Genomic analyses of wild argali, domestic sheep, and their hybrids provide insights into chromosome evolution, phenotypic variation, and germplasm innovation

Xin Li, San-Gang He, Wen-Rong Li, et al.

Genome Res. 2022 32: 1669-1684 originally published online August 10, 2022
Access the most recent version at doi:10.1101/gr.276769.122

Research

Genomic analyses of wild argali, domestic sheep, and their hybrids provide insights into chromosome evolution, phenotypic variation, and germplasm innovation

Xin Li,[1,2,3,13] San-Gang He,[4,13] Wen-Rong Li,[4,13] Ling-Yun Luo,[1] Ze Yan,[1] Dong-Xin Mo,[1] Xing Wan,[1] Feng-Hua Lv,[1] Ji Yang,[1] Ya-Xi Xu,[1] Juan Deng,[1,5] Qiang-Hui Zhu,[1,2,3] Xing-Long Xie,[2,3] Song-Song Xu,[2,3] Chen-Xi Liu,[4] Xin-Rong Peng,[4] Bin Han,[4] Zhong-Hui Li,[4] Lei Chen,[4] Jian-Lin Han,[6,7] Xue-Zhi Ding,[8] Renqing Dingkao,[9] Yue-Feng Chu,[10] Jin-Yan Wu,[10] Li-Min Wang,[11,12] Ping Zhou,[11,12] Ming-Jun Liu,[4] and Meng-Hua Li[1]

[1]College of Animal Science and Technology, China Agricultural University, Beijing 100193, China; [2]CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing 100101, China; [3]College of Life Sciences, University of Chinese Academy of Sciences (UCAS), Beijing 100049, China; [4]MOA Key Laboratory of Ruminant Genetics, Breeding and Reproduction, Ministry of Agriculture (MOA); Key Laboratory of Animal Technology of Xinjiang, Xinjiang Academy of Animal Science, Urumqi, 830000, China; [5]College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China; [6]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, 100193, China; [7]Livestock Genetics Program, International Livestock Research Institute (ILRI), Nairobi, 00100, Kenya; [8]MOA Key Laboratory of Veterinary Pharmaceutical Development of Ministry of Agriculture (MOA), Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou, 730050, China; [9]Institute of Animal Science and Veterinary Medicine, Gannan Tibetan Autonomous Prefecture, Hezuo, 747000, China; [10]State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, 730046, China; [11]Institute of Animal Husbandry and Veterinary Medicine, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi 832000, China; [12]State Key Laboratory of Sheep Genetic Improvement and Healthy Production, Xinjiang Academy of Agricultural and Reclamation Sciences, Shihezi 832000, China

Understanding the genetic mechanisms of phenotypic variation in hybrids between domestic animals and their wild relatives may aid germplasm innovation. Here, we report the high-quality genome assemblies of a male Pamir argali (*O. ammon polii*, $2n = 56$), a female Tibetan sheep (*O. aries*, $2n = 54$), and a male hybrid of Pamir argali and domestic sheep, and the high-throughput sequencing of 425 ovine animals, including the hybrids of argali and domestic sheep. We detected genomic synteny between Chromosome 2 of sheep and two acrocentric chromosomes of argali. We revealed consistent satellite repeats around the chromosome breakpoints, which could have resulted in chromosome fusion. We observed many more hybrids with karyotype $2n = 54$ than with $2n = 55$, which could be explained by the selfish centromeres, the possible decreased rate of normal/balanced sperm, and the increased incidence of early pregnancy loss in the aneuploid ewes or rams. We identified genes and variants associated with important morphological and production traits (e.g., body weight, cannon circumference, hip height, and tail length) that show significant variations. We revealed a strong selective signature at the mutation (c.334C > A, p.Gl12W) in *TBXT* and confirmed its association with tail length among sheep populations of wide geographic and genetic origins. We produced an intercross population of 110 F_2 offspring with varied number of vertebrae and validated the causal mutation by whole-genome association analysis. We verified its function using CRISPR-Cas9 genome editing. Our results provide insights into chromosomal speciation and phenotypic evolution and a foundation of genetic variants for the breeding of sheep and other animals.

[Supplemental material is available for this article.]

Hybridization is widespread in diverse groups of species across the tree of life (Moran et al. 2021), which results in genomic recombination, mosaics of ancestral segments, and phenotypic diversity

[13]These authors contributed equally to this work.
Corresponding authors: menghua.li@cau.edu.cn, liumingjun@xjaas.net

Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.276769.122. Freely available online through the *Genome Research* Open Access option.

(e.g., heterosis) and can lead to speciation (Abbott et al. 2013; Goulet et al. 2017; Moran et al. 2021). However, the specific molecular mechanisms underlying these changes remain to be elucidated. Interspecific hybridization between domestic animals and their wild relatives (including the wild ancestors and other wild related species) occurs in several genera (e.g., *Ovis*, *Capra*, *Canis*, *Felis*, *Sus*, *Gallus*, *Anas*, and *Camelus*), producing fertile (e.g., domestic animals vs. wild ancestors) or completely/partially sterile (e.g., domestic animals vs. some wild relatives except the wild ancestors) progenies with possible phenotypic heterosis and different numbers of chromosomes (Bunch and Foote 1977; Arnold 2004; Randi 2008; Chen 2013; Zhou et al. 2018; Lawal et al. 2020; Luo et al. 2020; Vidale et al. 2022). Thus, the intercross between most domestic animals and their wild relatives other than the wild ancestors provides a useful resource for unveiling the molecular and genetic mechanisms underlying chromosomal, genomic, and phenotypic evolution and reproductive isolation (Stelkens and Seehausen 2009; Thomsen et al. 2011; Salviano et al. 2017). Additionally, hybrids also provide a powerful system for dissecting the genetic basis of complex phenotypes in domestic animals (Zhou et al. 2018).

To date, eight species of the genus *Ovis* have been identified with varying diploid numbers of chromosomes ($2n = 52–58$), consisting of domestic sheep (*Ovis aries*) and their seven extant wild relatives (argali, *Ovis ammon*; Asiatic mouflon, *Ovis orientalis*; European mouflon, *Ovis musimon*; urial, *Ovis vignei*; bighorn sheep, *Ovis canadensis*; thinhorn sheep, *Ovis dalli*; and snow sheep, *Ovis nivicola*) (Rezaei et al. 2010; Cao et al. 2021; Chen et al. 2021b). Hybridization between wild relatives—in particular, the Pamir argali (also called Marco Polo sheep, *O. ammon polii*), which has larger biomass, less sexual dimorphism, and the shortest tail of any wild goat-antelope or sheep (Fedosenko and Blank 2005)—and domestic sheep has been documented to produce viable and fertile hybrids with significant heterotic effects even though some of them were aneuploid with odd numbers of chromosomes ($2n = 55$ and $2n = 57$) (Woronzow et al. 1972; Bunch and Foote 1977; Arnold 2004; Schröder et al. 2016; Alberto et al. 2018). For example, a Chinese native sheep breed, Bashibai, has been developed by intercrossing local sheep with argali in Xinjiang since the beginning of the twentieth century (Du 2011). Additionally, genomic tracts of wild introgression have been revealed in domestic sheep, contributing to their adaptive, morphological, and production traits (Barbato et al. 2017; Hu et al. 2019; Deng et al. 2020; Cao et al. 2021). Research on the genetic mechanisms and evolutionary dynamics of fertility and heterosis for the hybrids between domestic sheep and their wild relatives has not been explored owing to the lack of a chromosome-level reference genome for wild sheep and adequate hybrid samples.

Here, we constructed a large argali-sheep hybrid population ($n = 402$) and a Texel × Kazakh F_2 intercross population ($n = 110$) with segregation of gene variants and phenotypes (Fig. 1A,B; Supplemental Fig. S1). We produced high-quality chromosome-level genomes of argali, Tibetan sheep, and two haplotype-resolved assemblies of the F_1-hybrid. We generated a large set of genomes including populations of the hybrids, the F_2, and sheep with wide geographic origins and variable tail length. We implemented cytogenetic examination, population genomic analyses, genome-wide association studies (GWAS), and genome editing. We mainly aimed at (1) revealing genes and variants for several important morphological and production traits with heterotic effects and (2) exploring the possible molecular basis underlying the genomic recombination and chromosomal evolution.

Results

Genome assembly and annotation

We assembled three genomes of a Tibetan sheep ewe, an argali ram, and an F_1-hybrid ram using a sequencing strategy combining Pacific Biosciences (PacBio) high-throughput sequencing and Illumina systems (Supplemental Table S1; Supplemental Note S1). The draft and polished assemblies comprise 128 contigs with a contig N50 length of 77.47 Mb for Tibetan sheep and 174 contigs with a contig N50 length of 77.94 Mb for argali (Table 1; Supplemental Note S2). After further scaffolding and clustering using Bionano and Hi-C data, a final genome of Tibetan sheep (2.65 Gb) was de novo assembled, consisting of 27 pseudochromosomes with varying sizes from 45.01–280.64 Mb, whereas the de novo genome assembly of argali ram had a total length of 2.66 Gb across 29 chromosomes (i.e., 27 autosomes, X Chromosome, and Y Chromosome) with variable sizes of 45.10–280.78 Mb (Supplemental Fig. S2; Supplemental Table S2). For the F_1-hybrid assay, we first assembled a genome of ~5.21 Gb consisting of 458 contigs with a contig N50 length of 81.34 Mb after correction by the Illumina paired-end data. We obtained 54 clusters of contigs and manually adjusting Hi-C maps (Supplemental Figs. S3, S4). By performing intragenomic alignment, we preliminarily identified two pseudohaploid genome drafts with 28 (F_1-1) and 26 (F_1-2) pseudochromosomes (Supplemental Fig. S4A). The Hi-C map of F_1-2 showed that cluster LG02 shows false joining compared with LG02 in F_1-1 (Supplemental Fig. S4B,C). After mapping high-fidelity (HiFi) reads to LG02 in F_1-2 and further reorienting sequence scaffolds based on read coverage, we successfully divided LG02 into two acrocentric pseudoautosomes. Finally, two high-quality haploid genomes (i.e., 26 autosomes and one X Chromosome from sheep; 27 autosomes and one Y Chromosome from argali) were obtained, with contig and scaffold N50 values of 77.91 Mb and 101.52 Mb for the maternal assembly and 80.50 Mb and 102.90 Mb for the paternal assembly (Table 1; Supplemental Fig. S5; Supplemental Table S2; Supplemental Note S3).

Benchmarking universal single-copy ortholog (BUSCO) analysis showed a high degree of completeness of these genomes containing >95% of the complete eukaryotic universal genes (Supplemental Table S3; Yang et al. 2017; Luo et al. 2020). The mapping rate and coverage from Illumina reads reached >99%, and each base in the assembly showed a high accuracy of >99.9% (Supplemental Table S3). GC-depth and BLAST analyses did not detect contamination of the genomes (Supplemental Fig. S6; Supplemental Table S3). The distribution of transposable elements (TEs) and tandem repeats (TRs) across the whole genomes showed similarity among them (Supplemental Table S4; Supplemental Figs. S7, S8). The four assemblies showed similar numbers (around 20,000) of annotated protein-coding genes, with an average length of ~47.89 kb for a gene, ~1645.20 bp for a coding sequence, and ~176.88 bp for an exon (Supplemental Table S5). Additionally, we predicted a number of different types of RNA for the four assemblies (Supplemental Table S6).

Genomic characteristics

Of a total of 189,981 protein sequences in sheep, argali, and seven other species (i.e., human, dog, horse, pig, cattle, yak, and goat), we identified 19,917 gene families (orthogroups) based on homologous proteins. Of these gene families, 14,401 were shared among sheep, argali, and goats (Fig. 1D), and a majority of these genes

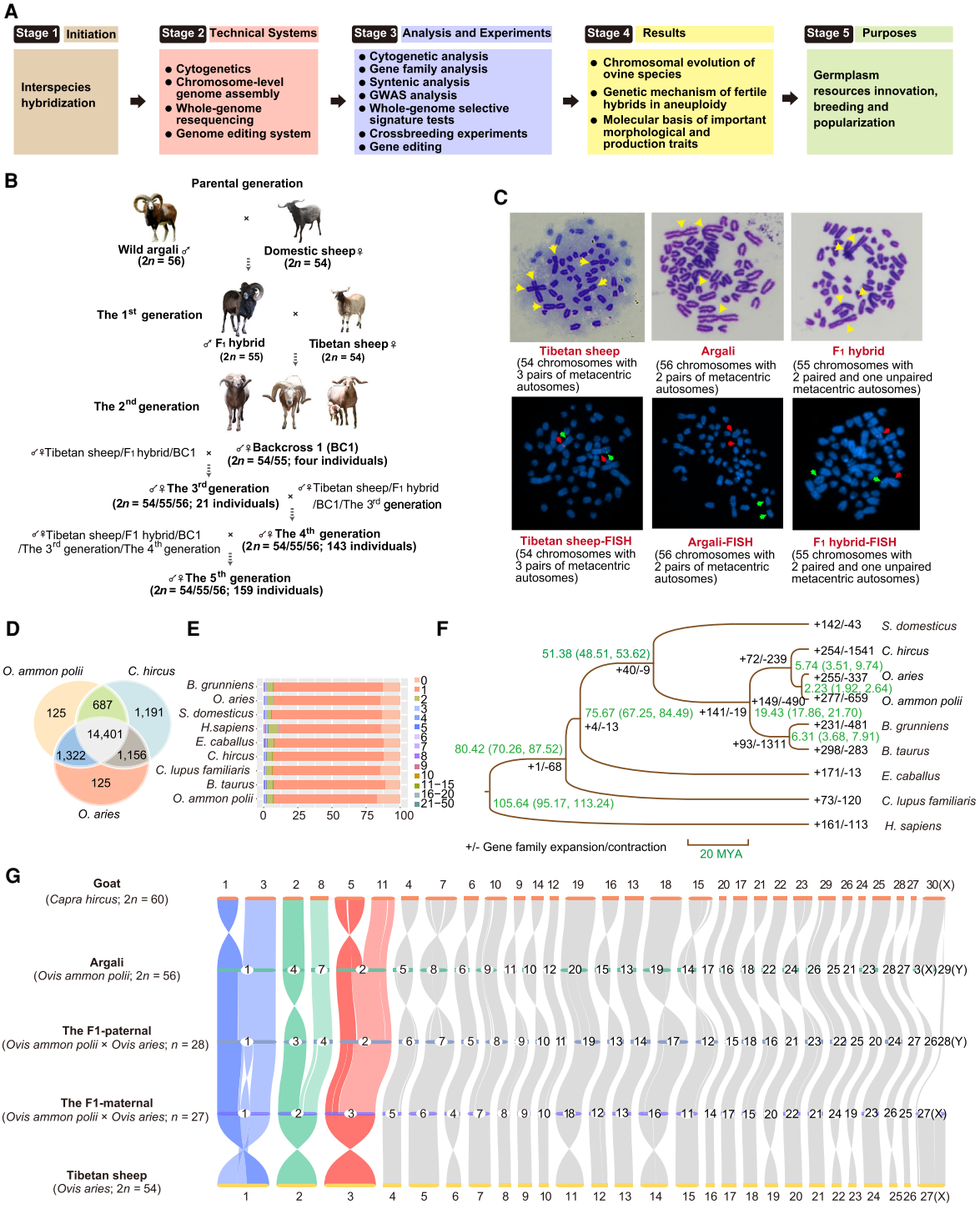

Figure 1. Research framework, hybridization scheme, karyotype and chromosomal painting, statistics of homologs and gene families, phylogenetic tree, and synteny landscape. (*A*) Roadmap of initiation, technical systems, analysis and experiments, the results, and purposes of this study. (*B*) The wild (argali, *O. ammon*)-domestic sheep hybridization scheme. The F_1-hybrid is subsequently backcrossed to Tibetan sheep to produce the progenies, and later hybrid generations are produced by intercrossing and backcrossing. (*C*) Karyotype and chromosomal painting of Tibetan sheep, argali, and the F_1-hybrid. DNA probes (*GNAQ*: green signal; *STK39*: red signal) were used to label the long arm and short arm of Chromosome 2 and two acrocentric chromosomes. DNA was stained with 4,6-diamidino-2-phenylindole (DAPI; blue). (*D*) Number of homologs among the Marco Polo subspecies of argali (*O. ammon polii*), sheep (*O. aries*), and goat (*Capra hircus*). (*E*) Proportions of different gene numbers in 19,917 gene families in the nine mammal species. (*F*) Phylogenetic tree of nine mammal species based on 10,043 single-copy orthologous genes. Numbers in green to the *right* of nodes are the divergence times and their 95% confidence intervals (95% CIs). Values next to the branches represent the numbers of gene family expansions/contractions. (MYA) Million years ago. (*G*) Genome-wide syntenic relationship among goat (*C. hircus*), sheep (*O. aries*), the Marco Polo subspecies of argali (*O. ammon polii*), and the argali-domestic sheep F_1-hybrid. Syntenic blocks involved in chromosome fusion are marked with colored ribbons.

Table 1. Genome assembly and annotation statistics for Tibetan sheep (*O. aries*), argali (*O. ammon polii*), and the F_1-hybrid (*O. ammon polii* × *O. aries*)

Genomic features	*O. aries*	*O. ammon polii*	*O. ammon polii* × *O. aries* (the F_1-hybrid; $n = 27$)	*O. ammon polii* × *O. aries* (the F_1-hybrid; $n = 28$)
Assembly size (bp)	2,653,843,355	2,664,280,018	2,898,997,297	2,710,908,005
No. of chromosomes	27	29	27	28
Genome in chromosomes	99.85%	99.79%	95.85%	97.64%
No. of scaffolds	106	136	165	114
Scaffold N50 (bp)	95,403,819	92,548,278	101,515,600	102,897,427
No. of contigs	128	174	210	177
Contig N50 (bp)	77,474,290	77,943,461	77,913,213	80,500,000
GC content	41.95%	41.96%	43.96%	42.83%
Annotated protein-coding genes (*n*)	20,919	20,916	21,078	20,363
Mean gene length (kb)	47.73	48.63	47.65	47.17
Repeat sequences	45.16%	45.42%	50.38%	46.98%

(75.7%–82.1%) within families were single copies across the species (Fig. 1E).

Phylogenetic reconstruction of the nine mammalian species indicated that argali and Tibetan sheep shared a common ancestor ~2.2 (95% CI, 1.9–2.6) million years ago (MYA) based on the 10,043 single-copy orthologous genes (Fig. 1F). Argali showed more events of gene-family contraction ($n = 659$) and expansion ($n = 277$) than those of Tibetan sheep (contraction: $n = 377$; expansion: $n = 255$), respectively.

Chromosomal evolution

Cytogenetic analysis showed that the argali's karyotype ($2n = 56$) contains two pairs of metacentric chromosomes and 26 pairs of acrocentric chromosomes; the karyotype of Tibetan sheep ($2n = 54$) consists of three pairs of metacentric chromosomes and 24 pairs of acrocentric chromosomes; and the F_1-hybrid ($2n = 55$) has two pairs of metacentric chromosomes, one unpaired biarmed chromosome, two unpaired acrocentric autosomes, and 24 pairs of acrocentric chromosomes (Fig. 1C). Of 39 F_1 backcrosses ($2n = 54/55$) with Tibetan sheep ($2n = 54$), we detected 32 individuals with 54 chromosomes and seven individuals with 55 chromosomes. We did not observe animals with 56 chromosomes (Supplemental Fig. S9; Supplemental Table S7), which could only be generated by two parental individuals with 55 chromosomes and had previously been reported to occur at a very low proportion (~5.56%) (Bunch and Foote 1977).

Chromosomal synteny analysis between argali and domestic sheep showed that pseudochromosomes LG04 and LG07 of argali are syntenic to the long and short arms of Chromosome 2 of domestic sheep, respectively (Fig. 1G; Supplemental Fig. S7A; Supplemental Table S8). Analysis of chromosome collinearity between two subgenomes of the F_1-hybrid also revealed that the unpaired biarmed Chromosome 2 of domestic sheep represents a high degree of sequence synteny with two unpaired acrocentric pseudochromosomes LG04 and LG07 of argali (Fig. 1G; Supplemental Fig. S7B; Supplemental Note S4). In the fluorescent in situ hybridization (FISH) analysis, the probes guanine nucleotide-binding protein alpha-q (*GNAQ*; green signal) on the short arm and serine threonine kinase 39 (*STK39*; red signal) on the long arm of Chromosome 2 were mapped onto two acrocentric chromosomes of argali (Fig. 1C). We further examined the chromosomal collinearity between goat (Saanen_v1, GCA_01544 3085.1) and ovine species (i.e., Tibetan sheep, argali, and the F_1-hybrid). Two biarmed pseudochromosomes (i.e., LG01 and LG02) of argali showed the synteny between the fusions of acrocentric Chromosomes 1 and 3 and Chromosomes 5 and 11 of goat, respectively. The two biarmed chromosomes have been maintained in domestic sheep, corresponding to Chromosomes 1 and 3. The third biarmed chromosome (i.e., Chromosome 2) observed in domestic sheep was syntenic to the fusion of acrocentric pseudochromosomes LG04 and LG07 in argali, which corresponds to acrocentric Chromosomes 2 and 8 of goat (Fig. 1G; Supplemental Table S9).

Robertsonian translocations (ROBs) (e.g., nonallelic homologous recombination [NAHR]) involved the breakpoints on Chromosomes 1, 3, 2, 8, 5, and 11 in goat; pseudochromosomes LG01, LG02, LG04, LG07 in argali; and Chromosomes 1, 2, and 3 in domestic sheep. We found specific repeat sequences in the regions adjacent to the centromeres of these three species (Fig. 2A). For ROB(G1qG3q) and ROB(G5qG11q) of goat, the common specific repeats of ROB(G1qG3q) (i.e., CGTGGGAAAGCCT) and ROB(G5qG11q) (i.e., GTGGGAAAGCCTC) occurred three to four times on Chromosomes 3 and 5 but were repeated many more times (79–476) on Chromosomes 1 and 11. A large number (968–1786) of repeat sequences, "GGGGGCCCACGTGATTG(T)CCCCT," were located on Chromosomes 3 and 5 but not on Chromosomes 1 or 11. For ROB(G8qG2q), the sequences "GGGGGCCCACGTGATTG(T)CCCCT" and "TTTCCCACGAGGC" were repeated several hundreds of times (13–469) on Chromosomes 2 and 8 of goat, but they were lost, and only the repeat "TTTCCCACGAGGC" was retained in argali and domestic sheep (Fig. 2A; Supplemental Fig. S10; Supplemental Table S10). This observation indicated that the sequence-specific recognition between homologous DNA elements on nonhomologous acrocentric chromosomes could be the underlying molecular basis of such chromosome fusions in ovine species. During the backcrossing between the F_1-hybrid and domestic sheep, a high sequence similarity between the metacentric chromosome and two acrocentric autosomes could account for the successful homologous chromosome pairing during meiosis, mainly giving rise to normal/balanced embryos after fertilization with the normal gametes of domestic sheep (Fig. 2B).

Pattern of variations and genetic differentiation

Whole-genome sequences of 425 individuals (i.e., 89 Tibetan sheep, eight argali, and 328 hybrids) (Fig. 1B) showed an average depth of ~17.88× and a genome coverage of 99.29% (Supplemental Table S11). After variant calling and filtering, we obtained a total of 31.86 million SNPs, including 17.59 million SNPs (14.71–14.99 million per individual) in argali, 31.86 million

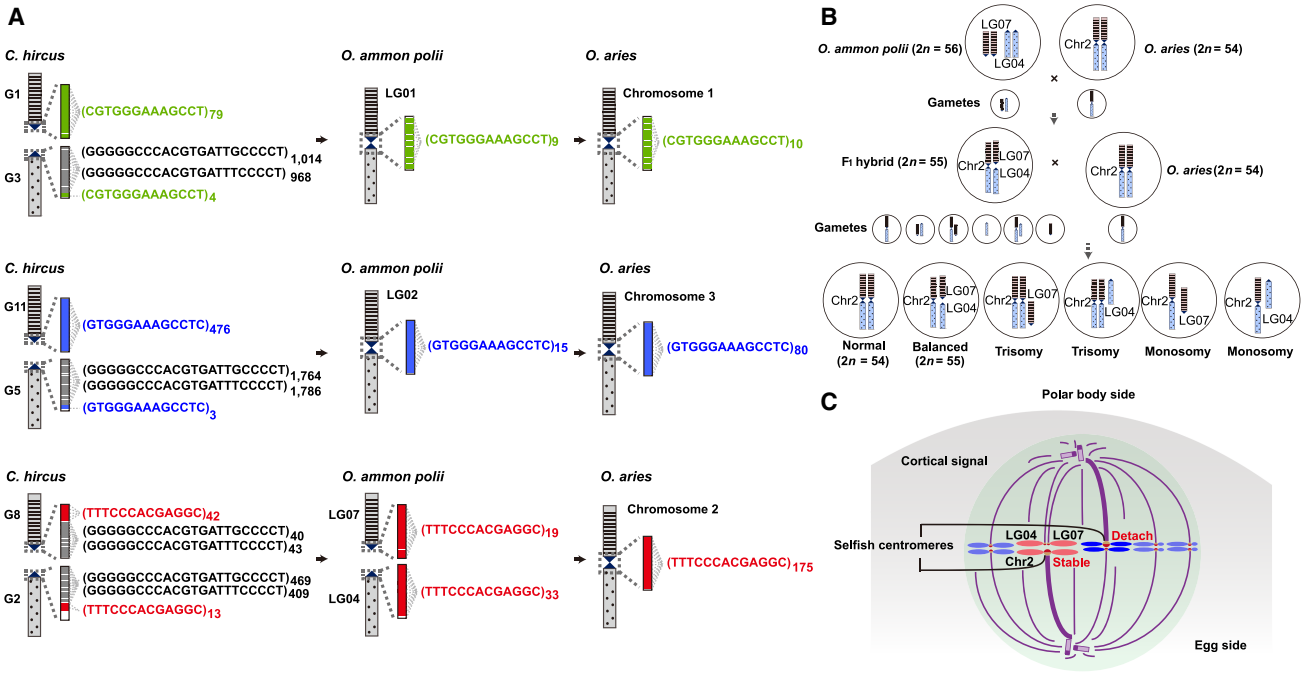

Figure 2. Evolution of three metacentric chromosomes of domestic sheep and the proposed molecular basis of fertility for the F_1-hybrid. (*A*) Repeated sequences surrounding the breakpoints on the chromosomes involved in chromosome fusions in goat (*C. hircus*), argali (*O. ammon polii*), and domestic sheep (*O. aries*). (*B*) Schematic representation of chromosomal segregation and pairing (metacentric Chromosome 2 of domestic sheep and two acrocentric pseudochromosomes LG04 and LG07 from argali) during hybridization between argali (*O. ammon polii*) and domestic sheep (*O. aries*). Six different types of zygotes can be produced by the F_1-hybrid, giving only one normal and one balanced embryo after fertilization with domestic sheep. (*C*) Proposed molecular mechanisms underlying the preference of F_1-hybrid oocytes to generate gametes with 27 chromosomes rather than 28 chromosomes. That is, selfish centromeres (red/yellow circles) prefer the egg side when attached to the metaphase I spindle (purple) (Nikalayevich and Verlhac 2021). Chromosomes are in blue/pink, oocyte cytoplasm is in gray, and the zone of cortical proximity is in darker gray.

SNPs (12.00–18.82 million per individual) in the hybrids, and 26.64 million SNPs (10.29–12.15 million per individual) in Tibetan sheep (Supplemental Fig. S11A,B; Supplemental Table S12). In addition, we observed 5.34 million indels (0.88–1.60 million per individual) (Supplemental Table S12).

Principal component analysis (PCA) and clustering analyses based on maximum likelihood estimation separated the samples into three separate clusters, argali, Tibetan sheep, and the hybrids (Fig. 3A,C). Genomic diversity (π) based on SNPs was 1.13×10^{-3} for argali, 3.11×10^{-3} for Tibetan sheep, and 3.78×10^{-3} for the hybrids (Fig. 3B). Pairwise genome-wide F_{ST} values calculated by SNPs were 0.38 between argali and Tibetan sheep, 0.26 between argali and the hybrids, and 0.03 between Tibetan sheep and the hybrids.

GWAS of morphological and production traits

In general, argali have a larger body size and shorter tail than those of Tibetan sheep (Fedosenko and Blank 2005). In Tibetan sheep and the hybrids, the Pearson correlation coefficient for morphological and production traits such as body weight, body height, body slanting length, chest circumference, cannon circumference, and hip height showed a strong ($r^2 = 0.6$–0.8) and significant ($P < 0.05$) correlation between them (Supplemental Table S13). The hybrids showed significant heterosis over Tibetan sheep for traits such as body weight and body height (Supplemental Fig. S12). In the backcross/intercross hybrid population, GWAS analyses revealed multiple significant SNPs for body weight, body height, body slanting length, chest circumference, cannon circumference, hip height, hip width, and tail length, respectively (Supplemental Table S16). Within each set of significant SNPs, we estimated the levels of linkage disequilibrium (LD; r^2) among the loci on each chromosome. At the LD threshold value of $r^2 = 0.6$ and an interval distance ≥ 1 Mb, we concluded that the significant SNPs represent 13, five, four, five, seven, and six independent loci for body weight, body height, body slanting length, chest circumference, cannon circumference, and hip height, and one and one quantitative trait locus (QTL) for hip width and tail length, respectively (Supplemental Figs. S13–S15; Supplemental Tables S16, S17; Supplemental Note S5). Through examining the independent QTLs for each trait on each chromosome (Supplemental Figs. S14, S15), we found that the QTLs (Chr 2: 90,071,756–98,428,523 bp, Chr 6: 28,993,028–30,457,109 bp, Chr 7: 11,574,342–14,364,056 bp) for body weight, the QTLs (Chr 7: 11,757,199–14,356,384 bp) for body height, the QTLs (Chr 2: 96,177,779–98,428,159 bp, Chr 6: 28,993,028–30,401,993 bp) for body slanting length, the QTLs (Chr 7: 11,595,855–14,356,384 bp) for cannon circumference, and the QTLs (Chr 7: 11,563,650–14,461,622 bp) for chest circumference have overlapping and broad signals, which could be owing to structural variants (e.g., an inversion). Further annotation of the significant GWAS signals in their upstream and downstream genomic regions revealed a number of genes associated with these morphological and production traits (Supplemental Table S18). Within these functional genes, we calculated the r^2 between the top signals with the other signals and selected those independent loci under the LD criteria of $r^2 < 0.6$ and an interval distance ≥ 1 Mb, such as

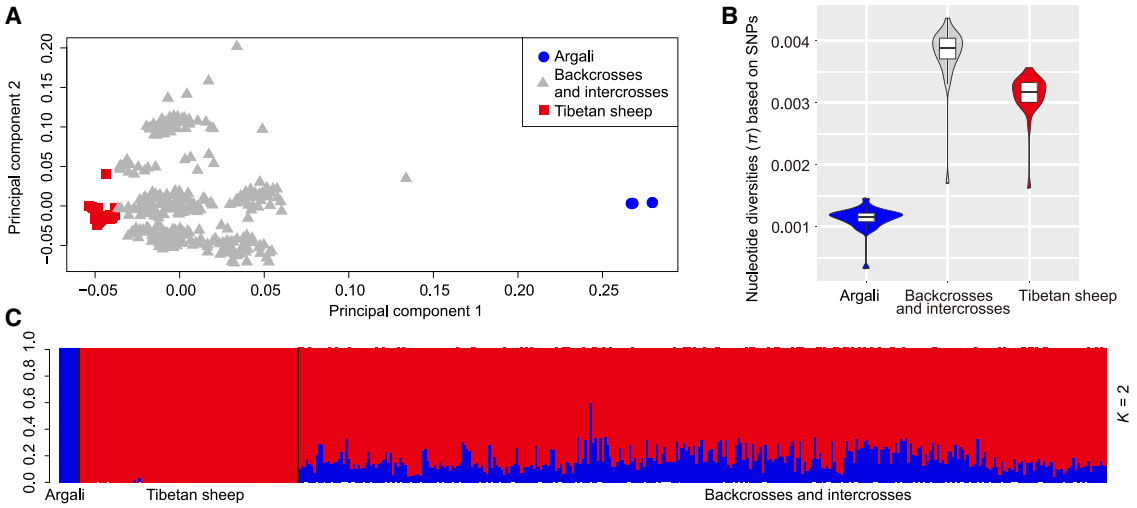

Figure 3. Genetic structure and genetic diversity of Tibetan sheep, argali, and the hybrids. (*A*) Plots of principal component analysis (principal components 1 and 2) for 425 individuals. (*B*) Nucleotide diversity of argali, Tibetan sheep, and the hybrids. (*C*) Population genetic structure of argali, Tibetan sheep, and the hybrids inferred using the ADMIXTURE program. The vertical solid line indicated the separations between argali, Tibetan sheep, and the hybrids.

TEK (Chr 2: 96,177,779), *FLRT2* (Chr 7: 96,426,040), *IQCH* (Chr 7: 14,356,384), *AUTS2, CASTOR2* (Chr 24: 35,060,092) for body weight; *MSRA* (Chr 2: 104,441,386), *IQCH* (Chr 7: 14,356,384), *UBASH3B* (Chr 15: 33,383,374) for body height; *TEK* (Chr 2: 96,177,779), *LINGO2* (Chr 2: 98,428,159), *BMRP1B*, *PDLIM5* (Chr 6: 30,401,933), *IQCH* (Chr 7: 14,356,384) for body slanting length; *PCDH10*, *HMGN1* (Chr 17: 26,391,533) for chest circumference; *LGALSL* (Chr 3: 44,121,028), *IQCH* (Chr 7: 14,356,384), *TFB2M* (Chr 12: 30,675,722) for cannon circumference; *MRS2* (Chr 20: 32,952,100) for hip height; and *ACTR3B, DPP6* (Chr 4: 117,196,289) for hip width (Supplemental Table S18; Supplemental Fig. S16). Of the genes identified here, several (e.g., *PCDH10*, *HMGN*, *TEK*, *BMPR1B*, and *IQCH*) were associated with two or more of the morphological and production traits (Supplemental Table S16), probably owing to strong genetic correlations between the traits (Sieber et al. 1988). Overall, the genes play important roles in obesity, regulation of feeding, spine morphogenesis, bone formation, and skeletal muscle development (Supplemental Table S18).

In particular, we detected 43 significant SNPs in and between the *PCDH10* and *HMGN1* genes associated with body weight traits. *PCDH10* has been reported to be associated with carcass weight and bone weight in cattle (Wang et al. 2018b). *HMGN1* functions in promoting astrocyte differentiation and regulating feeding (Fig. 4D; Hsuchou et al. 2009; Nagao et al. 2014; Xu and Xie 2016). These signals showed strong LD with the most significant SNP (Fig. 4E), and the phenotypic data of body weight associated with the three genotypes (TT, TG, and GG) of the top SNP (Chr 17: 26,391,397) showed a significant (Mann–Whitney *U* test, $P<0.001$) difference between each other (Fig. 4F). For tail length, we identified a single contiguous genomic region on Chromosome 8 (88.34–88.43 Mb), including two of the most significant signals located in the *TBXT* gene (Fig. 4A,B; Supplemental Note S5). *TBXT* is known as a founding member of the T-box transcription factor family and is expressed only in the early stages of notochord development, reportedly affecting tail length and sacral vertebrae in heterozygous animals by regulating the transforming growth factor and Wnt signaling pathway during vertebrae development (Supplemental Fig. S17A; Smith et al. 1991; Martin and Kimelman 2008; Han et al. 2019). The two top linked signals in *TBXT* contain one missense transversion mutation (Chr 8: 88,341,610 C/A), c.334C>A (CCC>ACC), leading to the conversion of glycine to tryptophan. Phenotypic data indicated that individuals with the homozygous reference genotype (CC) had a significantly (Mann–Whitney *U* test, $P<0.001$) longer tail length than individuals with the heterozygous variant genotype (CA), whereas individuals with the homozygous variant genotype (AA), which showed specific geographic and genetic origins (see below), were not observed (Fig. 4C).

Selective and association signatures associated with tail length and validation of the *TBXT* gene

We focused on the tail length trait from a wide range of origins. We integrated the whole genomes of 189 sheep (~8.96×) from 10 domestic populations (Pan et al. 2018; Wang et al. 2018a; Cheng et al. 2022) with variable tail length. Compared with fat-tailed sheep, fat-rumped sheep showed shorter tails (Fig. 5A,B). In the genome-wide scans for selective signatures between 103 fat-rumped sheep (i.e., 28 Kazakh sheep, 19 Bashibai sheep, 30 Bayinbuluke sheep, and 26 Duolang sheep) and 86 fat-tailed sheep (i.e., 30 Cele Black sheep, nine Hetian sheep, 10 Hu sheep, 10 Wuzhumuqin sheep, nine small-tailed Han sheep, and 18 Tan sheep) (Supplemental Table S19), we identified 168 and 211 candidate selected regions (top 0.5% outliers) by the F_{ST} and π-ratio methods, respectively. The strongest signal was located on Chromosome 8, ranging from 87.56 Mb to 87.88 Mb (Fig. 5C; Supplemental Table S20). In this candidate selective region, we further estimated genetic differentiation at the individual SNP level by estimating the π and Tajima's *D*-values between fat-rumped and fat-tailed populations. We observed the two most significantly differentiated SNPs (c.333C>G and c.334C>A) in the *TBXT* gene (Fig. 5D). Moreover, we genotyped the nonsynonymous SNP (c.334C>A, p. G112 W) in *TBXT* in 867 sheep representing 19 populations from a worldwide origin. We found that mutant allele A occurred at high frequency in fat-rumped sheep, whereas the

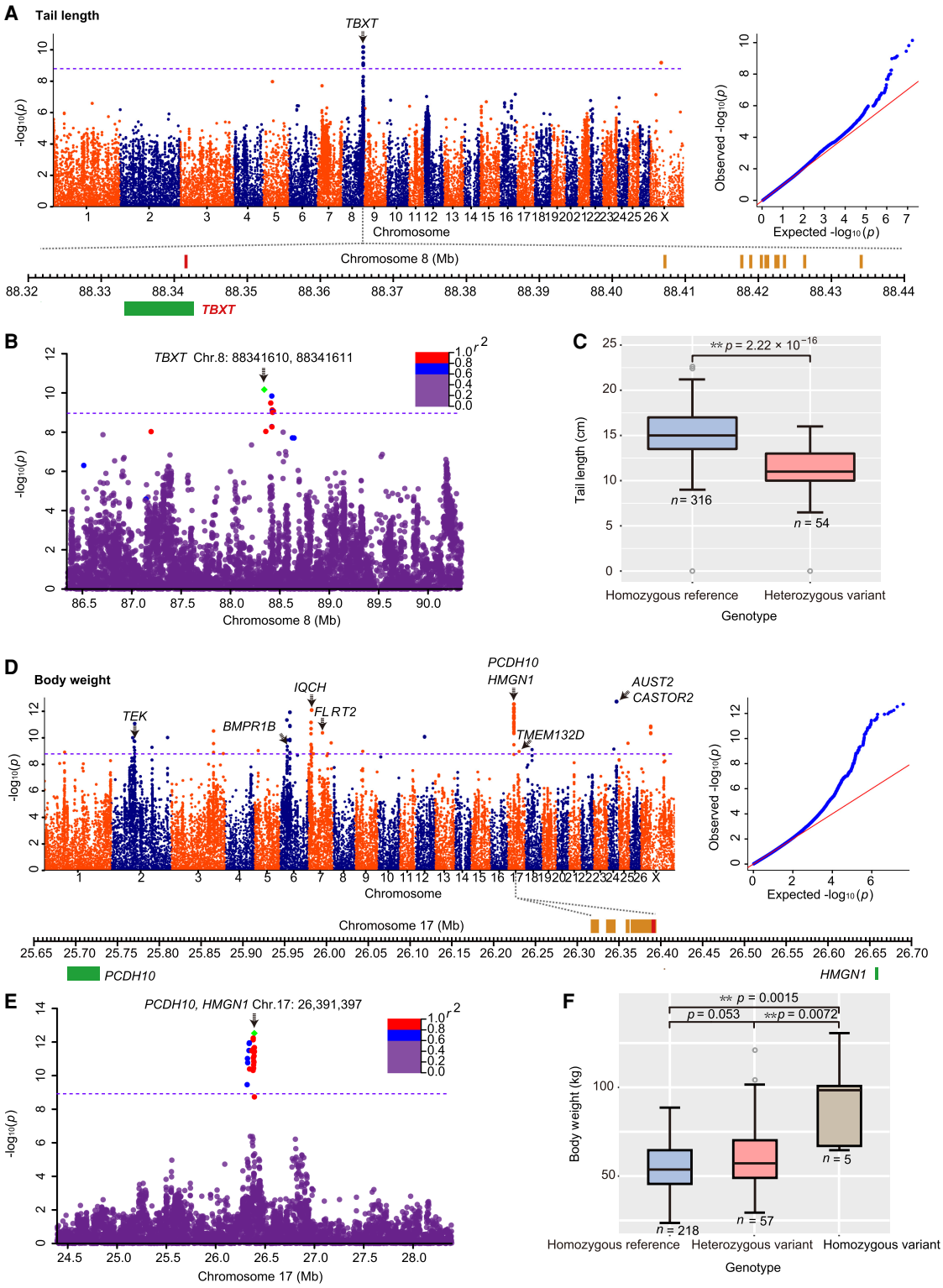

Figure 4. Identification of candidate genes related to the tail length and body weight of domestic sheep. (*A*,*D*) Manhattan and quantile–quantile (Q-Q) plots of association signals for the traits of tail length and body weight. The dashed lines colored purple represent the significant thresholds with –log(*P*-value) = 8.88. SNPs near the peaks with different significant values are marked in red (maximum $-\log_{10}(P)$) and brown ($-\log_{10}(P) \geq 7$). Functional genes surrounding the peaks are indicated by green boxes. (*B*,*E*) Linkage disequilibrium (LD) plots in the regions surrounding the *TBXT* and *HMGN1* genes. (*C*,*F*) Boxplots for the tail length associated with SNP (Chr 8: 88,341,610 C/A) and for the body weight associated with SNP (Chr 17: 26,391,397 T/G). The lines in boxes denote the median values; box limits are the upper and lower quartiles, and whiskers show the range of the data; *n* indicates the number of individuals with the same genotype. Significance of differences between the phenotypic values: (*) $P < 0.05$, (**) $P < 0.01$, determined by the Mann–Whitney *U* test.

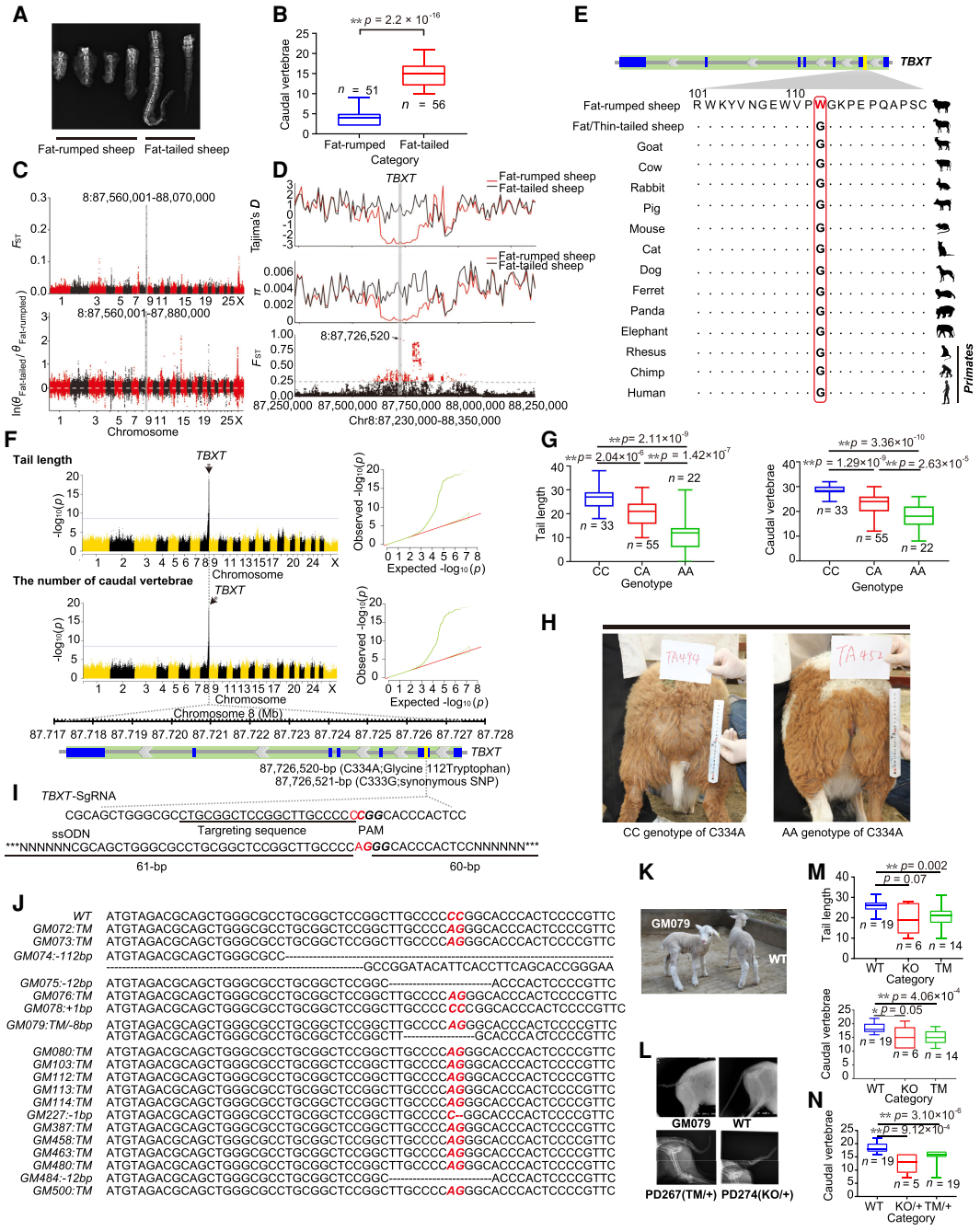

Figure 5. Genome-wide selection sweep test, genome-wide association study (GWAS) analysis, and functional validation of the *TBXT* gene. (*A*) Computerized tomography (CT) scanning of the caudal vertebrae number for fat-rumped (Kazakh, Duolang, Bashibai, and Bayinbuluke) and fat-tailed (Hetian and Cele Black) sheep. (*B*) Boxplots for the number of caudal vertebrae between the fat-rumped and fat-tailed sheep breeds. (*C*) Genome-wide selection sweep tests (the F_{ST}-based and π-ratio-based methods) for the tail configuration between fat-rumped and fat-tailed sheep breeds. (*D*) Calculation of Tajima's *D*-values and π and F_{ST} values for SNPs in the candidate genomic region Chr 8: 87.56–87.88 Mb between the fat-rumped and fat-tailed sheep breeds. (*E*) Amino acid (i.e., p. G112W in the *TBXT* gene) alternation among different mammalian species. (*F*) Manhattan and Q-Q plots of association signals for the tail length and the number of caudal vertebrae in domestic sheep. The dashed line represents the significance threshold ($-\log_{10}(0.05/\text{total SNPs}) = 8.56$). The gene structure of the *TBXT* gene is shown in green, and the exon regions are shown in blue at the *bottom*. (*G*) Boxplots for the tail length and the number of caudal vertebrae associated with the three genotypes of c.334C > A in *TBXT* in the F_2 intercross population of 110 individuals. (*H*) Different phenotypes in the tail configuration for individuals with the CC and AA genotypes of c.334C > A in *TBXT*; picture credit: Wen-Rong Li. (*I*) Sequences of sgRNA targeting exon 2 of the *TBXT* gene and 123-bp single-strand DNA oligonucleotides (ssODNs) for homologous recombination-mediated repair in CRISPR-Cas9 experimentation. (*J*) Target sequences in wild-type (WT) and 19 mosaic mutant merino sheep. Target mutations are indicated in the red bold font. (*K*) Different phenotypes of the tail configuration for WT and mutant merino sheep (GM079); picture credit: Wen-Rong Li. (*L*) CT scanning of the tail configuration for WT, GM079, and the offspring of GM079. (*M*) Boxplots for the tail length and the number of caudal vertebrae of the 19 WT sheep, 14 C334A target mutation (TM) merino sheep, and six sheep with a short indel (KO) in the *TBXT* gene. (*N*) Boxplots for the number of caudal vertebrae of the 19 WT sheep and 24 offspring of GM079 (i.e., 19 heterozygotes with the C334A target [TM/+] and five heterozygotes with an 8-bp deletion [KO/+]).

wild allele C appeared with high frequency in fat-tailed and thin-tailed sheep (Supplemental Table S21). In addition, we generated a Texel (thin-tailed with more vertebrae) × Kazakh (fat-rumpled with fewer vertebrae) F_2 intercross population of 110 sheep (Supplemental Fig. S1; Supplemental Table S22) and performed GWAS analysis of the traits of tail length and number of caudal vertebrae (Fig. 5F). Both the nonsynonymous SNP c.334C > A and the synonymous SNP c.333C > G in the *TBXT* gene showed the highest significance for the two traits (Supplemental Table S23). At the SNP c.334C > A, the F_2 offspring with the AA genotype showed the tailless phenotype, whereas offspring with the CC genotype had thin and long tails (Fig. 5G,H). Both genotypes showed significant (Mann–Whitney *U* test, $P < 0.001$) differences between tailed and tailless sheep in the F_2 intercross population. Thus, our data of the worldwide populations and the F_2 intercross confirmed the previously reported association between the two mutations in *TBXT* and the number of caudal vertebrae (Han et al. 2019).

We validated the phenotypic effect of the C334A mutation in sheep by CRISPR-Cas9 (Fig. 5I). After microinjection, 338 injected embryos were transplanted into 216 surrogate ewes of Chinese merino sheep. Finally, 31 surrogate ewes successfully produced 28 offspring. By screening the mutation in lambs, we identified 19 mosaic merino mutants, including 13 sheep with the C334A target mutation (TM), five sheep with a short indel (KO), and one sheep (ID: GM079) with both the target mutation and an 8-bp deletion (Fig. 5J). The sheep GM079 had the shortest tail with a tail length of 10 cm and 11 caudal vertebrae compared with wild-type merino sheep, with an average tail length of ~25.71 cm and caudal vertebrae length of ~18.47 (Fig. 5K; Supplemental Table S24). The average tail length and number of caudal vertebrae for the KO (tail length: 19.33 cm; number of caudal vertebrae: 15.17) and TM (tail length: 20.82 cm; number of caudal vertebrae: 15.07) merino animals were significantly shorter/lower than those for the wild-type merino sheep (tail length: 25.71 cm; number of caudal vertebrae: 18.47; Mann–Whitney *U* test, $P < 0.05$) (Fig. 5M; Supplemental Table S24). Additionally, we crossed the individual GM079 with normal merino sheep and generated 19 heterozygous progenies for the SNP mutation C334A and five heterozygous progenies for the 8-bp deletion. The number of caudal vertebrae trait showed a significant difference between these heterozygous progenies and wild-type merino sheep (Mann–Whitney *U* test, $P < 0.001$) (Fig. 5L,N; Supplemental Table S25).

Discussion

Robertsonian translocation occurs in a few mammals, such as humans (Page et al. 1996), dogs (Mayr et al. 1986), and water buffaloes (Luo et al. 2020). In humans, it involves five acrocentric chromosomes, 13, 14, 15, 21, and 22, all of which have very small short arms containing no unique genes (Song et al. 2016). Recombination between homologous sequences on nonhomologous chromosomes, which leads to the preferential formation of rob(13q14q) and rob(14q21q), has been proposed as the primary reason for this translocation in humans (Page et al. 1996). In the *Ovis* genus ($2n = 52$–58), previous cytogenetic analysis revealed four chromosomal fusions from the metacentric chromosomes of goats, corresponding to centric fusions of goat ($2n = 60$), namely, Chromosomes 1/3, 2/8, 5/11, and 9/19 sequentially (Bunch et al. 2006). However, we observed that the second pair of metacentric chromosomes in argali ($2n = 56$) is the result of the fusion of Chromosomes 5 and 11 of goat, whereas Chromosomes 2 and 8 from goat fused to give rise to the third pair of metacentric chromosomes of domestic sheep ($2n = 54$).

Of 39 F_1 backcrosses ($2n = 54/55$) with Tibetan sheep ($2n = 54$), we observed a much higher frequency of animals with 54 (number of animals = 32, 82.05%) chromosomes than with 55 chromosomes (number of animals = 7, 17.95%). This observation suggested that the descendants were inclined to generate a more stable karyotype with 27 pairs of chromosomes but not 26 pairs of chromosomes together with one unpaired biarmed autosome and two unpaired acrocentric autosomes (Bunch and Foote 1977). Notably, we found that Chromosome 2 of domestic sheep contains more satellite repeats (175 repeats) than the other two acrocentric chromosomes (LG04 and LG07) of argali (19 and 33 repeats, respectively) (Fig. 2A), which may make Chromosome 2 more "selfish" to assemble a larger kinetochore and have a higher chance of facing the "egg side" during the first meiotic division of the F_1-hybrid oocyte ($2n = 55$) (Fig. 2C; Nikalayevich and Verlhac 2021). Previous studies reported that when "selfish" chromosomes face the polar body side and sense the proximity of the cortex, they will dissociate the attachment to the spindle, which makes it possible for them to reorient and reattach until they face the egg side (Fig. 2C; Akera et al. 2017; Akera et al. 2019). Thus, this genetic mechanism may explain the priority of generating gametes with 27 chromosomes over 28 chromosomes for the F_1-hybrids ($2n = 55$), which could have contributed to the chromosomal speciation.

In humans, 99.7% of sperm from the homozygous Robertsonian translocation carrier (44, XY) was normal/balanced, which is much higher than that (79.9%) from the heterozygous Robertsonian translocation carrier (45, XY) (Song et al. 2016). Most conceptuses with autosomal aneuploid perish in utero, which makes aneuploidy the leading genetic cause of early pregnancy loss (Hassold and Hunt 2001; Kurahashi et al. 2012). Therefore, the possible decreased rate of normal/balanced sperm and increased incidence of early pregnancy loss in aneuploid ewes or rams ($2n = 55$) could also account for many more hybrids with $2n = 54$ than with $2n = 55$ observed here.

We have generated high-quality assemblies of Tibetan sheep, argali, and the F_1-hybrid, which have provided important insights into the molecular mechanisms of chromosomal fission/fusion among ovine species. Additionally, the assemblies can be integrated with those of other species to fully exploit the genome divergence and chromosomal evolution among species of the *Ovis* genus and the subfamily Caprinae. Because the assemblies are among only few high-quality chromosome-level reference genomes of high-altitude vertebrates (but see yak, *BosGru3.0*, GCA_005887515.2), they will be of great importance in exploring the genetic mechanisms underlying high-altitude genetic adaptation. Moreover, understanding the molecular basis of heterosis and phenotypic variation requires knowledge of diploid alleles (Sun et al. 2020). The high-resolution species-specific haploids of the F_1-hybrid can provide the diploid chromatin architecture for understanding the parent-of-origin effects and the regulation (e.g., *cis*- or *trans*-acting regulatory variation) of allele-specific expression (ASE) between homologous chromosomes. Additionally, haplotype-resolved diploid genomes of the F_1-hybrid will help to elucidate the landscape of genome rearrangements (SVs, indels, and TEs) and the epigenetic regulation of haploid-specific gene expression during hybridization, which could account for heterosis and ASE (McClintock 1984; Han et al. 2020). Assemblies of interspecies F_1-hybrid diploid genomes have been reported in cultured fish (*Silurus asotus*, $2n = 58$) ×

(*Silurus meridionalis*, $2n = 58$) (Chen et al. 2021a), yak (*Bos grunniens*, $2n = 60$) × cattle (*Bos taurus*, $2n = 60$) (Rice et al. 2020), and domestic cat (*Felis catus*, $2n = 38$) × Asian leopard cat (*Prionailurus bengalensis*, $2n = 38$) (Bredemeyer et al. 2021). However, representative assemblies of aneuploidy ($2n = 55$) with odd numbers of chromosomes have rarely been reported, but these resources would aid in dissecting the genetic basis of reduced fertility, interspecific hybridization compatibility, and speciation (Pauciullo et al. 2016; de Vos et al. 2020).

Body weight, body height, body slanting length, cannon circumference, chest circumference, hip height, hip width, and tail length are important morphological and production traits that remain the main selective targets of current livestock breeding programs. Our results unraveled many previously unreported trait-associated genes and variants as well as those with reported functions (e.g., *PTPRG*, *CACNB2*, *GALNTL6*, and *TMEM132D*). A majority of these quantitative traits such as body weight and body height are complex and multigenic traits (Yang et al. 2011b), which, however, are difficult to modulate with genome editing (Blighe et al. 2018). Tails are a vital part of the evolutionary package and play important roles in many mammals, such as counterbalancing (e.g., cats and kangaroos) (Walker et al. 1998; O'Connor et al. 2014), defense (porcupine) (Mori et al. 2014), warning signals (e.g., danger in deer) (Caro et al. 1995), social and emotional signals (canids) (Siniscalchi et al. 2013), hunting (alligator) (Willey et al. 2004), and fly swatting behavior in livestock such as cows, horses, and sheep. The prehensile tails of some mammals (e.g., monkeys) are used to increase mobility and stability (Young et al. 2015). Tails on sheep lambs are typically docked in most breeds for health reasons (Smith et al. 1997). Nevertheless, docking is not necessary in short-tailed breeds, and tails are not usually docked in breeds in which a long tail is valued, such as the Zwartbles breed with white socks and white tips on their tails (Lauvergne and Hoogschagen 1978). Docking is often considered cruel and unnatural by animal welfare activists, but it is considered essential in maintaining the health of sheep by sheep farmers (Brown and Meadowcroft 1996; Weaver 2005). Additionally, long and woolly tails in sheep make shearing more difficult, interfere with mating, and make the animals extremely susceptible to parasites, especially those that cause flystrike (Wooster 2005). These genes and alleles, along with the identified new genes and variants, provide useful resources for understanding the genetic basis of various phenotypic traits.

Early studies showed that individuals with homozygous *TBXT* mutations always show severe developmental disorders (Meisler 1997), and even specific amino acid changes are lethal in the early fetal life of animals (e.g., mouse, cat, and dog) (Haworth et al. 2001; Wu et al. 2010; Buckingham et al. 2013). However, the fat-rumped sheep population and intercross population with homozygous AA at nt 334 of exon 2 of *TBXT* showed no cases of lethality or other birth defects (Fig. 5E). A very recent study reported that the simian-specific *Alu*Sx1 element and hominoid-specific *Alu*Y element in the *TBXT* gene can form an inverted repeat pair, leading to an alternative splicing isoform without exon 6 for the *TBXT* gene and further causing reduction or loss of an external tail in hominoids (Xia et al. 2021). In mice, deletion of exon 6 in the *TBXT* gene could induce tail loss (Xia et al. 2021). In humans, the rs2305089 polymorphism in the *TBXT* gene and the epigenetic inactivation of *TBXT* by H3K27 were reported to be associated with chordoma (Cottone et al. 2020; Jalessi et al. 2022). The missense mutation (c. G47T, p. R16L) in *TBXT* is related to the development of congenital scoliosis (Feng et al. 2021). This and a previous study suggested different genetic mechanisms for vertebrae variation in mammals (Supplemental Fig. S17B). Therefore, analysis of the origins of these genes and variants will provide new insights into the evolution of trait-associated alleles during domestication and artificial selection. We detected a few trait-associated genes/variants that were nearly fixed in wild introgressive regions in the hybrids. This distinct feature suggests that introgression of new genes/alleles from their wild relatives might be a hallmark of livestock breeding through hybridization. In addition, our results indicated that introgression from wild relatives is a very important source of new genetic variation when adapting to a new environment (Wu et al. 2018; Mabry et al. 2021). Our results suggest that these species do not evolve in genetic isolation and point out future directions in the germplasm collection and utilization of wild relatives, which could broaden the genomic diversity of domestic animals (Zhou et al. 2018).

In conclusion, we have generated high-quality complete assemblies for argali, Tibetan sheep, and the F_1-hybrid with two sets of different haploid chromosomes as well as high-depth WGS of the interspecific hybrids, intercross population, and native sheep breeds. We show a case for creating new traits in domestic animals on a short timescale by combining the desirable attributes of wild relatives. Genomic recombination may help to maintain a repository of epigenes or epialleles generated by interspecific hybridization and intercrossing. The resources generated here provide a genomic framework for future germplasm use and expedite breeding and genetic improvement of sheep and possibly other livestock.

Methods

Sampling and DNA extraction

Genome assembly

A blood sample was collected from a Tibetan sheep ewe (*O. aries*, $2n = 54$), and fibroblasts were collected from the skin of a wild Marco Polo ram (a subspecies of argali, *O. ammon polii*, $2n = 56$) and an F_1-hybrid ram of Marco Polo sheep and domestic sheep (*O. ammon polii* × *O. aries*, $2n = 55$) for genome assembly (Supplemental Methods). Tibetan sheep were ethically sacrificed, and various tissues, such as heart, liver, spleen, lung, skeletal muscle, esophagus, uterus, trachea, small-intestine, and skin tissues, were snap-frozen in liquid nitrogen and stored at −80°C until further use. The fibroblasts were cultured in DMEM (Gibco) supplemented with 10% FBS (Gibco) and 1% antibiotic-antimycotic (Gibco) at 37°C under 5% CO_2.

Population genomics analysis and GWAS

For the GWAS analysis of morphological and production traits, a total of 402 blood samples were collected, including 74 Tibetan sheep and 328 hybrid descendants of domestic sheep and wild argali produced in five generations and >10 yr. In addition to the whole-genome sequences generated here, genomes of 15 Tibetan sheep and eight argali from our previous studies (Deng et al. 2020; Cao et al. 2021; Lv et al. 2022) were integrated in the population genomics analyses, totaling 425 genomes with an average depth of 17.88×. Furthermore, we produced an intercross population between Texel (thin-tailed with more vertebrae) and Kazakh (fat-rumped with fewer vertebrae) sheep, consisting of 110 F_2 offspring (average depth = 10.98×). We created the F_1 population by mating four Texel rams with 220 Kazakh ewes. Of the F_1 population, four rams were mated with 116 ewes, producing the F_2

population of 110 offspring with varied numbers of vertebrae. We measured the tail length and number of vertebrae via CT scanning and included them in the GWAS analysis of the two traits.

Whole-genome selection test

A total of 189 publicly available genomes from 10 domestic populations were integrated in the selection tests related to the tail-length trait, with a mean coverage of 8.96× (Supplemental Methods; Supplemental Table S19).

Genome and transcriptome sequencing

After SMRT sequencing libraries construction, DNA fragment selection, and purification, for Tibetan sheep and argali, sequencing was performed on a PacBio Sequel II instrument using continuous long-read sequencing (CLR) mode, whereas for the F_1-hybrid sample, sequencing was performed on a PacBio Sequel II instrument using circular consensus sequencing (CCS/HiFi) mode (Supplemental Methods). Finally, three SMRT cells for Tibetan sheep, two cells for argali, and five cells for the F_1-hybrid were used to generate raw data, respectively. Raw sequencing data were processed to remove low-quality reads and adapters using the PacBio SMRTlink v8.0 pipeline.

For genome assembly, libraries for Illumina sequencing were prepared using TruSeq PCR-free preparation kits (Illumina) with an insert size of ∼350 bp and sequenced on the Illumina NovaSeq/MGI-2000 platform. For whole-genome resequencing, paired-end sequencing libraries were constructed and sequenced on the Illumina HiSeq X Ten platform. After mapping 150-bp paired-end reads onto the assembly genomes of domestic sheep using the BWA-MEM v.0.7.17 (Li and Durbin 2009), SNP calling was performed using HaplotypeCaller in GATK (McKenna et al. 2010) for each individual (Supplemental Methods).

After RNA extraction, measurement of the concentration and integrity, and library construction, RNA sequencing was performed on the Illumina NovaSeq/MGI-2000 platform and PacBio Sequel II machine, respectively (Supplemental Methods).

Genome assembly and scaffolding

For Tibetan sheep and argali, we used NextDenovo v2.3.1 (https://github.com/Nextomics/NextDenovo) to generate the draft assembly. For the F_1-hybrid, the software hifisam v0.12 (Cheng et al. 2021) was used for the preliminary assembly, and NextPolish (Hu et al. 2020) was then used to polish all assemblies (Supplemental Methods). After evaluation of the draft assemblies (Supplemental Methods), we generated optical maps. DNA molecules were recorded in the Bionano Genomics Saphyr system (Supplemental Methods). Contigs were assembled into scaffolds using the Hybrid Scaffolding pipeline of the Bionano Solve v3.3 package (Mostovoy et al. 2016).

To further order and orient sequence scaffolds along chromosomes, we used genome-wide chromatin interaction data (Hi-C reads) sequenced by the Illumina NovaSeq/MGI-2000 platform. After filtering, valid interactive paired reads were identified and retained by HiC-Pro v2.8.1 (Supplemental Methods; Servant et al. 2015). We then used the LACHESIS software package (Burton et al. 2013; https://github.com/shendurelab/LACHESIS) to cluster, order, and orient scaffolds on chromosomes (Supplemental Methods). Finally, placement and orientation errors showing obvious discrete chromatin interaction patterns were manually adjusted.

To generate two pseudohaploid genome drafts of the F_1-hybrid, we retained the chromosomes with an overlapping region and manually copied this region to the corresponding position of the homologous chromosome based on the Hi-C maps of the normalized contact matrix for the overlapping regions in homologous chromosomes (Supplemental Fig. S3). We then identified homologous chromosomes by performing intragenomic alignment and aligning the two pseudohaploid assemblies to the argali and Tibetan sheep genome assemblies using the NUCmer program in MUMmer v.4.0.0 (Marçais et al. 2018). To divide the LG02 cluster in one of the pseudohaploid genome drafts (F_1-2) into two acrocentric pseudochromosomes, we mapped HiFi reads back to F_1-2 using minimap2 (Li 2018) and calculated the mapping depth and coverage for each base on LG02 through SAMtools v1.4 (Li et al. 2009). We then reordered and reoriented sequence scaffolds of LG02 using Hi-C data based on the coverage. Finally, we redefined two highly resolved pseudohaploid genome drafts as species-specific haploids based on the coverage rate between two homologous chromosomes and their parental assemblies (i.e., argali and Tibetan sheep). For the homologous chromosomes (i.e., LG09 in F_1-1 and LG11 in F_1-2) that failed to be assigned, we used the argali-specific SNPs inferred from the 425 argali and domestic sheep whole-genome data as markers to define species-specific chromosomes.

Genome annotation

Repeat prediction and noncoding RNA annotation

We used GMATA v2.2 (Wang and Wang 2016) and Tandem Repeats Finder v4.07b (Benson 1999) to search for TRs in the assembled genomes. RepeatMasker v r1.331(Bedell et al. 2000) was applied to search for TEs. To obtain a reliable profile of noncoding RNA, we aligned the assembly against the Rfam database (Griffiths-Jones et al. 2005) using the program cmscan in the software Infernal v1.1.2 (Nawrocki and Eddy 2013). Furthermore, we used tRNAscan-SE v2.0 (Lowe and Eddy 1997) and RNAmmer v1.2 (Lagesen et al. 2007) for tRNA and rRNA predictions, respectively.

Gene annotation

The gene models generated from the following three approaches were integrated with EVidenceModeler v1.1.1 (Supplemental Methods; Haas et al. 2008). First, for annotation using RNA-seq, we combined transcripts generated from Illumina RNA-seq and PacBio Iso-Seq to predict open reading frames (ORFs) using PASA v2.3.3 (Haas et al. 2008). Second, for de novo gene prediction, AUGUSTUS v3.3.1 (Stanke et al. 2008) was applied to predict genes using a generalized hidden Markov model (GHMM) (Stanke et al. 2006). Third, for the homology-based prediction, it was implemented by aligning the protein sequences of other species to the assembled genome using GeMaMo v1.6.1 (Keilwagen et al. 2016). Finally, we integrated these three predicted gene sets based on the relative weights (10:1:5) for transcriptomic, de novo, and protein homology-based evidence using EVidenceModeler v1.1.1. The genes with TEs were removed by TransposonPSI (Urasaki et al. 2017), and the miscoded genes were further filtered. Functional annotation of the predicted genes was obtained by aligning the protein sequences of these genes against the sequences in the NCBI nonredundant protein database and four public protein databases (Supplemental Methods). All the gene symbols were named according to the nomenclature guidelines from the Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC; https://www.genenames.org).

Comparative analysis of genome assemblies

To analyze the chromosomal collinearity between species, we first aligned all the assemblies of species using the NUCmer program in

the MUMmer v.4.0.0 package (Marçais et al. 2018). The positions of syntenic and rearranged regions and local sequence variations in the regions were mapped using the show-snps program in the MUMmer and Assemblytics software v1.2.1 (Nattestad and Schatz 2016). Alignment coordinates were then extracted using show-coords in MUMmer with the -THrd option, and chromosomal features and synteny of the assemblies were plotted using Circos (Krzywinski et al. 2009). TRs in the region surrounding the breakpoints were identified by TRF software (Benson 1999). In addition, we plotted macrosynteny for the genome assemblies of goat Saanen_v1 (GCA_015443085.1), argali (JAKZEL000000000), Tibetan sheep CAU_O.aries_1.0 (GCA_017524585.1), and the F_1-hybrid (JALAIW000000000 and JALAIX000000000) using MCscan (Python version) (Wang et al. 2012).

To identify the number of gene families and number of gene copies in specific families, annotated genomic information for various species was retrieved (Supplemental Methods). Classification of all the annotated protein-coding sequences into gene families was performed using OrthoFinder (Emms and Kelly 2019). Phylogenetic trees were built by IQ-TREE (Minh et al. 2020) with 1000 bootstrap replicates. The divergence time was estimated with PAML MCMCTree (Yang 2007) based on the JC69 model, and the nodes were calibrated using five reported divergence times: 1.94–2.61 MYA for *O. aries* and *O. ammon polii* (Yang et al. 2017), 2.9–7.6 MYA for *B. taurus* and *B. grunniens* (Luo et al. 2020), 18–22 MYA for *Bovinae* and *Caprinae*, 48.3–53.5 MYA for *Ruminantia* and *Suina*, and 95.3–113 MYA for *Euarchontoglires* and *Laurasiatheria* (Benton and Donoghue 2007). We then applied the CAFE v4.2.1 program (Han et al. 2013) to examine the expanded and contracted gene families along different nodes of a given phylogeny.

Cell cultures and karyotyping

For argali and the F_1-hybrid assays, chromosome preparations were made from a fibroblast culture mix composed of 10 mL of DMEM with L-glutamine, fetal calf serum (10%), and a penicillin–streptomycin mixture (1%). For Tibetan sheep and the 68 hybrid descendants, whole-blood samples were collected from the jugular vein, and chromosome preparations were made from a leukocyte culture mix composed of 7 mL of RPMI medium with L-glutamine, fetal calf serum (20%), penicillin–streptomycin mixture (1%), and concanavalin A with a final concentration of 5 μg/mL. After incubation for 45 h at 37.5°C, cell cultures were subjected to colchicine (0.07 μg/mL) treatment for 5 h. Following centrifugation, the cells were swollen in hypotonic 75 mM KCl saline solution and fixed in fixative (methanol:glacial acetic acid = 3:1) (Iannuzzi and Di Berardino 2008). Cell suspensions were dropped onto clean and wet slides and then air-dried. After Giemsa staining, the slides were observed under a Leica DM6B microscope equipped with a camera.

Probe production and FISH

Two chromosome painting probes for the *STK39* gene (on the long arm) and *GNAQ* gene (on the short arm) located on metacentric Chromosome 2 of the Tibetan sheep assembly were produced by PCR and labeled with TexasRed-dUTP and FITC-dUTP through nick translation. The information for the PCR primers and the length of amplified DNA fragments was detailed in Supplemental Table S26. In FISH, metaphase spreads were first permeabilized using Triton X-100 for 10 min and rinsed in PBS for 5 min, treated with pepsin at room temperature for 5 min, rinsed in PBS for 5 min twice and in 70% ethanol for 30 sec. After the slides were air-dried and dehydrated with a gradient of 70%, 85%, and 100% of cold ethanol for 2 min, respectively, a 10-μL probe solution was used for each slide, and then denaturing was performed for 2 min at 72°C–73°C. The hybridization was performed in a humid chamber for at least 16 h at 37°C. After hybridization, the slides were washed in 0.3% IGEPAL CA-630, Sigma-Aldrich (or NP-40)/0.4XSSC for 2 min at 73°C and then washed in 0.1% IGEPAL CA-630 Sigma-Aldrich (or NP-40)/2XSSC for 90 sec at room temperature. Finally, slides were counterstained with DAPI (4′,6-diamidino-2-phenylindole) solution (0.24 μg/mL, Sigma-Aldrich), and the images were captured with an Olympus BX-51 microscope equipped with a DP-70 CCD camera.

Population genomics analysis

In population genomics analyses, SNPs and SVs that did not meet one of the following criteria were excluded: (1) a minor allele frequency >0.05 and (2) maximum missing rate <0.1. PCA of whole-genome SNPs and SVs for all 425 individuals was performed with GCTA v.1.24.2 (Yang et al. 2011a). Population structure was examined using ADMIXTURE v.1.23 with the default settings (Alexander et al. 2009). The number of assumed genetic clusters, *K*, was set as two. An individual-based neighbor-joining (NJ) tree was constructed for all the samples based on the nucleotide *p*-distance matrix using TreeBeST v 1.9.2 (Vilella et al. 2009) and visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The parameter r^2 for LD was calculated for pairwise SNPs within each chromosome using PLINK v.1.90 (Purcell et al. 2007). Nucleotide diversity (π) and pairwise genome-wide F_{ST} values were calculated using VCFtools v.0.1.16 (Danecek et al. 2011).

GWAS analysis of morphological and production traits

Genome-wide association analyses of phenotypic traits such as body weight, body height, body slanting length, cannon circumference, chest circumference, hip height, hip width, and tail length were performed in a total of 402 animals (74 Tibetan sheep and 328 domestic sheep × argali hybrid descendants) (Supplemental Table S14) using the efficient mixed-model association expedited (EMMAX) statistical test approach (Kang et al. 2010), which is based on a mixed linear model and accounts for population stratification and relatedness (Supplemental Note S6). To estimate the effect of age on phenotypic data, we implemented analysis of variance (ANOVA) and selected various sets of samples for each different trait in GWAS analysis based on a Duncan's multiple range test (Supplemental Table S15). The covariates used to adjust for each phenotype simultaneously are also listed in Supplemental Table S15. For body weight, body slanting length, and tail length, the genome-wide significance thresholds were set as $-\log_{10}(0.05/\text{total SNPs}) = 8.88$ after the Bonferroni correction. We set the thresholds as $-\log_{10}(P\text{-value}) = 8$ for body height and $-\log_{10}(P\text{-value}) = 7$ for cannon circumference, chest circumference, hip height, and hip width (Xiao et al. 2021).

Selection test, GWAS, and functional validation of TBXT

Genome-wide selection tests for the tail length trait

We implemented selective sweep tests between the genomes of 103 fat-rumped sheep and 86 fat-tailed sheep (Supplemental Table S19) by calculating population differentiation values (F_{ST}) and nucleotide diversity ratios ($\ln(\pi_{\text{Fat-tailed}}/\pi_{\text{Fat-rumped}})$) using VCFtools with a window of 50 kb and a step size of 10 kb (Supplemental Methods).

GWAS of the number of caudal vertebrae in an F_2 population

After mapping and variation calling as detailed above, we performed GWAS analysis of the number of caudal vertebrae using a mixed model in GEMMA software (Supplemental Methods; Zhou and Stephens 2012). Sex and the first three PCA values calculated by PLINK were taken as the covariates. The genome-wide significance threshold was set at $-\log_{10}(0.05/\text{total SNPs}) = 8.561$.

Genetic modification in TBXT using CRISPR-Cas9

The sgRNA targeting exon 2 of *TBXT* (Fig. 5F,I) was designed using the online tool (http://www.e-crisp.org/E-CRISP/reannotate_crispr.html) and then cloned into the pX330 vector containing two expression cassettes, a human *Streptococcus pyogenes* (hSpy) Cas9 and the chimeric gRNA. To synthesize sgRNA in vitro, expression vectors were first linearized by digestion with DraI and then used as templates to produce sgRNA and Cas9 mRNA using the MEGAshortscript T7 transcription kit (Ambion) and mMESSAGE mMACHINE T7 ultra transcription kit (Ambion), respectively. Cas9 mRNA and sgRNA were purified using a MEGAclear transcription clean-up kit (Ambion) and dissolved in RNase-free water. RNA concentrations were measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific). RNA quality was assessed by agarose gel electrophoresis.

To obtain the designed mutations in *TBXT* through homologous recombination (HR)–mediated repair, we synthesized 123-bp long single-strand DNA oligonucleotide (ssODN) donors, allowing knock-in of two specific mutations: Gly112Trp substitution (C334A) and synonymous mutation (C333G) (Fig. 5I). Finally, 50 ng/μL sgRNA, 100 ng/μL Cas9 mRNA mixture, and 100 ng/μL ssODN mixture were prepared using RNase-free water and microinjected into the cytoplasm of Chinese merino zygotes, which were collected through surgical oviduct flushing of donors by estrus synchronization and superovulation. A 1-yr-old GM079 (male) was selected and mated with the wild-type Chinese merino ewe, producing 19 offspring with the target mutation C334A and five offspring with an 8-bp indel (Supplemental Table S24).

Validation of the C334A mutation in TBXT

We further genotyped the *TBXT* mutation C334A in the genetically modified sheep and their offspring as well as in 867 sheep from 19 populations with different tail configurations (Supplemental Table S21) by Sanger sequencing. We performed PCR amplifications with the primers *TBXT*-F (5′-ACAAGAAGGTGCAGAGTC ACAGGCCCCTC) and *TBXT*-R (GAGCTTCCTGCCCCAAATGAC AGATGCC-3′) (Supplemental Methods).

Data access

Raw sequencing reads from PacBio, Illumina, Hi-C, RNA-seq, and Iso-Seq generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under the accession numbers PRJNA685905 (Tibetan sheep assembly), PRJNA807094 (Marco Polo sheep assembly), PRJNA807886 (maternal assembly of the F_1-hybrid), and PRJNA808184 (paternal assembly of the F_1-hybrid). The assembly and annotation have been submitted to the NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/) under accession number GCA_017524585.1 (Tibetan sheep) and the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) under accession numbers JAKZEL000000000 (Marco Polo sheep), JALAIW000000000 (maternal haplotype of the F_1-hybrid), and JALAIX000000000 (paternal haplotype of the F_1-hybrid). Raw Illumina resequencing data for 402 individuals (i.e., 74 Tibetan sheep and 328 hybrids) and 110 individuals of F_2 population have been submitted to the NCBI BioProject database (PRJNA835294 and PRJNA838735) under NCBI Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) accession numbers SRR19392209–SRR19392610 and SRR19414940–SRR19415049.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank the farmers in Gannan Tibetan Autonomous Prefecture (Hezuo City, Gansu Province) and Associate Prof. Weimin Wang for their help in sampling and Prof. Yu Jiang and several members of his research group for advice and help in part of the data analyses. This study was financially supported by grants from the National Key Research and Development Program-Key Projects (2021YFD1200900), National Natural Science Foundation of China (nos. 31825024, 31661143014, 31972527, and 32061133010), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP; no. 2019QZKK0501), Initiation Fund of Sanya Institute of China Agricultural University (SYND-2022-12), the National Key Project for Cultivation of New Varieties of Genetically Modified Organisms by the Ministry of Agriculture (MOA) of China (no. 2016ZX08008001-002-001), and the Xinjiang Natural Science Foundation-Key Projects (no. 2021D01D13).

Author contributions: M.-H.L. and M.-J.L. designed and supervised the study. X.L., S.-G.H., X.W., and D.-X.M. performed the genome data analyses. X.L., W.-R.L., L.-Y.L., Z.Y., C.-X.L., X.-R.P., B.H., Z.-H.L., and L.C. performed the laboratory and/or farm work. M.-H.L., M.-J.L., X.L., S.-G.H, W.-R.L., L.-Y.L., Z.Y., F.-H.L., J.Y., Y.-X.X., J.D., Q.-H.Z., X.-L.X., S.-S.X., J.-L.H., X.-Z.D., R.D., Y.-F.C., J.-Y.W., L.-M.W., and P.Z. prepared the samples or provided help during the sample collection. X.L. and M.-H.L. wrote the manuscript with main contributions from M.-J.L., S.-G.H., and W.-R.L. All authors reviewed and approved the final version of the manuscript.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. *J Evol Biol* **26:** 229–246. doi:10.1111/j.1420-9101.2012.02599.x

Akera T, Chmátal L, Trimm E, Yang K, Aonbangkhen C, Chenoweth DM, Janke C, Schultz RM, Lampson MA. 2017. Spindle asymmetry drives non-Mendelian chromosome segregation. *Science* **358:** 668–672. doi:10.1126/science.aan0092

Akera T, Trimm E, Lampson MA. 2019. Molecular strategies of meiotic cheating by selfish centromeres. *Cell* **178:** 1132–1144.e10. doi:10.1016/j.cell.2019.07.001

Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, Benjelloun B, Librado P, Biscarini F, Colli L, et al. 2018. Convergent genomic signatures of domestication in sheep and goats. *Nat Commun* **9:** 813. doi:10.1038/s41467-018-03206-y

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19:** 1655–1664. doi:10.1101/gr.094052.109

Arnold ML. 2004. Natural hybridization and the evolution of domesticated, pest and disease organisms. *Mol Ecol* **13:** 997–1007. doi:10.1111/j.1365-294X.2004.02145.x

Barbato M, Hailer F, Orozco-terWengel P, Kijas J, Mereu P, Cabras P, Mazza R, Pirastru M, Bruford MW. 2017. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci Rep* **7:** 7623. doi:10.1038/s41598-017-07382-7

Bedell JA, Korf I, Gish W. 2000. MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics* **16:** 1040–1041. doi:10.1093/bioinformatics/16.11.1040
Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27:** 573–580. doi:10.1093/nar/27.2.573
Benton MJ, Donoghue PC. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol* **24:** 26–53. doi:10.1093/molbev/msl150
Blighe K, DeDionisio L, Christie K, Chawes B, Shareef S, Kakouli-Duarte T, Chao-Shern C, Harding V, Kelly R, Castellano L, et al. 2018. Gene editing in the context of an increasingly complex genome. *BMC Genomics* **19:** 595. doi:10.1186/s12864-018-4963-8
Bredemeyer KR, Harris AJ, Li G, Zhao L, Foley NM, Roelke-Parker M, O'Brien SJ, Lyons LA, Warren WC, Murphy WJ. 2021. Ultracontinuous single haplotype genome assemblies for the domestic cat (*Felis catus*) and Asian leopard cat (*Prionailurus bengalensis*). *J Hered* **112:** 165–173. doi:10.1093/jhered/esaa057
Brown D, Meadowcroft S. 1996. *The modern shepherd*. Farming Press, Ipswich, UK.
Buckingham KJ, McMillin MJ, Brassil MM, Shively KM, Magnaye KM, Cortes A, Weinmann AS, Lyons LA, Bamshad MJ. 2013. Multiple mutant T alleles cause haploinsufficiency of brachyury and short tails in Manx cats. *Mamm Genome* **24:** 400–408. doi:10.1007/s00335-013-9471-1
Bunch T, Foote W. 1977. Evolution of the $2n = 54$ karyotype of Domestic sheep (*Ovis aries*). *Ann Genet Sel Anim* **9:** 509–515. doi:10.1186/1297-9686-9-4-509
Bunch TD, Wu C, Zhang Y-P, Wang S. 2006. Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. *J Hered* **97:** 21–30. doi:10.1093/jhered/esi127
Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat Biotechnol* **31:** 1119–1125. doi:10.1038/nbt.2727
Cao Y-H, Xu S-S, Shen M, Chen Z-H, Gao L, Lv F-H, Xie X-L, Wang X-H, Yang H, Liu C-B, et al. 2021. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Mol Biol Evol* **38:** 838–855. doi:10.1093/molbev/msaa236
Caro T, Lombardo L, Goldizen A, Kelly M. 1995. Tail-flagging and other antipredator signals in white-tailed deer: new data and synthesis. *Behav Ecol* **6:** 442–450. doi:10.1093/beheco/6.4.442
Chen ZJ. 2013. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat Rev Genet* **14:** 471–482. doi:10.1038/nrg3503
Chen W, Zou M, Li Y, Zhu S, Li X, Li J. 2021a. Sequencing an F1 hybrid of *Silurus asotus* and *S. meridionalis* enabled the assembly of high-quality parental genomes. *Sci Rep* **11:** 13797. doi:10.1038/s41598-021-93257-x
Chen Z-H, Xu Y-X, Xie X-L, Wang D-F, Aguilar-Gómez D, Liu G-J, Li X, Esmailizadeh A, Rezaei V, Kantanen J, et al. 2021b. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Commun Biol* **4:** 1307. doi:10.1038/s42003-021-02817-4
Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18:** 170–175. doi:10.1038/s41592-020-01056-5
Cheng H, Zhang Z, Wen J, Lenstra JA, Heller R, Cai Y, Guo Y, Li M, Li R, Li W, et al. 2022. Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep. bioRxiv doi:10.1101/2022.05.17.492311
Cottone L, Cribbs AP, Khandelwal G, Wells G, Ligammari L, Philpott M, Tumber A, Lombard P, Hookway ES, Szommer T, et al. 2020. Inhibition of histone H3K27 demethylases inactivates brachyury (TBXT) and promotes chordoma cell death. *Cancer Res* **80:** 4540–4551. doi:10.1158/0008-5472.CAN-20-1387
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158. doi:10.1093/bioinformatics/btr330
Deng J, Xie X-L, Wang D-F, Zhao C, Lv F-H, Li X, Yang J, Yu J-L, Shen M, Gao L, et al. 2020. Paternal origins and migratory episodes of domestic sheep. *Curr Biol* **30:** 4085–4095.e6. doi:10.1016/j.cub.2020.07.077
de Vos JM, Augustijnen H, Bätscher L, Lucek K. 2020. Speciation through chromosomal fusion and fission in *Lepidoptera*. *Philos Trans R Soc Lond B Biol Sci* **375:** 20190539. doi:10.1098/rstb.2019.0539
Du LX. 2011. *Animal genetic resources in China*. China Agriculture Press, Beijing.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20:** 238. doi:10.1186/s13059-019-1832-y
Fedosenko AK, Blank DA. 2005. Ovis ammon. *Mamm Species* **773:** 1–15. doi:10.1644/1545-1410(2005)773[0001:OA]2.0.CO;2
Feng X, Cheung JPY, Je JS, Cheung PW, Chen S, Yue M, Wang N, Choi VN, Yang X, Song YQ, et al. 2021. Genetic variants of *TBX6* and *TBXT* identified in patients with congenital scoliosis in Southern China. *J Orthop Res* **39:** 971–988. doi:10.1002/jor.24805
Goulet BE, Roda F, Hopkins R. 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiol* **173:** 65–78. doi:10.1104/pp.16.01340
Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33:** D121–D124. doi:10.1093/nar/gki081
Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol* **9:** R7. doi:10.1186/gb-2008-9-1-r7
Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* **30:** 1987–1997. doi:10.1093/molbev/mst100
Han J, Yang M, Guo T, Niu C, Liu J, Yue Y, Yuan C, Yang B. 2019. Two linked *TBXT* (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Anim Genet* **50:** 772–777. doi:10.1111/age.12852
Han Z, Cui K, Placek K, Hong N, Lin C, Chen W, Zhao K, Jin W. 2020. Diploid genome architecture revealed by multi-omic data of hybrid mice. *Genome Res* **30:** 1097–1106. doi:10.1101/gr.257568.119
Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* **2:** 280–291. doi:10.1038/35066065
Haworth K, Putt W, Cattanach B, Breen M, Binns M, Lingaas F, Edwards YH. 2001. Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mamm Genome* **12:** 212–218. doi:10.1007/s003350010253
Hsuchou H, He Y, Kastin AJ, Tu H, Markadakis EN, Rogers RC, Fossier PB, Pan W. 2009. Obesity induces functional astrocytic leptin receptors in hypothalamus. *Brain* **132:** 889–902. doi:10.1093/brain/awp029
Hu X-J, Yang J, Xie X-L, Lv F-H, Cao Y-H, Li W-R, Liu M-J, Wang Y-T, Li J-Q, Liu Y-G, et al. 2019. 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. doi:10.1093/molbev/msy208
Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36:** 2253–2255. doi:10.1093/bioinformatics/btz891
Iannuzzi L, Di Berardino D. 2008. Tools of the trade: diagnostics and research in domestic animal cytogenetics. *J Appl Genet* **49:** 357–366. doi:10.1007/BF03195634
Jalessi M, Gholami MS, Razmara E, Hassanzadeh S, Sadeghipour A, Jahanbakhshi A, Tabibkhooei A, Bahrami E, Falah M. 2022. Association between *TBXT* rs2305089 polymorphism and chordoma in Iranian patients identified by a developed T-ARMS-PCR assay. *J Clin Lab Anal* **36:** e24150. doi:10.1002/jcla.24150
Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42:** 348–354. doi:10.1038/ng.548
Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. 2016. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res* **44:** e89. doi:10.1093/nar/gkw092
Kurahashi H, Tsutsumi M, Nishiyama S, Kogo H, Inagaki H, Ohye T. 2012. Molecular basis of maternal age-related increase in oocyte aneuploidy. *Congenit Anom (Kyoto)* **52:** 8–15. doi:10.1111/j.1741-4520.2011.00350.x
Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645. doi:10.1101/gr.092759.109
Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35:** 3100–3108. doi:10.1093/nar/gkm160
Lauvergne J, Hoogschagen P. 1978. Genetic formulas for the colour in the *Texel*, the *Dutch* and the *Zwartbles* sheep in the Netherlands. *Ann Genet Sel Anim* **10:** 343–351. doi:10.1186/1297-9686-10-3-343
Lawal RA, Martin SH, Vanmechelen K, Vereijken A, Silva P, Al-Atiyat RM, Aljumaah RS, Mwacharo JM, Wu D-D, Zhang Y-P, et al. 2020. The wild species genome ancestry of domestic chickens. *BMC Biol* **18:** 13. doi:10.1186/s12915-020-0738-1
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34:** 3094–3100. doi:10.1093/bioinformatics/bty191
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760. doi:10.1093/bioinformatics/btp324
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955–964. doi:10.1093/nar/25.5.955

Luo X, Zhou Y, Zhang B, Zhang Y, Wang X, Feng T, Li Z, Cui K, Wang Z, Luo C, et al. 2020. Understanding divergent domestication traits from the whole-genome sequencing of swamp- and river-buffalo populations. *Natl Sci Rev* **7:** 686–701. doi:10.1093/nsr/nwaa024

Lv F-H, Cao Y-H, Liu G-J, Luo L-Y, Lu R, Liu M-J, Li W-R, Zhou P, Wang X-H, Shen M, et al. 2022. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. *Mol Biol Evol* **39:** msab353. doi:10.1093/molbev/msab353

Mabry ME, Rowan TN, Pires JC, Decker JE. 2021. Feralization: confronting the complexity of domestication and evolution. *Trends Genet* **37:** 302–305. doi:10.1016/j.tig.2021.01.005

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol* **14:** e1005944. doi:10.1371/journal.pcbi.1005944

Martin BL, Kimelman D. 2008. Regulation of canonical Wnt signaling by brachyury is essential for posterior mesoderm formation. *Dev Cell* **15:** 121–133. doi:10.1016/j.devcel.2008.04.013

Mayr B, Krutzler J, Schleger W, Auer H. 1986. A new type of Robertsonian translocation in the domestic dog. *J Hered* **77:** 127. doi:10.1093/oxfordjournals.jhered.a110185

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226:** 792–801. doi:10.1126/science.15739260

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303. doi:10.1101/gr.107524.110

Meisler MH. 1997. Mutation watch: mouse brachyury (T), the T-box gene family, and human disease. *Mamm Genome* **8:** 799–800. doi:10.1007/s003359900581

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37:** 1530–1534. doi:10.1093/molbev/msaa015

Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. 2021. The genomic consequences of hybridization. *eLife* **10:** e69016. doi:10.7554/eLife.69016

Mori E, Maggini I, Menchetti M. 2014. When quills kill: the defense strategy of the crested porcupine *Hystrix cristata* L., 1758. *Mammalia* **78:** 229–234. doi:10.1515/mammalia-2013-0126

Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, Lee J, Chu C, Lin C, Džakula Ž, et al. 2016. A hybrid approach for *de novo* human genome sequence assembly and phasing. *Nat Methods* **13:** 587–590. doi:10.1038/nmeth.3865

Nagao M, Lanjakornsiripan D, Itoh Y, Kishi Y, Ogata T, Gotoh Y. 2014. High mobility group nucleosome-binding family proteins promote astrocyte differentiation of neural precursor cells. *Stem Cells* **32:** 2983–2997. doi:10.1002/stem.1787

Nattestad M, Schatz MC. 2016. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32:** 3021–3023. doi:10.1093/bioinformatics/btw369

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29:** 2933–2935. doi:10.1093/bioinformatics/btt509

Nikalayevich E, Verlhac M-H. 2021. Selfish centromeres, selfless heterochromatin. *Cell* **184:** 4843–4844. doi:10.1016/j.cell.2021.08.027

O'Connor SM, Dawson TJ, Kram R, Donelan JM. 2014. The kangaroo's tail propels and powers pentapedal locomotion. *Biol Lett* **10:** 20140381. doi:10.1098/rsbl.2014.0381

Page SL, Shin J-C, Han J-Y, Andy Choo K, Shaffer LG. 1996. Breakpoint diversity illustrates distinct mechanisms for Robertsonian translocation formation. *Hum Mol Genet* **5:** 1279–1288. doi:10.1093/hmg/5.9.1279

Pan Z, Li S, Liu Q, Wang Z, Zhou Z, Di R, Miao B, Hu W, Wang X, Hu X, et al. 2018. Whole-genome sequences of 89 Chinese sheep suggest role of *RXFP2* in the development of unique horn phenotype as response to semi-feralization. *Gigascience* **7:** giy019. doi:10.1093/gigascience/giy019

Pauciullo A, Knorr C, Perucatti A, Iannuzzi A, Iannuzzi L, Erhardt G. 2016. Characterization of a very rare case of living ewe-buck hybrid using classical and molecular cytogenetics. *Sci Rep* **6:** 34781. doi:10.1038/srep34781

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559–575. doi:10.1086/519795

Randi E. 2008. Detecting hybridization between wild species and their domesticated relatives. *Mol Ecol* **17:** 285–293. doi:10.1111/j.1365-294X.2007.03417.x

Rezaei HR, Naderi S, Chintauan-Marquier IC, Taberlet P, Virk AT, Naghash HR, Rioux D, Kaboli M, Luikart G, et al. 2010. Evolution and taxonomy of the wild species of the genus *Ovis* (Mammalia, Artiodactyla, Bovidae). *Mol Phylogenet Evol* **54:** 315–326. doi:10.1016/j.ympev.2009.10.037

Rice ES, Koren S, Rhie A, Heaton MP, Kalbfleisch TS, Hardy T, Hackett PH, Bickhart DM, Rosen BD, Ley BV, et al. 2020. Continuous chromosome-scale haplotypes assembled from a single interspecies F1 hybrid of yak and cattle. *Gigascience* **9:** giaa029. doi:10.1093/gigascience/giaa029

Salviano MB, Cursino MS, Zanetti EDS, Abril VV, Duarte JMB. 2017. Intraspecific chromosome polymorphisms can lead to reproductive isolation and speciation: an example in red brocket deer (*Mazama americana*). *Biol Reprod* **96:** 1279–1287. doi:10.1093/biolre/iox041

Schröder O, Lieckfeldt D, Lutz W, Rudloff C, Frölich K, Ludwig A. 2016. Limited hybridization between domestic sheep and the European mouflon in Western Germany. *Eur J Wildl Res* **62:** 307–314. doi:10.1007/s10344-016-1003-3

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16:** 259. doi:10.1186/s13059-015-0831-x

Sieber M, Freeman AE, Kelley DH. 1988. Relationships between body measurements, body weight, and productivity in Holstein dairy cows. *J Dairy Sci* **71:** 3437–3445. doi:10.3168/jds.S0022-0302(88)79949-X

Siniscalchi M, Lusito R, Vallortigara G, Quaranta A. 2013. Seeing left- or right-asymmetric tail wagging produces different emotional responses in dogs. *Curr Biol* **23:** 2279–2282. doi:10.1016/j.cub.2013.09.027

Smith J, Price B, Green J, Weigel D, Herrmann B. 1991. Expression of a *Xenopus* homolog of brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67:** 79–87. doi:10.1016/0092-8674(91)90573-H

Smith B, Aseltine M, Kennedy G. 1997. *Beginning shepherd's manual*, 2nd ed. Iowa State University Press, Ames, IA.

Song J, Sun L, Xu S, Liu N, Yao Y, Liu Z, Wang W, Rong H, Wang B. 2016. A family with Robertsonian translocation: a potential mechanism of speciation in humans. *Mol Cytogenet* **9:** 48. doi:10.1186/s13039-016-0255-7

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform* **7:** 62. doi:10.1186/1471-2105-7-62

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24:** 637–644. doi:10.1093/bioinformatics/btn013

Stelkens R, Seehausen O. 2009. Genetic distance between species predicts novel trait expression in their hybrids. *Evolution (N Y)* **63:** 884–897. doi:10.1111/j.1558-5646.2008.00599.x

Sun Y, Lu Z, Zhu X, Ma H. 2020. Genomic basis of homoploid hybrid speciation within chestnut trees. *Nat Commun* **11:** 3375. doi:10.1038/s41467-020-17111-w

Thomsen PD, Schauser K, Bertelsen MF, Vejlsted M, Grøndahl C, Christensen K. 2011. Meiotic studies in infertile domestic pig-babirusa hybrids. *Cytogenet Genome Res* **132:** 124–128. doi:10.1159/000320421

Urasaki N, Takagi H, Natsume S, Uemura A, Taniai N, Miyagi N, Fukushima M, Suzuki S, Tarora K, Tamaki M, et al. 2017. Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res* **24:** 51–58. doi:10.1093/dnares/dsw047

Vidale M, Berlioz S, Mohammed R. 2022. Iconographic evidence of hybridisation between *Camelus bactrianus* and *Camelus dromedarius* at second-century AD Hatra, Iraq. *Antiquity* **96:** 201–207. doi:10.15184/aqy.2021.169

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19:** 327–335. doi:10.1101/gr.073585.107

Walker C, Vierck CJ Jr, Ritz LA. 1998. Balance in the cat: role of the tail and effects of sacrocaudal transection. *Behav Brain Res* **91:** 41–47. doi:10.1016/S0166-4328(97)00101-0

Wang X, Wang L. 2016. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Front Plant Sci* **7:** 1350. doi:10.3389/fpls.2016.01350

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H, et al. 2012. *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40:** e49. doi:10.1093/nar/gkr1293

Wang X, Liu J, Niu Y, Li Y, Zhou S, Li C, Ma B, Kou Q, Petersen B, Sonstegard T, et al. 2018a. Low incidence of SNVs and indels in trio genomes of

Cas9-mediated multiplex edited sheep. *BMC Genomics* **19:** 397. doi:10.1186/s12864-018-4712-z

Wang X, Miao J, Xia J, Chang T, Guangxin E, Bao J, Jin S, Xu L, Zhang L, Zhu B, et al. 2018b. Identifying novel genes for carcass traits by testing G×E interaction through genome-wide meta-analysis in Chinese Simmental beef cattle. *Livest Sci* **212:** 75–82. doi:10.1016/j.livsci.2018.04.001

Weaver S. 2005. *Sheep: small-scale sheep keeping for pleasure and profit*. Hobby Farm Press, Irvine, CA.

Willey JS, Biknevicius AR, Reilly SM, Earls KD. 2004. The tale of the tail: limb function and locomotor mechanics in *Alligator mississippiensis*. *J Exp Biol* **207:** 553–563. doi:10.1242/jeb.00774

Wooster C. 2005. *Living with sheep: everything you need to know to raise your own flock*. The Lyons Press, Guilford, CT.

Woronzow N, Korobizgna K, Nadler C, Hofman R, Esalojnitskow T, Gorelow J. 1972. Chromossomi dikich baranow i proisschojdjenije domaschnich owjez. *Lriroda* **3:** 74–81.

Wu B, Shao Y, Chen B, Liu C, Xue Z, Wu P, Li H. 2010. Identification of a novel mouse brachyury (T) allele causing a short tail mutation in mice. *Cell Biochem Biophys* **58:** 129–135. doi:10.1007/s12013-010-9097-9

Wu D-D, Ding X-D, Wang S, Wójcik JM, Zhang Y, Tokarska M, Li Y, Wang M-S, Faruque O, Nielsen R, et al. 2018. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat Ecol Evol* **2:** 1139–1145. doi:10.1038/s41559-018-0562-y

Xia B, Zhang W, Wudzinska A, Huang E, Brosh R, Pour M, Miller A, Dasen JS, Maurano MT, Kim SY, et al. 2021. The genetic basis of tail-loss evolution in humans and apes. bioRxiv doi:10.1101/2021.09.14.460388

Xiao Y, Jiang S, Cheng Q, Wang X, Yan J, Zhang R, Qiao F, Ma C, Luo J, Li W, et al. 2021. The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol* **22:** 148. doi:10.1186/s13059-021-02370-7

Xu B, Xie X. 2016. Neurotrophic factor control of satiety and body weight. *Nat Rev Neurosci* **17:** 282–292. doi:10.1038/nrn.2016.24

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24:** 1586–1591. doi:10.1093/molbev/msm088

Yang J, Lee SH, Goddard ME, Visscher PM. 2011a. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88:** 76–82. doi:10.1016/j.ajhg.2010.11.011

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, De Andrade M, Feenstra B, Feingold E, Hayes MG, et al. 2011b. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* **43:** 519–525. doi:10.1038/ng.823

Yang Y, Wang Y, Zhao Y, Zhang X, Li R, Chen L, Zhang G, Jiang Y, Qiu Q, Wang W, et al. 2017. Draft genome of the Marco Polo Sheep (*Ovis ammon polii*). *Gigascience* **6:** 1–7. doi:10.1093/gigascience/gix106

Young JW, Russo GA, Fellmann CD, Thatikunta MA, Chadwell BA. 2015. Tail function during arboreal quadrupedalism in squirrel monkeys (*Saimiri boliviensis*) and tamarins (*Saguinus oedipus*). *J Exp Zool A Ecol Genet Physiol* **323:** 556–566. doi:10.1002/jez.1948

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44:** 821–824. doi:10.1038/ng.2310

Zhou Z, Li M, Cheng H, Fan W, Yuan Z, Gao Q, Xu Y, Guo Z, Zhang Y, Hu J, et al. 2018. An intercross population study reveals genes associated with body size and plumage color in ducks. *Nat Commun* **9:** 2648. doi:10.1038/s41467-018-04868-4

Received April 1, 2022; accepted in revised form July 29, 2022.