocco	7,324,580	2,657,714	4,666,866	155,081	127,176	27,905	
MCCH16	ERR246143	12.20	female	AFR	Morocco	7,296,198	2,669,730	4,626,468	208,158	72,158	136,000	
MCCH17	ERR246152	13.76	male	AFR	Morocco	6,487,369	3,585,512	2,901,857	158,680	130,304	28,376	
MCCH18	ERR246153	14.63	female	AFR	Morocco	7,203,867	2,749,227	4,454,640	197,786	86,568	111,218	yes
MCCH19	ERR248926	13.47	female	AFR	Morocco	7,281,347	2,657,817	4,623,530	202,357	74,339	128,018	
MCCH20	ERR248928	14.55	female	AFR	Morocco	7,364,541	2,612,649	4,751,892	211,925	73,826	138,099	
MCCH21	ERR248929	13.63	female	AFR	Morocco	7,309,797	2,645,795	4,664,002	205,985	72,631	133,354	
MCCH22	ERR248933	13.49	female	AFR	Morocco	7,217,658	2,731,835	4,485,823	205,630	76,846	128,784	yes
MCCH23	ERR299283	13.62	male	AFR	Morocco	7,286,846	2,732,274	4,554,572	158,426	130,303	28,123	
MCCH24	ERR313257	12.23	male	AFR	Morocco	7,324,295	2,621,591	4,702,704	153,217	125,250	27,967	
MCCH25	ERR313258	13.21	male	AFR	Morocco	7,173,855	2,791,622	4,382,233	153,676	123,103	30,573	
MCCH26	ERR313259	15.06	male	AFR	Morocco	7,396,448	2,648,724	4,747,724	151,601	123,745	27,856	
MCCH27	ERR313264	15.22	female	AFR	Morocco	7,466,651	2,584,261	4,882,390	209,282	71,923	137,359	
MCCH28	ERR313266	12.35	male	AFR	Morocco	6,663,936	3,352,994	3,310,942	152,683	125,103	27,580	

MCCH29	ERR313272	13.44	male	AFR	Morocco	7,380,812	2,621,383	4,759,429	151,146	122,201	28,945	
MCCH30	ERR315498	13.26	female	AFR	Morocco	7,407,369	2,563,753	4,843,616	213,265	73,327	139,938	
MCCH31	ERR315500	13.25	male	AFR	Morocco	7,328,719	2,639,615	4,689,104	154,432	126,267	28,165	
MCCH32	ERR315503	15.76	female	AFR	Morocco	7,279,280	2,720,081	4,559,199	197,414	89,171	108,243	yes
MCCH33	ERR315505	13.49	male	AFR	Morocco	7,363,209	2,604,940	4,758,269	147,725	121,037	26,688	
MCCH34	ERR315508	14.21	female	AFR	Morocco	7,229,463	2,705,212	4,524,251	212,255	73,330	138,925	
MCCH35	ERR315510	12.67	female	AFR	Morocco	7,309,978	2,656,436	4,653,542	207,506	72,793	134,713	
MCCH36	ERR315512	12.38	female	AFR	Morocco	7,329,860	2,658,346	4,671,514	202,202	71,180	131,022	
MCCH37	ERR315515	11.89	male	AFR	Morocco	7,290,004	2,643,202	4,646,802	154,258	126,193	28,065	
MCCH38	ERR315516	14.42	male	AFR	Morocco	6,630,928	3,308,863	3,322,065	151,202	122,764	28,438	
MCCH39	ERR315795	11.93	male	AFR	Morocco	7,256,439	2,682,006	4,574,433	154,540	127,187	27,353	
MCCH40	ERR318768	12.88	male	AFR	Morocco	6,479,974	3,445,106	3,034,868	150,599	117,504	33,095	
MCCH41	ERR332581	11.86	male	AFR	Morocco	7,344,370	2,560,213	4,784,157	153,840	117,216	36,624	
MCCH42	ERR332592	11.86	male	AFR	Morocco	7,262,471	2,646,847	4,615,624	147,913	124,022	23,891	
MCCH43	ERR340428	12.89	male	AFR	Morocco	7,193,037	2,751,892	4,441,145	149,852	122,516	27,336	
MCCH44	ERR345973	11.53	male	AFR	Morocco	7,331,815	2,594,560	4,737,255	152,135	124,996	27,139	
NGCH01	this study	8.23	female	AFR	Nigeria	6,397,757	2,835,779	3,561,978	179,388	78,347	101,041	
NGCH02	this study	10.84	female	AFR	Nigeria	6,501,718	3,034,356	3,467,362	184,740	85,183	99,557	
SDCH01	this study	9.47	female	AFR	Sudan	7,183,958	2,853,722	4,330,236	209,741	82,038	127,703	
SDCH02	this study	10.18	female	AFR	Sudan	7,219,253	2,869,335	4,349,918	208,313	82,487	125,826	
SDCH03	this study	8.86	male	AFR	Sudan	7,079,349	2,855,975	4,223,374	154,350	129,343	25,007	
SDCH04	this study	9.84	female	AFR	Sudan	7,087,863	2,832,044	4,255,819	206,826	82,311	124,515	
SDCH05	this study	9.68	male	AFR	Sudan	6,829,064	3,069,343	3,759,721	147,901	124,403	23,498	
SDCH06	this study	9.69	female	AFR	Sudan	6,944,596	2,904,455	4,040,141	196,871	86,089	110,782	
CHCH01	SRR3144625	12.69	female	EUR	Swiss	6,516,499	3,622,050	2,894,449	201,658	72,561	129,097	
CHCH02	SRX1560780	11.59	female	EUR	Swiss	7,102,163	2,950,380	4,151,783	203,984	76,111	127,873	
ESCH01	this study	13.80	female	EUR	Spain	7,094,439	2,722,987	4,371,452	202,446	72,881	129,565	
ESCH02	this study	10.44	female	EUR	Spain	6,776,393	2,862,886	3,913,507	186,849	88,804	98,045	
FRCH01	ERR470101	13.45	female	EUR	France	7,263,557	2,875,521	4,388,036	197,043	68,559	128,484	yes
FRCH02	ERR470102	13.25	female	EUR	France	7,305,827	2,847,375	4,458,452	203,195	68,196	134,999	yes
FRCH03	ERR470103	14.23	female	EUR	France	7,277,129	2,885,918	4,391,211	194,402	76,455	117,947	yes
FRCH04	ERR470105	13.89	female	EUR	France	7,208,321	2,959,577	4,248,744	201,640	67,864	133,776	yes

FRCH05	this study	13.83	male	EUR	France	7,280,889	2,812,689	4,468,200	145,578	114,250	31,328	yes
FRCH06	this study	14.40	male	EUR	France	7,269,032	2,734,792	4,534,240	143,886	113,166	30,720	
ITCH01	ERR405774	14.11	female	EUR	Italy	7,279,760	2,864,788	4,414,972	199,622	72,223	127,399	yes
ITCH02	ERR405775	14.42	female	EUR	Italy	7,358,898	2,805,803	4,553,095	199,346	77,143	122,203	yes
ITCH03	ERR405776	13.69	female	EUR	Italy	7,138,373	3,005,874	4,132,499	195,982	77,016	118,966	yes
ITCH04	ERR405777	13.85	female	EUR	Italy	7,266,163	2,867,616	4,398,547	200,698	67,299	133,399	yes
ITCH05	ERR405778	13.67	female	EUR	Italy	7,261,563	2,843,544	4,418,019	192,262	89,497	102,765	yes
NLCH01	this study	13.56	male	EUR	Netherlands	7,021,507	3,046,971	3,974,536	152,960	122,306	30,654	
NLCH02	this study	14.01	female	EUR	Netherlands	7,030,444	3,170,155	3,860,289	203,971	85,443	118,528	
NLCH03	this study	11.11	male	EUR	Netherlands	7,266,780	2,760,307	4,506,473	146,866	117,905	28,961	yes
IRCA01	ERR340328	6.51	female	bezoar (Alborz)	Iran	6,933,969	3,245,765	3,688,204	216,962	109,936	107,026	yes
IRCA02	ERR340329	5.56	male	bezoar (Zagros)	Iran	6,170,163	3,949,216	2,220,947	138,379	123,321	15,058	yes
IRCA03	ERR340330	12.33	male	bezoar (Alborz)	Iran	7,904,428	3,218,364	4,686,064	190,748	164,052	26,696	yes
IRCA04	ERR340331	11.65	male	bezoar (Alborz)	Iran	7,433,499	3,618,012	3,815,487	184,905	158,686	26,219	yes
IRCA05	ERR340333	6.35	male	bezoar (Zagros)	Iran	6,497,702	3,806,648	2,691,054	146,176	127,910	18,266	yes
IRCA06	ERR340334	12.11	male	bezoar (Azerbaijan)	Iran	7,361,515	3,740,976	3,620,539	189,086	163,400	25,686	yes
IRCA07	ERR340335	6.62	male	bezoar (Zagros)	Iran	6,680,833	3,752,362	2,928,471	150,144	131,409	18,735	yes
IRCA08	ERR340336	6.51	male	bezoar (Azerbaijan)	Iran	6,807,849	3,578,533	3,229,316	154,642	136,448	18,194	yes
IRCA09	ERR340338	11.37	male	bezoar (Alborz)	Iran	7,717,451	3,349,567	4,367,884	184,299	158,811	25,488	yes
IRCA10	ERR340340	12.07	male	bezoar (Azerbaijan)	Iran	7,579,000	3,529,477	4,049,523	187,877	162,001	25,876	yes
IRCA11	ERR340341	5.17	male	bezoar (Zagros)	Iran	6,143,265	3,758,863	2,384,402	129,340	113,606	15,734	yes
IRCA12	ERR340342	6.42	male	bezoar (Azerbaijan)	Iran	6,773,004	3,585,467	3,187,537	159,027	141,117	17,910	yes
IRCA13	ERR340343	7.26	male	bezoar (Azerbaijan)	Iran	7,076,423	3,587,532	3,488,891	163,110	141,654	21,456	yes
IRCA14	ERR340344	11.20	male	bezoar (Azerbaijan)	Iran	7,536,358	3,504,944	4,031,414	183,880	158,196	25,684	yes
IRCA15	ERR340345	11.60	male	bezoar (Azerbaijan)	Iran	7,435,301	3,639,107	3,796,194	186,304	163,340	22,964	yes
IRCA16	ERR340347	12.70	male	bezoar (Azerbaijan)	Iran	7,443,507	3,649,340	3,794,167	188,096	161,692	26,404	yes
IRCA17	ERR340348	10.72	male	bezoar (Zagros)	Iran	7,671,169	3,236,902	4,434,267	176,888	149,130	27,758	yes
IRCA18	ERR340426	12.36	male	bezoar (Azerbaijan)	Iran	7,573,668	3,521,243	4,052,425	179,094	162,220	16,874	yes

IRCA19	ERR470100	15.04	male	bezoar (Alborz)	Iran	7,861,964	3,135,552	4,726,412	181,814	153,368	28,446	yes
IRCA20	ERR470104	14.44	male	bezoar (Alborz)	Iran	7,944,135	3,216,458	4,727,677	189,740	160,241	29,499	yes
IRCA21	ERR470106	14.66	male	bezoar (Alborz)	Iran	7,825,865	3,156,458	4,669,407	184,925	155,133	29,792	yes
IRCA22	ERS154870	43.70	female	bezoar (Azerbaijan)	Iran	7,849,370	3,342,918	4,506,452	246,434	114,090	132,344	yes
IRCA23	SRR1576679	16.81	male	bezoar (Zagros)	Iran	7,936,661	3,143,116	4,793,545	181,297	151,139	30,158	yes
IRCA24	this study	11.56	male	bezoar (Zagros)	Iran	7,805,135	3,092,291	4,712,844	175,346	144,459	30,887	yes
NUHY01	this study	7.60	female	hybrid[2]	Sudan	16,650,876	3,699,308	12,951,568	453,422	99,071	354,351	
NUHY02	this study	7.27	male	hybrid[2]	Sudan	16,413,852	3,718,584	12,695,268	139,777	101,594	38,183	
NUHY03	this study	3.88	male	hybrid[2]	Sudan	12,271,503	4,148,725	8,122,778	102,648	79,126	23,522	
NUHY04	this study	4.15	male	hybrid[2]	Sudan	11,905,405	3,571,672	8,333,733	101,433	73,917	27,516	
C.falconeri01	this study	8.94	male	*Capra falconeri*	Central Asia	17,528,518	13,344,279	4,184,239	568,353	539,390	28,963	
C.falconeri02	this study	27.60	male	*Capra falconeri*	Central Asia	18,822,547	13,508,682	5,313,865	655,012	602,579	52,433	
C.falconeri03	this study	10.40	male	*Capra falconeri*	Central Asia	18,332,587	13,487,789	4,844,798	608,669	576,862	31,807	
C.sibirica01	this study	47.43	male	*Capra sibirica*	Central Asia	18,448,986	13,914,991	4,533,995	652,906	614,919	37,987	
C.sibirica02	this study	9.75	female	*Capra sibirica*	Central Asia	17,260,145	13,760,875	3,499,270	667,551	554,102	113,449	
C.sibirica03	this study	12.31	female	*Capra sibirica*	Central Asia	17,564,840	13,451,003	4,113,837	688,586	529,597	158,989	
C.ibex	this study	10.08	female	*Capra ibex*	Europe	14,946,935	14,014,603	932,332	600,948	571,073	29,875	

[1] The samples used in the ADMIXTURE and SMC++ analysis after downsampling, [2] *Capra nubiana* × domestic goat.

Table S2.
Summary statistics of the historical sample.

Sample ID	MNHN Identifier	Species	Skeletal Material	Mass of bone powder used (mg)	Gender	Mean depth of autosomal
Tur1	MNHN ZM AC 1982-1092	*Capra caucasica*	Tooth root	148	male	3.52

Table S3.
Summary statistics of the ancient goats. Samples marked with an asterisk indicate that the sample has been directly radiocarbon.

Sample ID	Age (years ago)	Location	Gender	Endogenous content	Breadth of coverage in autosome	Mean depth of coverage in autosome	mtDNA depth	Haplotype
YJL02G*	~650	Yanjialiang, Inner Mongolia	male	78.58%	99.75%	13.44	455	A
WDH06S	~2,500	Wangdahu, Ningxia	female	88.05%	97.79%	8.19	565	A
SMG07	~3,900	Shimao, Shaanxi	male	3.96%	3.46%	0.04	9	B
SMG11	~3,900	Shimao, Shaanxi	male	1.56%	1.94%	0.04	10	C
KA01G*	~1,300	Northern Caucasus	male	16.33%	0.14%	0.32	53	A

Table S4.
Ancient genomes used in this study. Samples marked with an asterisk were sequenced in this study.

Sample ID	Coverage (q30)	mtDNA	Context	Grouping for autosomal analysis
Hovk1	3.08	F	Paleolithic	Armenia
Direkli1-2	11.55	T	Epipaleolithic	Anatolia
Direkli6	1.93	T	Epipaleolithic	Anatolia
Direkli5	0.27	T	Epipaleolithic	Anatolia
Lur12	1.05	G	Neolithic	Neolithic Iran
Ainghazal2	0.06	F	Neolithic	Neolithic Levant
Ainghazal1	0.03	F	Neolithic	Neolithic Levant
Semnan1-2	6.85	B	Neolithic	Neolithic Iran
Fars2-5	0.03	B	Neolithic	Neolithic Iran
Blagotin3	11.47	A	Neolithic	Neolithic Balkans
Blagotin2	4.02	A	Neolithic	Neolithic Balkans
Blagotin1	6.99	A	Neolithic	Neolithic Balkans
Blagotin16	3.51	A	Neolithic	Neolithic Balkans
Semnan9	3.05	G	Neolithic	Neolithic Iran
Semnan8	0.21	D	Neolithic	Neolithic Iran
Semnan7	3.28	D	Neolithic	Neolithic Iran
Semnan3	14.89	D	Neolithic	Neolithic Iran
Semnan17	0.12	D	Neolithic	Neolithic Iran
Semnan13	2.54	D	Neolithic	Neolithic Iran
Semnan10	1.43	G	Neolithic	Neolithic Iran
Monjukli8	2.57	D	Neolithic	Neolithic East
Kov57	0.07	A	Neolithic	Neolithic Balkans
Fars4	1.05	A	Chalcolithic	Chalcolithic Iran
AP49	0.02	A	Neolithic	Neolithic Balkans
AP45	0.02	A	Neolithic	Neolithic Balkans
Fars1	0.02	A	Chalcolithic	Chalcolithic Iran
Monjukli6	0.03	D	Chalcolithic	Chalcolithic Turkmenistan
Monjukli4	0.6	A	Chalcolithic	Chalcolithic Turkmenistan
Monjukli2	0.21	D	Chalcolithic	Chalcolithic Turkmenistan
Monjukli1	0.24	A	Chalcolithic	Chalcolithic Turkmenistan
Darre1	0.04	A	Chalcolithic	Chalcolithic Iran
Gilat8	0.02	D	Chalcolithic	Chalcolithic Levant
Gilat2	0.012	A	Chalcolithic	-
Azer6	0.28	A	Chalcolithic	Chalcolithic Caucasus
Safi2	0.04	A	Bronze Age	Bronze Age Levant
Kohneh2	0.04	A	Bronze Age	Bronze Age Caucasus
Chalow1	0.05	D	Bronze Age	Bronze Age Iran
Acem1	4.76	A	Bronze Age	Bronze Age Anatolia
Qazvin1	3.16	A	Bronze Age	Bronze Age Iran
Bulak5	0.27	D	Bronze Age	Bronze Age Uzbekistan
Bulak2	2.67	A	Bronze Age	Bronze Age Uzbekistan
Bulak1	0.87	A	Bronze Age	Bronze Age Uzbekistan
Azer3-5	4.66	A	Bronze Age	Bronze Age Caucasus
SMG11*	0.04	C	Neolithic	late Neolithic China
SMG07*	0.04	B	Neolithic	late Neolithic China

Acem2	8.67	A	Bronze Age	Bronze Age Anatolia
Yoqneam2	2.2	A	Bronze Age	Bronze Age Levant
Tac3	0.13	A	Bronze Age	Bronze Age Caucasus
Potterne1	3.67	A	Bronze Age	Bronze Age Britain
WDH06S*	8.19	A	Bronze Age	Bronze Age China
Azer4	2.57	A	Iron Age	Iron Age/Medieval Caucasus
Geor2	1.5	A	Medieval	Iron Age/Medieval Caucasus
KA01G*	0.32	A	Iron Age	Iron Age/Medieval Caucasus
Kazbeg1	3.84	A	Medieval	Iron Age/Medieval Caucasus
YJL02G*	13.44	A	Iron Age	Iron Age China
Darre2	3.93	A	Medieval	Medieval Iran

Table S5.
Distribution of SNPs identified in bezoars and domestic goats within various genomic regions annotated by ANNOVAR.

Category	SNP Count
intergenic	28,785,194
intronic	16,686,992
ncRNA_intronic	485,410
downstream	311,151
UTR3	298,612
upstream	298,295
synonymous	203,187
nonsynonymous	169,878
UTR5	87,386
ncRNA_exonic	54,835
unknown	12,917
upstream; downstream	11,394
stopgain	1,905
splicing	1,031
ncRNA_splicing	213
stoploss	168
UTR5; UTR3	150
exonic; splicing	47
ncRNA_exonic; splicing	10

Table S6.
Pairwise F_{ST} values calculated at the continent scale. On the continent scale, F_{ST} values correlated with geographical distances between populations when the center of origin was considered to be Southwest Asia.

Population	bezoar	SWA	EUR	AFR	SAS
SWA	0.071				
EUR	0.101	0.069			
AFR	0.100	0.049	0.062		
SAS	0.095	0.030	0.107	0.095	
EAS	0.091	0.031	0.098	0.089	0.038

Table S7.
Summary of population histories calculated from 2D-SFS. Confidence intervals were obtained by bootstrapping all sites and performing parameter inference on each bootstrap dataset with 100 runs.

Group	Model	Parameter	Description	Point estimation	Confidence interval (95%)	
					Lower bound	Upper bound
EUR-AFR	model6	Nanc	Ancestral population size	347,424	297,882	540,824
		T1	The scale time before split	161,566	111,647	169,156
		Napop1	Size of population 1 after split	52,538	48,650	57,944
		Napop2	Size of population 2 after split	23,496	21,779	29,256
		Ncpop1	Present size of population 1	12,158	11,473	13,585
		Ncpop2	Present size of population 2	87,314	87,159	103,875
		T2	The scale time between the split and population 2 size change	24,264	18,003	35,080
		T3	The scale time between the population 2 size change and population 1 size change	14,369	8,975	15,677
		T4	The scale time between the population 1 size change and the present	3,499	3,109	4,127
		m12_1	Migration from population 2 to population 1 during T2	8.64E-07	3.27E-07	1.26E-06
		m21_1	Migration from population 1 to population 2 during T2	4.61E-06	3.40E-06	1.78E-05
		m12_2	Migration from population 2 to population 1 during T3	2.94E-05	2.48E-05	5.25E-05
		m21_2	Migration from population 1 to population 2 during T3	5.13E-05	3.20E-05	6.86E-05
		m12_3	Migration from population 2 to population 1 during T4	1.48E-05	5.19E-06	1.41E-05
		m21_3	Migration from population 1 to population 2 during T4	7.51E-06	1.86E-06	7.87E-06
EUR-SWA	model5	Nanc	Ancestral population size	1,054,443	659,416	1,939,890
		T1	The scale time before split	67,785	43,369	68,293
		Napop1	Size of population 1 after split	142,879	123,544	144,188
		Napop2	Size of population 2 after split	105,474	102,453	124,650
		Ncpop1	Present size of population 1	15,002	12,650	16,754
		Ncpop2	Present size of population 2	128,478	92,545	130,776
		T2	The scale time between the split and population 1 size change	178,163	148,397	225,631
		T3	The scale time between the population 1 size change and population 2 size change	10,376	7,509	11,832
		T4	The scale time between the population 2 size change and the present	2,647	2,490	5,155
		m12_1	Migration from population 2 to population 1 during T2	5.86E-06	3.73E-06	6.90E-06

		m21_1	Migration from population 1 to population 2 during T2	1.10E-05	6.76E-06	1.12E-05
		m12_2	Migration from population 2 to population 1 during T3	6.18E-05	6.10E-05	1.09E-04
		m21_2	Migration from population 1 to population 2 during T3	1.98E-05	2.16E-05	3.85E-05
		m12_3	Migration from population 2 to population 1 during T4	2.67E-06	4.74E-08	1.32E-05
		m21_3	Migration from population 1 to population 2 during T4	8.98E-07	5.47E-09	9.00E-07
AFR-SWA	model4	Nanc	Ancestral population size	2,599,134	2,566,167	2,861,780
		T1	The scale time before split	103,095	102,467	104,954
		Napop1	Size of population 1 after split	35,040	33,446	36,663
		Napop2	Size of population 2 after split	144,192	134,486	155,022
		Ncpop1	Present size of population 1	63,084	59,762	66,535
		Ncpop2	Present size of population 2	100,117	93,284	108,257
		T2	The scaled time between the split and the secondary contact	147,379	122,095	183,998
		T3	The scaled time between the secondary contact and the present	11,689	10,563	13,507
		m12_1	Migration from population 2 to population 1 during T2	2.72E-05	2.55E-05	2.88E-05
		m21_1	Migration from population 1 to population 2 during T2	1.61E-06	7.46E-07	2.29E-06
		m12_2	Migration from population 2 to population 1 during T3	3.73E-05	3.52E-05	3.92E-05
		m21_2	Migration from population 1 to population 2 during T3	3.36E-05	3.15E-05	3.61E-05
SWA-EAS	model4	Nanc	Ancestral population size	198,725	198,090	202,464
		T1	The scale time before split	516,982	431,042	525,841
		Napop1	Size of population 1 after split	16,780	10,554	25,752
		Napop2	Size of population 2 after split	1,824	1,052	3,130
		Ncpop1	Present size of population 1	201,549	182,630	258,568
		Ncpop2	Present size of population 2	12,403	12,092	13,199
		T2	The scaled time between the split and the secondary contact	1,001	615	2,064
		T3	The scaled time between the secondary contact and the present	9,373	8,105	9,641
		m12_1	Migration from population 2 to population 1 during T2	3.77E-05	2.88E-05	7.94E-05
		m21_1	Migration from population 1 to population 2 during T2	1.32E-04	7.87E-05	1.31E-04
		m12_2	Migration from population 2 to population 1 during T3	9.53E-06	8.26E-06	1.06E-05
		m21_2	Migration from population 1 to population 2 during T3	7.05E-05	6.71E-05	7.29E-05

Table S8.
***D*-statistics for population relatedness between wild *Capra* species and domestic goat populations.**

Statistic	Populations X	Z-score(s)	Interpretation
D(*C. ibex*, *C. caucasica*; X, argali); *D*(*C. falconeri*, *C. caucasica*; X, argali); *D*(*C. sibirica*, *C. caucasica*; X, argali)	All ancient and modern domestic goats and bezoars	All > 16.8	Domestic goats and bezoars share more alleles with the *C. caucasica* than with the other three ibex-like species
D(*C. falconer*i, *C. ibex*; X, argali); *D*(*C. sibirica*, *C. ibex*; X, argali)	All ancient and modern domestic goats and bezoars	All > 66	Domestic goats and bezoars share more alleles with the *C. ibex* than with the *C. falconeri* and *C. sibirica*
D(*C. sibirica*, *C. falconeri*; X, argali)	All ancient and modern domestic goats and bezoars	All > 23	Domestic goats and bezoars share more alleles with the *C. falconeri* than with the *C. sibirica*
D(X, Armenian Bezoar; *C. caucasica*, argali); *D*(X, Anatolian Bezoar; *C. caucasica*, argali)	All ancient domestic goats	All > 6.9	*C. caucasica* shares more alleles with pre-domestication bezoars than with ancient domestic goats
D(Neolithic Iran, Neolithic Balkans; *C. caucasica*, argali); *D*(Neolithic China, Neolithic Balkans; *C. caucasica*, argali)	–	All > 5	*C. caucasica* shares more alleles with the Neolithic Balkan domestic goats than with the Neolithic Iranian and the Neolithic Chinese domestic goats
D(X_1, X_2; *C. caucasica*, argali)	X_1: modern Asian (SWA, SAS, and EAS) domestic goats; X_2: modern European/African (EUR and AFR) domestic goats	All > 11.4	*C. caucasica* shares more alleles with modern European/African domestic goats than with modern Asian domestic goats

Here the ancient samples with an average of coverage equal to or higher than three were used in this analysis.

Table S9.

f_3-statistics results showing AFR gene flow into SWA and *C. caucasica* gene flow into Armenian bezoar (Hovk1) dating to > 47,000. As positive Z-scores are not meaningful in f_3-statisctics, we only reported the case with the negative Z-score.

Statistic	Source1	Source2	Target	f_3	std.err	Z-score	SNPs
modern	AFR	EAS (South China)	EAS (North China)	-0.00625	0.000281	-22.255	15,730,078
	AFR	SAS (Bangladesh)	SWA	-0.00518	0.000289	-17.961	16,299,325
	AFR	SAS (Pakistan)	SWA	-0.00494	0.000229	-21.547	17,454,670
	AFR	EAS (South China)	SWA	-0.00244	0.00028	-8.711	16,832,722
(modern bezoar, *C. caucasica*; target)	bezoar (Azerbaijan)	*C. caucasica*	Hovk1	-0.02363	0.000965	-24.487	4,771,412
	bezoar (Alborz)	*C. caucasica*	Hovk1	-0.01791	0.000973	-18.417	4,746,128
	bezoar (Zagros)	*C. caucasica*	Qazvin1	-0.01626	0.000627	-25.915	5,314,101
	bezoar (Alborz)	*C. caucasica*	Qazvin1	-0.01487	0.000766	-19.425	4,881,248
	bezoar (Azerbaijan)	*C. caucasica*	Qazvin1	-0.01221	0.000801	-15.241	4,921,386
	bezoar (Alborz)	*C. caucasica*	Kazbeg1	-0.00086	0.00103	-0.835	4,730,479

Table S10.

KEGG pathway enrichment analysis of the genes contained in the candidate introgressed regions.

KEGG pathway	Input number	Background number	Corrected *P*-Value	genes
Amoebiasis	4	111	0.005283635	*SERPINB4, CD1B, COL4A4, SERPINB3*
Neuroactive ligand-receptor interaction	5	285	0.00931116	*LOC102191616, LOC102170425, P2RY13, TAAR8, PTH2R*

Table S11.
Summary of 24 selection sweeps on the X chromosome. The maximum F_{ST} and ln π ratio for each region are shown.

No.	Chr	Start (bp)	End (bp)	Max F_{ST}	Max ln ratio ($\theta_{\pi, bezoar}$ / $\theta_{\pi, domestic}$)	genes
1	NW_017189516.1	60,001	150,000	0.46	1.23	
2	NW_017189516.1	860,001	910,000	0.45	1.15	*LOC102175103*
3	NW_017189516.1	4,640,001	4,690,000	0.49	1.05	*ZNF711, POF1B*
4	NW_017189516.1	5,720,001	5,890,000	0.55	1.31	*CYLC1*
5	NW_017189516.1	5,900,001	6,090,000	0.48	1.26	
6	NW_017189516.1	6,100,001	6,210,000	0.47	1.11	
7	NW_017189516.1	6,680,001	6,750,000	0.46	1.31	
8	NW_017189516.1	6,800,001	6,970,000	0.47	1.34	
9	NW_017189516.1	6,980,001	7,030,000	0.47	1.00	
10	NW_017189516.1	7,040,001	7,110,000	0.53	1.18	*RPS6KA6*
11	NW_017189516.1	7,140,001	7,230,000	0.55	1.21	*RPS6KA6, HDX*
12	NW_017189516.1	7,260,001	7,310,000	0.54	1.07	
13	NW_017189516.1	13,720,001	14,450,000	**0.75**	**3.08**	***AR***
14	NW_017189516.1	16,920,001	17,010,000	0.47	1.27	
15	NW_017189516.1	18,840,001	18,910,000	0.55	1.58	
16	NW_017189516.1	22,100,001	22,270,000	0.66	1.20	
17	NW_017189516.1	23,620,001	23,690,000	0.49	1.34	*PHF8, FAM120C*
18	NW_017189516.1	24,340,001	24,450,000	0.56	1.23	*LOC102181853*
19	NW_017189516.1	24,480,001	24,550,000	0.54	1.05	
20	NW_017189516.1	24,560,001	24,630,000	0.56	1.07	
21	NW_017189516.1	25,140,001	25,250,000	0.53	1.70	*RRAGB*
22	NW_017189516.1	25,600,001	25,670,000	0.60	1.06	*PFKFB1*
23	NW_017189516.1	32,480,001	32,590,000	0.64	1.23	
24	NW_017189516.1	50,000,001	50,110,000	0.63	1.15	*MAP7D2, CXHXorf23*

Table S12.
Enriched KEGG pathway among candidate selection sweep genes.

KEGG pathway	Input number	Background number	Corrected *P*-value	genes
Olfactory transduction	26	420	2.80E-11	*LOC102190689*, *LOC102177547*, *LOC102184830*, *LOC102175911*, *LOC102178263*, *LOC102177261*, *LOC102185386*, *LOC108636866*, *LOC102183422*, *LOC102182057*, *LOC102182595*, *LOC102185677*, *LOC108637609*, *LOC102175539*, *LOC102176982*, *LOC102183709*, *LOC102185107*, *LOC102177709*, *LOC102181521*, *LOC102177974*, *LOC102182330*, *LOC102184268*, *LOC102183163*, *LOC102183437*, *LOC102177429*, *LOC102175633*
Influenza A	14	176	2.35E-07	*IRF3*, *LOC106503997*, *IFNGR2*, *NUP98*, *LOC102179184*, *LOC102190983*, *LOC102178719*, *IFNAR1*, *IFNGR1*, *MAPK12*, *LOC108637852*, *EP300*, *LOC108634246*, *MAPK11*
Malaria	7	49	2.77E-05	*HBBC*, *LOC102175600*, *SDC1*, *LOC102168959*, *LOC102176710*, *LOC102175876*, *LOC102168680*
African trypanosomiasis	6	35	5.85E-05	*LOC102176710*, *LOC102175600*, *LOC102168959*, *HBBC*, *LOC102175876*, *LOC102168680*
Natural killer cell mediated cytotoxicity	9	135	0.000240677	*LOC106503997*, *CD48*, *FCER1G*, *LOC102175817*, *IFNAR1*, *IFNGR1*, *IFNGR2*, *LOC108634246*, *LOC102182460*
Measles	9	136	0.00025292	*IRF3*, *LOC106503997*, *LOC102190983*, *IFNAR1*, *IFNGR1*, *IFNGR2*, *LOC108637852*, *LOC108634246*, *SLAMF1*
Neuroactive ligand-receptor interaction	12	278	0.00054143	*F2*, *LOC102177821*, *MC5R*, *CHRNB1*, *LOC102179184*, *MC2R*, *LEPR*, *LOC102178719*, *NPBWR2*, *LOC108637670*, *OPRL1*, *MC4R*
Signaling pathways regulating pluripotency of stem cells	8	142	0.001701777	*IGF1*, *RIF1*, *LHX5*, *AXIN1*, *JARID2*, *DVL2*, *MAPK11*, *MAPK12*
Herpes simplex infection	9	186	0.001931567	*IRF3*, *LOC102175817*, *IFNAR1*, *POLR2A*, *IFNGR1*, *IFNGR2*, *LOC108637852*, *EP300*, *LOC102190983*
Cytokine-cytokine receptor interaction	10	265	0.005033411	*TNFSF13*, *TNFSF12*, *LOC106503997*, *CSF1*, *LEPR*, *IFNAR1*, *IFNGR1*, *IFNGR2*, *LOC108634246*, *TNFRSF6B*

Osteoclast differentiation	7	132	0.005359714	*IFNGR2*, *CSF1*, *IFNAR1*, *MAPK11*, *MAPK12*, *MITF*, *IFNGR1*
Hepatitis C	7	133	0.005567875	*IRF3*, *IFNAR1*, *MAPK11*, *MAPK12*, *LOC108637852*, *LOC102190983*, *CLDN7*
FoxO signaling pathway	7	134	0.005760138	*PRMT1*, *SLC2A4*, *GABARAP*, *IGF1*, *MAPK11*, *MAPK12*, *EP300*
Adipocytokine signaling pathway	5	70	0.008846879	*CPT1C*, *CPT1B*, *NPY*, *SLC2A4*, *LEPR*

Table S13.
Additional genes with a function in immunity referring to GeneCards/NCBI.

Gene name	Potential function/effect
RIF1	Immunoglobulin class-switch recombination (CSR) during antibody genesis
LOC102183650 (*SLAMF7*)	Humoral immunity and antibody response
LOC102182927 (*SLAMF8*)	Humoral immunity and antibody response
SLAMF6	Humoral immunity and antibody response
CD84	Signaling lymphocyte activation
F11R	Immune System
ARHGAP30	Regulation of IgA production in intestinal immune network
PFDN2	Adaptive immune system
NIT1	A negative regulator of primary T-cells
FBXL13	Innate Immune System
IRAK3	Immunity
KLHL42	Innate Immune System
TNIP2	Immune response IL-23 signaling pathway
TNFAIP6	Innate Immune System
CYSTM1	Innate Immune System
IGIP	IgA Inducing Protein
ARG1	Inflammation and immunity
REL	Regulation of the survival and proliferation of B lymphocytes
OPRL1	Inflammatory and immune responses
POP1	Regulation of excessive inflammatory responses
ATG13	Autophagy Pathway
ART1	Innate Immune System
RHOG	Regulation of trans-endothelial migration of leukocytes
GABARAP	Inflammation
PSMD8	Adaptive Immune System
SPRED3	Innate Immune System
RASGRP4	Innate Immune System
FCGRT	Selective Igg Deficiency Disease and Immunodeficiency 43
PHF23	Autophagy pathway
GPS2	Key regulator of inflammation, lipid metabolism and mitochondrion homeostasis
KDM6B	Association with inflammatory diseases
AHRR	Suppress inflammation
MC5R	Pigmentation and inflammation
MC2R	Pigmentation and inflammation
MITF	Innate Immune System
MC4R	Pigmentation and inflammation
MGRN1	Innate Immune System
PPL	Innate Immune System
MUC6	Gastrointestinal parasite resistance and expulsion
ELF2	B and T cell development, cell cycle progression, and angiogenesis

Table S14.
Summary of nonsynonymous SNPs showing F_{ST} > 0.88 in *MUC6* between modern bezoars and domestic goats.

Chr	Pos	Ref	Alt	F_{ST}
29	46,245,173	C	T	0.888
29	46,247,389	T	A	0.896
29	46,247,410	T	C	0.896
29	46,247,683	T	C	0.891
29	46,248,226	A	G	0.891
29	46,248,397	G	T	0.887
29	46,248,537	C	G	0.895
29	46,248,619	T	C	0.891
29	46,248,646	C	T	0.891
29	46,254,119	G	A	0.903
29	46,259,627	G	A	0.887
29	46,264,308	T	C	0.886
29	46,264,872	T	C	0.891
29	46,265,931	C	T	0.891
29	46,266,673	C	T	0.942
29	46,267,289	C	T	0.888

Table S15.
Additional genes with a function in nervous system referring to GeneCards/NCBI.

Gene name	Potential function/effect
ROR1	Neurite growth in the central nervous system
NPY	Central nervous system
HMGA2	Neural stem cell development
RBFOX2	A key regulator of alternative exon splicing in the nervous system and other cell types
KCTD8	Regulation in inhibitory neurotransmitter in the mammalian central nervous system
HTT	Striatal neurons development
PURA	Nervous system development
EP300	Nervous system development
NRG2	Neuregulin family of growth and differentiation factors
MFSD14B	Intracellular neuronal membrane-bound proteins
ALK	Regulation in brain development and specific neurons development in the nervous system
PUM2	Nervous system development
NBEA	Central nervous system
VPS13B	Central nervous system development
STIM1	Regulation in the differentiation from mouse embryonic stem cells to neural cells
RRM1	Nervous system development
SBF2	Association with neuropathy, congenital hypomyelinating or amyelinating through infecting peripheral nervous
RPH3A	Regulation in synaptic transmission
LHX5	Differentiation and development of the forebrain

ELF2	Nervous system development
FREM3	Regulation of reactivity to environmental stimuli and perceptual processing speed
DVL2	Nervous system development
NOSIP	Development of neurogenesis
RRAS	Regulation of axon formation and guidance
DLG4	Regulation of the ratio of excitatory to inhibitory synapses in hippocampal neurons
NLGN2	Formation and remodeling of central nervous system synapses
FGF11	Nervous system development
KDM6B	Association with neurodegenerative diseases
CHD3	Nervous system development
JARID2	Regulation in neurulation in mouse models
DTNBP1	Neurite extension
ANKS1A	Association with Adjustment Disorder
PACSIN1	Neuron morphogenesis
ROGDI	Nervous system development
GDF10	A stroke-induced signal for axonal sprouting and functional recovery
GDF2	Regulation in cartilage and bone development, angiogenesis and differentiation of cholinergic central nervous system neurons
RYR1	Association with the release of Ca^{2+} from intracellular stores in neurons, promoting prolonged Ca^{2+} signaling in the brain

Data file S1. Overview of the 112 candidate regions (introgressed haplotype frequency > 0.1) introgressed from ibex-like species into domestic goats.

Data file S2. Autosomal regions identified as candidate selection sweeps by comparing genetic variants between 24 modern bezoars and 164 modern domestic goats. The statistics significant regions for F_{ST}/ln-ratio (θπ, bezoar/θπ, domestic)/XP-EHH in all four main groups (EUR, AFR, SWA-SAS, and EAS) were calculated separately. The "overlapped study" indicates the regions reside in 100 kb up- or down-stream of the regions reported in the corresponding literature. The functions related to these genes were annotated through literature mining.

Data file S3. Variants with high derived allele frequency (> 0.95) in domestic goats. The derived allele was identified as being absent in 24 modern bezoars and four ancient bezoars (Hovk1, Direkli1-2, Direkli6 and Direkli5). "Domestic Freq" represents the derived allele frequency in 164 modern domestic goats.

Data file S4. The fecal egg counts (FEC) for gastrointestinal nematodes in 268 animals from a polymorphic goat population with *MUC6*^{B} haplotype. Three replicate measurements were made for each individual. Genotype key: dom = *MUC6*^{D}/*MUC6*^{D}, het = *MUC6*^{B}/*MUC6*^{D}, wild = *MUC6*^{B}/*MUC6*^{B}. Animals used for genome wide association study are indicated in bold.

Data file S5. Genotyping at the *STIM1-RRM1* locus for the ancient and modern bezoars and goats. The SNPs (represent by "position, ancestral allele/derived allele") denote the sites with nearly fixed derived allele (frequency > 0.95) in 164 modern domestic goats but as being absent in 24 modern bezoars and four ancient bezoars. The columns for each SNP show the number of reads that pass GATK quality filters and support each allele at those positions. Genotype key: "bezoar-like" = homozygous for bezoar haplotype, "heterozygous" = heterozygous for bezoar haplotype, and "domestic-like" = homozygous for domestic haplotype. Three recombinant individuals are marked in bold.

Data file S6. Genotyping at the *MUC6* locus for the ancient and modern bezoars and goats. The SNPs (represent by "position, ancestral allele/derived allele") denote the sites with nearly fixed derived allele (frequency > 0.95) in 164 modern domestic goats but as being absent in 24 modern bezoars and four ancient bezoars. The columns for each SNP show the number of reads that pass GATK quality filters and support each allele at those positions. Genotype key: "bezoar-like" = homozygous for bezoar haplotype, "heterozygous" = heterozygous for bezoar haplotype, and "domestic-like" = homozygous for domestic haplotype.

Data file S7. Scaffolds that were putatively inferred as representing goat Y chromosome.

REFERENCES

1. G. Larson, D. Q. Fuller, The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).

2. M. A. Zeder, Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11597–11604 (2008).

3. F. Pereira, A. Amorim, in *Origin and Spread of Goat Pastoralism* (John Wiley & Sons, 2010).

4. L. Colli, M. Milanesi, A. Talenti, F. Bertolini, M. Chen, A. Crisà, K. G. Daly, M. Del Corvo, B. Guldbrandtsen, J. A. Lenstra, B. D. Rosen, E. Vajana, G. Catillo, S. Joost, E. L. Nicolazzi, E. Rochat, M. F. Rothschild, B. Servin, T. S. Sonstegard, R. Steri, C. P. Van Tassell, P. Ajmone-Marsan, P. Crepaldi, A. Stella, AdaptMap Consortium, Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet. Sel. Evol.* **50**, 58 (2018).

5. F. J. Alberto, F. Boyer, P. Orozco-terWengel, I. Streeter, B. Servin, P. de Villemereuil, B. Benjelloun, P. Librado, F. Biscarini, L. Colli, M. Barbato, W. Zamani, A. Alberti, S. Engelen, A. Stella, S. Joost, P. Ajmone-Marsan, R. Negrini, L. Orlando, H. R. Rezaei, S. Naderi, L. Clarke, P. Flicek, P. Wincker, E. Coissac, J. Kijas, G. Tosser-Klopp, A. Chikhi, M. W. Bruford, P. Taberlet, F. Pompanon, Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* **9**, 813 (2018).

6. K. G. Daly, P. M. Delser, V. E. Mullin, A. Scheu, V. Mattiangeli, M. D. Teasdale, A. J. Hare, J. Burger, M. P. Verdugo, M. J. Collins, R. Kehati, C. M. Erek, G. Bar-Oz, F. Pompanon, T. Cumer, C. Çakirlar, A. F. Mohaseb, D. Decruyenaere, H. Davoudi, O. Çevik, G. Rollefson, J.-D. Vigne, R. Khazaeli, H. Fathi, S. B. Doost, R. R. Sorkhani, A. A. Vahdati, E. W. Sauer, H. A. Kharanaghi, S. Maziar, B. Gasparian, R. Pinhasi, L. Martin, D. Orton, B. S. Arbuckle, N. Benecke, A. Manica, L. K. Horwitz, M. Mashkour, D. G. Bradley, Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* **361**, 85–88 (2018).

7. N. Pidancier, S. Jordan, G. Luikart, P. Taberlet, Evolutionary history of the genus *Capra* (Mammalia, Artiodactyla): Discordance between mitochondrial DNA and Y-chromosome phylogenies. *Mol. Phylogenet. Evol.* **40**, 739–749 (2006).

8. S. E. Hammer, H. M. Schwammer, F. Suchentrunk, Evidence for introgressive hybridization of captive markhor (*Capra falconeri*) with domestic goat: Cautions for reintroduction. *Biochem. Genet.* **46**, 216–226 (2008).

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

10. F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sánchez, Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015).

11. X.-J. Hu, J. Yang, X.-L. Xie, F.-H. Lv, Y.-H. Cao, W.-R. Li, M.-J. Liu, Y.-T. Wang, J.-Q. Li, Y.-G. Liu, Y.-L. Ren, Z.-Q. Shen, F. Wang, E. Hehua, J.-L. Han, M. H. Li, The genome landscape of tibetan sheep reveals adaptive introgression from argali and the history of early human settlements on the qinghai-tibetan plateau. *Mol. Biol. Evol.* **36**, 283–303 (2019).

12. D.-D. Wu, X.-D. Ding, S. Wang, J. M. Wójcik, Y. Zhang, M. Tokarska, Y. Li, M.-S. Wang, O. Faruque, R. Nielsen, Q. Zhang, Y.-P. Zhang, Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nat. Ecol. Evol.* **2**, 1139–1145 (2018).

13. P. W. Hedrick, Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–4618 (2013).

14. Z. Sun, J. Shao, L. Liu, J. Cui, M. F. Bonomo, Q. Guo, X. Wu, J. Wang, The first Neolithic urban center on China's north Loess Plateau: The rise and fall of Shimao. *Archaeol. Res. Asia* **14**, 33–45 (2018).

15. Y. Dong, X. Zhang, M. Xie, B. Arefnezhad, Z. Wang, W. Wang, S. Feng, G. Huang, R. Guan, W. Shen, R. Bunch, R. M. Culloch, Q. Li, B. Li, G. Zhang, X. Xu, J. W. Kijas, G. H. Salekdeh, W. Wang, Y. Jiang, Reference genome of wild goat (*Capra aegagrus*) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics* **16**, 431–431 (2015).

16. V. Manceau, L. Després, J. Bouvet, P. Taberlet, Systematics of the genus *Capra* inferred from mitochondrial DNA sequence data. *Mol. Phylogenet. Evol.* **13**, 504–510 (1999).

17. D. M. Shackleton, *Wild sheep and goats and their relatives : Status survey and conservation action plan for Caprinae* (International Union for Conservation of Nature and Natural Resources, 1997).

18. S. Naderi, H.-R. Rezaei, F. Pompanon, M. G. B. Blum, R. Negrini, H.-R. Naghash, Ö. Balkiz, M. Mashkour, O. E. Gaggiotti, P. Ajmone-Marsan, A. Kence, J.-D. Vigne, P. Taberlet, The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17659–17664 (2008).

19. M. A. Miguel, C. N. Mingala, Screening of pig (*Sus scrofa*) bactericidal permeability-increasing protein (BPI) gene as marker for disease resistance. *Anim. Biotechnol.* **30**, 146–150 (2019).

20. W. Yan, C. Zheng, J. He, W. Zhang, X.-A. Huang, X. Li, Y. Wang, X. Wang, Eleutheroside B1 mediates its anti-influenza activity through POLR2A and N-glycosylation. *Int. J. Mol. Med.* **42**, 2776–2792 (2018).

21. A. Omoumi, A. Fok, G.-Y. Hsiung, P4- 246: Analysis of two immune function genes (CR1 and CD2AP) in Alzheimer's disease in two large Canadian cohorts. *Alzheimers Dement.* **8**, P722 (2012).

22. S. D. Babu, V. Jayanthi, N. Devaraj, C. A. Reis, H. Devaraj, Expression profile of mucins (MUC2, MUC5AC and MUC6) in *Helicobacter pylori* infected pre-neoplastic and neoplastic human gastric epithelium. *Mol. Cancer* **5**, 10 (2006).

23. M. A. McGuckin, S. K. Lindén, P. Sutton, T. H. Florin, Mucin dynamics and enteric pathogens. *Nat. Rev. Microbiol.* **9**, 265–278 (2011).

24. M. V. Benavides, T. S. Sonstegard, C. Van Tassell, Genomic regions associated with sheep resistance to gastrointestinal nematodes. *Trends Parasitol.* **32**, 470–480 (2016).

25. J. Hartmann, R. M. Karl, R. P. D. Alexander, H. Adelsberger, M. S. Brill, C. Rühlmann, A. Ansel, K. Sakimura, Y. Baba, T. Kurosaki, T. Misgeld, A. Konnerth, STIM1 controls neuronal Ca^{2+} signaling, mGluR1-dependent synaptic transmission, and cerebellar motor behavior. *Neuron* **82**, 635–644 (2014).

26. Ł. Majewski, F. Maciag, P. M. Boguszewski, I. Wasilewska, G. Wiera, T. Wójtowicz, J. Mozrzymas, J. Kuznicki, Overexpression of STIM1 in neurons in mouse brain improves contextual learning and impairs long-term depression. *Biochim. Biophys. Acta. Mol. Cell Res.* **1864**, 1071–1087 (2017).

27. C. Ryu, D. C. Jang, D. Jung, Y. G. Kim, H. G. Shim, H.-H. Ryu, Y.-S. Lee, D. J. Linden, P. F. Worley, S. J. Kim, STIM1 regulates somatic Ca^{2+} Signals and intrinsic firing properties of cerebellar purkinje neurons. *J. Neurosci.* **37**, 8876–8894 (2017).

28. J. C. Craig, G. D. Bennett, R. C. Miranda, S. A. Mackler, R. H. Finnell, Ribonucleotide reductase subunit r1: A gene conferring sensitivity to valproic acid-induced neural tube defects in mice. *Teratology* **61**, 305–313 (2000).

29. M. Rinaldi, L. Dreesen, P. R. Hoorens, R. W. Li, E. Claerebout, B. Goddeeris, J. Vercruysse, W. Van Den Broek, P. Geldhof, Infection with the gastrointestinal nematode *Ostertagia ostertagi* in cattle affects mucus biosynthesis in the abomasum. *Vet. Res.* **42**, 61 (2011).

30. H. V. Simpson, S. Umair, V. C. Hoang, M. S. Savoian, Histochemical study of the effects on abomasal mucins of *Haemonchus contortus* or *Teladorsagia circumcincta* infection in lambs. *Vet. Parasitol.* **226**, 210–221 (2016).

31. N. Moniaux, F. Escande, N. Porchet, J.-P. Aubert, S. K. Batra, Structural organization and classification of the human mucin genes. *Front. Biosci.* **6**, D1192-D1206 (2001).

32. S. Z. Hasnain, A. L. Gallagher, R. K. Grencis, D. J. Thornton, A new role for mucins in immunity: Insights from gastrointestinal nematode infection. *Int. J. Biochem. Cell Biol.* **45**, 364–374 (2013).

33. D. Lawrence, G. Philip, H. Hunt, L. Snape-Kennedy, T. J. Wilkinson, Long term population, city size and climate trends in the fertile crescent: A first approximation. *PLOS ONE* **11**, e0152563 (2016).

34. K. N. Harper, G. J. Armelagos, Genomics, the origins of agriculture, and our changing microbe-scape: Time to revisit some old tales and tell some new ones. *Am. J. Phys. Anthropol.* **152**, 135–152 (2013).

35. F. Pereira, S. J. M. Davis, L. Pereira, B. McEvoy, D. G. Bradley, A. Amorim, Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Mol. Biol. Evol.* **23**, 1420–1426 (2006).

36. B. S. Arbuckle, Pace and process in the origins of animal husbandry in Neolithic Southwest Asia. *Bioarchaeol. Near East* **8**, 53–81 (2014).

37. C. Vilà, J. Seddon, H. Ellegren, Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet.* **21**, 214–218 (2005).

38. E. Huerta-Sánchez, X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, R. Nielsen, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).

39. M. R. Jones, L. S. Mills, P. C. Alves, C. M. Callahan, J. M. Alves, D. J. R. Lafferty, F. M. Jiggins, J. D. Jensen, J. Melo-Ferreira, J. M. Good, Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**, 1355–1358 (2018).

40. M. Dannemann, A. M. Andrés, J. Kelso, Introgression of neandertal- and denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *Am. J. Hum. Genet.* **98**, 22–33 (2016).

41. L. Abi-Rached, M. J. Jobin, S. Kulkarni, A. McWhinnie, K. Dalva, L. Gragert, F. Babrzadeh, B. Gharizadeh, M. Luo, F. A. Plummer, J. Kimani, M. Carrington, D. Middleton, R. Rajalingam, M. Beksac, S. G. E. Marsh, M. Maiers, L. A. Guethlein, S. Tavoularis, A.-M. Little, R. E. Green, P. J. Norman, P. Parham, The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334**, 89–94 (2011).

42. F. L. Mendez, J. C. Watkins, M. F. Hammer, Neandertal origin of genetic variation at the cluster of *OAS* immunity genes. *Mol. Biol. Evol.* **30**, 798–801 (2013).

43. M. Giacometti, R. Roganti, D. D. Tann, N. Stahlberger-Saitbekova, G. Obexer-Ruff, Alpine ibex *Capra ibex ibex* x domestic goat *C. aegagrus* domesticahybrids in a restricted area of southern Switzerland. *Wildl. Biol.* **10**, 137–143 (2004).

44. K. Junker, I. G. Horak, B. Penzhorn, History and development of research on wildlife parasites in southern Africa, with emphasis on terrestrial mammals, especially ungulates. *Int. J. Parasitol. Par.* **4**, 50–70 (2015).

45. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).

46. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

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

48. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).

49. B. L. Browning, S. R. Browning, Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).

50. S. Zhao, P. Zheng, S. Dong, X. Zhan, Q. Wu, X. Guo, Y. Hu, W. He, S. Zhang, W. Fan, L. Zhu, D. Li, X. Zhang, Q. Chen, H. Zhang, Z. Zhang, X. Jin, J. Zhang, H. Yang, J. Wang, J. Wang, F. Wei, Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat. Genet.* **45**, 67–71 (2013).

51. J. P. de Magalhães, J. Costa, A database of vertebrate longevity records and their relation to other life-history traits. *J. Evol. Biol.* **22**, 1770–1774 (2009).

52. L. Chen, Q. Qiu, Y. Jiang, K. Wang, Z. Lin, Z. Li, F. Bibi, Y. Yang, J. Wang, W. Nie, W. Su, G. Liu, Q. Li, W. Fu, X. Pan, C. Liu, J. Yang, C. Zhang, Y. Yin, Y. Wang, Y. Zhao, C. Zhang, Z. Wang, Y. Qin, W. Liu, B. Wang, Y. Ren, R. Zhang, Y. Zeng, R. R. da Fonseca, B. Wei, R. Li, W. Wan, R. Zhao, W. Zhu, Y. Wang, S. Duan, Y. Gao, Y. E. Zhang, C. Chen, C. Hvilsom, C. W. Epps, L. G. Chemnick, Y. Dong, S. Mirarab, H. R. Siegismund, O. A. Ryder, M. T. P. Gilbert, H. A. Lewin, G. Zhang, R. Heller, W. Wang, Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* **364**, eaav6202 (2019).

53. S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, J. M. Akey, Analysis of human sequence data reveals two pulses of archaic denisovan admixture. *Cell* **173**, 53–61.e59 (2018).

54. C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C.-Y. Li, L. Wei, KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **39**, W316–W322 (2011).

55. H. Li, Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).

56. A. Heger, L. Holm, Rapid automatic detection and alignment of repeats in protein sequences. *Proteins* **41**, 224–237 (2000).

57. H. V. Whitlock, Some modifications of the McMaster helminth egg-counting technique and apparatus. *J. Council Sci. Ind. Res. Au.* **21**, 177–180 (1948).

58. J. Meisner, A. Albrechtsen, Inferring population structure and admixture proportions in low depth NGS Data. *Genetics* **210**, 719–731 (2018).

59. H. M. Kang, J. H. Sul, S. K. Service, N. A. Zaitlen, S.-Y. Kong, N. B. Freimer, C. Sabatti, E. Eskin, Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).

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

61. D. M. Bickhart, B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie, S. Chan, J. Lee, E. T. Lam, I. Liachko, S. T. Sullivan, J. N. Burton, H. J. Huson, J. C. Nystrom, C. M. Kelley, J. L. Hutchison, Y. Zhou, J. Sun, A. Crisà, F. A. P. de León, J. C. Schwartz, J. A. Hammond, G. C. Waldbieser, S. G. Schroeder, G. E. Liu, M. J. Dunham, J. Shendure, T. S. Sonstegard, A. M. Phillippy, C. P. Van Tassell, T. P. L. Smith, Single-molecule sequencing and chromatin conformation capture enable de *novo* reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650 (2017).

62. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN], (26 May 2013).

63. K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).

64. N. Chen, Y. Cai, Q. Chen, R. Li, K. Wang, Y. Huang, S. Hu, S. Huang, H. Zhang, Z. Zheng, W. Song, Z. Ma, Y. Ma, R. Dang, Z. Zhang, L. Xu, Y. Jia, S. Liu, X. Yue, W. Deng, X. Zhang, Z. Sun, X. Lan, J. Han, H. Chen, D. G. Bradley, Y. Jiang, C. Lei, Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 2337 (2018).

65. P. J. Reimer, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. B. Ramsey, C. E. Buck, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, H. Haflidison, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, K Felix Kaiser, B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E Marion Scott, J. R. Southon, R. A. Staff, C. S. M. Turney, J. van der Plicht, IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 Years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

66. D. Y. Yang, B. Eng, J. S. Waye, J. C. Dudar, S. R. Saunders, Technical note: Improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105**, 539–543 (1998).

67. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC. Res. Notes* **9**, 88 (2016).

68. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

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

70. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2. 0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

71. B. S. Pedersen, A. R. Quinlan, Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2017).

72. A. Peltzer, G. Jäger, A. Herbig, A. Seitz, C. Kniep, J. Krause, K. Nieselt, EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).

73. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).

74. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I W de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

75. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–729 (2013).

76. I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).

77. A. Cornish-Bowden, Nomenclature for incompletely specified bases in nucleic acid sequences: rcommendations 1984. *Nucleic Acids Res.* **13**, 3021–3030 (1985).

78. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

79. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).

80. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

81. R. D. C. Team, R: A language and environment for statistical computing. R foundation for statistical computing. *Computing* **1**, 12–21 (2016).

82. S. J. Puechmaille, The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* **16**, 608–627 (2016).

83. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, M. Foll, Robust demographic inference from genomic and SNP data. *PLOS Genet.* **9**, e1003905 (2013).

84. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967 (2012).

85. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLOS Genet.* **8**, e1002453 (2012).

86. J. C. Barrett, B. Fry, J. Maller, M. J. Daly, Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).

87. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).

88. R. E. Handsaker, V. Van Doren, J. R. Berman, G. Genovese, S. Kashin, L. M. Boettger, S. A. McCarroll, Large multiallelic copy number variations in humans. *Nat. Genet.* **47**, 296–303 (2015).

89. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).

90. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genet.* **5**, e1000695 (2009).

91. R. Nielsen, T. Korneliussen, A. Albrechtsen, Y. Li, J. Wang, SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. *PLOS ONE* **7**, e37558 (2012).

92. J. B. Johnson, K. S. Omland, Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**, 101–108 (2004).

93. A. W. Briggs, J. M. Good, R. E. Green, J. Krause, T. Maricic, U. Stenzel, C. Lalueza-Fox, P. Rudan, D. Brajković, Ž. Kućan, I. Gušić, R. Schmitz, V. B. Doronichev, L. V. Golovanova, M. de la Rasilla, J. Fortea, A. Rosas, S. Pääbo, Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321 (2009).

94. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

95. X. Wang, Z. Zheng, Y. Cai, T. Chen, C. Li, W. Fu, Y. Jiang, CNVcaller: Highly efficient and widely applicable software for detecting copy number variations in large populations. *GigaScience* **6**, 1–12 (2017).

96. E. Paradis, pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).

97. J. A. Lenstra, Evolutionary and demographic history of sheep and goats suggested by nuclear, mtDNA and Y chromosome markers, in *The Role of Biotechnology for the Characterization and Conservation of Crop, Forestry, Animal and Fishery Geneic Resources* (International Workshop, Turin, Italy, 5 to 7 March 2005), pp. 1–4.

98. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, A. J. Drummond, BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).

99. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).

100. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

101. C. Gaunitz, A. Fages, K. Hanghøj, A. Albrechtsen, N. Khan, M. Schubert, A. Seguin-Orlando, I. J. Owens, S. Felkel, O. Bignon-Lau, P. de Barros Damgaard, A. Mittnik, A. F. Mohaseb, H. Davoudi, S. Alquraishi, A. H. Alfarhan, K. A. S. Al-Rasheid, E. Crubézy, N. Benecke, S. Olsen, D. Brown, D. Anthony, K. Massy, V. Pitulko, A. Kasparov, G. Brem, M. Hofreiter, G. Mukhtarova, N. Baimukhanov, L. Lõugas, V. Onar, P. W. Stockhammer, J. Krause, B. Boldgiv, S. Undrakhbold, D. Erdenebaatar, S. Lepetz, M. Mashkour, A. Ludwig, B. Wallner, V. Merz, I. Merz, V. Zaibert, E. Willerslev, P. Librado, A. K. Outram, L. Orlando, Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* **360**, 111–114 (2018).

102. S. Soraggi, C. Wiuf, A. Albrechtsen, Powerful inference with the D-statistic on low-coverage whole-genome data. *G3* **8**, 551–566 (2018).

103. H. Tang, D. O. Siegmund, P. Shen, P. J. Oefner, M. W. Feldman, Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. *Genetics* **161**, 447–459 (2002).

104. G. Ewing, J. Hermisson, MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **26**, 2064–2065 (2010).

105. F. Gao, C. Ming, W. Hu, H. Li, New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. *G3* **6**, 1563–1571 (2016).

106. T. S. Korneliussen, I. Moltke, A. Albrechtsen, R. Nielsen, Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *Bmc Bioinformatics* **14**, 289 (2013).

107. Z. A. Szpiech, R. D. Hernandez, selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014).

108. M. Degiorgio, C. D. Huber, M. J. Hubisz, I. Hellmann, R. Nielsen, S WEEP F INDER 2: Increased sensitivity, robustness and flexibility. *Bioinformatics* **32**, 1895–1897 (2015).

109. C. A. Heinlein, C. Chang, The roles of androgen receptors and androgen-binding proteins in nongenomic androgen actions. *Mol. Endocrinol.* **16**, 2181–2187 (2002).

110. P. Veinberg, Analysis of horn shape and coat coloration in *Capra* (Artiodactyla). *Byulleten Moskovskogo Obshchestva Ispytatelei Prirody, Otdel biologicheskii* **98**, 3–14 (1993).

111. S. Jordan, C. Miquel, P. Taberlet, G. Luikart, Sequencing primers and SNPs for rapidly evolving reproductive loci in endangered ibex and their kin (Bovidae, *Capra* spp.). *Mol. Ecol. Notes* **6**, 776–779 (2006).

112. F. G. Vieira, F. Lassalle, T. S. Korneliussen, M. Fumagalli, Improving the estimation of genetic distances from next-generation sequencing data. *Biol. J. Linn. Soc.* **117**, 139–149 (2015).

113. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).

114. B. L. Browning, S. R. Browning, Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).

115. H. Hu, N. Petousi, G. Glusman, Y. Yu, R. Bohlender, T. Tashi, J. M. Downie, J. C. Roach, A. M. Cole, F. R. Lorenzo, A. R. Rogers, M. E. Brunkow, G. Cavalleri, L. Hood, S. M. Alpatty, J. T. Prchal, L. B. Jorde, P. A. Robbins, T. S. Simonson, C. D. Huff, Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLOS Genet.* **13**, e1006675 (2017).

116. Y. Dong, M. Xie, Y. Jiang, N. Xiao, X. Du, W. Zhang, G. Tosser-Klopp, J. Wang, S. Yang, J. Liang, W. Chen, J. Chen, P. Zeng, Y. Hou, C. Bian, S. Pan, Y. Li, X. Liu, W. Wang, B. Servin, B. Sayre, B. Zhu, D. Sweeney, R. Moore, W. Nie, Y. Shen, R. Zhao, G. Zhang, J. Li, T. Faraut, J. Womack, Y. Zhang, J. Kijas, N. Cockett, X. Xu, S. Zhao, J. Wang, W. Wang, Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* **31**, 135–141 (2013).

117. I. Letunic, T. Doerks, P. Bork, SMART 7: Recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **40**, D302–D305 (2012).

118. E. L. Clark, S. J. Bush, M. E. B. McCulloch, I. L. Farquhar, R. Young, L. Lefevre, C. Pridans, H. G. Tsang, C. Wu, C. Afrasiabi, M. Watson, C. B. Whitelaw, T. C. Freeman, K. M. Summers, A. L. Archibald, D. A. Hume, A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLOS Genet.* **13**, e1006997 (2017).